FRANCISCO MONTES SUAY

PROBABILIDAD

UNIVERSITAT DE VALÈNCIA

Índice general

Capítulo 1

El concepto de Probabilidad

EL CONCEPTO DE PROBABILIDAD

SUCESOS CIERTO, IMPOSIBLE Y ALEATORIO.- Sobre la base de la observación y la experimentación las ciencias enuncian leyes que explican y gobiernan los fenómenos bajo estudio. La más elemental y general de estas leyes podría enunciarse de la siguiente manera:

1. En cada realización de un conjunto de condiciones $\mathcal{G}$ ocurre el suceso A.

En efecto, para ilustrar esta afirmación general supongamos calentamos a 100° C y a la presión atmosférica, 760 mm (conjunto de condiciones $\mathcal{G}$), agua, el resultado será que dicha agua se transformará en vapor (suceso A). Análogamente, para cualquier reacción química de substancias sin intercambio con el medio que le rodea (condiciones $\mathcal{G}$), la cantidad total de materia permanece constante (suceso A). Esta afirmación es conocida como la ley de conservación de la materia. Fenómenos como los descritos son sobradamente conocidos en Química, Física, Biología y otras ciencias experimentales.

Los sucesos que resultan de la realización de determinado conjunto de condiciones pueden agruparse en tres tipos:

a) Aquellos sucesos que inevitablemente ocurren para cada realización del conjunto de condiciones $\mathcal{G}$ y que se denominan sucesos cierto o seguros

b) Si A no puede ocurrir cuando se realizan las condiciones $\mathcal{G}$, se dice que A es un suceso imposible, finalmente

c) Si realizadas las condiciones $\mathcal{G}$, A puede o no puede ocurrir, se dice entonces que A es un suceso aleatorio.

** Obsérvese que de estas definiciones los conceptos de certeza, imposibilidad y aleatoriedad están siempre referidos a un conjunto definido de condiciones $\mathcal{G}$.

La afirmación de aleatoriedad para un suceso A aporta muy poca información para el conocimiento del suceso, establece simplemente que el conjunto de condiciones $\mathcal{G}$ no refleja la colección entera de razones necesarias y suficientes para la realización de A. En nuestro interés por conocer más acerca de este tipo de sucesos, observamos que existen una gran cantidad de fenómenos en los que, dada una serie de repeticiones de las realizaciones de la condiciones $\mathcal{G}$, la proporción de ocurrencias de A sólo ocasionalmente se desvía de forma significativa de un cierto valor medio, valor que puede servirnos para caracterizar de alguna manera el suceso examinado. Para este tipo de fenómenos podemos, no sólo establecer su aleatoriedad, sino estimar cuantitativamente la posibilidad de su ocurrencia. Esta estimación podemos expresarla mediante una proposición del tipo:

2. La probabilidad de que un suceso A ocurra durante la realización de un conjunto de condiciones $\mathcal{G}$ es igual a p.

Regularidades de este tipo son conocidas como probabilísticas o estocásticas y juegan un importante papel en diversos campos de la ciencia. Por ejemplo, no hay forma de predecir si un determinado átomo de radio se desintegrará o no en un período de tiempo dado, pero es posible, sobre la base de experimentos de ajuste, determinar la probabilidad de tal desintegración y decir: un átomo de radio se desintegra en un intervalo de tiempo de t años con una probabilidad

$$p = 1 - e^{-0.000436t}$$

El conjunto de condiciones $\mathcal{G}$ consiste, aquí, en el hecho de considerar un átomo de radio, durante un período de t años, no sujeto a ninguna acción externa inusual (como bombardeo con partículas de alta velocidad, por ejemplo), cualesquiera otras condiciones de existencia son irrelevantes en el proceso que nos ocupa (presión, temperatura, etc..). El suceso A consiste en el hecho de que el átomo se desintegre en un período dado de t años.

La idea que parece ahora muy natural, de que la probabilidad de un suceso aleatorio A, bajo condiciones conocidas, admite una cuantificación mediante algún número $p = P(A)$ fue elaborada de manera sistemática por primera vez en el siglo 17 en los trabajos de Fermat (1601-1665), Pascal (1623-1662), Huygens (1629-1695) y en particular por Jacobo Bernoulli (1654-1705). Sus investigaciones colocaron los fundamentos de la teoría de la probabilidad. Desde aquel tiempo, esta teoría ha estado desarrollándose como una disciplina matemática y ha sido enriquecida con importantes resultados. Su aplicación al estudio de fenómenos reales de la más diversa naturaleza continúa encontrando nuevas e importantes confirmaciones.

No existe ninguna duda en cuanto a que el concepto de probabilidad matemática comporta connotaciones filosóficas que justifican su estudio desde esta perspectiva. El problema básico, desde la perspectiva filosófica, sería el siguiente: bajo qué condiciones existe significación objetiva en la estimación cuantitativa de la probabilidad de un suceso aleatorio A con la ayuda de un número fijo $P(A)$, llamado probabilidad matemática del suceso A, y cuál es la significación objetiva de esta estimación?

Cada investigador que aplica la teoría de la probabilidad a su área específica, procede realmente con la convicción de que los juicios probabilísticos expresan ciertas propiedades objetivas del fenómeno bajo estudio. Afirmar que bajo ciertas condiciones $\mathcal{G}$ la ocurrencia del suceso A tiene una probabilidad p es afirmar que entre el conjunto de condiciones y el suceso existe una cierta relación perfectamente definida, aunque peculiar y no por ello menos objetiva, relación que es independiente del investigador. El problema es determinar la naturaleza de esta relación. La dificultad de contestar a este último interrogante ha hecho posible la circunstancia paradójica de que algunos científicos hayan afirmado que los juicios probabilísticos dependen sólo del estado del investigador. Hay que señalar, que si bien esto no es así, la experiencia demuestra que el significado objetivo de la probabilidad matemática tiene sentido en el contexto de un conjunto de condiciones perfectamente definido.

En cualquier caso cuestiones de este tipo no deben ser objeto de estudio en un curso como el presente y nos ocuparemos tan sólo de los aspectos matemáticos del problema al margen de la determinación de la naturaleza del concepto.

DIFERENTES APROXIMACIONES A LA DEFINICIÓN DE PROBABILIDAD

Un gran número de definiciones de la probabilidad matemática han sido propuestas por diferentes autores. No nos ocuparemos aquí de todas ellas. Hay que señalar que el interés por una definición lógicamente irreprochable del concepto de probabilidad surgió más tarde, históricamente hablando, que la capacidad para determinar la probabilidad de los sucesos, para llevar a cabo cálculos con estas probabilidades y también para utilizar los resultados de estos cálculos en asuntos prácticos y en investigación científica. Por esta razón, en gran parte de los intentos por establecer una definición científica del concepto general de probabilidad es fácil percibir varios aspectos del proceso de conocimiento concreto que en cada caso específico conduce a una determinación real de la probabilidad de un suceso dado, ya se trate de la probabilidad de obtener un seis en cuatro lanzamientos de un dado o la probabilidad de desintegración de un átomo de una materia radiactiva.

En la gran mayoría de las definiciones podemos establecer los tres grupos siguientes:

1.- Definiciones de la probabilidad matemática como una medida cuantitativa del <u>grado de certidumbre</u> del observador.

2.- Definiciones que reducen el concepto al de <u>igual posibilidad</u>, el más primitivo de todos (se las conoce como <u>definiciones clásicas</u>).

3.- Definiciones que proceden a partir de la <u>frecuencia</u> de ocurrencia de un suceso en un elevado número de pruebas (<u>definiciones estadísticas</u>).

Nos ocuparemos de las del tipo 2 y 3 no sin antes criticar las definiciones agrupadas en el tipo 1. En efecto, el hecho de admitir la existencia de la realidad al margen de nosotros y del conocimiento que de ella tengamos y de tener en cuenta el éxito de los ejercicios probabilísticos en el conocimiento y aprendizaje del mundo exterior, debería dejar suficientemente claro que las definiciones puramente subjetivas de la probabilidad matemática son totalmente insostenibles. El problema es que en el lenguaje ordinario las expresiones "probablemente", "muy probablemente", "es altamente improbable", etc. expresan simplemente la actitud del que habla respecto de la verdad o la falsedad de algún juicio <u>simple</u>, en contra de otras interpretaciones que justificarían definiciones subjetivas como las del tipo 1. Debemos por tanto poner el énfasis en una circunstancia que todavía no ha sido señalada: <u>mientras en todas las aplicaciones científicas</u> de la teoría de la probabilidad "probabilidad" es la probabilidad de la ocurrencia de algún suceso A, cuando un determinado conjunto de condiciones $\mathcal{G}$, que es <u>fundamentalmente reproducible un infinito número de veces</u>, ha sido realizado (es sólo en tal reiteración que la afirmación $p = P(A)$ expresa una cierta regularidad con significado objetivo), <u>en el habla ordinaria</u> es costumbre hablar de una mayor o menor probabilidad de algún juicio <u>muy definido</u>. Veamos un ejemplo a través de los dos juicios siguientes:

a) cualquier número natural, par, mayor que 2 puede ser representado mediante la suma de dos números primos $(4=2+2,\ 6=3+3,\ 8=5+3,\ \ldots)$

b) nevará en Valencia el 3 de Febrero de 1982.

Respecto de la afirmación hecha en a) si bien no tenemos un conocimiento completo muchos se inclinarían a creerlo como extremadamente probable, en cuanto a la afirmación hecha en b) habrá que esperar al 3 de Febrero de 1982 para tener una respuesta exacta. Sin embargo y dado que no suele nevar en Valencia nunca, esta afirmación será considerada altamente improbable. El concepto de probabilidad utilizado en el análisis de estos dos juicios no tiene nada que ver con el correspondiente concepto matemático. En ambos casos se trata de reflejar la actitud del que habla respecto de la cuestión planteada, actitud que tendrá su confirmación o negación cuando a) sea probado o en el caso de b) el día 2 de Febrero de 1982 y por tanto carecería de sentido pretender estimar su probabilidad matemática ahora porque, insistimos, se trata de hechos aislados (afirmación b)) o juicios específicos (afirmación a)).

ESPACIO MUESTRAL

Vamos a ocuparnos en primer lugar y de manera exhaustiva de la definición clásica de probabilidad, y de su utilidad para resolver determinados tipos de problemas con unas características que en su momento describiremos. Previo a ello hemos de introducir una serie de conceptos preliminares que nos serán útiles posteriormente.

Consideremos un conjunto fijo de condiciones $\mathcal{G}$ y una familia $\mathcal{S}$ de sucesos, A, B, C... que pueden ocurrir o no cuando se realiza $\mathcal{G}$. Existen entre los elementos de $\mathcal{S}$ ciertas relaciones de interés que vamos a poner de manifiesto.

1.- Si para cada realización del conjunto de condiciones $\mathcal{G}$ en la que ocurre el suceso A ocurre también el suceso B, decimos que A <u>implica</u> B y lo denotaremos mediante

$$A \subset B \quad \text{o} \quad B \supset A.$$

2.- El hecho de que A y B se impliquen mutuamente supone que ambos sucesos son <u>equivalentes</u>, lo que denotamos mediante $A = B$. Hemos de señalar que por lo que respecta a la teoría de la probabilidad, los sucesos equivalentes pueden reemplazarse uno por otro y por tanto en adelante los consideraremos simplemente como idénticos.

3.- Un suceso que consiste en la ocurrencia simultánea de otros dos A y B se llamará <u>producto</u> de A y B y lo denotaremos AB (o $A \cap B$).

4.- Un suceso que consiste en la ocurrencia de al menos uno de los sucesos A y B se llama <u>suma</u> de A y B, su representación es $A+B$ (o $A \cup B$).

5.- Un suceso que consiste en el hecho de que ocurra A pero no B se llama <u>diferencia</u>

de A y B y se designa mediante $A-B$.

Ilustremos mediante un ejemplo estos nuevos conceptos. Supongamos que G consiste en el hecho de lanzar un dado una vez. Sea A obtener un seis en el lanzamiento, B obtener un tres, C obtener un número par de puntos y D obtener un múltiplo de 3. Los sucesos A, B, C y D están relacionados de la siguiente manera

$$A \subset C \ , \ A \subset D \ , \ B \subset D$$

$$A + B = D \ , \ CD = A$$

La definición de suma y producto de dos sucesos puede generalizarse a un número finito de sucesos mediante

$$A+B+\ldots+N \quad (A\cup B\cup \ldots \cup N)$$

que significa un suceso que consiste en la ocurrencia de al menos uno de los sucesos $A, B, \ldots, N$. Análogamente

$$AB\ldots N \quad (A\cap B\cap \ldots \cap N)$$

significa un suceso consistente en la ocurrencia de todos los sucesos $A, B, \ldots, N$.

6.- Un suceso se llama *cierto* si ocurre necesariamente al realizar G. Por ejemplo al lanzar dos dados la suma de sus caras es ciertamente mayor o igual a 2. Cuando el suceso no ocurre bajo las condiciones utilizadas, se le llama suceso *imposible*. Por ejemplo obtener 13 puntos en el lanzamiento de los dos dados. Como todos los sucesos ciertos e imposibles son equivalentes entre sí es costumbre designarlos genéricamente mediante las letras Ω y ϕ respectivamente (o cualquier otras, pero siempre la misma para todos ellos).

7.- Dos sucesos A y $\bar{A}$ se llaman *contrarios* si se verifican simultáneamente las dos relaciones

$$A+\bar{A} = \Omega \quad \text{y} \quad A\bar{A} = \phi$$

Por ejemplo, si C representa obtener un número par de puntos al lanzar un dado, entonces

$\bar{C} = \Omega - C$ representa obtener una suma impar de puntos.

8.- Dos sucesos A y B se dice que son *mutuamente excluyentes* o *incompatibles* si no pueden ocurrir simultáneamente, es decir, si

$$AB = \phi$$

Si

$$A = A_1 + A_2 + \ldots + A_n$$

con los A_i incompatibles entre sí, es decir, $A_iA_j = \phi$, $i \neq j$, decimos que A puede descomponerse en sucesos particulares $A_1 \ldots A_n$, o simplemente que los $A_1 \ldots A_n$ constituyen *una partición de* A. Por ejemplo, si C es el número par de puntos y C_2, C_4 y C_6 representan obtener 2, 4 o 6 en el lanzamiento respectivamente,

entonces $C = C_2 + C_4 + C_6$ y además son incompatibles entre sí.

Los sucesos $B_1, B_2, \ldots, B_n$ forman un *grupo completo de sucesos* [redacted] si al menos uno de ellos ocurre en la realización de G. Es decir

$$B_1 + B_2 + \ldots + B_n = \Omega .$$

Especialmente interesantes son los grupos completos de sucesos incompatibles dos a dos también llamados *particiones*.

9.- Cualquier problema en teoría de la probabilidad supone la existencia de un conjunto de condiciones G y de una familia de sucesos S que ocurren o no cuando se realizan las condiciones de G. La experiencia demuestra que es aconsejable hacer las siguientes suposiciones respecto del sistema descrito:

a) Si la familia S incluye los sucesos A y B, incluye también los sucesos AB, $A+B$, $A-B$.

b) La familia S contiene a los sucesos cierto e imposible.

Una familia de sucesos que satisface estas condiciones se denomina *álgebra de sucesos*.

Observamos que en los ejemplos utilizados como ilustración, siempre es posible aislar sucesos que no admiten descomposición en otros más simples: obtener un determinado número de puntos, entre 1 y 6 al lanzar un dado, obtener cara o cruz al lanzar una moneda, etc. A este tipo de sucesos que no admiten descomposición les llamaremos *sucesos simples* o *elementales*.

En la construcción de una teoría matemática de las probabilidades parece necesaria una formalización mayor de la llevada a cabo en los párrafos precedentes. En las exposiciones modernas de la teoría de la probabilidad el punto de partida es un conjunto de sucesos elementales al que denominamos espacio de sucesos simples o *espacio muestral*. La naturaleza de los elementos de dicho espacio no se especifica de antemano de manera que permita la suficiente libertad de elección como para abarcar todas las situaciones posibles. A los subconjuntos del espacio muestral se les conoce como *sucesos aleatorios*. Un suceso consistente en todos los puntos del espacio muestral es conocido como el *suceso cierto* o *seguro*. Todas las relaciones descritas anteriormente entre los sucesos se verifican también cuando llevamos a cabo una construcción formal de la teoría. En su momento volveremos sobre el tema para llevar a cabo de forma exhaustiva dicha construcción. Por ahora continuaremos con la situación descrita en los párrafos anteriores y en particular, para la definición clásica del concepto de probabilidad tendremos en cuenta tan solo los espacios muestrales con un número finito de elementos.

Digamos finalmente que entre los sucesos y las relaciones establecidas se verifican las siguientes leyes, que enunciamos sin demostración:

Ley Conmutativa: $A+B = B+A$, $AB = BA$

Ley Asociativa: $A+(B+C) = (A+B)+C$, $A(BC) = (AB)C$

Ley Distributiva: $A(B+C) = AB+AC$, $A+(BC) = (A+B)(A+C)$

Ley Idempotente: $A+A = A$, $AA = A$

LA DEFINICION CLASICA DE PROBABILIDAD

La definición clásica de probabilidad reduce el concepto de probabilidad a la noción de equiprobabilidad entre los sucesos elementales, noción que se considera básica y no está sujeta a definición formal.

Sea Ω un conjunto de n sucesos elementales todos ellos con igual probabilidad de ocurrir,

$$\Omega = \{E_1, \dots, E_n\}.$$

Designemos por $\mathcal{S}$ la familia formada por Ω, el suceso cierto, ϕ, el suceso imposible, los E_k, $k = 1, \dots, n$ y cualquier otro suceso A que admita una descomposición en elementos de Ω. Por ejemplo, si $\Omega = \{E_1, E_2, E_3\}$, entonces $\mathcal{S} = \{\Omega, E_1, E_2, E_3, E_1+E_2, E_1+E_3, E_2+E_3, \phi\}$. Se comprueba fácilmente que la familia $\mathcal{S}$ constituye una álgebra de sucesos. La definición clásica de probabilidad puede formularse como sigue:

Definición: Si A, un suceso de $\mathcal{S}$, está compuesto por m de los sucesos elementales que componen Ω, entonces la probabilidad del suceso A es

$$P(A) = \frac{m}{n}.$$

Esta fórmula es conocida con el nombre de fórmula de Laplace por ser este científico quien la formuló por primera vez en el siglo XVIII. Se enuncia también en ocasiones diciendo que "la probabilidad de un suceso A es el cociente entre el número de resultados del experimento favorables a la realización del mismo y el número de resultados posibles".

Ejemplo.- Supongamos que lanzamos un dado, el espacio muestral estará formado por $\Omega = \{E_1, \dots, E_6\}$ donde E_i representa el hecho de que haya aparecido una cara con i puntos. Definimos

$$G = E_2 + E_4 + E_6$$

que es un suceso que representa el hecho de que haya aparecido un número de puntos par en la cara, de acuerdo con la definición dada

$$P(G) = \frac{3}{6} = \frac{1}{2}.$$

Para cada uno de los sucesos elementales y siempre de acuerdo con la definición $P(E_i) = \frac{1}{6}$, $i = 1, \dots, 6$

Volvamos nuevamente a la definición de probabilidad que acabamos de dar. Observemos que $P(A)$ puede considerarse como una función definida sobre el álgebra de conjuntos, más exactamente como una función de conjunto definida sobre el álgebra $\mathcal{S}$ de los sucesos. Como tal función y a la vista de su definición posee las siguientes propiedades.

Propiedades de la probabilidad

1) $\forall A \in \mathcal{S} \quad P(A) \geq 0$

2) Para el suceso cierto, Ω, $P(\Omega) = 1$

3) Si A puede descomponerse de la forma $A = B + C$, con $BC = \phi$, entonces

$$P(A) = P(B) + P(C).$$

4) La probabilidad de $\bar{A}$, el suceso contrario de A, viene dada por

$$P(\bar{A}) = 1 - P(A).$$

5) La probabilidad del suceso imposible, ϕ, es cero.

6) Si un suceso A implica un suceso B, entonces

$$P(A) \leq P(B).$$

7) La probabilidad de cualquier suceso está comprendida entre 0 y 1, es decir

$$0 \leq P(A) \leq 1, \quad \forall A \in \mathcal{S}.$$

Las demostraciones de estas propiedades es sencilla y puede hacerse como ejercicio.

EJEMPLOS DE APLICACION DE LA DEFINICION CLASICA DE PROBABILIDAD

Consideraremos a continuación algunos casos en los que el cálculo de las correspondientes probabilidades puede llevarse a cabo mediante la utilización de la definición estudiada, por cuanto las condiciones del problema satisfacen las condiciones exigidas en aquella. Nos supondremos también de algunos resultados de combinatoria que son interesantes y útiles para la solución de este tipo de problemas.

EJEMPLO 1.- De una baraja de 36 cartas se extraen 3 al azar. Encontrar la probabilidad de que haya exactamente un as entre ellas.

El espacio muestral, todas las posibles extracciones de 3 cartas, está formado por $\binom{36}{3}$ puntos y además todos ellos son igualmente posibles. El número de estas extracciones que contienen un solo as puede determinarse de la siguiente manera: una de las cartas ha de ser un as y podrá ser uno de los cuatro, podrá ser elegido de $\binom{4}{1}$ maneras distintas. Las otras dos cartas deberán elegirse al azar entre las 32 cartas restantes del mazo y sólo podrá hacerse de $\binom{32}{2}$ maneras, en definitiva los casos favorables serán $\binom{4}{1} \cdot \binom{32}{2}$ y la probabilidad del suceso A, hay una carta entre las tres extraídas, será

$$P(A) = \frac{\binom{4}{1}\binom{32}{2}}{\binom{36}{3}} = \frac{496}{1785} \simeq 0.2778.$$

EJEMPLO 2.- De una baraja de 36 cartas se extraen 3 al azar. Probabilidad de que haya al menos un as entre ellas.

Sea A el suceso que nos interesa. Este suceso puede descomponerse en otros 3 incompatibles entre sí, a saber, A_1, hay un solo as en las tres cartas extraídas, A_2, hay 2 ases en las cartas extraídas y A_3, las tres son ases. De acuerdo con una propiedad de la probabilidad tendremos

$$P(A) = P(A_1) + P(A_2) + P(A_3).$$

Razonando como en el ejemplo anterior tendremos

$$P(A_i) = \binom{4}{i}\cdot\binom{32}{3-i}\Big/\binom{36}{3} \qquad , i=1,2,3.$$

y de aquí

$$P(A) = \frac{109}{3.119} \simeq 0.3053.$$

Puede resolverse este problema de otra forma. En efecto, podemos calcular la probabilidad de $\bar{A}$, suceso que representa el hecho de que no haya ningún as entre las cartas extraídas, los casos favorables se obtienen eliminando los 4 ases del mazo y eligiendo las tres cartas entre las 32 cartas restantes, de aquí

$$P(\bar{A}) = \frac{\binom{32}{3}}{\binom{36}{3}} = \frac{31.8}{3.17.7} \simeq 0.6947$$

y de aquí $P(A) = 1 - P(\bar{A}) \simeq 1 - 0.6947 \simeq 0.3053$.

Estos dos ejemplos ponen de manifiesto la utilidad de la combinatoria en la solución de este tipo de problemas. Vamos a dar a continuación algunos resultados de combinatoria que serán de gran ayuda.

Algunos resultados de combinatoria.

El primero de estos resultados, que se conoce en ocasiones con el nombre del Principio Fundamental del conteo, es de demostración inmediata y afirma:

Teorema 1.- Sea T un trabajo o tarea que requiere, para llevarse a cabo, de la presencia de ~~un~~ total de k trabajos, T_i, $i=1,\dots,k$, de manera que cada uno de estos puede realizarse de n_i, $i=1,\dots,k$ formas distintas. Entonces, el número de formas distintas de ejecutar T viene dado por $\prod_{i=1}^{k} n_i$.

A continuación vamos a considerar problemas de seleccionar bolas de una urna o de colocar bolas en celdas, problemas que sirven como modelos generales para gran número de problemas de la vida real. En todos ellos se aplicará implícita o explícitamente el principio fundamental que acabamos de enunciar.

Consideremos una urna con n bolas numeradas, todas ellas idénticas. Si extraemos k bolas de la urna, diremos que hemos extraído una muestra de tamaño k. La muestra decimos que es ordenada si el orden en que las bolas fueron extraídas es tenido en cuenta, en cualquier otro caso diremos que la muestra es desordenada. Veamos el siguiente resultado.

Teorema 2.- El número de muestras diferentes de tamaño k, ordenadas, es el número de variaciones de n elementos tomados de k en k, $V_{n,k} = n(n-1)\dots(n-k+1)$, si las extracciones han sido hechas sin reemplazamiento y es igual a n^k si las extracciones se llevaron a cabo con reemplazamiento.

Para muestras desordenadas dichos números son respectivamente

$\binom{n}{k}$ para muestras sin reemplazamiento, y

$\binom{n+k-1}{k}$ para muestras con reemplazamiento.

Demostración.- Para las muestras ordenadas bastará aplicar el principio antes enunciado con $n_j = (n-j+1)$ y $n_j = n$, $j=1,\dots,k$, respectivamente.

Para las muestras desordenadas extraídas sin reemplazamiento, observamos que cada una de ellas da lugar a $k!$ muestras ordenadas, por tanto si su número es x, tendremos

$$V_{n,k} = k!\,x \qquad \text{y de aquí} \qquad x = \frac{V_{n,k}}{k!} = \binom{n}{k}$$

para las extracciones con reemplazamiento puede recurrirse al método de inducción para comprobar la fórmula dada, no obstante vamos a presentar un método alternativo más sencillo. Supongamos que las n bolas son n cartas numeradas del 1 al n y que les añadimos $k-1$ cartas, numeradas del $n+1$ a $n+k-1$ conteniendo cada una una inscripción que dice: "repita la carta menor", "repita la 2ª carta menor", ——, "repita la $k-1$ ésima carta menor". Extraer una muestra de tamaño k con reemplazamiento de entre las n cartas iniciales es equivalente a extraer una muestra de tamaño k sin reemplazamiento de entre las $n+k-1$ cartas, así pues su número vendrá dado por $\binom{n+k-1}{k}$.

Una aplicación de estas propiedades podemos verla en el siguiente ejemplo:

Ejemplo.- ¿Cuál es la probabilidad de que una mano de poker contenga exactamente un par?

Una mano de poker es un subconjunto de 5 cartas elegidas entre las 52 del mazo de cartas. El espacio muestral de las posibles manos de poker es finito y contiene exactamente $\binom{52}{5}$ puntos. Por otra parte, una mano de poker con un par contiene dos cartas iguales y las otras tres diferentes entre sí y distintas de las que forman el par. Podemos obtener semejante mano mediante los siguientes pasos:

1) Elegir la carta que formará el par entre las 13 distintas de cada palo. Esto puede hacerse de $\binom{13}{1}$ formas

2) Elegida la carta en 1), elegir las dos cartas de ese número entre las cuatro (una de cada palo) que van a constituir la pareja. Lo haremos de $\binom{4}{2}$ formas distintas

3) Elegir ahora las tres cartas restantes, los números correspondientes, entre los 12 números que quedan después de eliminar el correspondiente a la pareja. Hay un total de $\binom{12}{3}$ formas de hacerlo

4) Finalmente determinar para cada uno de los tres números elegidos, cuál es el palo elegido lo podemos hacer de $4.4.4 = 4^3$ formas distintas.

Los casos favorables son pues $\binom{13}{1}\cdot\binom{4}{2}\cdot\binom{12}{3}\cdot 4^3$. En definitiva la probabilidad pedida vendrá dada por

$$\frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}4^3}{\binom{52}{5}} = \frac{1098240}{2598960} \simeq 0.42$$

Otro problema interesante que puede plantearse es el de la ocupación de celdas mediante bolas. Genéricamente este tipo de problemas son conocidos con el nombre de problemas de ocupación. El teorema siguiente es útil en la resolución de los mismos.

TEOREMA 2.- i) El número de maneras distintas en que n bolas <u>distintas</u> pueden distribuirse en k celdas distintas es k^n.

ii) Si deseamos que cada celda contenga un determinado número de bolas, es decir, la j-ésima celda contenga n_j bolas, $j=1,\dots,k$ $(n_j \geq 0,\ j=1,\dots,k,\ \sum n_j = n)$, entonces las maneras de llevar a cabo su distribución son

$$\frac{n!}{n_1!\,n_2!\cdots n_k!} = \binom{n}{n_1, n_2, \dots, n_k}.$$

iii) Si las n bolas son <u>indistinguibles</u> pero las k celdas son distintas, entonces el número de distribuciones es

$$\binom{k+n-1}{n}.$$

Además, si $n \geq k$ y ninguna celda debe de estar vacía, entonces esta cantidad se convierte en

$$\binom{n-1}{k-1}.$$

Demostración.- i) Cada una de las n bolas puede ocupar k posiciones distintas, por tanto k^n es el número de posibles distribuciones.

ii) El problema puede asimilarse al de establecer k grupos con las n bolas de manera que cada uno de ellos contenga un número determinado de bolas, exactamente n_j el j-ésimo grupo. Esto podemos hacerlo como sigue

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{k-1}}{n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!} = \binom{n}{n_1, n_2, \dots, n_k}.$$

iii) El problema puede reducirse a una de las situaciones descritas y resueltas en el teorema anterior. Supongamos una urna con k bolas numeradas de 1 a k, llevamos a cabo n extracciones con reemplazamiento y colocamos una bola en la celda que indica el número de la bola extraída, obviamente el orden de aparición es irrelevante. Bastará pues con determinar el número de muestras desordenadas de tamaño n, extraídas con reemplazamiento, que pueden obtenerse de una urna con k bolas numeradas de 1 a k. De acuerdo con el anterior teorema su número es

$$\binom{k+n-1}{n}$$

Si $n \geq k$ y queremos que todas las celdas contengan una bola al menos, depositamos la bola correspondiente en cada una y distribuimos como antes las $n-k$ restantes, tendremos

$$\binom{k+(n-k)-1}{n-k} = \binom{n-1}{n-k} = \binom{n-1}{k-1}.$$

EJEMPLO 3.- Tenemos n partículas, cada una de las cuales puede ocupar cada una de N $(N \geq n)$ celdas, con la misma probabilidad, $1/N$. Encontrar la probabilidad de que:

a) Haya una partícula en cada una de n celdas fijadas de antemano,

b) Haya una partícula en cada una de n celdas elegidas al azar.

Este problema es un problema clásico de ocupación que juega un papel importante en Mecánica Estadística. Como las condiciones del problema no aparecen determinadas de manera única existen varias respuestas al mismo, tantas como distintas de maneras de definir el espacio muestral (conjunto de los posibles resultados), que dan lugar a las distintas estadísticas, conocidas en Mecánica por el nombre de los autores que las desarrollaron. Exactamente tres son las respuestas que podemos dar a esta cuestión, son

1) <u>Estadística de Boltzmann</u>. Supongamos que las bolas y las celdas son <u>distinguibles</u>, y haciendo uso del teorema 2, apartado i) los casos posibles vendrán dados por N^n. Las respuestas a las dos probabilidades buscadas son

1.a) $p_1 = \dfrac{n!}{N^n}$, hay $n!$ formas de disponer las partículas distinguibles en las n celdas distintas.

1.b) $p_2 = \dfrac{\binom{N}{n} n!}{N^n} = \dfrac{N!}{N^n (N-n)!}$, $\binom{N}{n}$ son las formas de elegir las n ~~celdas~~ entre las N disponibles.

2) <u>Estadística de Bose-Einstein</u>. Si bien las celdas continúan siendo distintas, las n partículas son ahora <u>indistinguibles</u>, por tanto, volviendo recurriendo al teorema 2 el número de casos posibles es ahora $\binom{n+N-1}{n}$, y las probabilidades buscadas

2.a) $p_1 = \dfrac{1}{\binom{n+N-1}{n}}$, los casos favorables son 1 al no distinguirse las partículas

2.b) $p_2 = \dfrac{\binom{N}{n}}{\binom{n+N-1}{n}} = \dfrac{N!\,(N-1)!}{(N-n)!\,(N+n-1)!}$

3) <u>Estadística de Fermi-Dirac</u>. En esta estadística se supone que las celdas puede contener a lo sumo 1 partícula, partícula que además no se distingue de las otras. Los casos posibles son pues $\binom{N}{n}$, y las probabilidades

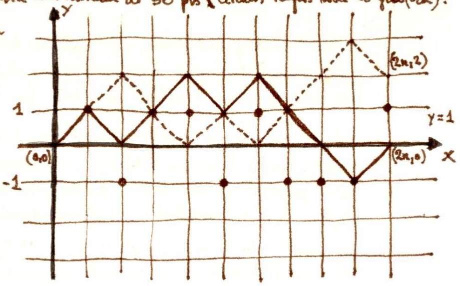
3.a) $p_1 = \dfrac{1}{\binom{N}{n}}$

3.b) $p_2 = \dfrac{\binom{N}{n}}{\binom{N}{n}} = 1.$

EJEMPLO 4.- En la taquilla de un matinal de cine hay una cola de $2n$ niños. El precio de la entrada son 50 pesetas y de los $2n$ niños, n llevan monedas de 50 ptas y los n restantes llevan billetes de 100 pesetas. Si el cajero no tiene cambio cuando comienza el despacho de entradas, ¿cual es la probabilidad de que ningún niño deba esperar para el cambio?

Resulta dificultoso en este caso la determinación del número de casos favorables por métodos directos, para averiguar tales casos procedemos de la siguiente forma:

Representaremos el orden de llegada a la taquilla en el eje de las x's. Sobre cada uno de los puntos levantaremos una ordenada que valdrá 1 si el niño va a pagar con un billete de 100 pts y -1 si lleva una moneda de 50 pts (Circulos negros sobre la gráfica).



A continuación trazamos una poligonal, partiendo de $(0,0)$, cuyos vertices son aquellos puntos que tienen por abcisa el orden de llegada y por ordenada la acumulación de todas las ordenadas anteriores más la correspondiente al punto en cuestión (vease la figura adjunta). A esta poligonal le llamaremos trayectoria. Dos propiedades interesantes tiene la trayectoria así definida:

a) Hay tantas trayectorias como ordenes distintos de llegada, a saber $\binom{2n}{n}$.

b) Todas las trayectorias comienzan en $(0,0)$ y acaban en $(2n,0)$.

De todas las trayectorias, las que se corresponden con ordenes de llegadas favorables a las condiciones requeridas en el enunciado son aquellas que no cruzan el eje de las x's (hacia las y's positivas). Ello supone que en cada instante, es decir a la llegada de cada niño, los que llevaban monedas de 50 pts han sido al menos 1 más que los que llevaban 100 pts para pagar su entrada, es decir, tenemos así garantizada la existencia de cambio para el mero comprador. Vamos a determinar el número de estas trayectorias.

Mediante la recta $y=1$, utilizada como eje de simetría, construiremos nuevas trayectorias a partir de las iniciales. Estas nuevas trayectorias comienzan también en $(0,0)$, coinciden con las antiguas mientras estas no alcanzan la recta $y=1$, cuando esto ocurre la recta hace de eje de simetría de la trayectoria. Esto supone:

a) Las trayectorias que no alcanzan $y=1$ conservan su forma inicial y no dan lugar a nuevas trayectorias ficticias. Observese que estas trayectorias son precisamente aquellas que no cruzan el eje de las x's, es decir, las favorables a la situación pedida.

b) Las demas trayectorias, las desfavorables, dan lugar a nuevas trayectorias (linea discontinua en la figura) que todas finalizan en el punto $(2n,2)$. Ello quiere decir que tienen 2 subidas más que bajadas, es decir, 2 individuos más con billetes de 100 pts

que los que pagan con monedas de 50 pts. Su número será pues $\binom{2n}{n+1}$.

En definitiva el número de casos favorables vendrá dado por

$$\binom{2n}{n} - \binom{2n}{n+1}$$

y la probabilidad pedida será

$$p = \frac{\binom{2n}{n} - \binom{2n}{n+1}}{\binom{2n}{n}} = \frac{\frac{2n!}{n!\,n!} - \frac{2n!}{(n+1)!\,(n-1)!}}{\frac{2n!}{n!\,n!}} = 1 - \frac{n!\,n!}{(n+1)!\,(n-1)!} =$$

$$= 1 - \frac{n}{n+1} = \frac{n+1-n}{n+1} = \frac{1}{n+1}.$$

FRECUENCIA Y PROBABILIDAD

La definición clásica de probabilidad encuentra en algunas situaciones, aquellos problemas que se derivan de las ciencias experimentales, por ejemplo, dificultades insalvables para su aplicación correctamente. Es muy dificil aislar y determinar con exactitud los casos igualmente probables o bien las condiciones de simetria que la equiprobabilidad comporta no se verifican en las situaciones estudiadas (desintegración de un átomo en un intervalo de tiempo dado, nacimiento de un niño, etc.)

Buscando soluciones a estas situaciones, se comprueba que las observaciones de la ocurrencia o no ocurrencia de un suceso A, para un gran número de experimentos repetidos bajo un mismo conjunto de condiciones G, muestran que gran cantidad de fenómenos son de tal naturaleza que el número de dichas ocurrencias (no ocurrencias) del suceso A obedece una ley estable. Concretando, si m es el número de veces que ocurre A en n experiencias independientes (los resultados de unas y otras no se influyen), se encuentra que la relación $\frac{m}{n}$, para n suficientemente grande, es practicamente constante, siendo sus desviaciones cada vez menos frecuentes a medida que aumenta n.

Esta especie de estabilidad de las frecuencias (es decir, la relación m/n) fue puesta de manifiesto primeramente para fenomenos demográficos y ello desde tiempos muy antiguos. Parece ser que en la China del año 2000 antes de Cristo ya había costumbre de confeccionar censos y a partir de las observaciones que ellos permitían se consideraba que la frecuencia de nacimientos de niños era una cantidad practicamente constante alrededor de 1/2. La estabilidad de estos nacimientos es un caso especialmente interesante, podría decirse que incluso anecdótico. En efecto, Laplace en su libro Essai philosophique sur les probabilités relata la siguiente anecdota:

Observaciones llevadas a cabo durante varias decadas en varias ciudades europeas, Londres, San Petersburgo, Berlin y en toda Francia le llevaron a la conclusión que el nacimiento de niños presentaba una frecuencia extraordinariamente coincidente en el tiempo y en el espacio, siendo dicho valor 22/43. Estudio Laplace dichos nacimientos durante un un periodo de 40 años (de 1745 a 1784) para la ciudad de Paris y encontró que su frecuencia era de 25/49, cantidad ligeramente diferente de la anterior, pero lo suficiente para sorprender a Laplace. Profundizando más en su estudio descubrió que en los censos utilizados figuraban tambien los niños expósitos y que además las poblaciones de los suburbios de Paris tenian preferencia por abandonar niños de un solo sexo. En aquella época este fenómeno era bastante corriente y lo suficientemente importante como para alterar los resultados de la encuesta que se estaba llevando a cabo. Corregido el censo de estos niños se encontró para Paris una frecuencia de nacimientos de niños de 22/43.

Desde la epoca de Laplace son muchos los experimentos llevados a cabo en este sentido. En la tabla se recogen los resultados de distintas experiencias llevadas a cabo en la determinación de la frecuencia con que aparecía cara en una serie de lanzamientos

Realizó la experiencia	Nº de lanzamientos	Nº de caras	Frecuencia
BUFFON	4040	2028	0.5080
KARL PEARSON	12000	6019	0.5016
KARL PEARSON	24000	12012	0.5005

La tecnología actual permite llevar a cabo de manera sencilla y rápida otras comprobaciones de este hecho experimental (por ejemplo mediante la generación de números aleatorios con un ordenador).

La verificación de este hecho induce a pensar en la existencia de ciertas regularidades en determinados fenómenos, inherentes a los mismos y por tanto independientes del experimentador. El hecho de que para los sucesos a los que es aplicable la definición clasica de probabilidad, la estabilidad de sus frecuencias se sitúe alrededor y muy cerca de dicha probabilidad, nos conduce, de manera lógica, a pensar que algo similar ocurre para sucesos de caracter general. No es pues extraño que decidamos denominar probabilidad del suceso A a la constante numérica alrededor de la cual oscilan las distintas frecuencias observadas para la ocurrencia de A.

Definición.- Diremos que el suceso A tiene una probabilidad si posee las siguientes peculiaridades:

a) Es posible, al menos teoricamente, llevar a cabo bajo las mismas condiciones G un número ilimitado de experimentos independientes, en cada uno de los cuales A puede o no ocurrir.

b) Cuando el número de experimentos es suficientemente grande se observa que la frecuencia de aparición de A se desvía muy poco de una constante (generalmente desconocida).

Esta constante desconocida llamada probabilidad estadística de A, para un gran número de experiencias puede aproximarse mediante la frecuencia observada.

De esta definición y teniendo en cuenta las propiedades de las frecuencias se derivan las siguientes propiedades para la probabilidad.

Propiedades de la probabilidad

1) La probabilidad de un suceso que es cierto, es la unidad.
2) La probabilidad del suceso imposible es cero.
3) Si un suceso aleatorio C es la suma de un número finito de sucesos incompatibles dos a dos, $A_1, A_2, \ldots, A_n$ que tienen probabilidad, entonces su probabilidad existe y es igual a la suma de las probabilidades de los componentes, es decir

$$P(C) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

Es interesante hacer algunas observaciones referentes a esta definición de probabilidad.

Observaciones

1) La definición de probabilidad que acabamos de dar es descriptiva, no se trata de una definición matemáticamente formal tan solo hemos postulado la existencia de la probabilidad bajo ciertas condiciones y hemos indicado un método para aproximar su valor.

2) Señalar además que la definición no pone de manifiesto las peculiaridades de aquellos fenomenos para los que la frecuencia es estable, si quisieramos averiguarlas deberiamos llevar a cabo investigaciones posteriores en este sentido. En cualquier caso la definición si pone de manifiesto el caracter objetivo de la probabilidad y su independencia del investigador que lleva a cabo la experiencia.

3) La importancia de la definición frecuencialista de la probabilidad (tambien se la conoce con este nombre) estriba en el hecho de servir de punto de partida para la construcción de una teoria formal de la probabilidad que considera a esta como una parte de la Matematicas. Tanto es así, que en ocasiones la Teoria de la Probabilidad se define "como la parte de las Matemáticas que trata de los modelos matemáticos de los fenómenos aleatorios que poseen la propiedad de la estabilidad de las frecuencias"

CONSTRUCCION AXIOMATICA DE LA TEORIA DE LAS PROBABILIDADES

En las dos definiciones estudiadas hasta ahora ha quedado puesto de manifiesto la limitación de las mismas. Esta limitación es doble, por un lado ambas adolecen de una ~~[illegible]~~ falta de rigor matemático o de formalización matemática si se prefiere, por otra parte ninguna de las dos es exhaustiva por cuanto dejan fuera de su campo de aplicación muchos fenómenos. Esto, unido a las exigencias cada vez mayores de las ciencias experimentales, como primeras usuarias del concepto, planteó a principios de siglo la necesidad de dar una solución general al problema. La solución pasaba, desde el punto de vista matemático, por la elaboración de una teoría formal mediante una axiomatica adecuada, tal como se hace en otras partes de la Matemática. Los axiomas iniciales no suponian, o debian suponer, un inicio en el desarrollo de la teoria, es decir, un estado inicial sin información previa, sino que por el contrario debian derivarse de los conocimientos que los resultados previos habian proporcionado (vease definiciones clasica y frecuencialista).

Así las cosas S.N. Bernstein llevó a cabo un primer intento, en este sentido, hacia 1917. Construyó una axiomatica sobre la base de comparar sucesos aleatorios en función de su mayor o menor probabilidad.

La solución definitiva al problema fue propuesta por A.N. Kolmogorov que definió una axiomática, hacia 1930, que relacionaba la teoria de la probabilidad con la teoria de funciones y la teoría de conjuntos. Partia Kolmogorov de una situación abstracta, para darle la mayor generalidad posible, considerando inicialmente un espacio Ω de <u>sucesos elementales</u> y una familia $\mathcal{S}$ de partes de Ω a cuyos elementos denominaba <u>sucesos aleatorios</u>. La familia de sucesos $\mathcal{S}$ verifica las siguientes condiciones

a) $\Omega \in \mathcal{S}$

b) Si $A \in \mathcal{S}$ entonces $\bar{A} \in \mathcal{S}$

c) Si $\{A_n\}_{n \geq 1} \subset \mathcal{S}$, entonces $\bigcup_{n \geq 1} A_n \in \mathcal{S}$.

Una familia con estas características es conocida con el nombre de <u>σ-álgebra de sucesos</u> o <u>tribu de sucesos</u>.

Los axiomas que introdujo Kolmogorov sobre Ω y los elementos de $\mathcal{S}$ fueron los siguientes:

A.I) A cada elemento de la σ-álgebra $\mathcal{S}$ le asociamos una cantidad no negativa a la que llamaremos su probabilidad. $P(A) \geq 0$, $\forall A \in \mathcal{S}$.

A.II) $P(\Omega) = 1$

A.III) Si $\{A_n\}_{n \geq 1} \subset \mathcal{S}$ y $A_i \cap A_j = \phi$, $i \neq j$, entonces

$$P\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} P(A_n).$$

A partir de esta axiomática se desarrolla toda la Teoría Moderna de la Probabilidad, pero no lo haremos así nosotros porque la probabilidad, definida mediante estos axiomas, se inscribe en el contexto más general de la Teoría de la Medida y es así como vamos a actuar en el próximo capítulo.

Antes de terminar con este tema señalar dos propiedades que el sistema de axiomas de Kolmogorov posee:

1) El sistema de axiomas es <u>consistente</u>, puesto que existen objetos reales que satisfacen todos los axiomas. Si tomamos $\Omega = \{a_1, a_2, \ldots, a_n\}$, los a_i cualesquiera y si $\mathcal{S} = \mathcal{P}(\Omega)$, haciendo $P(a_i) = p_i$, con $p_i \geq 0$ y tales que $p_1 + p_2 + \ldots + p_n = 1$ y si para $A \in \mathcal{S}$, tal que $A = \{a_{i_1}, a_{i_2}, \ldots, a_{i_n}\}$, definimos $P(A) = p_{i_1} + \ldots + p_{i_n}$, entonces los axiomas de Kolmogorov se satisfacen.

2) El sistema es <u>incompleto</u> puesto que para un mismo Ω podemos elegir las probabilidades en $\mathcal{S}$ de formas diferentes. Por ejemplo para el caso del lanzamiento de un dado, $\Omega = \{E_1, \ldots, E_6\}$ podemos definir $P(E_i) = 1/6$, $i = 1, \ldots, 6$, o bien hacerlo de la forma

$$P(E_1) = P(E_2) = P(E_3) = 1/4 \quad y \quad P(E_4) = P(E_5) = P(E_6) = 1/12,$$

por cuanto de ambas formas el sistema de axiomas es satisfecho.

La <u>incompletitud</u> del sistema de axiomas no significa en absoluto su inadecuación para explicar o desarrollar la teoría en estudio, en este caso la teoría de la Probabilidad, al contrario es debido a la esencia de la materia en estudio. En efecto, en el caso del dado, si éste es un dado perfecto el modelo probabilístico adecuado para describirlo será el primero, aquel en el que todas las caras tienen probabilidad 1/6, en caso de tratarse de un dado irregular quizás el segundo modelo se ajuste más a la realidad.

Capítulo 2

Probabilidad y medida

PROBABILIDAD Y MEDIDA

ALGEBRAS Y σ-ALGEBRAS

Definición.- Sea Ω un conjunto y sea $\mathcal{A}$ una subfamilia de $\mathcal{P}(\Omega)$, decimos que $\mathcal{A}$ tiene estructura de álgebra si se verifica:

1) $\Omega \in \mathcal{A}$
2) Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$
3) Si A y $B \in \mathcal{A}$ entonces $A \cup B \in \mathcal{A}$.

De esta definición se derivan de inmediato las siguientes propiedades

Propiedades

1) $\phi \in \mathcal{A}$
2) Si $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$, entonces $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ y $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$

Ejemplos de álgebras

1) $\mathcal{A} = \{\phi, \Omega\}$ es un álgebra (el álgebra trivial)
2) $\mathcal{A} = \{\text{subconjuntos de } \Omega\} = \mathcal{P}(\Omega)$ es también un álgebra
3) Para $A \subset \Omega$, $\mathcal{A} = \{\phi, A, A^c, \Omega\}$ es un álgebra.
4) Sea Ω un conjunto infinito, entonces $\mathcal{A} = \{A, A \subset \Omega \,/\, A \text{ o } A^c \text{ es finito}\}$ es un álgebra.

Las álgebras aseguran la estabilidad para las uniones e intersecciones de sus elementos en número finito. Nada puede asegurarse acerca de estas operaciones cuando se llevan a cabo en número infinito numerable. La solución a este problema llega por vía de la generalización mediante la siguiente estructura.

Definición.- Sea Ω un conjunto y sea $\mathcal{B}$ una subfamilia de $\mathcal{P}(\Omega)$, decimos que $\mathcal{B}$ es una σ-álgebra si verifica:

1) $\mathcal{B}$ ~~es~~ un álgebra
2) Si $\{A_n\}_{n\geq 1} \subset \mathcal{B}$, entonces $\bigcup_{n\geq 1} A_n \in \mathcal{B}$.

Al igual que para las álgebras también ahora $\bigcap_{n\geq 1} A_n \in \mathcal{B}$, propiedad que se deriva directamente de la definición. Como ejemplos de σ-álgebras podemos citar:

Ejemplos de σ-álgebras

1) $\mathcal{B} = \{\phi, \Omega\}$ es la σ-álgebra trivial
2) $\mathcal{B} = \mathcal{P}(\Omega)$ es una σ-álgebra.

Una estructura interesante y útil es la engendrada a partir de una familia de subconjuntos del espacio Ω inicial. Veámoslo.

Observación.- Las σ-álgebras se denominan también, en ocasiones, tribus.

Estructuras engendradas

Sea Ω un conjunto y sea $\mathcal{C}$ un subconjunto de $\mathcal{P}(\Omega)$ que posee o no alguna estructura concreta. Sean $\{\mathcal{A}_i\}_{i\in I}$, $\{\mathcal{B}_j\}_{j\in J}$ las familias de álgebras y σ-álgebras que contienen a $\mathcal{C}$. Hay que decir que dichas familias son no vacías por cuanto ambas contienen a $\mathcal{P}(\Omega)$. A partir de ellas obtenemos

$$\mathcal{A}(\mathcal{C}) = \bigcap_{i\in I} \mathcal{A}_i \quad , \quad \mathcal{B}(\mathcal{C}) = \bigcap_{j\in J} \mathcal{B}_j .$$

Es fácil comprobar que estas familias de conjuntos así definidas verifican:

1.a) $\mathcal{A}(\mathcal{C})$ es un álgebra	1.b) $\mathcal{B}(\mathcal{C})$ es una σ-álgebra
2.a) $\mathcal{C} \subset \mathcal{A}(\mathcal{C})$	2.b) $\mathcal{C} \subset \mathcal{B}(\mathcal{C})$
3.a) $\mathcal{A}(\mathcal{C})$ es la mínima álgebra que contiene a $\mathcal{C}$.	3.b) $\mathcal{B}(\mathcal{C})$ es la mínima σ-álgebra que contiene a $\mathcal{C}$.

$\mathcal{A}(\mathcal{C})$ y $\mathcal{B}(\mathcal{C})$ se llaman, respectivamente, álgebra y σ-álgebra engendrada por la familia $\mathcal{C}$.

La utilidad de esta construcción estriba en el hecho de permitir definir estructuras a partir de familias de conjuntos específicas. Sin ninguna duda el caso más interesante es cuando Ω no es un conjunto cualquiera sino que se trata de un espacio topológico. Aparece entonces la tribu de Borel, de la que nos vamos a ocupar a continuación.

La tribu de Borel, β.- Si Ω es un espacio topológico posee un sistema de abiertos y cerrados que le caracterizan. Sea Θ el sistema de abiertos y $\mathcal{F}$ el de cerrados. Definimos, mediante el proceso antes descrito, la σ-álgebra engendrada por el sistema de abiertos, $\mathcal{B}(\Theta)$, dicha σ-álgebra es conocida como la σ-álgebra de Borel del espacio topológico Ω y se la designa mediante β. Esta σ-álgebra tiene entre otras las siguientes propiedades dignas de mención:

a) $\mathcal{B}(\mathcal{F}) \equiv \beta$

b) Si S es una familia de abiertos que, siendo numerable, engendra el sistema de abiertos, es decir, si se trata de una base numerable del espacio topológico Ω, entonces $\mathcal{B}(S) = \beta$.

Un caso de tribu de Borel especialmente interesante para nosotros es aquel en que $\Omega = \mathbb{R}$. Haciendo uso de la propiedad b) podemos afirmar que, como $\mathbb{R}$ posee una base numerable, la tribu de Borel de la recta real viene engendrada por una familia de intervalos cualesquiera. Así por ejemplo, si tomamos $S = \{\,]x,y],\ x,y \in \mathbb{R},\ x<y\}$, entonces $\mathcal{B}(S) = \beta$. Si hubiéramos tomado los extremos racionales para definir los intervalos el resultado hubiera sido el mismo. Esta versatilidad en la construcción de β será útil posteriormente.

La tribu de Borel en $\mathbb{R}^2$, o en general en $\mathbb{R}^k$, β^2 o β^k respectivamente, se define análogamente a partir de familias de rectángulos (productos de intervalos) cuyas proyecciones son intervalos con extremos cualesquiera. Por ejemplo, para R^2, β^2 puede venir engendrada por la familia

$$S = \{]x_1, y_1] \times]x_2, y_2] \text{ , con } x_1, y_1, x_2, y_2 \in R \; ; \; x_1 < y_1 \, , \, x_2 < y_2 \} .$$

Definición.- Al par formado por un conjunto Ω y una σ-algebra de sus partes se le denomina espacio medible y se designa mediante $(\Omega, \mathcal{A})$.

MEDIDA

Sea $(\Omega, \mathcal{A})$ un espacio medible, definimos sobre $\mathcal{A}$ una función de conjunto, μ, que verifica

1) $\mu(\phi) = 0$

2) $\mu(A) \geq 0$, $\forall A \in \mathcal{A}$

3) si $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ y $A_n \cap A_m = \phi$, $n \neq m$, entonces

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n).$$

Decimos entonces que μ es una medida sobre $\mathcal{A}$. A la tripleta $(\Omega, \mathcal{A}, \mu)$ la denominamos espacio de medida. Una consecuencia inmediata de esta definición es la monotonía de la medida, en efecto

si $A \subset B$, entonces $B = A + (B - A)$

aplicando 3) tenemos

$$\mu(B) = \mu(A) + \mu(B - A)$$

y como por 2) $\mu(B - A) \geq 0$, de aquí $\mu(B) \geq \mu(A)$. Es decir, la medida es monótona creciente.

Son muchas las propiedades de interés que posee una medida, no nos ocuparemos de ellas directamente ahora sino que las estudiaremos a medida que sean necesarias en el transcurso del desarrollo del tema. Sí tiene interés resaltar ~~ahora~~ algunos de los distintos tipos de medidas con las que podemos encontrarnos.

Medida σ-finita - Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, decimos que μ es σ-finita si $\forall A \in \mathcal{A}$,

$$\exists \{A_n\}_{n\geq 1} \subset \mathcal{A} \, / \, A \subset \bigcup_{n\geq 1} A_n \text{ y } \mu(A_n) < +\infty, \forall n.$$

Medida finita - Sea $(\Omega, \mathcal{A}, \mu)$ un espacio de medida, decimos que μ es finita si $\mu(\Omega) < +\infty$.

Ello implica, dada la monotonía de μ, que $\mu(A) < +\infty$, $\forall A \in \mathcal{A}$.

Ejemplo.- Una medida de uso muy corriente en Análisis Matemático es la que se conoce con el nombre de Medida de Lebesgue, porque fue Lebesgue quien la introdujo a principios de siglo para desarrollar su famosa teoría de la integración, teoría que por otra parte va directamente ligada a la teoría de la medida. En su versión univariante, la medida de Lebesgue se define sobre el espacio medible (R, β) como aquella que ~~a~~ cada intervalo le hace corresponder su longitud, cuando los extremos son finitos, o vale infinito sobre los intervalos con algún extremo infinito. No es lugar aquí para entrar en detalles

acerca de la definición y propiedades de esta medida.

En ~~su~~ versión k-dimensional, la medida se define sobre el espacio medible (R^k, β^k), $k \geq 2$ haciendo que sobre los hiperrectángulos de k-dimensiones, $\prod^k]x_i, y_i]$ valga el hipervolumen correspondiente (area para $k=2$, volumen para $k=3$, etc.) si los $2k$ extremos de los intervalos factores son finitos, o bien valga infinito en cualquier otro caso.

Dadas las características topológicas de R y R^k, $k \geq 2$ podemos afirmar que las distintas medidas de Lebesgue son σ-finitas.

Ejemplo.- Otro ejemplo de medida definida sobre un espacio medible $(\Omega, \mathcal{A})$, es aquella definida como sigue:

$$\forall A \in \mathcal{A}, \quad \mu(A) = \{\text{número de elementos de } A\} = \text{card}(A).$$

Se comprueba fácilmente que es una medida a la que se conoce como medida de cardinalidad o medida de conteo (counting measure). Si Ω es finito, obviamente μ será finita. En contra de lo que ocurre con la medida de Lebesgue, esta medida no es σ-finita, si Ω es infinito no numerable, porque si fuera infinito numerable,

$$\Omega = \{w_1, \ldots w_n \ldots\} \text{ , entonces } \Omega = \bigcup_{n\geq 1} \{w_n\} \text{ y como } \mu(\{w_n\}) = 1$$

entonces μ es σ-finita pues $\forall A \in \mathcal{A}$, $A \subset \Omega$.

Vamos ahora ya a ocuparnos del objeto de nuestro estudio, a saber, las medidas de probabilidad.

MEDIDA DE PROBABILIDAD

Una medida de probabilidad no es más que una medida para la que $\mu(\Omega) = 1$. Es costumbre en este caso designar la medida mediante la letra P y hablar de $(\Omega, \mathcal{A}, P)$ como de un espacio de probabilidad, y denominar a Ω espacio muestral y a $\mathcal{A}$ σ-algebra de sucesos.

Ejemplo 1.- Sea $\Omega = \{w_1, \ldots, w_n\}$ y sea $\mathcal{A} = \mathcal{P}(\Omega)$, definimos P de la siguiente manera

$$P(A) = \sum_{j=1}^{p} P(\{w_{i_j}\}) \text{ , con } A = \{w_{i_1}, \ldots, w_{i_p}\}$$

además $P(\{w_1\}) = \ldots = P(\{w_n\}) = 1/n$.

La función así definida es una medida de probabilidad, más concretamente se la conoce como probabilidad uniforme y coincide con la definición clásica debida a Laplace. Queda puesto así de manifiesto que esta nueva definición de la probabilidad es más general y contiene aquella como caso particular.

Ejemplo 2.- Sea $\Omega = [0,1]$ y $\mathcal{A} = \beta_{[0,1]}$ tribu de Borel sobre el espacio $[0,1]$. Definimos P mediante

$$P(A) = \text{longitud}(A) \text{ , donde } A \text{ es un subintervalo de } \Omega.$$

P define una probabilidad que es conocida como la probabilidad uniforme sobre el intervalo $[0,1]$.

Ejemplo 3.- Sea $\Omega=\{w_1,w_2\}$ y sea $\mathcal{A}=\mathcal{P}(\Omega)$, definimos P mediante

$$P(\{w_1\})=p \quad y \quad P(\{w_2\})=1-p \quad , (0\leq p\leq 1)$$

$(\Omega,\mathcal{A},P)$ constituye un espacio de probabilidad.

Propiedades de la medida de probabilidad

1) P es finitamente aditiva.

En efecto, sabemos que si $\{A_n\}_{n\geq 1}\subset\mathcal{A}$ con $A_n\cap A_m=\phi$, $n\neq m$ entonces

$$P\Big(\bigcup_{n\geq 1}A_n\Big)=\sum_{n\geq 1}P(A_n) \quad (\sigma\text{-aditividad de la medida de probabilidad}).$$

Si el número de sucesos es finito, a saber, $A_1 \dots A_n$, entonces podemos suponer $A_k=\phi$, $k\geq n+1$ y de aquí, aplicar la anterior condición y obtener

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=P\Big(\bigcup_{n\geq 1}A_n\Big)=\sum_{n\geq 1}P(A_n)=\sum_{i=1}^{n}P(A_i) \text{ , por ser } P(\phi)=0.$$

2) $\forall A,B\in\mathcal{A}$, $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.

$A\cup B=A\cup(B-A)$ y por la aditividad finita $P(A\cup B)=P(A\cup(B-A))=P(A)+P(B-A)$

pero $B=(B-A)\cup(A\cap B)$ y $P(B)=P(B-A)+P(A\cap B)$, es decir $P(B-A)=P(B)-P(A\cap B)$

sustituyendo

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

3) P es subaditiva, es decir, para $\{A_n\}_{n\geq 1}\subset\mathcal{A}$, cualquiera $P\Big(\bigcup_{n\geq 1}A_n\Big)\leq\sum_{n\geq 1}P(A_n)$.

A partir de la sucesión $\{A_n\}_{n\geq 1}$ podemos definir una nueva sucesión como sigue

$$B_n=A_n-\bigcup_{i=1}^{n-1}A_i$$

estos B_n verifican, $A_n\supset B_n$ y además $B_n\cap B_m=\phi$, $n\neq m$ con $\bigcup_{n\geq 1}B_n=\bigcup_{n\geq 1}A_n$, entonces

$$P\Big(\bigcup_{n\geq 1}A_n\Big)=P\Big(\bigcup_{n\geq 1}B_n\Big)=\sum_{n\geq 1}P(B_n)\leq[\text{por monotonía}]\leq\sum_{n\geq 1}P(A_n).$$

La desigualdad es también válida para un número finito de sucesos.

TEOREMA.- Sea $\{A_i\}_{i=1}^{n}\subset\mathcal{A}$, entonces

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leq i_1\leq i_2\leq n}P(A_{i_1}\cap A_{i_2})+\sum_{1\leq i_1\leq i_2\leq i_3\leq n}P(A_{i_1}\cap A_{i_2}\cap A_{i_3})-\dots$$
$$\dots+(-1)^{n+1}P(A_1\cap A_2\cap\dots\cap A_n).$$

Demostración.- Se comprueba fácilmente por inducción sobre n.

Sucesiones monótonas de conjuntos (sucesos)

Sea $\{A_n\}_{n\geq 1}$ una sucesión de conjuntos, se definen los límites superior e inferior de la siguiente forma

$$\overline{\lim_n}\, A_n=\bigcap_{n\geq 1}\bigcup_{i\geq n}A_i \quad , \text{ límite superior}$$

$$\underline{\lim_n}\, A_n=\bigcup_{n\geq 1}\bigcap_{i\geq n}A_i \quad , \text{ límite inferior}$$

cuando ambos límites coinciden, decimos que la sucesión tiene límite.

Hay una situación en la que está siempre garantizada la existencia del límite. Este es el caso de las llamadas sucesiones monótonas, que pueden ser de dos tipos:

a) Una sucesión de conjuntos se dice que es monótona decreciente o contractiva si $A_n\supset A_{n+1}, \forall n$

b) Una sucesión de conjuntos se dice que es monótona creciente o expansiva si $A_n\subset A_{n+1}, \forall n$.

Es fácil comprobar para este tipo de sucesiones que,

Si $\{A_n\}_{n\geq 1}$ es contractiva $\overline{\lim_n}\, A_n=\underline{\lim_n}\, A_n=\bigcap_{n\geq 1}A_n=\lim_n A_n$

Si $\{A_n\}_{n\geq 1}$ es expansiva $\overline{\lim_n}\, A_n=\underline{\lim_n}\, A_n=\bigcup_{n\geq 1}A_n=\lim_n A_n$

Cuando se trata de sucesiones de sucesos existe una importante propiedad de las probabilidades basadas sobre sucesiones monótonas. Es la siguiente

TEOREMA (Continuidad respecto del paso al límite).- Sea $\{A_n\}_{n\geq 1}$ una sucesión monótona de sucesos, entonces

$$P\big(\lim_n A_n\big)=\lim_n P(A_n).$$

Demostración.- a) Supongamos que $\{A_n\}_{n\geq 1}$ es expansiva, entonces $\lim_n A_n=\bigcup_{n\geq 1}A_n$. Definimos una nueva sucesión

$$A'_n=A_n-A_{n-1} \quad , n\geq 1. \quad A_0=\phi,$$

sus elementos verifican $A'_n\cap A'_m=\phi$ y $\bigcup_{n\geq 1}A'_n=\bigcup_{n\geq 1}A_n$. A partir de aquí

$$P\big(\lim_n A_n\big)=P\Big(\bigcup_{n\geq 1}A_n\Big)=P\Big(\bigcup_{n\geq 1}A'_n\Big)=\sum_{n\geq 1}P(A'_n)=\lim_{n\to\infty}\sum_{i=1}^{n}P(A'_i)=$$
$$=\lim_{n\to\infty}P\Big(\bigcup_{i=1}^{n}A'_i\Big)=\lim_{n\to\infty}P(A_n).$$

b) Si $\{A_n\}_{n\geq 1}$ es contractiva, la sucesión $\{A_n^c\}_{n\geq 1}$ será expansiva y $\lim_n A_n^c=\bigcup_{n\geq 1}A_n^c$.

Es fácil comprobar por otra parte que para un suceso A y su complementario A^c, se verifica

$P(A) + P(A^c) = 1.$

Por el apartado anterior sabemos que

$$P\left(\lim_n A_n^c\right) = \lim_n P(A_n^c) = P\left(\bigcup_{n\geq 1} A_n^c\right)$$

utilizando la propiedad dicha

$$\lim_n P(A_n^c) = \lim_n \left[1 - P(A_n)\right] = P\left(\left[\bigcap_{n\geq 1} A_n\right]^c\right) = 1 - P\left(\bigcap_{n\geq 1} A_n\right)$$

y de aquí

$$\lim_n P(A_n) = P\left(\bigcap_{n\geq 1} A_n\right) = P\left(\lim_n A_n\right).$$

Se deriva de aquí la llamada continuidad en el ϕ de la probabilidad y que damos en forma de

COROLARIO.- Si $\{A_n\}_{n\geq 1}$ es una sucesión monótona decreciente, tal que $\lim_n A_n = \phi$, entonces $\lim_n P(A_n) = 0$.

La importancia de este corolario estriba en el hecho de que este resultado es equivalente a la aditividad completa o σ-aditividad postulada para la medida (en este caso para la probabilidad). La implicación en el sentido σ-aditividad → continuidad en ϕ está ya comprobada en el teorema anterior por cuanto en él hemos hecho uso de la citada σ-aditividad. Comprobémosla en sentido inverso.

TEOREMA.- Sustituyendo en la definición de medida de probabilidad la σ-aditividad por la continuidad en ϕ, tenemos que,

$$\text{si } \{A_n\}_{n\geq 1} \subset \mathcal{A} \;/\; A_n \cap A_m = \phi, \text{ entonces } P\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} P(A_n).$$

NOTA.- Se supone la aditividad finita.

Demostración.- Definamos, $A = \bigcup_{n\geq 1} A_n$, $B_n = \bigcup_{i=1}^{n} A_i$

$B_n' = A - B_n$.

La sucesión $\{B_n'\}_{n\geq 1}$ es contractiva y $\lim B_n' = \phi$, aplicando la continuidad en ϕ de la probabilidad, tendremos

$$P\left(\lim_n B_n'\right) = \lim_n P(B_n') = 0$$

pero $\lim_n P(B_n') = \lim_n P(A - B_n)$.

Se comprueba fácilmente, que para dos sucesos cualesquiera E y F tales que $F \subset E$, se verifica que $P(E-F) = P(E) - P(F)$, aplicando esto

$$\lim_n P(B_n') = \lim_n \left[P(A) - P(B_n)\right] = P(A) - \lim_n \sum_{i=1}^{n} P(A_i) = P(A) - \sum_{n\geq 1} P(A_n) = 0$$

la serie es convergente

de aquí $P(A) = \sum_{n\geq 1} P(A_n)$.

Observaciones.- Debemos señalar que en ocasiones nos encontraremos con sucesos tales que su probabilidad será 0, esto no implica en modo alguno que se trate del ϕ, de la misma manera que hay sucesos tales que su probabilidad es 1 y no por ello coinciden con el espacio muestral.

Veamos algunos ejemplos de medidas de probabilidad.

Ejemplos

① Sea Ω el conjunto de los naturales con el cero. Sea $\mathcal{A}$ la σ-algebra de las partes de Ω. Sobre el espacio medible $(\Omega, \mathcal{A})$ definimos una función de conjunto P, mediante

$$\forall A \in \mathcal{A}, \quad P(A) = \sum_{x \in A} \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad \lambda > 0.$$

La función P verifica

a) $P(\phi) = 0$

b) $P(A) \geq 0$, $\forall A \in \mathcal{A}$

c) Para $\{A_n\}_{n\geq 1} \subset \mathcal{A}$, tal que $A_n \cap A_m = \phi$, $n \neq m$, tenemos

$$A = \bigcup_{n\geq 1} A_n \quad y \quad P(A) = \sum_{x\in A} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \sum_n \sum_{x \in A_n} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \sum_n P(A_n).$$

se trata pues de una medida. Además

$$P(\Omega) = \sum_{x \in \Omega} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \sum_{x \geq 0} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = e^{-\lambda} \cdot \sum_{x\geq 0} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

y es por tanto una medida de probabilidad.

② Sobre el mismo espacio medible del ejemplo anterior, definimos P, mediante

$\forall A \in \mathcal{A}$, $P(A) = 1$, si A tiene un número finito de puntos

$P(A) = 0$, en caso contrario.

La P así definida no es una medida de probabilidad porque ni siquiera es una medida. En efecto, no verifica la σ-aditividad, pues tomando $A_n = \{n\}$, $n \geq 1$, tendremos

$$\bigcup_{n\geq 1} A_n = N, \quad P(N) = 0 \text{ por contener un nº infinito de puntos}$$

mientras que $P(A_n) = 1$, $\forall n$ y por tanto $\sum_{n\geq 1} P(A_n) = +\infty$.

③ Existe una medida de probabilidad especialmente importante que da lugar a lo que genéricamente se conoce como probabilidad geométrica, por cuanto la obtención de la probabilidad de cualquier suceso está directamente relacionada con las características geométricas del mismo. Veamos cuál es.

Tomemos R^n como espacio muestral y sea β^n la tribu de Borel de R^n. Sobre el espacio medible (R^n, β^n) consideremos la medida de Lebesgue en R^n, μ (Recordemos que esta medida a cada elemento de β^n le hace corresponder su hipervolumen).

Sea ahora $\Omega \in \beta^n$ tal que $\mu(\Omega) < +\infty$. Restringimos nuestra atención a este conjunto y lo dotamos de la σ-algebra de sucesos $\beta^n_\Omega = \{A \,/\, A \subset \Omega \text{ y } A \in \beta^n\}$, definimos P de la siguiente manera

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}, \quad \forall A \in \beta^n_\Omega.$$

Es muy fácil comprobar que P es una medida, teniendo en cuenta que μ lo es. Es además una <u>medida de probabilidad</u> porque efectivamente está <u>normalizada</u> (vale 1 sobre Ω), ya que

$$P(\Omega) = \frac{\mu(\Omega)}{\mu(\Omega)} = 1.$$

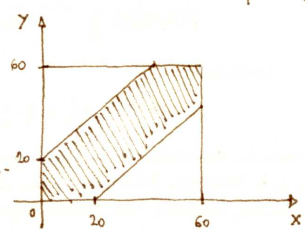
Para $n=2$ entendemos de inmediato el significado físico de este tipo de medida de probabilidad. En R^2, la medida de Lebesgue es el área del conjunto de Borel considerado y la probabilidad definida supone que para un suceso A, la probabilidad es proporcional a la superficie que dicho suceso representa. Veamos un ejemplo que nos aclare estos conceptos.

<u>Ejemplo de probabilidades geométricas</u>.- Dos personas A y B acuerdan una cita en un lugar determinado entre las doce y la una del mediodía. El primero en llegar esperará 20 minutos al otro y después se irá. ¿Cuál es la probabilidad de que A y B se encuentren si la llegada de ambos es aleatoria dentro de la hora fijada?

En la definición de probabilidad geométrica que hemos dado hay implícita una condición de uniformidad en cuanto a la distribución de la probabilidad en todo el hipervolumen que Ω representa. Es decir que la elección de puntos en Ω es completamente aleatoria. El problema que nos ocupa explicita esta condición y podemos intentar contestarlo de acuerdo con el tipo de probabilidad admitida. Designemos por x e y los tiempos de llegada de A y B respectivamente. De acuerdo con las hipótesis que el enunciado plantea, A y B se encontrarán si

$$|x-y| \leq 20$$

(x e y se expresan en minutos)



Si representamos gráficamente la desigualdad observamos que el suceso E, se encuentran, viene representado geométricamente por la parte rayada de la figura. Mientras que todos los posibles tiempos de llegada serían la superficie del cuadrado de 60 unidades de lado y que representa el suceso Ω de la definición de probabilidad geométrica. La probabilidad buscada viene pues dada por

$$P(E) = \frac{\text{Superficie }(E)}{\text{Superficie }(\Omega)} = \frac{60^2 - 2 \times 40^2/2}{60^2} = \frac{3600-1600}{3600} = \frac{2000}{3600} = \frac{5}{9}.$$

El concepto de probabilidad geométrica fue introducido por BUFFON, en el siglo 18, mediante un problema ya famoso en la historia de la Teoría de la Probabilidad, el problema de la <u>aguja de BUFFON</u>. Consiste en determinar la probabilidad de que al lanzar aleatoriamente una aguja de longitud $2l$, sobre un plano en el que se han pintado paralelas que distan entre sí $2a$ $(l<a)$, la aguja intersecte alguna de estas líneas. Nos ocuparemos de la solución de este problema en su momento.

Como en el siglo 18 los conceptos de medida y medida de Lebesgue no eran conocidos la solución se planteaba directamente en términos de superficie y volúmenes.

Señalar finalmente dos aspectos interesantes:

a) El concepto de probabilidad geométrica supone una generalización de la definición clásica de probabilidad. Allí en realidad utilizábamos para medir los conjuntos una medida adecuada para aquellas circunstancias, a saber, el cardinal de los conjuntos. Aquí, puesto que se trata de elementos de otra naturaleza recurrimos a la medida de Lebesgue correspondiente para hacerlo.

b) Cuando introduzcamos el concepto de variable aleatoria y de independencia entre variables veremos como este tipo de problemas puede resolverse dentro de un contexto más general, determinar la probabilidad de que las variables tomen valores en determinados conjuntos conociendo la ley que rige su probabilidad.

PROBABILIDAD CONDICIONAL

El concepto de probabilidad manejado hasta ahora es conocido también, en ocasiones, como <u>probabilidad incondicional</u> o <u>absoluta</u> por contraposición al que vamos a introducir a continuación. Antes, no obstante, de definirlo formalmente vamos a intentar dar alguna motivación intuitiva de su razón de ser.

Supongamos que tenemos un dado correcto en el que hemos pintado de <u>rojo</u> las caras que tienen 1, 4 y 6 puntos y de <u>azul</u> las restantes. Si lanzamos el dado una vez, sabemos que la probabilidad de que aparezca un 6 es $1/6$ (definición clásica), ahora bien, si después del lanzamiento nos es permitido mirar el dado de tal forma que podamos distinguir el color de cara, pero no su número y si observamos que dicha cara es roja, entonces podemos asegurar que la probabilidad de obtener 6 es ahora $1/3$. Esta última probabilidad viene condicionada por el hecho de conocer de antemano el color que ha aparecido en el lanzamiento y además se observa que su valor viene dado por

$$P(\text{Salga 6 condicionado a que haya salido rojo}) = \frac{P(\text{salga 6 y rojo})}{P(\text{rojo})} = \frac{1/6}{1/2} = \frac{1}{3},$$

lo que utilizando una notación más simplificada y si $A=\{\text{salir } 6\}$, $B=\{\text{salir rojo}\}$, queda

$$P(A/B) = \frac{P(A\cap B)}{P(B)}$$

donde $P(A/B)$ designa la probabilidad del suceso A condicionado a que ha ocurrido B.

Daremos ahora una definición formal del concepto.

<u>Definición</u>.- Sea $(\Omega, \mathcal{B}, P)$ un espacio de probabilidad, sea A un suceso tal que $P(A)>0$. Definimos una función de conjunto sobre los elementos de $\mathcal{B}$, mediante

$$P(B/A) = \frac{P(A\cap B)}{P(A)}, \quad \forall B \in \mathcal{B}.$$

a la que denominamos probabilidad del suceso B condicionada a la ocurrencia del suceso A, o simplemente <u>probabilidad condicional</u> de B, dado A.

<u>Observación</u>.- No es una casualidad designar a la función de conjunto así definida con el nombre de probabilidad condicional. Este nombre se justifica plenamente cuando comprobamos (cosa que se hace muy fácilmente) que la función así definida es una medida de probabilidad y posee por tanto todas las propiedades inherentes al concepto.

TEOREMA (Regla de multiplicación).- Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sean $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$, tales que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Entonces

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdots\cdots P\left(A_n / \bigcap_{i=1}^{n-1} A_i\right).$$

La demostración es sencilla y se deja al cuidado del interesado (se supone).

El interés de la fórmula que nos proporciona reside en el hecho de que en ocasiones es mucho más sencillo obtener las probabilidades del 2º miembro de la igualdad que la probabilidad del primer miembro. Veámoslo con un ejemplo.

Ejemplo.- Una urna contiene 10 bolas idénticas, de las cuales 5 son negras, 3 rojas y 2 blancas. Se extraen cuatro bolas sin reemplazamiento. ¿Probabilidad de que la primera sea negra, roja la segunda, blanca la tercera y la cuarta negra.

Sean $N_1 \to$ la 1ª bola es negra $\quad B_3 \to$ la 3ª bola es blanca

$R_2 \to$ la 2ª bola es roja $\quad N_4 \to$ la 4ª bola es negra.

Nos piden obtener

$$P(N_1 \cap R_2 \cap B_3 \cap N_4)$$

haciendo uso del teorema tenemos

$$P(N_1 \cap R_2 \cap B_3 \cap N_4) = P(N_1) \cdot P(R_2/N_1) \cdot P(B_3/R_2 \cap N_1) \cdot P(N_4/B_3 \cap R_2 \cap N_1)$$

y

$$P(N_1) = \frac{1}{2}, \quad P(R_2/N_1) = \frac{1}{3}, \quad P(B_3/R_2 \cap N_1) = \frac{1}{4}, \quad P(N_4/B_3 \cap R_2 \cap N_1) = \frac{4}{7}$$

De aquí

$$P(N_1 \cap R_2 \cap B_3 \cap N_4) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{4}{7} = \frac{1}{42}$$

EL TEOREMA DE LA PROBABILIDAD TOTAL Y LA FORMULA DE BAYES

Supongamos que en el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, la familia $\{A_i\}_{i \geq 1}$ constituye una partición del espacio muestral. Es decir

$$\Omega = \bigcup_{i \geq 1} A_i \quad \text{y} \quad A_i \cap A_j = \phi, \ i \neq j$$

La familia además es finita o infinita numerable. Supongamos además que los sucesos que componen la partición son tales que $P(A_i) > 0$, $\forall i$. Sea ahora B un suceso cualquiera, sabemos que

$$B = B \cap \Omega = B \cap \left(\bigcup_{i \geq 1} A_i\right) = \biguplus_{i \geq 1} (B \cap A_i)$$

tomando probabilidades y por la σ-aditividad

$$P(B) = P\left(\bigcup_{i \geq 1} (B \cap A_i)\right) = \sum_{i \geq 1} P(B \cap A_i)$$

pero sabemos que $P(B/A_i) = \dfrac{P(A_i \cap B)}{P(A_i)} \to P(A_i \cap B) = P(B/A_i) \cdot P(A_i)$, sustituyendo

$$\boxed{P(B) = \sum_{i \geq 1} P(B/A_i) \cdot P(A_i)}$$

Fórmula que es conocida con el nombre de <u>fórmula de la probabilidad total</u>. Comprobemos su <u>utilidad</u>.

Ejemplo.- La urna 1 contiene 2 bolas negras, la urna 2 contiene una bola negra y 2 blancas y la urna 3 contiene tres bolas negras y tres bolas blancas. Se lanza un dado, si sale 1, 2 o 3 seleccionamos la urna 1, si sale un 4 seleccionamos la urna 2 y con 5 o 6 seleccionamos la 3. Extraemos una bola de la urna seleccionada. Probabilidad de que sea blanca.

Sean $A \to$ la bola extraída es blanca

$U_i \to$ la urna elegida es la i, $i = 1, 2, 3$.

$$P(A) = \sum_{i=1}^{3} P(A \cap U_i) = \sum_{i=1}^{3} P(A/U_i) \cdot P(U_i)$$

y

$$P(A/U_1) = 0, \quad P(U_1) = \frac{1}{2}$$

$$P(A/U_2) = \frac{2}{3}, \quad P(U_2) = \frac{1}{6}$$

$$P(A/U_3) = \frac{1}{2}, \quad P(U_3) = \frac{1}{3}$$

Finalmente

$$P(A) = 0 \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{9} + \frac{1}{6} = \frac{5}{18}.$$

Una consecuencia inmediata de la fórmula que acabamos de obtener es otra conocida fórmula de gran trascendencia en el desarrollo de la teoría de la Probabilidad o, más exactamente, en una nueva concepción de la misma.

Bajo las mismas condiciones antes descritas, supongamos además que B es tal que $P(B) > 0$. Entonces

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{i \geq 1} P(B/A_i) \cdot P(A_i)}, \quad \forall i$$

Esta fórmula es conocida como la <u>FORMULA DE BAYES</u>. Su importancia estriba en que aprovecha la información conocida para modificar las probabilidades de los sucesos. En efecto, si observamos la fórmula en ella suponemos conocidas las probabilidades $P(A_i)$ y $P(B/A_i)$, $\forall i$. Podemos suponer que B es un suceso que tiene lugar bajo diversas condiciones, respecto a cuya naturaleza podemos establecer una serie de hipótesis, A_i. Las probabilidades $P(A_i)$ se denominan probabilidades <u>a priori</u>. Cuando llevamos a cabo el experimento ocurre B y el conocimiento de este hecho nos permite obtener $P(A_i/B)$, conocidas como probabilidades <u>a posteriori</u>, que corrigen las $P(A_i)$ a través de la información previa adquirida (conocimiento del resultado del experimento). Veamos un ejemplo

Ejemplo.- Volvamos nuevamente a las urnas del ejemplo anterior. Calculemos $P(U_i/A)$, $i = 1, 2, 3$

$$P(U_1/A) = \frac{P(A/U_1) \cdot P(U_1)}{\sum_{i=1}^{3} P(A/U_i) \cdot P(U_i)} = 0$$

$$P(U_2/A) = \frac{\frac{2}{3}\cdot\frac{1}{6}}{\frac{5}{18}} = \frac{2}{5} \quad , \quad P(U_3/A) = \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{5}{18}} = \frac{3}{5}$$

Obsérvese que al saber que la bola extraída es blanca las probabilidades de que la urna elegida haya sido la 2 o la 3 se han modificado y en concreto han mejorado por cuanto

Elegir la urna	A priori	A posteriori
2	1/6	2/5
3	1/3	3/5

La probabilidad de elegir la urna 1 se altera y como dicha urna no tiene bolas blancas dicha probabilidad, lógicamente, vale 0 (se convierte en algo imposible de ocurrir).

La importancia de la Fórmula de Bayes, insistimos, estriba en este hecho: Siempre que poseemos información a priori del experimento nos permite aprovecharla para modificar adecuadamente las probabilidades que manejamos.

INDEPENDENCIA DE SUCESOS

Supongamos una urna con 10 bolas, 6 de ellas rojas y 4 blancas. Llevamos a cabo dos extracciones con reemplazamiento y sean A_1 y A_2 los sucesos, "obtener bola roja en la primera y la segunda extracción" respectivamente. Sabemos que $P(A_1) = 6/10$, $P(A_2) = 6/10$ y además $P(A_2/A_1) = 6/10 = P(A_2)$. Es decir, el conocimiento de lo que ocurrió en la primera extracción no es de ninguna utilidad a la hora de saber qué pasará con la segunda. Decimos, ante una situación como la presente, que ambos sucesos son independientes.

Este ejemplo nos ayudará mejor a comprender el por qué de la siguiente definición

<u>Definición 1</u>.- Sean A y B sucesos del espacio de probabilidad $(\Omega, \mathcal{A}, P)$, tales que $P(A)>0$. Decimos que A y B son <u>independientes</u> si $P(B/A) = P(B)$.

Obsérvese que si además $P(B)>0$, entonces

$$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B/A).P(A)}{P(B)} = \frac{P(B).P(A)}{P(B)} = P(A)$$

La independencia es una relación mutua entre los sucesos. Señalemos que en el transcurso de la comprobación anterior, y como resultado intermedio, hemos obtenido la siguiente propiedad:

<u>Propiedad 1</u>.- Si A y B son independientes, entonces $P(A\cap B) = P(A).P(B)$.

Ocurre en ocasiones que muchos autores invierten los términos de este punto y dan como definición la propiedad, obteniendo como propiedad lo que hemos presentado como definición.

<u>Definición 2</u>.- Dos sucesos A y B del espacio de probabilidad $(\Omega, \mathcal{A}, P)$ decimos que son <u>independientes</u> si

$$P(A\cap B) = P(A).P(B)$$

De aquí y si $P(A)>0$,
$$P(B/A) = \frac{P(B\cap A)}{P(A)} = \frac{P(B).P(A)}{P(A)} = P(B). \quad \underline{\text{Propiedad 2}}$$

Esta segunda definición tiene la ventaja de no exigir condición alguna a los sucesos A y B. Así, mientras sucesos A y B con $P(A)$ o $P(B)$ nulas no permiten una definición correcta de la probabilidad condicional, ello no impide la definición de independencia que acabamos de dar.

<u>TEOREMA</u>.- Si A y B son independientes también lo son A y B^c, A^c y B o A^c y B^c.

<u>Demostración</u>.- Inmediata.

Conviene señalar que los conceptos de <u>independencia</u> e <u>incompatibilidad</u> no deben ser confundidos. De hecho, cuando dos sucesos incompatibles A y B son tales que $P(A)>0$ y $P(B)>0$, ambos son siempre <u>dependientes</u> por cuanto $P(A\cap B) = P(\emptyset) = 0$, mientras que $P(A).P(B) \neq 0$.

<u>Ejemplo</u>.- Elegimos al azar una carta de una baraja de 52. Sea A, la carta extraída es un as y sea B, la carta extraída es una copa. Tenemos

$$P(A) = \frac{4}{52} = \frac{1}{13} \quad , \quad P(B) = \frac{13}{52} = \frac{1}{4}$$

$$P(A\cap B) = P\,(\text{Extraer el as de copas}) = \frac{1}{52}$$

pero $P(A).P(B) = \frac{1}{13}\cdot\frac{1}{4} = \frac{1}{52} = P(A\cap B)$, luego ambos sucesos son independientes.

La definición puede generalizarse para un número finito de sucesos, pero hay entonces que introducir un matiz en la misma.

<u>Definición</u>.- Sean $\{A_1, \ldots, A_n\}$ una familia finita de sucesos del espacio de probabilidad $(\Omega, \mathcal{A}, P)$

a) Decimos que son <u>mutuamente</u> o <u>completamente independientes</u> si se verifican

$$P(A_{i_1}\cap A_{i_2}\cap \ldots \cap A_{i_k}) = P(A_{i_1}). \ldots .P(A_{i_k})$$

$$k = 2, \ldots, n \qquad \{i_1, \ldots, i_k\} \subset \{1, 2, \ldots, n\}.$$

b) Decimos que son <u>independientes dos a dos</u> si

$$P(A_i\cap A_j) = P(A_i).P(A_j) \qquad i,j = 1, \ldots, n \ , \ i\neq j.$$

Observemos que la independencia mutua exige que se cumplan

$$\binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n} = 2^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1$$

relaciones entre los sucesos, $A_1, \ldots, A_n$. Además, de la definición se sigue de inmediato que esta independencia implica la independencia dos a dos. Por el contrario, el contraejemplo que damos a continuación demuestra que no es cierta la implicación contraria.

<u>Ejemplo</u>.- Supongamos un tetraedro cuyas caras están pintadas, una de rojo, otra de azul, otra de verde y la cuarta con los tres colores. Sean

A → la cara sobre la que se apoya el tetraedro presenta color rojo
B → id. id. azul
C → id. id. verde

si el tetraedro está bien construido, tendremos

$$P(A)=P(B)=P(C)=\frac{1}{2}$$

$$\left.\begin{array}{l} P(A\cap B)=\frac{1}{4}=P(A).P(B) \\ P(A\cap C)=\frac{1}{4}=P(A).P(C) \\ P(B\cap C)=\frac{1}{4}=P(B).P(C) \end{array}\right\}\text{ hay independencia dos a dos.}$$

pero $\quad P(A\cap B\cap C)=\frac{1}{4}\neq P(A).P(B).P(C)$ y no hay independencia mutua.

Independencia entre σ-algebras.- El concepto de independencia puede hacerse extensivo a las σ-algebras en lugar de atribuirlo tan solo a los sucesos. Así, para un espacio de probabilidad $(\Omega,\mathcal{A},P)$ supongamos que existen $\mathcal{A}_1,\dots,\mathcal{A}_n$, subfamilias de $\mathcal{A}$ con estructura de σ-algebra. Decimos que las σ-algebras $\mathcal{A}_i$, $i=1,\dots,n$ son independientes si

$$P(A_1\cap A_2\cap\dots\cap A_n)=P(A_1).P(A_2)\dots P(A_n)\quad,\ \forall A_i\in\mathcal{A}_i,\ i=1,\dots,n.$$

El problema de la independencia tanto de sucesos como de σ-algebras se sitúa en el contexto de espacios medibles productos y la definición sobre ellos de medidas a partir de los espacios coordenados, y que son conocidas como medidas producto. En cualquier caso este problema sobrepasa, sobradamente, los límites de este curso y no nos ocuparemos de él.

APENDICE . EL TEOREMA DE EXTENSION

Supongamos el espacio medible $(\mathbb{R},\beta)$, la recta real provista de la tribu de Borel. Para cada intervalo I, definimos

$$P(I)=\int_I \frac{1}{\pi}\cdot\frac{1}{1+x^2}\,dx.$$

La función de conjunto así definida verifica:

a) $P(I)\geq 0$, $\forall I\in\beta$

b) $P(R)=1$

c) Dados $\{I_j\}_{j=1}^{n}$ con $I_i\cap I_j=\phi$, $i\neq j$, entonces

$$P\Big(\bigcup_{j=1}^{n} I_j\Big)=\sum_{j=1}^{n}P(I_j).$$

Podemos decir que P es una función de conjunto que se asemeja mucho a una medida de probabilidad, pero no lo es por cuanto:

a) No sabemos si es σ-aditiva

b) La hemos definido sobre los intervalos de R que son una subfamilia de β, pero no sobre β.

En realidad a) es soslayable pues las propiedades de la integral nos permiten afirmar que la σ-aditividad (para intervalos) también ha a verificarse. El verdadero problema lo plantea la posible extensión de P, desde la familia de intervalos a la tribu de Borel. Este es un problema común a la construcción de muchas medidas y tiene una adecuada respuesta en la teoría de la Medida. Es muy sencillo, en ocasiones, definir una función de conjunto sobre una familia de la σ-algebra con alguna característica especial. Por ejemplo, sobre la familia de intervalos de la forma $]a,b]$, ~~con a y b finitos o infinitos pero no simultáneamente~~, podemos definir una función tal que $\mu(]a,b])=b-a$, ~~o bien $\mu(]a,b])=+\infty$ si a=±∞~~, si consideramos además intervalos de la forma $]-\infty,b]$, $]a,+\infty[$ y añadimos R y ϕ y definimos $\mu(]-\infty,b])=+\infty$, $\mu(]a,+\infty[)=+\infty$. $\mu(R)=+\infty$, $\mu(\phi)=0$. Esta función es la que dará lugar a la medida de Lebesgue antes citada. Pero, ¿cómo?

La familia de intervalos que acabamos de considerar tiene dos características a resaltar

1) La σ-algebra que engendra es la tribu de Borel.

2) Puede engendrar una estructura intermedia, llamada algebra de conjuntos y que difiere de la σ-algebra en que es estable tan solo para uniones finitas, de manera sencilla. En efecto, ~~la familia~~ el algebra $\mathcal{Q}(\mathcal{I})$, si por $\mathcal{I}$ designamos la familia de intervalos, se caracteriza por que

$$\text{si } A\in\mathcal{Q}(\mathcal{I})\ ,\quad A=\bigcup_{i=1}^{n}I_i\quad\text{con}\quad I_i\cap I_j=\phi\ ,\ i\neq j.$$

3) Es fácil comprobar que la σ-algebra que engendraríamos con $\mathcal{Q}(\mathcal{I})$, con $\mathcal{I}$ es la misma, es decir, la tribu de Borel.

La característica 2) nos permite extender de forma inmediata y natural cualquier función de conjunto definida sobre $\mathcal{I}$ a $\mathcal{Q}(\mathcal{I})$, sin mas que escribiendo $\mu(A)=\sum_{i=1}^{n}\mu(I_i)$. Obsérvese que, así definida, μ verifica sobre $\mathcal{Q}(\mathcal{I})$

a) $\mu(A)\geq 0$, $\forall A\in\mathcal{Q}(\mathcal{I})$

b) $\mu(A\cup B)=\mu(A)+\mu(B)$ si $A\cap B=\phi$

Llegados a este punto y con las propiedades que μ verifica podemos aplicar un conocido teorema de Teoría de la Medida, el teorema de extensión de Hahn-Carathéodory, que afirma que si sobre un algebra $\mathcal{Q}$, tenemos definida una medida μ que es σ-finita, ésta puede extenderse a la σ-algebra que $\mathcal{Q}$ engendra, de manera única.

La demostración del teorema requiere profundos conocimientos de teoría de la Medida. Era conveniente, no obstante, hacer referencia al mismo porque el proceso de construcción de medidas de probabilidad sobre el espacio medible (R,β), que hemos presentado al comienzo del apendice, es muy frecuente en teoría de la Probabilidad y sólo ahora sabemos que es lícito afirmar que P es una medida de probabilidad, porque el teorema de extensión nos asegura su definición sobre β.

Capítulo 3

Aplicaciones medibles. Variables aleatorias

APLICACIONES MEDIBLES . VARIABLES ALEATORIAS

APLICACIONES MEDIBLES . PROPIEDADES

Definición.- Sean $(\Omega,\mathcal{A})$ y $(\mathcal{X},\beta)$ dos espacios medibles. Sea h una aplicación de Ω en $\mathcal{X}$, decimos que h es una aplicación medible si

$$\forall B\in\beta,\ h^{-1}(B)\in\mathcal{A}.$$

Veamos a continuación algunas propiedades de interés.

PROPOSICION I.- Sea h una aplicación medible de $(\Omega,\mathcal{A})$ en $(\mathcal{X},\beta)$ y g una aplicación medible de $(\mathcal{X},\beta)$ en $(Z,\mathcal{C})$, entonces $g\circ h$ es una aplicación medible.

Demostración.- En efecto, $\forall C\in\mathcal{C}$, tenemos $g^{-1}(C)\in\beta$ y por tanto $h^{-1}[g^{-1}(C)]\in\mathcal{A}$, pero

$$[g\circ h]^{-1}(C) = h^{-1}[g^{-1}(C)]$$ y de aquí el resultado.

PROPOSICION II.- Sea h una aplicación medible entre los espacios medibles $(\Omega,\mathcal{A})$ y $(\mathcal{X},\beta)$, entonces

a) Definimos $\mathcal{A}_h=\{B,\ B\subset\mathcal{X}/ h^{-1}(B)\in\mathcal{A}\}$. Entonces $\mathcal{A}_h$ es una σ-álgebra, a la que llamaremos σ-álgebra imagen de $\mathcal{A}$ por h y que tiene la propiedad de ser la mayor σ-álgebra (para la relación de orden de inclusión) que hace medible la función h (entre los espacios medibles $(\Omega,\mathcal{A})$ y $(\mathcal{X},\mathcal{A}_h)$).

b) Definimos $\beta_h=\{A,\ A\subset\Omega/ A=h^{-1}(B)$ con $B\in\beta\}$. β_h es una σ-álgebra, la σ-álgebra engendrada por h a partir de β. Tiene la propiedad de ser la menor σ-álgebra con la que podemos dotar a Ω para que h sea medible (entre (Ω,β_h) y $(\mathcal{X},\beta)$).

Demostración.- Es fácil comprobar que tanto $\mathcal{A}_h$ como β_h son σ-álgebras, por ejemplo, si $\{B_n\}_{n\ge1}\subset\mathcal{A}_h$, entonces $h^{-1}(B_n)\in\mathcal{A}$, $\forall n$ y de aquí $\bigcup_{n\ge1} h^{-1}(B_n)\in\mathcal{A}$

pero

$$\bigcup_{n\ge1} h^{-1}(B_n) = h^{-1}\Big(\bigcup_{n\ge1}B_n\Big),\ \text{luego}\ \bigcup_{n\ge1}B_n\in\mathcal{A}_h.$$

Análogamente se comprueba para β_h y también las otras propiedades que definen el concepto de σ-álgebra. Por otra parte, de acuerdo con las definiciones de aplicación medible y la de $\mathcal{A}_h$, tenemos que

$$h:(\Omega,\mathcal{A})\longrightarrow(\mathcal{X},\mathcal{A}_h)\quad\text{es medible}$$

pero si β es tal que $h:(\Omega,\mathcal{A})\to(\mathcal{X},\beta)$ es medible, entonces $h^{-1}(B)\in\mathcal{A}$, $\forall B\in\beta$ luego $\beta\subset\mathcal{A}_h$ y ésta es la mayor de cuantas hacen medible a h ($(\Omega,\mathcal{A})$ fijo).

Razonando de forma análoga comprobaríamos que β_h es la menor de las que hace medible h, cuando actúa entre $(\Omega,\beta_h)\to(\mathcal{X},\beta)$ y este último espacio medible es fijo.

PROPOSICION III.- 1) Sean $(\Omega,\mathcal{A})$ y $(\mathcal{X},\beta)$ dos espacios medibles. Sea $\mathcal{C}$ una familia de partes de $\mathcal{X}$ tal que $\beta=\sigma(\mathcal{C})$. Una aplicación h entre Ω y $\mathcal{X}$ es medible si y sólo si,

$\forall C\in\mathcal{C},\ h^{-1}(C)\in\mathcal{A}$. [$\sigma(\mathcal{C})$. σ-álgebra engendrada por $\mathcal{C}$]

2) En particular, si $\mathcal{X}$ es un espacio topológico y si β es su tribu de Borel, h es medible si la antiimagen de cualquier abierto es un elemento de $\mathcal{A}$.

3) Para espacios topológicos dotados de sus sistemas de abiertos y de sus tribus de Borel respectivas, toda aplicación continua es medible.

Demostración.- 1) Si h es medible obviamente $h^{-1}(C)\in\mathcal{A}$, $\forall C\in\mathcal{C}$.
Recíprocamente, si $h^{-1}(C)\in\mathcal{A}$, $\forall C\in\mathcal{C}$, ello supone que $\mathcal{A}_h$, de la proposición anterior, contiene a $\mathcal{C}$ y por tanto contendrá a la σ-álgebra que éste engendra, es decir, $\beta\subset\mathcal{A}_h$, pero esto supone que $h^{-1}(B)\in\mathcal{A}$, $\forall B\in\beta$, y h será medible.

2) Se deriva inmediatamente de 1)

3) Se deriva del hecho de que las aplicaciones continuas verifican que las antiimágenes de abiertos son abiertos.

FUNCIONES MEDIBLES. PROPIEDADES

Definición.- Una función medible es una aplicación medible, f, entre el espacio medible $(\Omega,\mathcal{A})$ y (R,β), donde R es la recta real y β la tribu de Borel correspondiente.

Enunciaremos a continuación algunas propiedades cuyas demostraciones omitiremos por escapar de las posibilidades y pretensiones de este curso. En cualquier caso convendrá tenerlas presente por el uso posterior que de ellas hagamos.

Propiedades.- 1) La suma de funciones medibles, si está definida, es también medible

2) El producto de funciones medibles es medible, si está definido.

3) Los supremos e ínfimos de funciones medibles son medibles

En estas 3 propiedades el número de funciones al que hacemos referencia es finito y, obviamente, todas ellas están definidas sobre el mismo espacio medible $(\Omega,\mathcal{A})$.

4) El $\overline{\lim}$, $\underline{\lim}$ y $\lim$ (cuando existe) de una sucesión de funciones medibles es también una función medible.

Las funciones medibles admiten una caracterización al margen de la definición dada y es:

Caracterización de las funciones medibles: Una función f, de $(\Omega,\mathcal{A})$ en (R,β) es medible si y sólo si los conjuntos de la forma $\{\omega:\ f(\omega)\le a\}\in\mathcal{A}$, $\forall a\in R$.

ALGUNAS FUNCIONES MEDIBLES DE INTERES

1) Función constante. Sea $f: \Omega \to R$ tal que $f(\omega) = k$, $\forall \omega \in \Omega$. Esta función es medible. Observemos que los conjuntos de la forma $\{\omega;\ f(\omega) \leq a\}$ son

$$\{\omega;\ f(\omega) \leq a\} = \begin{cases} \phi, & \text{si } a < k \\ \Omega, & \text{si } a \geq k \end{cases}$$

y por tanto, como ϕ y Ω son elementos de $\mathcal{B}$, f será medible.

2) Función característica de un conjunto. Sea $A \subset \Omega$, definimos $f_A: \Omega \to R$ mediante

$$f_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A. \end{cases}$$

La función así definida no siempre será medible si dotamos a Ω de una σ-algebra, su medibilidad va ligada al conjunto al cual caracteriza. Señalemos, antes de comprobar la afirmación que acabamos de hacer, que se designan como conjuntos medibles a los elementos de la σ-algebra del espacio medible. Es decir, $(\Omega, \mathcal{B})$ medible, $A \in \mathcal{B}$, entonces A es un conjunto medible.

Proposición.- La función característica de un conjunto medible es medible.

En efecto, si A es medible, entonces

$$\{\omega : f_A(\omega) \leq a\} = \begin{cases} \phi, & a < 0 \\ A^c, & 0 \leq a < 1 \\ \Omega, & a \geq 1 \end{cases}$$

y como todos ellos son medibles, entonces f_A es medible.

3) Función simple. Una función simple es una función medible con rango finito. Es decir

$$\forall \omega \in \Omega,\ f(\omega) \in \{a_1, \ldots, a_n\}$$

La medibilidad de la función supone que $f^{-1}(\{a_i\}) = A_i \in \mathcal{B}$, $i = 1, \ldots, n$ y además los A_i constituyen una partición de Ω. A saber

$$\bigcup_{i=1}^{n} A_i = \Omega, \quad A_i \cap A_j = \phi,\ i \neq j \quad i,j = 1, \ldots, n.$$

Esta propiedad sugiere una representación de f mediante las funciones características de los A_i. En efecto, fácilmente se comprueba que toda función simple puede representarse de la forma

$$f = \sum_{i=1}^{n} a_i f_{A_i}.$$

Pero por otra parte observemos que una función, s, definida mediante la expresión $\sum_{i=1}^{n} a_i f_{A_i}$ con $A_i = s^{-1}(\{a_i\})$ medible, es una función simple. Por eso en ocasiones encontramos la siguiente definición para las funciones simples

Definición.- Sea s una aplicación de $\Omega \to R$, decimos que s es una función simple si $s = \sum_{i=1}^{n} a_i f_{A_i}$, donde los A_i constituyen una partición medible de Ω y $a_i \in R$, $i = 1, \ldots, n$.

La importancia de las funciones simples quedará puesta de manifiesto más adelante. Ocupemonos ahora de unas funciones medibles que son una generalización inmediata de las simples.

4) Función discreta. Una función d, definida del espacio medible $(\Omega, \mathcal{B})$ en (R, β) decimos que es discreta si

$$d = \sum_{n \geq 1} a_n f_{A_n},$$ donde los $a_n \in R$ y los A_n constituyen una partición medible numerable de Ω.

Obviamente si la partición numerable la tomamos finita aparece la función simple como caso particular de las funciones discretas.

VARIABLES ALEATORIAS

Cuando estamos estudiando el experimento consistente en lanzar un dado representamos el suceso A_i, la cara que nos muestra el dado presenta i puntos, simplemente mediante el número i, $i = 1, \ldots, 6$. Nos limitamos a decir, ha salido un 6, o ha salido un 2, etc... Quizás no seamos consciente de ello, pero estamos pasando del mundo abstracto de los sucesos al mundo concreto de los números, mediante una aplicación que a cada cara le hace corresponder el número de puntos que contiene. Esta actuación es muy frecuente y va de hecho implícita en el tratamiento que damos a todos los procesos aleatorios. Nos permite, en definitiva, abandonar las abstracciones que los sucesos suponen y actuar en la recta real, medio en el que nos desenvolvemos con facilidad y cuyas propiedades hemos estudiado exhaustivamente.

No obstante este cambio de escenario no puede llevarse a cabo de cualquier manera. Parece lógico pensar que la aplicación que nos traslada de Ω a R goce de unas mínimas propiedades que garanticen el manejo de las imágenes de los sucesos de la misma forma en que manejabamos a estos. Sería bueno que dicha aplicación conservara las estructuras que poseíamos sobre Ω, por ejemplo, y puesto que la única estructura que allí tenemos es la σ-algebra de sucesos, en realidad estamos pidiendo que la aplicación sea una función medible. De aquí la definición

Definición.- Una variable aleatoria es una función medible definida sobre un espacio de probabilidad.

Es costumbre designar las variables aleatorias mediante las letras mayúsculas X, Y, Z, etc... Así pues, si X es una variable aleatoria (v.a.), tenemos

$$X: \Omega \to R \quad \text{y además} \quad X^{-1}(B) \in \mathcal{B},\ \text{con } B \in \beta,$$ $\mathcal{B}$ la tribu de sucesos.

Así, cada suceso se ha convertido, mediante X, en la antiimagen de un conjunto de Borel y en definitiva en un subconjunto de R.

Ejemplo.- Consideremos el experimento consistente en lanzar una moneda. El espacio muestral Ω está constituido por

$$\Omega = \{\text{cara}, \text{cruz}\}$$

podemos definir $X: \Omega \to R$ mediante $X(\{\text{cara}\}) = 1$, $X(\{\text{cruz}\}) = 0$. X es una variable aleatoria.

Ejemplo.- Lanzamos un dado dos veces consecutivas y estamos interesados en la suma de puntos que sus caras presentan. El espacio muestral está constituido por

$$\Omega = \{(i,j),\ i = 1, \ldots, 6,\ j = 1, \ldots, 6\}$$ i, j representan los puntos que muestran la primera y la segunda cara, respectivamente

Definimos sobre Ω la aplicación

$$X(\omega) = i+j \quad , \omega \in \Omega, \quad \omega=(i,j), \quad i,j=1,\dots,6$$

Esta aplicación es una aplicación medible y por tanto una variable aleatoria. Más concretamente se trata de una función simple por cuanto su rango es finito, a saber $\{2,3,\dots,12\}$, las posibles sumas que pueden obtenerse. Si denominamos S_d, el conjunto de puntos que constituyen el rango de X, tendremos que para cualquier conjunto de Borel, B

$$X^{-1}(B) = \phi \quad , \text{si} \quad B\cap S_d = \phi$$

Si $\quad B\cap S_d=\{a_1,\dots,a_k\}$, entonces

$$X^{-1}(B) = \{(i,j)=\omega\in\Omega \,/\, i+j=a_l, \ l=1,\dots,k\}$$

conjunto que evidentemente es un suceso por cuanto veremos que lo usual es dotar a los espacios muestrales finitos de la σ-álgebra de sus partes.

Por ejemplo, si $B\cap S_d=\{8\}$, tendremos $X^{-1}(B)=\{(2,6),(6,2),(3,5),(5,3),(4,4)\}$.

Probabilidad inducida. Sobre el espacio de probabilidad $(\Omega,\mathcal{A},P)$ tenemos una medida de probabilidad P que nos permite determinar la probabilidad de cualquier suceso. Mediante la variable aleatoria X hemos trasladado nuestra acción a (R,β) y requeriremos de una medida de probabilidad que nos permita determinar la probabilidad de cualquier conjunto de Borel, en tanto en cuanto representa la transformación de algún suceso o conjunto de sucesos del espacio muestral. Veamos como hacerlo.

Sea $B\in\beta$ sabemos que $X^{-1}(B)\in\mathcal{A}$, vamos a definir sobre β la siguiente aplicación

$$\forall B\in\beta, \quad P'(B)=P(X^{-1}(B)).$$

Esta aplicación está correctamente definida sobre todos los elementos de β y además:

I) $P'(\phi)=P(X^{-1}(\phi))=P(\phi)=0$.

II) $P'(B)\geq 0$, $\forall B\in\beta$.

III) $\{B_n\}_{n\geq1}\subset\beta$, $B_n\cap B_m=\phi$ $n\neq m$, $P'(\bigcup_{n\geq1}B_n)=P(X^{-1}(\bigcup_{n\geq1}B_n))=P(\bigcup_{n\geq1}X^{-1}(B_n))$

y por las propiedades de P y de la imagen inversa,

$$P'(\bigcup_{n\geq1}B_n)=\bigcup_{n\geq1}P(X^{-1}(B_n))=\bigcup_{n\geq1}P'(B_n)$$

IV) $P'(R)=P(X^{-1}(R))=P(\Omega)=1$

P' es pues una probabilidad sobre R, más exactamente sobre la tribu de Borel, a la que denominaremos probabilidad inducida y que actúa trasladando P sobre β mediante X.

Ejemplo.- En el ejemplo anterior, si $B\cap S_d=\{8\}$, $P'(B)=P(X^{-1}(B))=P(\{(2,6),(6,2),(3,5),(5,3),(4,4)\})$ y si el dado es equilibrado, la probabilidad se obtendrá mediante la fórmula de Laplace y de aquí

$$P'(B)=\frac{5}{36}.$$

Observación.- Obsérvese que la probabilidad inducida es algo característico de la variable aleatoria que la define. No es de extrañar que en ocasiones se la designe mediante P'_X, sobre todo si existe riesgo de confusión por el manejo simultáneo de varias variables aleatorias. Partiendo de un mismo espacio de probabilidad, distintas variables aleatorias dan lugar a distintas probabilidades inducidas. Veámoslo mediante un ejemplo.

Sobre el espacio muestral $\Omega=\{(i,j), \ i,j=1,\dots,6\}$ de los posibles resultados obtenidos al lanzar dos veces un dado (o una vez dos dados), definimos la v.a. X como antes lo hicimos y una nueva v.a. Y, como sigue

$$Y(\omega)=|i-j| \quad , \omega=(i,j), \ i,j=1,\dots,6.$$

Consideremos un mismo conjunto de Borel, $B=\{2\}$, entonces

$$P'_X(B)=P(X^{-1}(\{2\}))=P(\{(1,1)\})=\frac{1}{36}$$

$$P'_Y(B)=P(Y^{-1}(\{2\}))=P(\{(1,3),(3,1),(2,4),(4,2),(3,5),(5,3),(4,6),(6,4)\})=\frac{8}{36}.$$

Función de distribución de una variable aleatoria. Mediante P', probabilidad inducida por X, podemos definir una función de R en R con unas propiedades muy interesantes.

para $x\in R$, definimos $F_X(x)=P'(]-\infty,x])=P(X^{-1}\{]-\infty,x]\})=P(X\leq x)$.

Esta función tiene las siguientes propiedades

F.1) Si $x_1\leq x_2$, entonces $]-\infty,x_1]\subset]-\infty,x_2]$ y por la monotonía de P'

$$P'(]-\infty,x_1])\leq P'(]-\infty,x_2]) \rightarrow F_X(x_1)\leq F_X(x_2)$$

La función es monótona creciente.

F.2) Sea x un punto cualquiera de R. Consideremos la siguiente familia de intervalos

$$\{]-\infty,x_n], \ n\geq1, \ x_n\geq x_{n+1}, \forall n\}$$

que constituye una familia monótona decreciente de intervalos. Supongamos además que la sucesión de los x_n converge a x, entonces $\lim_n\{]-\infty,x_n]\}=\bigcap_{n\geq1}\{]-\infty,x_n]\}=]-\infty,x]$.

Por las propiedades de continuidad de P', tenemos

$$P'(]-\infty,x])=P'(\bigcap_{n\geq1}\{]-\infty,x_n]\})=\lim_n P'(]-\infty,x_n])$$

y en términos de la función que acabamos de definir

$$F_X(x)=\lim_n F_X(x_n) \qquad x_n\downarrow x,$$

y dada la monotonía de F esto supone que F es continua por la derecha.

Estas dos propiedades caracterizan a F como una función de distribución, pero posee otras dos que la hacen especialmente interesante.

F.3) Si $\{x_n\}_{n\geq 1}$ es una sucesión de números reales que tiende a $-\infty$ y además $x_n \geq x_{n+1}$ entonces los $I_n =]-\infty, x_n]$ constituyen una sucesión monótona decreciente tal que

$$\bigcap_{n\geq 1} I_n = \phi$$

aplicando la continuidad en ϕ de P',

$$\lim_n P'(I_n) = 0 \quad , \text{pero} \quad P'(I_n) = F_X(x_n).$$

En definitiva, dada la monotonía de F, podemos afirmar

$$\lim_{x\to -\infty} F_X(x) = 0.$$

F.4) Razonando análogamente para una sucesión monótona de x_n que tiende a $+\infty$, podemos afirmar que

$$\lim_{x\to +\infty} F_X(x) = 1.$$

Se trata pues de una función de distribución acotada por la unidad a la que se conoce como función de distribución de probabilidad de la variable aleatoria X.

Siendo P' una característica de X y estando F_X derivada a partir de P', obviamente, será también una característica de la variable aleatoria que la origina. El interés de P' y F_X reside en que ambos conceptos permiten conocer la variable aleatoria que los ha originado. En este contexto "conocer" significa poder determinar la probabilidad de que la variable aleatoria tome valores en cualquier conjunto de Borel. Conocer esto de una variable aleatoria es saberlo todo de la misma.

Continuemos estudiando las propiedades de F_X. Hemos afirmado que la función era continua por la derecha, veamos ahora que ocurre cuando tratamos de comprobar su continuidad por la izquierda. Dada la monotonía de F_X bastará que trabajemos con sucesiones monótonas.

Sea $\{x_n\}_{n\geq 1}$ una sucesión monótona creciente que converge a x, ello supone que los $I_n =]-\infty, x_n]$ constituyen una sucesión monótona creciente de intervalos tales que

$$\lim_n I_n = \bigcup_n I_n =]-\infty, x[,$$

definamos una nueva familia de intervalos, $I'_n = I - I_n =]x_n, x]$, estos constituyen una familia monótona decreciente cuyo límite es

$$\lim_n I'_n = \bigcap_n I_n = \{x\},$$

tomando probabilidades,

$$P'(\lim_n I'_n) = P'(\{x\}) = \lim_n P'(I'_n) = \lim_n P'(I - I_n) = \lim_n \{P'(I) - P'(I_n)\} =$$

$$= \lim_n [F_X(x) - F_X(x_n)] = F_X(x) - \lim_n F_X(x_n).$$

Podemos pues afirmar que

$$F_X(x) - F_X(x-0) = P'(\{x\}) \ , \ \forall x \in R.$$

La continuidad por la izquierda no está garantizada, a menos que $P'(\{x\}) = 0$. Directamente relacionada con la discontinuidad que esto supone podemos enunciar el siguiente teorema

TEOREMA.- El número de discontinuidades (saltos) que presenta una función de distribución de probabilidad es a lo sumo numerable.

Demostración.- Sabemos que $0 \leq F_X(x) \leq 1, \ \forall x \in R$. Acabamos de comprobar que

$$F_X(x+0) - F_X(x-0) = P'(\{x\}),$$

siendo $P'(\{x\})$ la magnitud de la discontinuidad.

F_X no puede tener más de un salto (discontinuidad) con magnitud mayor que $\frac{1}{2}$, ni más de tres cuyas magnitudes estén comprendidas entre, $1/4 < P'(\{x\}) \leq \frac{1}{2}$. En general el número de saltos cuyas magnitudes estén comprendidas entre $1/2^n < P'(\{x\}) \leq 1/2^{n-1}$, no puede superar a $2^n - 1$. De acuerdo con ello su número total (unión numerable de cantidades finitas) será a lo sumo numerable.

El conjunto de puntos de discontinuidad de la función de distribución de una variable aleatoria juega un importante papel en la caracterización de ésta y sobre ello volveremos más tarde.

Señalemos finalmente la utilidad de la función de distribución de probabilidad a la hora de determinar la probabilidad de un intervalo de la forma $]x_1, x_2]$. En efecto, dada su definición

$$F_X(x) = P'(]-\infty, x])$$

y teniendo en cuenta que

$$]x_1, x_2] =]-\infty, x_2] -]-\infty, x_1]$$

y que

$$P'(]x_1, x_2]) = P'(]-\infty, x_2] -]-\infty, x_1]) = P'(]-\infty, x_2]) - P'(]-\infty, x_1]) = F_X(x_2) - F_X(x_1).$$

Obtención de una medida de probabilidad a partir de una función de distribución de probabilidad.

En el párrafo precedente hemos visto que partiendo del espacio de probabilidad (R, β, P') se define una función cuyas propiedades se han puesto de manifiesto y cuya principal característica es la de ser una función de distribución tal y como se entiende en el Análisis Matemático clásico.

Observemos la validez del desarrollo independientemente del origen de P'. Es decir, con cualquier medida de probabilidad definida inicialmente sobre los borelianos de R hubiéramos llegado a una función de características similares.

Un importante aspecto de este proceso, quizás el más relevante, es que puede invertirse. A saber, partiendo de una función de distribución debidamente normalizada podemos definir sobre β una medida de probabilidad. Veamos como.

Sea F una función de distribución, es decir

P.1) Si $x_1 \leq x_2 \rightarrow F(x_1) \leq F(x_2)$

P.2) F es continua por la derecha

además verifica

P.3) $\lim_{x\to -\infty} F(x) = 0$

P.4) $\lim_{x\to +\infty} F(x) = 1$,

lo que se conoce como función de distribución de probabilidad.

Consideremos ahora la siguiente familia de intervalos

$$\mathcal{I} = \{\phi, R,]-\infty, x],]x,y],]x,+\infty[\}$$

y definamos sobre ella la siguiente función de conjunto

$$P(\phi)=0 \quad , \quad P(]x,y]) = F(y)-F(x) \quad , \quad P(]-\infty,x]) = F(x) - \lim_{x\to-\infty} F(x) = F(x)$$

$$P(]x,+\infty[) = \lim_{x\to+\infty} F(x) - F(x) = 1-F(x) \quad , \quad P(R) = \lim_{x\to+\infty} F(x) - \lim_{y\to-\infty} F(y) = 1.$$

La familia $\mathcal{I}$ posee además, dos interesantes propiedades:

$\mathcal{I}.1)$ la σ-algebra que engendra, $\mathcal{B}(\mathcal{I})$ es la tribu de Borel

$\mathcal{I}.2)$ El algebra que engendra, $\mathcal{A}(\mathcal{I})$, verifica

$$\forall T \in \mathcal{A}(\mathcal{I}), \quad T = \bigcup_{i=1}^{n} I_i \quad , \quad I_i \in \mathcal{I}, \quad I_i \cap I_j = \phi \;\; (i\neq j).$$

Esta última propiedad nos permite extender P de $\mathcal{I}$ a $\mathcal{A}(\mathcal{I})$ de forma inmediata,

$$P(T) = \sum_{i=1}^{n} P(I_i)$$

con lo que P sobre $\mathcal{A}(\mathcal{I})$ verifica

a) Es no negativa, $P(T) \geq 0$, $\forall T \in \mathcal{A}(\mathcal{I})$

b) Es aditiva, $\forall T_1, T_2 \in \mathcal{A}(\mathcal{I})$, con $T_1 \cap T_2 = \phi$, $P(T_1 \cup T_2) = P(T_1) + P(T_2)$.

En estas condiciones podemos aplicar el teorema de extensión de Hahn-Caratheodory y definir sobre β una medida, de probabilidad, pues $P(R)=1$, que resulta ser la extensión, única, de la inicialmente definida sobre $\mathcal{I}$ y $\mathcal{A}(\mathcal{I})$.

Queda así puesta en evidencia la correspondencia biunívoca existente entre la medida de probabilidad que caracteriza a una variable aleatoria y la función de distribución de probabilidad de la misma. La equivalencia que ello supone hace que sea indistinto para nosotros trabajar con una u otra, y escogeremos hacerlo con la más sencilla de manejar, la función de distribución.

Ejemplo.- Volvamos nuevamente al espacio muestral $\Omega = \{(i \text{ puntos en la cara del dado}), i=1,\ldots,6\}$. Mediante la aplicación

$$X\{(i \text{ puntos en la cara del dado})\} = i$$

definimos una variable aleatoria y a partir de ella, P', en la forma ya conocida, siendo P la probabilidad clásica sobre Ω. Estudiemos con un poco de detalle la función de distribución de X, F_X.

Sea $S_d = \{1,2,3,4,5,6\}$, el rango de X. Tendremos

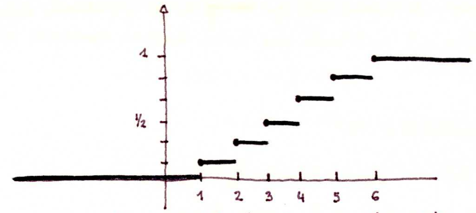
$$F_X(x) = 0, \quad \forall x < 1$$

para $x = i$, $i \in S_d$, tendremos

$$F_X(i) = P(X \leq i) = P\{(j \text{ puntos en la cara del dado}), j \leq i\} = \sum_{j=1}^{i} \frac{1}{6} = \frac{i}{6}$$

para $x > 6$, tendremos

$$F_X(x) = P(X \leq x) = P(\Omega) = 1$$



La función de distribución obtenida en el lanzamiento de un dado, mediante la variable aleatoria que a cada cara le hace corresponder el número de puntos que contiene.

Si el punto x está comprendido entre $[i, i+1[$, $i \in S_d$, entonces

$$F_X(x) = P(X \leq x) = P\{(j \text{ puntos en la cara del dado}), j \leq x\} = \sum_{j\leq x} \frac{1}{6} = \sum_{j\leq i} \frac{1}{6} = \frac{i}{6}.$$

De acuerdo con esto la gráfica de la función de distribución es la que presentamos. Es una gráfica en escalera, cuyos peldaños (tramos constantes) aparecen entre dos valores consecutivos de S_d, tal como hemos visto en la obtención de F_X.

Digamos finalmente algo sobre la continuidad de F_X. Recordemos que los puntos de discontinuidad de F_X son aquellos reales para los que,

$$x \in R, \quad P'(\{x\}) = P(X = x) \neq 0.$$

En el caso que estudiamos ello sólo ocurre para $x \in S_d$, es decir, F_X presenta discontinuidades en $x \in \{1,2,3,4,5,6\}$, todas ellas de igual magnitud, a saber

$$F_X(x+0) - F_X(x-0) = P(X = x) = \frac{1}{6}, \quad x = 1,2,3,4,5,6.$$

La función, como se pone de manifiesto en la gráfica, es continua en el complementario de S_d.

VARIABLES ALEATORIAS DISCRETAS

Definición.- Sea $(\Omega, \mathcal{B}, P)$ un espacio de probabilidad, definimos sobre él una variable aleatoria X. Decimos que X es una v.a. discreta si el conjunto de valores de su rango, S_d, es a lo sumo numerable.

La consecuencia inmediata de esta definición es que la variable X será de la forma

$$X = \sum_{i\geq 1} x_i \cdot \chi_{A_i} \quad , \text{ con } x_i \in S_d, \forall i \quad y \quad A_i = X^{-1}(\{x_i\}) \in \mathcal{B}, \forall i, \quad \bigcup A_i = \Omega, \quad A_i \cap A_j = \phi \;\; (i\neq j)$$

es decir, se trata de lo que definimos en su momento como una función discreta.

Por otra parte la función de distribución de X, F_X, será una función de salto, o en escalera, por cuanto

$$F_X(x) = 0, \quad \forall x < \min S_d = x_1, \quad [S_d = \{x_1, x_2, \ldots, x_n, \ldots\} \text{ lo suponemos ordenado}].$$

$$F_X(x) = P(X \leq x) = P(X \in \{x_1, \ldots, x_i\}) = \sum_{j=1}^{i} P(X = x_j) = \sum_{j=1}^{i} P(A_j), \quad x \in [x_i, x_{i+1}[.$$

Los puntos de discontinuidad de F_X son los de S_d y la magnitud de la discontinuidad en x_i es $P(A_i)$.

La función de cuantía. Podemos definir una nueva función, mediante X, que será muy útil para estudiar las peculiaridades de X.

Sea $f_X: R \to R$ definida mediante

$$f_X(x) = P(X = x) = P'(\{x\})$$

a esta función se la denomina función de cuantía de la variable aleatoria X. En ocasiones se la denomina también función de probabilidad, puesto que proporciona, en cada punto, la probabilidad de que la v.a. X tome dicho valor. ¿Cuáles son las propiedades de f_X? ¿Cuál su relación, si la hay, con F_X?

Propiedades de la función de cuantía:

1) $f_X(x) \geq 0, \ \forall x \in R$

En efecto, si $x \in S_d$, $f_X(x) = f_X(x_i) = P(A_i)$, $x = x_i$

si $x \in S_d^c$, $f_X(x) = P(X = x) = 0$.

2) $\sum_{x_i \in S_d} f_X(x_i) = 1$

$$\sum_{x_i \in S_d} f_X(x_i) = \sum_{x_i \in S_d} P(X = x_i) = \sum_{x_i \in S_d} P(A_i) = P\left(\bigcup_i A_i\right) = P(\Omega) = 1$$

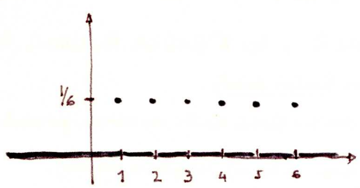
3) $f_X(x) = P(X = x) = P'(\{x\}) = F_X(x+0) - F_X(x-0) = F_X(x) - F_X(x-0)$.

La tercera propiedad establece la relación existente entre ambas funciones y más exactamente permite obtener la función de cuantía a partir de la función de distribución de la v.a. X. Pero el conocimiento de f_X permite también proceder en sentido contrario, tendremos para ello que

$$F_X(x) = P'(]-\infty, x]) = P(X \leq x) = P(X \in S_d \cap]-\infty, x]) = P(X \in \{x_1, \ldots, x_n\}, x_n \leq x) =$$
$$= \sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} f_X(x_i).$$

lo que permite construir F_X a partir de f_X. Esto supone que las características probabilísticas de X, que conocíamos a través de su distribución de probabilidad, es decir, F_X, pueden también conocerse mediante f_X y en ello estriba precisamente la importancia y utilidad de la función de cuantía.

Ejemplo.- Considerando nuevamente el ejemplo del dado.



La función de cuantía valdrá cero en toda la recta real excepto en los puntos de $S_d = \{1, 2, 3, 4, 5, 6\}$. En dichos puntos la función vale

$$f_X(i) = P(X = i) = \frac{1}{6}, \quad i = 1, 2, 3, 4, 5, 6.$$

La representación gráfica de $f_X(x)$ es la que se presenta.

Ejemplo.- Supongamos una familia con 4 hijos y supongamos también que las probabilidades de nacimiento de varón o hembra son iguales e iguales a $1/2$. El espacio muestral de las posibles combinaciones de sexos entre los hijos de la familia está constituido por 16 puntos, a saber:

$$\Omega = \{VVVV, VVVH, VVHV, VHVV, HVVV, VVHH, \ldots\ldots, HHHH\}$$

V - varón
H - hembra

Dotamos a Ω de la σ-álgebra de sus partes y definimos sobre ella la medida de probabilidad clásica, de manera que $P(A) = \frac{n}{16}$, n = número de puntos de A, $A \in \mathcal{A}$.

Sea ahora X una variable aleatoria definida mediante, $X: \Omega \to R$, $X(\omega)$ = nº de varones en ω. La variable así definida será discreta, y $S_d = \{0, 1, 2, 3, 4\}$. Veamos cómo son sus funciones de cuantía, f_X, y de distribución, F_X.

Función de cuantía: ya sabemos que $f_X(x) = 0$, si $x \notin S_d$. Para $x = n$, $n = 0, 1, 2, 3, 4$, tendremos

$$f_X(n) = P(X = n) = P(\text{haya } n \text{ varones en la familia}) = P(A_n)$$

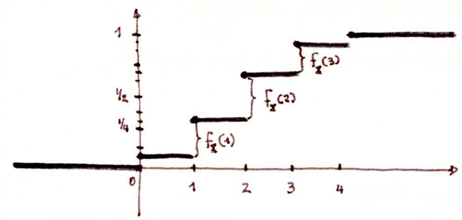
el número de casos posibles es 16, mientras que los casos favorables serán las posibles elecciones de n nacimientos (ocupados por varones) entre los 4 que han ocurrido, es decir, $\binom{4}{n}$. Así pues

$$f_X(n) = P(A_n) = \binom{4}{n} \cdot \frac{1}{16} = \binom{4}{n} \left(\frac{1}{2}\right)^n \cdot \left(\frac{1}{2}\right)^{4-n}, \quad n = 0, 1, 2, 3, 4$$

concretamente

$$f_X(0) = \frac{1}{16}, \quad f_X(1) = \frac{1}{4}, \quad f_X(2) = \frac{6}{16}, \quad f_X(3) = \frac{1}{4}, \quad f_X(4) = \frac{1}{16}.$$

Función de distribución: Utilizando la relación existente entre f_X y F_X, tendremos:



$F_X(x) = \sum_{n \leq x} f_X(n)$, así

$F_X(x) = 0$, si $x < 0$

$F_X(x) = \frac{1}{16}$, si $0 \leq x < 1$

$F_X(x) = \frac{5}{16}$, si $1 \leq x < 2$

$F_X(x) = \frac{11}{16}$, si $2 \leq x < 3$

$F_X(x) = \frac{15}{16}$, si $3 \leq x < 4$

$F_X(x) = 1$, si $x \geq 4$.

Esta variable aleatoria es un caso particular de una familia de variables aleatorias a las que se conoce generalmente como variable aleatoria binomial y son muy útiles en el estudio de fenómenos que obedecen a un determinado esquema, el modelo binomial. De una y otra nos ocuparemos a continuación cuando estudiemos algunos casos concretos de variables aleatorias discretas.

Capítulo 4

Variables aleatorias discretas

DISTRIBUCION DE BERNOULLI

Una prueba de Bernoulli es aquel experimento que tiene como resultado la realización (éxito) o no (fracaso) de determinado suceso A. El espacio muestral consta de dos puntos, $\omega_1 = A$, $\omega_2 = A^c$. Podemos definir una variable aleatoria, de la forma

$$X(\omega_1) = 1, \quad X(\omega_2) = 0,$$

si $P(A) = p$ y $P(A^c) = 1 - p = q$, entonces esta variable aleatoria, que se conoce como <u>v.a. de Bernoulli</u>, tiene por función de cuantía,

$$f_X(1) = p, \quad f_X(0) = q, \quad f_X(x) = 0, \ \forall x \in R - \{0,1\}.$$

Es muy fácil comprobar que se trata de una función de cuantía.

DISTRIBUCION BINOMIAL

Supongamos n pruebas de Bernoulli independientes, de manera que en cada una de ellas el resultado es éxito o fracaso y p y q = 1 - p son, respectivamente, las probabilidades de éxito o fracaso en cada una de las pruebas. El espacio muestral de las n pruebas, Ω, estará constituido por todas las posibles combinaciones de éxitos y fracasos que pueden ocurrir en las n pruebas, es decir

$$\Omega = \{\omega_1, \ldots, \omega_k\} \quad \text{con, por ejemplo,} \quad \omega_i = AAA^cAA^c \overset{n}{\frown} A, \quad A\text{-éxito}, \ A^c\text{-fracaso}.$$

tendremos pues que todos los puntos del espacio muestral no son igualmente probables. Así, por ejemplo, si ω_i es un punto de Ω que comporta x éxitos y n-x fracasos, tendremos

$$P(\omega_i) = p^x \cdot q^{n-x}.$$

Definamos, sobre Ω, la siguiente variable aleatoria

$$X(\omega_i) = \text{nº de éxitos que } \omega_i \text{ comporta},$$

se trata de una variable discreta y, concretamente, con un rango finito, a saber $S_d = \{0,1,2,\ldots,n\}$. [Es costumbre también escribir $X = \{0,1,\ldots,n\}$, para designar el rango de la función]. A esta variable se la conoce como <u>variable aleatoria Binomial</u>. Estudiemos su función de cuantía y veremos el porqué de tal nombre.

Ya sabemos que:

$$f_X(x) = P(X = x),$$

por tanto, si $x \in R - S_d$, entonces $f_X(x) = 0$, pero para $x \in S_d$, tendremos

$$x \in S_d, \quad f_X(x) = P(X = x) = P(\{\omega_i \text{, tales que en } \omega_i \text{ hay } x \text{ éxitos}\}) = P(B)$$

sabemos que cada uno de los $\omega_i \in B$, tienen una probabilidad, $P(\omega_i) = p^x q^{n-x}$, pero, ¿cuántos hay?. Los x éxitos pueden distribuirse de $\binom{n}{x}$ formas sobre los n resultados, así pues B contiene $\binom{n}{x}$ elementos, y en definitiva

$$f_X(x) = \binom{n}{x} p^x q^{n-x}, \quad x \in \{0,1,\ldots,n\}.$$

Efectivamente f_X es una función de cuantía pues

a) $f_X(x) \geq 0, \ \forall x \in R$

b) $\displaystyle\sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} (p+q)^n = 1^n = 1.$

El nombre de <u>binomial</u> se justifica si observamos que los valores de la función de cuantía son los elementos del desarrollo del binomio $(p+q)^n$.

Es costumbre designar abreviadamente la distribución binomial de la manera $B(n,p)$, donde n y p son llamados <u>parámetros</u> de la distribución. Observemos que conocidos n y p la distribución binomial queda perfectamente determinada por cuanto conocemos el nº de pruebas de Bernoulli que componen la experiencia y la probabilidad de éxito en cada una de ellas.

La v.a. Binomial es <u>de aplicación en todas aquellas experiencias</u> que consisten en:

a) La experiencia consta de un número finito de pruebas, independientes unas de otras.

b) En cada prueba ocurre o no un determinado suceso (éxito o fracaso)

c) La probabilidad de ocurrencia del suceso es la misma en cada prueba.

En estas condiciones la $B(n,p)$ describe adecuadamente la probabilidad de obtener x éxitos en la citada experiencia (o de obtener n-x fracasos).

EJEMPLOS DE UTILIZACION DE LA BINOMIAL

I) Los dados de Weldon. Consideremos un experimento consiste en lanzar un dado 12 veces y sea "éxito" la aparición de un cinco o un seis. Con un dado perfecto la probabilidad de éxito viene dada por $p = 1/3$ y el número de éxitos debería seguir una distribución Binomial $B(12, 1/3)$. Llevamos a cabo el experimento un total de 26306 veces y recogemos en una tabla, reproducida parcialmente a continuación, las frecuencias relativas observadas y las probabilidades teóricas. De su comparación (en su día veremos que metodología emplear para llevarla a cabo) y teniendo en cuenta el elevado número de repeticiones, puede deducirse que los datos teóricos no se ajustan adecuadamente a los experimentales. Puede suponerse, a la vista de este resultado, que el dado utilizado presenta un pequeño sesgo a favor de la aparición del cinco o el seis, de tal manera que $p = 0.3377$ sería una más razonable probabilidad de éxito.

DATOS DEL DADO DE WELDON

k	$B(12, 1/3)$	Frecuencia observada	$B(12, 0.3377)$
⋮	⋮	⋮	⋮
3	0.211952	0.208127	0.207736
4	0.238446	0.232418	0.238324
5	0.190757	0.197445	0.194429
6	0.111275	0.115589	0.115660
⋮	⋮	⋮	⋮

II) Supongamos que el número de "éxitos" en determinada experiencia sigue una ley binomial $B(n, 0.01)$, ¿cuántas repeticiones debemos llevar a cabo para asegurar que la probabilidad de al menos un éxito sea $\frac{1}{2}$ o más?

$$P(X \geq 1) = 1 - P(X=0) = 1 - (0.99)^n \geq \frac{1}{2}$$

y de aquí $\quad -n \cdot \log(0.99) \geq \log 2 \quad$ lo que da $\quad n \geq 70$.

III) Un problema de suministro eléctrico. Supongamos que $n = 10$ trabajadores usan una fuente de energía eléctrica intermitentemente y estamos interesados en estimar la carga total esperada. Como una aproximación al problema imaginemos que en un momento determinado cada trabajador tiene la misma probabilidad p de necesitar de utilizar una unidad de potencia. Si trabajan independientemente, la probabilidad de que exactamente k de ellos necesiten energía al mismo tiempo, vendrá dada por $B(n, p)$. Si por término medio un trabajador utiliza la fuente de energía 12 minutos por hora, $p = 1/5$. La probabilidad de que siete o más trabajadores requieran corriente en el mismo momento vendrá dada por:

$$P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x} = 0.0008643584 .$$

En otras palabras si la fuente de energía se ajusta a 6 unidades de potencia, la probabilidad de sobrecarga de uso será 0.00086.... Lo que en términos de tiempo, tomando el minuto como unidad básica, que 1 de cada 1157 minutos podría ocurrir dicha sobrecarga, es decir 1 minuto cada 20 horas de uso. La probabilidad para 8 o más trabajadores utilizando la corriente al mismo tiempo es 0.0000779264, es decir, unas 11 veces menor.

IV) Contraste de sueros o vacunas. Supongamos que la incidencia normal de determinada enfermedad en el ganado vacuno es 0.25. Para contrastar un suero descubierto recientemente se inyectan con él n animales sanos. ¿Cómo evaluar los resultados del experimento?. Para un suero sin efectividad ninguna la probabilidad de que exactamente k de los n animales inyectados permanezcan sin infección, viene dada por $B(n, 0.75)$. Para $k=n=10$ la probabilidad buscada es 0.056 y para $k=n=12$ es 0.032. Así, si entre 10 o 12 animales inyectados ninguno de ellos enferma puede interpretarse este hecho como una prueba de la efectividad del suero utilizado, aunque no se trate de una prueba concluyente. No obstante para $n = 17$, la $P(X \leq 1) = 0.0501$, siendo X el número de infectados, ello puede interpretarse en el sentido de que al inyectar 17 animales si sólo uno enferma la evidencia de la efectividad del suero es mayor que cuando inyectamos a 10 y ninguno enfermó. En la misma línea de razonamiento, para $n = 23$, $P(X \leq 2) = 0.0492$ y por tanto tendremos una evidencia más fuerte de la bondad del suero cuando al inyectar a 23 individuos sólo 2 de ellos enfermen.

LA DISTRIBUCIÓN DE POISSON (Como una aproximación de la binomial)

En la fabricación de botellas de vidrio, se consideran defectuosas aquellas en las que aparecen partículas sólidas, "piedras", que estaban previamente diluidas en la masa de vidrio fundido. Supongamos que tenemos 100 kgs. de masa, que cada botella requiere 1 kgs para su fabricación y que hay x "piedras" en los 100 kgs. de masa. ¿Cual será el porcentaje de botellas defectuosas?. La respuesta x % ignora la posibilidad de que en una botella aparezca más de una "piedra". Para contestar a la pregunta podemos recurrir a un <u>modelo probabilístico a base de urnas</u>.

Bajo el supuesto de que cada "piedra" tiene la misma probabilidad de aparecer en cada botella y esto independientemente de lo que ocurra con las otras "piedras", podemos reducir el problema a lo siguiente: se colocan aleatoriamente n bolas en N urnas, ¿cual es la probabilidad de que una urna elegida al azar contenga exactamente k bolas?. Como ya sabemos

casos posibles N^n

casos favorables $(N-1)^{n-k}$

esto para k bolas fijas, pero tengamos en cuenta que estas k bolas pueden elegirse de $\binom{n}{k}$ formas distintas, en consecuencia la probabilidad buscada es

$$W_k = \binom{n}{k}\frac{(N-1)^{n-k}}{N^n} = \binom{n}{k}\frac{1}{N^k}\cdot\left(1-\frac{1}{N}\right)^{n-k}$$

Si la producción de las N botellas requiere M quintals de masa de vidrio, tendremos, en este caso, $N = 100M$ y $n = xM$. Como estamos interesados en el porcentaje de botellas defectuosas en un largo periodo de fabricación, supondremos que M es muy grande. Sea $\lambda = \frac{x}{100}$, entonces

$$\left.\begin{matrix} N = 100M \\ n = xM \end{matrix}\right\} N = \frac{100}{x}\cdot n = \frac{n}{\lambda}$$

, sustituyendo en W_k, tendremos

$$W_k = \frac{1}{k!}\cdot\prod_{j=0}^{k-1}(n-j)\cdot\left(\frac{\lambda}{n}\right)^k\cdot\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}\cdot\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right).$$

que M sea grande, supone que n es grande y sabemos que

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

y que

$$\lim_{n\to\infty}\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right) = 1$$

entonces

$$\lim_{n\to\infty} W_k = \frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,\ldots$$

A la distribución

$$P_k = P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k=0,1,\ldots$$

se la conoce con el nombre de distribución de Poisson. Se trata de una probabilidad por cuanto

$$\sum_{k\geq 0} P_k = \sum_{k\geq 0}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\cdot\sum_{k\geq 0}\frac{\lambda^k}{k!} = e^{-\lambda}\cdot e^{\lambda} = 1.$$

Volvamos de nuevo al caso práctico inicial. La relación entre frecuencia relativa y probabilidad hace que la proporción de botellas defectuosas frente al total fabricado sea aproximadamente igual a la probabilidad de que haya al menos una botella defectuosa y esta probabilidad puede aproximarse mediante $1 - W_0 \simeq 1 - e^{-\lambda}$ y como $\lambda = \frac{x}{100}$, el porcentaje de piezas defectuosas será $100\left(1-\exp\left(-\frac{x}{100}\right)\right)$. Si x es muy pequeño, esta cantidad es prácticamente igual a x, pero para valores grandes x no ocurre así. Por ejemplo, para $x = 100$, el porcentaje de botellas defectuosas no es 100 sino $63.21\% = 100(1-e^{-1})$, cantidad que es de esperar que no ocurra evitando que x sea tan grande. Si por ejemplo $x = 30$, el porcentaje será 25.92%

<u>OBSERVACIÓN</u> puede asimilarse a una binomial razonando así. La probabilidad de que una bola sea introducida en una urna es $\frac{1}{N}$, y la contraria $\left(1-\frac{1}{N}\right)$ y por tanto la de obtener k-éxitos (éxito = bola en la urna en cuestión) será

$$\binom{n}{k}\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{n-k}$$

en lugar de 30% que obtendríamos razonando como al principio. Evidentemente si el número de "piedras" en la masa es grande es mucho más económico, si el proceso lo permite, fabricar botellas pequeñas, si no hay forma posible de limpiar la masa. Así por ejemplo, fabricando botellas que requieran sólo 0.25 kgs de masa en lugar de 1 Kg., el nuevo valor de λ es $\lambda' = x/400$ y la proporción de defectuosas, para $x=30$, baja del 25.92% al 7.23%. La teoría de la probabilidad proporciona, como hemos visto, sugerencias interesantes para problemas prácticos de producción. (**) ver APENDICE

LA DISTRIBUCIÓN DE POISSON (Como solución a un problema físico-químico)

Supongamos una cierta masa de sustancia radiactiva cuyo periodo de vida media es elevado. Supongamos además que las siguientes hipótesis son ciertas

1.- Si $t_1 < t_2 < t_3$ y $A_k(t_1,t_2)$ designa el suceso "durante el intervalo de tiempo (t_1,t_2) ocurren k desintegraciones", entonces los sucesos $A_k(t_1,t_2)$ y $A_\ell(t_2,t_3)$ son independientes para cualesquiera valores no negativos de k y ℓ.

2.- Los sucesos $A_k(t_1,t_2)$, $k=0,1,\ldots$ constituyen una partición del espacio muestral. Para k dado, $P(A_k(t_1,t_2))$ depende sólo de la diferencia t_2-t_1. En otras palabras, el proceso de desintegración radiactiva es homogéneo con respecto al tiempo. $W_k(t)$ designará la probabilidad de k desintegraciones durante un intervalo de tiempo de longitud t ($t_2-t_1=t$).

3.- Si t es suficientemente pequeño, la probabilidad de que durante un intervalo de tiempo t ocurra más de una desintegración es despreciablemente pequeña comparada con la probabilidad de que ocurra exactamente una. Es decir

$$\lim_{t\to 0} \frac{1-W_0(t)-W_1(t)}{W_1(t)} = 0, \qquad (1)$$

o lo que es lo mismo

$$\lim_{t\to 0} \frac{1-W_0(t)}{W_1(t)} = 1 \qquad (2)$$

En otras palabras: la probabilidad de que ocurra al menos una desintegración es, en el límite, igual a la probabilidad de que ocurra exactamente 1.

Claramente $W_0(0)=1$ y $W_k(0)=0$, $k\geq 1$. Además $W_0(t)$ es una función monótona decreciente. De aquí y de las condiciones 1 y 2 tenemos

$$W_0(t+s) = P(A_0(t_1,t_3)) = P(A_0(t_1,t_2)\cap A_0(t_2,t_3)) = P(A_0(t_1,t_2)).P(A_0(t_2,t_3)) =$$

$$= W_0(t).W_0(s) \qquad \text{para } t_2-t_1=t \text{ y } t_3-t_2=s.$$

se trata pues de una función aditiva, si exigimos la derivabilidad de la función en 0, tendremos

$$W_0(t) = e^{-\mu t} \qquad \text{con } \mu>0. \qquad (3)$$

En orden a determinar las funciones $W_k(t)$ observemos que

$$\lim_{\Delta t\to 0} \frac{W_k(\Delta t)}{\Delta t} = 0, \quad k\geq 2 \qquad (5)$$

En efecto

$$0 \leq W_k(t) \leq \sum_{k\geq 2} W_k(t) = 1-W_0(t)-W_1(t), \quad \forall k\geq 2$$

y de aquí

$$0 \leq \frac{W_k(\Delta t)}{\Delta t} \leq \frac{1-W_0(\Delta t)-W_1(\Delta t)}{\Delta t}\cdot\frac{W_1(\Delta t)}{W_1(\Delta t)} = \frac{1-W_0(\Delta t)-W_1(\Delta t)}{W_1(\Delta t)}\cdot\frac{W_1(\Delta t)}{\Delta t} \qquad (4)$$

de (2), tenemos

$$\lim_{t\to 0} \frac{(1-W_0(t))/t}{W_1(t)/t} = 1 \longrightarrow \lim_{t\to 0}\frac{1-W_0(t)}{t} = \lim_{t\to 0}\frac{W_1(t)}{t} = W_1'(0)$$

pero el primer miembro de esta última igualdad es $-W_0'(0)$ que existe y por tanto $\lim_{t\to 0}\frac{W_1(t)}{t}$ es una cantidad finita. Pasando al límite en (4) y teniendo en cuenta (1) llegamos al resultado buscado. Ahora bien, como para $k\geq 1$, $W_k(0)=0$, (5) puede escribirse

$$W_k'(0) = 0, \quad k\geq 2 \qquad (6)$$

lo que prueba la existencia de $W_k'(0)$.

Por otro lado consideremos ahora el suceso "k desintegraciones ocurren durante el intervalo de tiempo $(0,t+\Delta t)$". Este suceso puede darse de tres formas distintas:

a) $k-1$ desintegraciones entre 0 y t y una entre t y $t+\Delta t$,

b) k desintegraciones entre 0 y t y 0 entre t y $t+\Delta t$,

c) a lo sumo $k-2$ desintegraciones entre 0 y t y al menos 2 entre t y $t+\Delta t$.

por tanto, de acuerdo con las hipótesis 1 y 2, tendremos

$$W_k(t+\Delta t) = W_{k-1}(t).W_1(\Delta t) + W_k(t).W_0(\Delta t) + R \qquad (7)$$

donde $R = o(\Delta t)$, de acuerdo con la hipótesis 3 y la relación (5). De aquí

$$W_k(t+\Delta t) - W_k(t) = W_k(t)[W_0(\Delta t)-1] + W_{k-1}(t).W_1(\Delta t) + R$$

dividiendo por Δt y pasando al límite

$$\lim_{\Delta t\to 0}\frac{W_k(t+\Delta t)-W_k(t)}{\Delta t} = W_k(t)\lim_{\Delta t\to 0}\frac{W_0(\Delta t)-1}{\Delta t} + W_{k-1}(t)\lim_{\Delta t\to 0}\frac{W_1(\Delta t)}{\Delta t} + \lim_{\Delta t\to 0}\frac{R}{\Delta t}$$

0 por (5)

teniendo en cuenta (3) y el hecho de que $W_0'(0) = -W_1'(0)$, queda

$$W_k'(t) = \mu \left(W_{k-1}(t) - W_k(t) \right) \quad , \; k = 1, 2, \ldots \qquad (8)$$

para resolver esta ecuación diferencial hagamos

$$V_k(t) = W_k(t) \cdot e^{\mu t}$$

entonces derivando

$$V_k'(t) = W_k'(t) \cdot e^{\mu t} + W_k(t) \cdot \mu \cdot e^{\mu t}$$

y de aquí

$$W_k'(t) = \frac{V_k'(t)}{e^{\mu t}} - W_k(t) \cdot \mu$$

entonces, de (8)

$$\frac{V_k'(t)}{e^{\mu t}} = \mu \cdot W_{k-1}(t) \rightarrow V_k'(t) = \mu \cdot W_{k-1}(t) \cdot e^{\mu t} = \mu \cdot V_{k-1}(t) \quad , \; k = 1, 2, \ldots$$

De $W_0(t) = e^{-\mu t}$ se sigue $V_0(t) = 1$ y de aquí

$$V_1(t) = \mu t \quad , \quad V_2(t) = \frac{\mu^2 t^2}{2}$$

y en general

$$V_k(t) = \frac{(\mu t)^k}{k!} .$$

Por tanto

$$W_k(t) = \frac{(\mu t)^k}{k!} \cdot e^{-\mu t} \quad , \; k = 0, 1, \ldots$$

Es decir, que el número de desintegraciones durante un intervalo de tiempo t, dadas las hipótesis 1-3, tiene una distribución de Poisson con parámetro μt.

OBSERVACION

La distribución de Poisson que acabamos de obtener como solución al problema específico de las desintegraciones de sustancias radiactivas es aplicable, como modelo probabilístico, a un sinfín de situaciones a condición de que se verifiquen, debidamente modificadas, las hipótesis 1-3 citadas al comienzo del párrafo. Así por ejemplo, si en lugar de intervalos de longitud t tenemos dominios de area S, o volumen V y estamos interesados en calcular la

probabilidad de encontrar k partículas en un dominio determinado, podremos recurrir al modelo de Poisson a condición de que:

1) El proceso estudiado sea homogéneo. Es decir, la probabilidad de encontrar los k puntos en un dominio determinado depende sólo del area o volumen del dominio y no de su forma.
2) Si el dominio es pequeño (en cuanto a area o volumen), la probabilidad de encontrar más de un punto en dicho dominio es pequeña comparada con el tamaño del recinto o incluso si la comparamos con la probabilidad de encontrar una sola partícula.
3) Dominios disjuntos son mutuamente independientes.

NOTA HISTORICA. OIDO!!

Simeon D. Poisson vivió de 1781 a 1840. Su distribución de probabilidad la presentó por primera vez en 1837 en un libro titulado "Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités". Toma ya fardel!

ALGUNOS EJEMPLOS DE OBSERVACIONES QUE SE AJUSTAN A UNA DISTRIBUCION DE POISSON.

I) Desintegraciones radioactivas. Ya sabemos en este caso del ajuste del fenómeno a la distribución de Poisson, pero estudiemos con más detalle un experimento famoso. En 1920, Rutherford, Chadwick y Ellis estudiaron en Cambridge una sustancia radioactiva durante $N = 2608$ intervalos de 7.5 segundos cada uno, obteniendo el número de partículas que alcanzan un contador Geiger. En la tabla se recogen el número N_k de periodos con exactamente k partículas.

k	N_k	$N.P(X=k)$	k	N_k	$N.P(X=k)$
0	57	54.399	5	408	393.515
1	203	210.523	6	273	253.817
2	383	407.361	7	139	140.325
3	525	525.496	8	45	67.882
4	532	508.418	9	27	29.189
			$k \geq 10$	16	17.075
			Total	2608	2608.000

El número total de partículas fue de $T = \sum k N_k = 10094$ y por término medio en cada intervalo $T/N = 3.870$. Se obtuvieron entonces los valores teóricos $N.P(X=k)$, donde $P(X=k)$ se obtiene a partir de una distribución de Poisson con parámetro $\lambda = 3.870$, es decir partículas por unidad de tiempo empleada, a saber intervalos de 7.5 segundos. La comparación de los valores teóricos con los experimentales, con métodos que no hacen al caso ahora, sugieren que el experimento en cuestión puede explicarse correctamente mediante la correspondiente distribución de Poisson.

II) Bombas volantes sobre Londres. Como ejemplo de una distribución espacial de puntos aleatorios consideremos las estadísticas de bombas volantes que cayeron sobre Londres durante la Segunda Guerra Mundial. Se divide la ciudad en $N = 576$ areas pequeñas de superficie $S = \frac{1}{4}\ km^2$ cada una. La tabla presenta el número N_k de areas con exactamente k bombas. El número total de impactos fue de $T = \sum k . N_k = 537$ y la media de impactos por area $T/N = 0.9323$ que tomaremos

k	0	1	2	3	4	5 y más
N_k	229	211	93	35	7	1
$N . P(X=k)$	226.74	211.39	98.54	30.62	7.14	1.57

como valor de λ para la distribución de Poisson correspondiente. La concordancia entre los datos teóricos y los experimentales es más que aceptable.

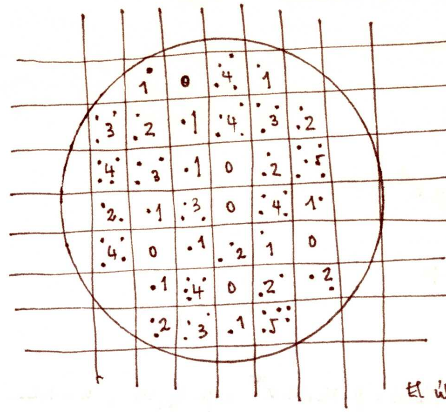
Es interesante señalar que existe una creencia general en el sentido de que los impactos tendían a agruparse. Si esto hubiera sido cierto hubieran existido gran cantidad de areas con muchos o muy pocos impactos y por contra gran deficiencia de areas con un número intermedio de impactos. La tabla presentada indica la homogeneidad y aleatoriedad de las areas y deshace la creencia general que tiende a identificar la aleatoriedad como una cierta tendencia al agrupamiento.

III) Conexiones telefónicas incorrectas. La tabla presenta estadísticas acerca de las conexiones telefónicas a un número equivocado. Un total de $N = 267$ números fueron observados; N_k indica cuántos

k	N_k	$N . P(X=k)$	k	N_k	$N . P(X=k)$
0-2	1	2.05	11	20	24.34
3	5	4.76	12	18	17.72
4	11	10.39	13	12	11.92
5	14	18.16	14	7	7.44
6	22	26.45	15	6	4.33
7	43	33.03	≥ 16	2	4.65
8	31	36.09	Total	267	267.00
9	40	35.04			
10	35	30.63			

números tuvieron exactamente k conexiones erróneas. El número medio de estas conexiones por número de teléfono controlado fue de 8.74 que se toma como valor de λ para la distribución de Poisson que nos permite obtener $P(X=k)$.

IV) Recuentos de hematíes y bacterias.



La figura reproduce una placa de Petri con colonias de bacterias, visibles al microscopio como puntos oscuros. La placa se divide en cuadrados pequeños, la tabla reproduce el número de cuadrados observados con exactamente k puntos, en varios experimentos llevados a cabo.

k	0	1	2	3	4	5	6
N_k	5	19	26	26	21	13	8
Poisson	6.1	18.0	26.7	26.4	19.6	11.7	9.5
N_k	8	16	18	15	9	7	
Poisson	6.8	16.2	19.2	15.1	9.0	6.7	
N_k	60	80	45	16	9		
Poisson	62.6	75.8	45.8	18.5	7.3		

El último valor incluye todos aquellos cuadrados con k o más colonias

LA DISTRIBUCIÓN BINOMIAL NEGATIVA

Consideremos n pruebas de Bernoulli, es decir, pruebas cuyo resultado es éxito o fracaso con probabilidades p y $q=1-p$, respectivamente. Supongamos además que se trata de pruebas independientes. Estamos interesados en conocer cuando ocurrirá el r-ésimo éxito. Aquí r es un entero positivo fijo. Con las condiciones planteadas, es obvio que la probabilidad de que el r-ésimo éxito ocurra en la prueba $\nu \leq n$ es independiente de n y depende tan sólo de ν, r y p. Puesto que necesariamente $\nu \geq r$, podemos escribirlo de la forma $\nu = r+k$. La probabilidad de que el r-ésimo éxito ocurra en el prueba $r+k$ $(k=0,1,\ldots)$ la denotaremos mediante $f(k;r,p)$. Esta probabilidad es igual a la probabilidad de que ocurran exactamente k fracasos antes del r-ésimo éxito. Este suceso ocurre si y sólo si, entre las $r+k-1$ pruebas hay exactamente k fracasos y en la siguiente, a saber la $(r+k)$-ésima, el resultado de la prueba es un éxito, la probabilidad de un suceso tal como el descrito viene dada por

$$\binom{r+k-1}{k} p^{r-1}\cdot q^k \quad \text{para los } k \text{ fracasos en las } r+k-1 \text{ pruebas, y}$$

$$p \qquad \text{para el último éxito}$$

en consecuencia

$$f(k;r,p) = \binom{r+k-1}{k} p^r \cdot q^k \qquad k=0,1,2,\ldots$$

Que suele también escribirse de la forma

$$f(k;r,p) = \binom{-r}{k} p^r \cdot (-q)^k \qquad k=0,1,2,\ldots$$

de donde le viene el nombre el Binomial negativa. Téngase en cuenta que:

$$\binom{-r}{k} = \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!} = \frac{(-r)(-(r+1))\cdots(-(r+k-1))}{k!} = \frac{(-1)^k r.(r+1)\cdots(r+k-1)}{k!} =$$

$$= \frac{(-1)^k (r-1)!\, r.(r+1)\cdots(r+k-1)}{(r-1)!\; k!} = (-1)^k \frac{(r+k-1)!}{(r-1)!\,k!} = (-1)^k \binom{r+k-1}{k}.$$

Supongamos ahora que la sucesión de pruebas de Bernoulli la continuamos el número de veces necesarias para conseguir los r éxitos. En teoría puede ocurrir que el éxito r-ésimo no aparezca nunca y debamos prolongar nuestra sucesión hasta infinito.

Tenemos así descrita una variable aleatoria discreta X, que toma los valores $\{0,1,2,\ldots\}$ y que representa el número de pruebas adicionales a r que debemos realizar para alcanzar

r éxitos si la probabilidad de cada éxito es p, $q=1-p$ la de obtener un fracaso, y las pruebas se llevan a cabo de forma independiente. La Probabilidad de que $X=k$, viene dada por

$$P(X=k) = f(k;r,p) = \binom{r+k-1}{k} p^r\cdot q^k = \binom{-r}{k} p^r (-q)^k.$$

No debe existir duda alguna respecto al hecho de que la función $f(k;r,p)$ es una función de cuantía, por cuanto:

1) $f(k;r,p) \geq 0$, $\forall k$

2) $f(k;r,p)$ es discreta por cuanto toma valores distintos de cero para $k=0,1,2,\ldots$ que es un conjunto discreto de valores

3) $f(k;r,p)$ es normada. En efecto

$$\sum_{k\geq 0} f(k;r,p) = 1$$

ya que si recordamos el desarrollo en serie de Mc Laurin de la función $(1-x)^{-n}$ tenemos:

$$\frac{1}{(1-x)^n} = \sum_{j\geq 0} \binom{n+j-1}{j} x^j \qquad |x|<1$$

en nuestro caso

$$\sum_{k\geq 0} \binom{r+k-1}{k} p^r q^k = p^r \sum_{k\geq 0} \binom{r+k-1}{k} q^k = p^r \cdot \frac{1}{(1-q)^r} = 1$$

$$|q|<1.$$

La distribución de probabilidad Binomial Negativa es conocida también como distribución de Pascal. Para $r=1$ la distribución se conoce con el nombre de distribución geométrica, y su función de cuantía se reduce a $f(k;1,p) = p.q^k$, $k=0,1,\ldots$.

EJEMPLOS

I) El problema de las cajas de cerillas de Banach. Un conocido matemático lleva siempre una caja de cerillas en cada uno de los dos bolsillos de su chaqueta. Cuando necesita una cerilla selecciona al azar uno de los bolsillos de manera que las sucesivas elecciones constituyen pruebas de Bernoulli con $p=\frac{1}{2}$. Supongamos que inicialmente cada caja contiene exactamente N cerillas y consideremos el momento en el cual, por primera vez, el matemático descubre que una de las cajas está vacía. En este momento la otra caja puede contener $0, 1, 2, \ldots, N$ cerillas, y vamos a designar mediante u_r la correspondiente probabilidad. Identificaremos el "éxito" con la elección del bolsillo izquierdo. Este bolsillo estará vacío, o mejor, descubriremos que está vacío justo en el momento que el bolsillo derecho contiene exactamente r cerillas, si y sólo si exactamente $(N-r)$ fracasos han precedido al $(N+1)$-simo éxito. La probabilidad de este suceso es $f(N-r; N+1, \frac{1}{2})$. El mismo argumento puede aplicarse respecto del bolsillo derecho y por tanto la probabilidad buscada es

$$u_r = 2f\left(N-r; N+1, \tfrac{1}{2}\right) = 2 \cdot \binom{2N-r}{N-r} \left(\frac{1}{2}\right)^{N+1} \cdot \left(\frac{1}{2}\right)^{N-r} = \binom{2N-r}{N} 2^{-2N+r}.$$

Por ejemplo, para $N = 50$ y $r = 4$, $u_r = 0.074790$ y para $r = 29$, $u_r = 0.000232$.

II) <u>Generalización: Tenis de mesa</u>. La naturaleza del problema anterior queda más de manifiesto cuando se atribuyen diferentes probabilidades a cada una de las cajas. Interpretaremos esto de forma diferente. Supongamos que dos amigos, Pedro y Pablo, juegan a un juego que puede ser tratado como una sucesión de pruebas de Bernoulli en las que las probabilidades p y q sirven como ~~para~~ medida de la destreza de los jugadores. En el tenis de mesa el jugador que primero llega a 21 gana la partida. Por comparación con el ejemplo precedente consideremos la situación general en la que $2\nu+1$ éxitos individuales se requieren. El juego finaliza, en el mejor de los casos en la prueba $2\nu+1$ (caso de 21 a 0) o bien, en la peor de las situaciones, en la prueba $4\nu+1$ (20 a 21). Designemos por a_r la probabilidad de que Pedro gane en la prueba (jugada) $4\nu+1-r$. Esto ocurrirá si y sólo si, en las $4\nu-r$ primeras jugadas Pedro ha obtenido 2ν tantos (éxitos) y además gana la jugada siguiente. Así pues

$$a_r = f(2\nu-r; 2\nu+1, p) = \binom{4\nu-r}{2\nu} p^{2\nu+1} q^{2\nu-r}. \qquad (1)$$

La probabilidad de que Pedro gane viene dada por $a_0 + a_1 + \cdots + a_{2\nu}$. La probabilidad de que el juego termine (gana uno u otro) en la jugada $4\nu+1-r$ viene dada por $a_r + b_r$, donde b_r, probabilidad de que gane Pablo en la jugada $4\nu+1-r$, viene dada por (1) con p y q intercambiados.

Para $2\nu = N$, $p = q = \frac{1}{2}$, las probabilidades $a_r + b_r$ se reducen a las u_r del ejemplo anterior.

<u>Nota</u>.- El ejemplo de las cerillas de Banach no se debe a Banach en contra de lo que pudiera ~~pensarse~~ parecer. Tiene su ~~origen~~ origen en una referencia humorística al hábito de fumador de Banach hecha por H. Steinhaus en un acto celebrado en honor de Banach. La anécdota se hizo famosa y de aquí que haya conservado el nombre hasta ahora.

LA DISTRIBUCIÓN HIPERGEOMÉTRICA

Supongamos n bolas de una urna de las cuales n_1 son rojas y $n_2 = n - n_1$ son negras. Se extraen r al azar. Buscamos la probabilidad de que en las r bolas extraídas haya exactamente k rojas y designaremos por q_k dicha probabilidad. k es cualquier valor entre 0 y n_1 o r si éste es más pequeño.

Para encontrar q_k, observemos que el grupo elegido contiene k bolas rojas y $r-k$ negras. Las rojas pueden elegirse de $\binom{n_1}{k}$ formas diferentes, mientras que las negras pueden serlo de $\binom{n-n_1}{r-k}$ formas distintas. Puesto que la elección de las k rojas puede combinarse con cualquiera de las posibles elecciones de las $r-k$ negras, tendremos

$$q_k = \frac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}}$$

Si con las condiciones enunciadas definimos una variable X cuyos valores representan el número de bolas rojas en la muestra extraída, esta variable será una variable aleatoria discreta cuyas probabilidades, $P(X=k)$ vienen dadas por

$$P(X=k) = q_k = \frac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}} \qquad \begin{array}{ll} k = 0, 1, \ldots, r & \text{si } r \leq n_1. \\ k = 0, 1, \ldots, n_1 & \text{si } n_1 < r \leq n. \end{array}$$

Obsérvese que

$$\sum_{k=0}^{r} P(X=k) = \sum_{k=0}^{r} q_k = \frac{1}{\binom{n}{r}} \sum_{k=0}^{r} \binom{n_1}{k}\binom{n-n_1}{r-k} = \frac{1}{\binom{n}{r}} \cdot \binom{n}{r} = 1.$$

<u>EJEMPLOS</u>.

I) El senado de los EE.UU. está compuesto por 100 senadores, 2 por cada uno de los 50 estados. Se elige un grupo de 50 senadores y se quiere conocer la probabilidad de que un estado determinado no esté representado en el grupo y la de que en el grupo haya 1 senador de este estado o los 2 senadores del estado en cuestión.

Aquí asimilamos senador del estado en cuestión a "rojo", con lo que $n_1 = 2$, $r = 50$ y $n = 100$ por tanto

$$q_0 = q_2 = \frac{50 \cdot 49}{100 \cdot 99} = 0.24747\ldots \quad \text{y} \quad q_1 = \frac{50}{99} = 0.50505\ldots$$

II) <u>Estimación del tamaño de una población animal a partir de datos de recaptura</u>. Supongamos que 1000 peces capturados en un lago son marcados mediante una mancha roja y se los arroja nuevamente al lago. Después de cierto tiempo se lleva a cabo una nueva captura de otros 1000 peces y 100 de entre ellos llevan la mancha roja. ¿Qué conclusiones podemos extraer respecto al número de peces en el lago? Este es un problema típico de <u>estimación estadística</u>. No es momento de profundizar y describir los distintos métodos que podrían utilizarse para su resolución, nos limitaremos a demostrar como la distribución hipergeométrica nos proporciona una clave para la solución de nuestro problema. Suponemos naturalmente que las dos capturas pueden ser consideradas como muestras aleatorias de la población total de peces del lago. (En la práctica semejante suposición excluye situaciones en las que las dos capturas se hacen en una misma localidad y en un corto período de tiempo). Suponemos también que el número de peces en el lago no cambia entre las dos capturas.

Generalizemos el problema admitiendo tamaños arbitrarios de las muestras. Sean

n = el número total de peces en el lago (desconocido)

n_1 = número de peces en la primera captura. Harán el papel de bolas rojas.

r = número de peces en la segunda captura.

k = número de peces con mancha roja en la segunda captura.

$q_k(n)$ = probabilidad de que la segunda captura contenga exactamente k peces con mancha roja.

Con esta formulación es obvio que $q_k(n)$ viene dada por la correspondiente probabilidad hipergeométrica, a saber

$$q_k(n) = \frac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}}.$$

En la práctica n_1, r y k pueden conocerse por observación, pero n es desconocida. Nótese que consideramos n como un número fijo desconocido cuyo valor en modo alguno depende del azar. Sabemos que n_1+r-k peces han sido capturados en el lago entre las dos capturas llevadas a cabo y que obviamente $n \geq n_1+r-k$. Esto es todo cuanto sabemos con certeza. En nuestro ejemplo $n_1 = r = 1000$ y $k = 100$; es teóricamente posible que $n = 1900$, pero si comenzamos con una hipótesis de este tipo, llegamos a la conclusión de que ha ocurrido un suceso con una probabilidad extraordinariamente pequeña. En efecto, suponiendo que $n = 1900$, la probabilidad $q_k(n)$ vendrá dada, con los datos concretos del problema, por

$$q_k(1900) = \frac{\binom{1000}{100}\binom{900}{900}}{\binom{1900}{1000}} = \frac{(1000!)^2}{100!\ 1900!}$$

utilizando la fórmula de Stirling, $n! \sim \sqrt{2\pi}.n^{n+\frac{1}{2}}.e^{-n}$, para aproximar los factoriales llegaremos a que $q_k(1900) \sim 10^{-430}$ y en una situación así el sentido común nos lleva a rechazar la hipótesis como carente de todo sentido. Un razonamiento similar nos induciría a rechazar hipótesis del tipo, n es muy grande, por ejemplo $n = 10^6$. Esta consideración nos conduce a buscar un valor de n para el cual $q_k(n)$ alcance su mayor valor, puesto que en este caso nuestra observación tendría la mayor probabilidad de producirse. Para cualquier conjunto particular de observaciones n_1, r, k, el valor de n para el que $q_k(n)$ es más grande se designa mediante $\hat{n}$ y se le llama <u>estimación máximo verosímil</u> de n. Con vistas a encontrar $\hat{n}$ consideremos el cociente

$$\frac{q_k(n)}{q_k(n-1)} = \frac{(n-n_1)(n-r)}{(n-n_1-r+k)n} = \frac{n^2-n_1n-nr+n_1r}{n^2-n_1n-nr+nk}$$

Este cociente será mayor o menor que la unidad según que $nk < n_1r$ ó $nk > n_1r$. Esto significa que a medida que n aumenta la sucesión $q_k(n)$ crece primero para decrecer después alcanzando su máximo cuando n sea igual al mayor entero menor que n_1r/k. En nuestro caso particular $\hat{n} = 10.000$.

El verdadero valor de n puede ser mayor o menor que el encontrado, y estamos interesados también en dar unos límites dentro de los cuales sea razonable encontrar el verdadero valor de n. Supongamos para ello que n es más pequeño que 8.500. Sustituimos en la correspondiente expresión de $q_k(n)$ y calculamos la probabilidad de que la segunda captura contenga a lo sumo 100 peces marcados. Esta probabilidad viene dada por

$$x = q_0 + q_1 + \text{——} + q_{100}$$

y es igual a 0.04. Análogamente para $n = 12000$, la probabilidad de que la segunda captura contenga al menos 100 peces marcados es alrededor de 0.03. Con semejantes probabilidades es razonable pensar que el verdadero valor de n será alguno de los comprendidos entre 8500 y 12000.

APÉNDICE

Cont. de la DISTRIBUCIÓN DE POISSON como aproximación a la Binomial

Hemos resuelto una situación muy concreta, a saber, la fabricación de botellas de vidrio y la determinación aproximada del número de botellas defectuosas. Ahora bien, ¿cómo podemos utilizar esta aproximación ante una situación genérica definida mediante una binomial cualquiera $B(n,p)$?. La forma de actuar quedará explicitada en cuanto pongamos de manifiesto la relación existente entre los parámetros n,p y λ de la Binomial y su aproximación de Poisson, respectivamente. Volvamos al ejemplo, recordemos que $\lambda = \frac{x}{100}$ y que por otra parte $N = n/\lambda$, o bien

$$\lambda = \frac{n}{N} = n\cdot\left(\frac{1}{N}\right)$$

en esta Binomial $p = \frac{1}{N}$ y por tanto $\lambda = n.p$, este resultado no se da por casualidad, sino que esta es la relación que existe, en general, entre el parámetro λ y los parámetros n, p de la Binomial. Destaquemos además el hecho de que n era grande y $p = \frac{1}{N}$ era pequeño, por cuanto que habíamos señalado que la aproximación era tanto mejor cuanto mayor fuera el número de botellas fabricadas, N.

Podíamos, de hecho, haber planteado el problema de forma genérica, a continuación lo haremos, pero puede resultar más ilustrativo hacerlo mediante un caso práctico.

Sea ahora $B(n,p)$ una Binomial tal que n es muy grande y p es muy pequeño, de manera que pueden planteársenos dificultades a la hora de determinar las probabilidades $P(X=k)$, $k=0,1,\dots,n$. Hagamos $\lambda = np$. Sabemos que

$$\left.\begin{aligned} &P(X=0) = (1-p)^n = \left(1-\frac{\lambda}{n}\right)^n \\ &\text{para } n \text{ grande} \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \end{aligned}\right\} \quad P(X=0) \simeq e^{-\lambda} \qquad (1)$$

Por otra parte

$$\frac{P(X=k)}{P(X=k-1)} = \frac{(n-k+1)p}{kq} = \frac{\lambda-(k-1)p}{k\cancel{q}}$$

que para p pequeño (entonces q estará muy próximo a la unidad), puede aproximarse mediante

$$\frac{P(X=k)}{P(X=k-1)} \simeq \frac{\lambda}{k}. \qquad (2)$$

de (1) y (2), tenemos

$$P(X=1) \simeq \lambda.P(X=0) \simeq \lambda.e^{-\lambda}$$

$$P(X=2) \simeq \frac{\lambda}{2}.P(X=1) \simeq \frac{\lambda^2}{2}.e^{-\lambda}$$

$$P(X=3) \simeq \frac{\lambda}{3}.P(X=2) \simeq \frac{\lambda^3}{2.3}.e^{-\lambda}$$

$$\cdots\cdots\cdots$$

$$P(X=k) \simeq \frac{\lambda}{k}.P(X=k-1) \simeq \frac{\lambda^k}{k!}e^{-\lambda}$$

En resumen las probabilidades binomiales $B(n.p)$ pueden aproximarse mediante las probabilidades de una distribución de Poisson, $P(\lambda)$, con $\lambda = n.p$, siendo la aproximación tanto mejor cuanto mayor sea n y menor p.

$$\boxed{P(X=k) = \binom{n}{k} p^k.(1-p)^{n-k} \simeq \frac{\lambda^k}{k!}e^{-\lambda}} \qquad \lambda = n.p$$

EJEMPLOS de aproximación

I) Cumpleaños. ¿Cuál es la probabilidad, p_k, de que en una compañía de 500 artistas exactamente k celebren su cumpleaños el día de Año Nuevo? En el supuesto de la elección aleatoria de los 500 artistas, la probabilidad podrá obtenerse a partir de una Binomial $B\left(500,\frac{1}{365}\right)$, o bien aproximarla mediante una Poisson, $P(\lambda)$, con $\lambda = \frac{500}{365} = 1.3699\dots$. A continuación se presentan las probabilidades correctas y sus aproximaciones.

k	0	1	2	3	4	5	6
Binomial	0.2537	0.3484	0.2388	0.1089	0.0372	0.0101	0.0023
Poisson	0.2541	0.3481	0.2385	0.1089	0.0373	0.0102	0.0023

II) Piezas defectuosas. Una máquina fabrica tornillos y por término medio 15 de cada mil son defectuosos. Estamos interesados en saber, ¿cuántos tornillos deben ponerse en una caja para que la probabilidad de encontrar al menos 100 tornillos correctos sea mayor o igual que 0.8?.

En el supuesto de presentarlos en cajas de 100 unidades la probabilidad de no encontrar tornillos defectuosos en una caja puede evaluarse a partir de una Binomial $B(100, 0.015)$, de manera que

$$P(X=0) = (0.985)^{100} = 0.22061$$

o bien mediante una aproximación de Poisson, $P(1.5)$, $P(X=0) = e^{-1.5} = 0.22313\dots$

aproximación que daremos por buena a efectos prácticos.

A la vista de las probabilidades obtenidas parece claro que las cajas deberán contener más de 100 tornillos, [illegible]. Supongamos pues que la caja contiene $100 + x$ tornillos, con x entero y pequeño. Aplicamos la aproximación de Poisson con $\lambda = n.p = (100+x) \cdot 0.015 \simeq 1.5$. La probabilidad de que haya al menos 100 tornillos buenos es la misma de encontrar a lo sumo x tornillos defectuosos, por tanto

$$P(X \leq x) = \sum_{k=0}^{x} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-1.5}\left[1 + \frac{1.5}{1} + \dots + \frac{(1.5)^x}{x!}\right] \geqslant 0.8$$

Existen tablas adecuadas para determinar el valor del sumatorio a partir del valor de x, encontramos en ellas que para $x=1$, $P(X \leq 1) \simeq 0.56$ y para $x=2$, $P(X \leq 2) = 0.809$. Habrá pues que embalar los tornillos en cajas de 102 unidades.

El verdadero valor de $P(X \leq 2)$ obtenido a partir de una Binomial $B(102, 0.015)$ es $0.8022\ldots$ muy próximo al encontrado con nuestra aproximación.

III) <u>Un ejemplo empírico</u>. El número de veces que aparece el par (7,7) entre 100 pares de números aleatorios suele de seguir una distribución binomial $B(100, 0.01)$. En la tabla se muestran las frecuencias absolutas, N_k, de k apariciones del par (7,7) en 100 grupos de 100 pares de dígitos aleatorios. Se presentan también en la tabla los valores teóricos de la distribución binomial y de la correspondiente aproximación de Poisson, $P(\lambda)$, con $\lambda = n.p = 1$. Las frecuencias relativas observadas, $N_k/100$, están razonablemente de acuerdo con las probabilidades teóricas de ambas distribuciones discretas.

k	$P(X=k)$ $B(100, 0.01)$	$P(X=k)$ $P(1)$	N_k
0	0.366032	0.367879	41
1	0.369730	0.367879	34
2	0.184865	0.183940	16
3	0.060999	0.061313	8
4	0.014942	0.015328	0
5	0.002898	0.003066	1
6	0.000463	0.000511	0
7	0.000063	0.000073	0
8	0.000007	0.000009	0
9	0.000001	0.000001	0

Capítulo 5

Variables aleatorias continuas

VARIABLES ALEATORIAS CONTINUAS

I. UNA CARACTERIZACION DE LAS FUNCIONES MEDIBLES

Recordemos que una función medible era una aplicación entre un espacio Ω, dotado de una σ-algebra de sus partes, $\mathcal{A}$ y la recta real $\mathbb{R}$, provista de la correspondiente tribu de Borel, que verificaba

$$f^{-1}(B) \in \mathcal{A}, \ \forall B \in \beta, \text{ donde } \beta \text{ es la tribu } (\sigma\text{-algebra}) \text{ de Borel.}$$

Por otra parte, una función simple era aquella que siendo medible tenía un rango finito, lo que equivale a decir que una función simple es de la forma

$$s = \sum_{i=1}^{n} a_i \cdot \chi_{A_i} \qquad a_i \in \mathbb{R}, \ i=1,\dots,n$$

donde $\{A_i\}_{i=1}^{n}$ es una partición medible finita del espacio Ω, es decir, $\Omega = \bigcup_{i=1}^{n} A_i$, $A_i \cap A_j = \phi$ $i \neq j$ y $A_i \in \mathcal{A}$, $\forall i$.

Existe una caracterización de las funciones medibles mediante funciones simples que damos a continuación en forma de teorema, cuya demostración omitiremos.

TEOREMA.- Una aplicación f del espacio medible $(\Omega, \mathcal{A})$ en el espacio medible $(\mathbb{R}, \beta)$, que sea no negativa, es medible $\Leftrightarrow$ $\exists$ una sucesión monótona de funciones simples que converge a f (las funciones simples son también no negativas)

Si pretendemos trabajar con una función ~~medible~~ cualquiera, no necesariamente no negativa, podemos dar una caracterización semejante, excepción hecha de la monotonía de la sucesión de funciones simples, basada en el hecho de que toda función admite una descomposición del tipo

$$f = f^{+} - f^{-}$$

donde

$$f^{+} = \begin{cases} f & \text{si } f \geq 0 \\ 0 & \text{si } f < 0 \end{cases} \qquad y \qquad f^{-} = \begin{cases} |f| & \text{si } f \leq 0 \\ 0 & \text{si } f > 0 \end{cases}$$

y por tanto f^{+} y f^{-} son no negativas. El teorema correspondiente dice

TEOREMA.- Una aplicación f de $(\Omega, \mathcal{A})$ en $(\mathbb{R}, \beta)$ es medible $\Leftrightarrow$ $\exists$ una sucesión de funciones simples que converge a f.

II. CONTINUIDAD ABSOLUTA

II.a) Medidas absolutamente continuas.

Sea $(\Omega, \mathcal{A})$ un espacio medible y dotémoslo de dos medidas, μ y λ.

Definición I: Decimos que μ es absolutamente continua respecto a λ si $\mu(A) = 0$ para cada conjunto A de $\mathcal{A}$ tal que $\lambda(A) = 0$.

Si la medida μ es finita la definición dada es equivalente a,

DEFINICIÓN II: μ es absolutamente continua respecto a λ si para cada $\varepsilon > 0$, $\exists \delta > 0$ tal que $\mu(E) < \varepsilon$, siempre que $\lambda(E) < \delta$, con $E \in \mathcal{A}$.

II.b) Funciones absolutamente continuas.

Existe un concepto designado de la misma manera que hace referencia a las funciones a valores reales. Luego veremos que esta igualdad de nombres no es casual.

Definición III: Una función h a valores en $\mathbb{R}$ se dice que es absolutamente continua si, dado $\varepsilon > 0$, $\exists \delta > 0$ / $\forall \{]a_i, b_i]\}_{i=1}^{n}$ intervalos dos a dos disjuntos verificando

$$\sum_{i=1}^{n} (b_i - a_i) < \delta \quad \text{entonces} \quad \sum_{i=1}^{n} |h(b_i) - h(a_i)| < \varepsilon.$$

Como consecuencia de esta definición es muy fácil comprobar que una función absolutamente continua y uniformemente continua.

II.c) Relación entre ambos conceptos.

Un caso especialmente interesante es aquel en que las medidas utilizadas son unas determinadas y además también lo es el espacio medible utilizado. Este caso nos permite además poner de manifiesto el porqué ambos conceptos se designan de la misma forma.

Supongamos que $(\Omega, \mathcal{A}) \equiv (\mathbb{R}, \beta)$ y además μ es una medida de probabilidad sobre $(\mathbb{R}, \beta)$ siendo λ la medida Lebesgue sobre los Borelianos de $\mathbb{R}$, es decir, aquella que sobre cada intervalo vale la longitud del mismo.

Ya sabemos que a partir de P, medida de probabilidad, podemos definir una función, que resulta ser de distribución y acotada, mediante la relación $F(x) = P(]-\infty, x])$. Observemos que si P es absolutamente continua respecto a λ, tendremos que dado $\varepsilon > 0$, construyamos $I = \bigcup_{i=1}^{n} \{]a_i, b_i]\}$ con $\{]a_i, b_i]\} \cap \{]a_j, b_j]\} = \phi$ $i \neq j$, de manera que si

$$\lambda(I) = \sum_{i=1}^{n} \lambda\{]a_i, b_i]\} = \sum_{i=1}^{n} (b_i - a_i) < \delta$$

, siendo δ elegido en función del ε, tendremos $P(I) < \varepsilon$, pero

$$P(I) = \sum_{i=1}^{n} P(]a_i, b_i]) = \sum_{i=1}^{n} [F(b_i) - F(a_i)] < \varepsilon$$

lo que quiere decir que F es una función absolutamente continua. Hemos visto pues que P absolutamente continua $\rightarrow$ F absolutamente continua. Puede comprobarse, no es aquí el momento, que la implicación contraria es también cierta, lo que nos permite establecer la siguiente importante relación:

TEOREMA.- Sea F una función de distribución de probabilidad. F es absolutamente continua si y sólo si la correspondiente ~~m~~medida de probabilidad, P, que da origen a F es absolutamente continua respecto de λ, medida de Lebesgue.

III. INTEGRACION RESPECTO DE UNA MEDIDA

Consideremos el espacio de medida $(\Omega, \mathcal{B}, \mu)$ y una función simple, no negativa, s, definida sobre él.

Definición: La integral sobre A de s respecto de la medida μ, se designa mediante el símbolo $\int_A s.d\mu$ y se define mediante la relación.

$$\int_A s.d\mu = \sum_{i=1}^{n} a_i.\mu(A\cap A_i), \text{ donde } A\in\mathcal{B}.$$

Esta relación está siempre bien definida por cuanto $a_i.\mu(A\cap A_i)$ se toma igual a 0, cuando $a_i=0$ y $\mu(A\cap A_i)=+\infty$.

La definición puede extenderse de inmediato a cualquier función medible no negativa que tengamos sobre Ω, de la siguiente forma.

Definición: La integral sobre A de f respecto de la medida μ se define

$$\int_A f\,d\mu = \sup\left\{\int_A s\,d\mu \;;\; 0\le s\le f\right\}.$$

El conjunto sobre el cual tomamos el supremo es no vacío de acuerdo con la caracterización dada para las funciones medibles no negativas, lo que garantiza la coherencia de la definición.

Propiedades de la integral. La aplicación así definida, en definitiva no es más que una aplicación de Ω en R, goza de una serie de propiedades que no vamos a demostrar pero que sí enunciaremos en algún caso concreto:

a) Si $a\ge 0$. $\int_A af\,d\mu = a\int_A f\,d\mu$.

b) Si $A_1\subset A_2$. $\int_{A_1} f\,d\mu \le \int_{A_2} f\,d\mu$.

c) Si $f\le g$. $\int_A f\,d\mu \le \int_A g\,d\mu$.

d) La aplicación λ, definida de $\mathcal{B}$ en R, mediante $\lambda(A)=\int_A f\,d\mu$, $A\in\mathcal{B}$ y f una función medible no negativa dada, es una medida.

e) Para f y g medibles y no negativas, tenemos $\int_A (f+g)\,d\mu = \int_A f\,d\mu + \int_A g\,d\mu$.

a) + e) suponen que para f y g medibles y no negativas $\int_A (af+bg)\,d\mu = a\int_A f\,d\mu + b\int_A g\,d\mu$ siendo los coeficientes de la combinación lineal no negativos.

f) Si A es tal que $\mu(A)=0$, entonces $\int_A f\,d\mu=0$, $\forall f$ medible y no negativa.

g) Si A es tal que $\mu(A)>0$, y f, medible y no negativa es tal que $\int_A f\,d\mu=0$ entonces $f=0$ c.p.p. (casi por todas partes, es decir salvo en un conjunto que tiene medida nula).

Integración de funciones medibles cualesquiera. Hemos introducido el concepto de integral únicamente para el caso de funciones medibles no negativas, el último paso consiste en generalizar el concepto para funciones medibles cualesquiera, y simultáneamente introducir el concepto de integrabilidad.

Definición: Sea f una función medible cualquiera sobre el espacio de medida $(\Omega, \mathcal{B}, \mu)$, diremos que f es integrable [redacted] sobre $A\in\mathcal{B}$, si $\int_A f^+ d\mu$ y $\int_A f^- d\mu$ son finitas. En este caso definimos la integral de f mediante

$$\int_A f\,d\mu = \int_A f^+ d\mu - \int_A f^- d\mu.$$

La integral así definida goza prácticamente de las mismas propiedades antes citadas, por ejemplo es un operador lineal, es decir

$$\int_A (af+bg)\,d\mu = a\int_A f\,d\mu + b\int_A g\,d\mu \qquad \forall a,b\in R \qquad f,g \text{ medibles}.$$

si bien algunas de ellas se pierden, por ejemplo $\lambda(A)=\int_A f\,d\mu$ no es una medida, por cuanto no puede garantizarse que sea siempre una cantidad no negativa. Surgen, sin embargo, propiedades nuevas, como por ejemplo una muy conocida:

$$\left|\int_A f\,d\mu\right| \le \int_A |f|\,d\mu.$$

Observación.- El concepto de integrabilidad es aplicable a cualquier función medible y por tanto también a las no negativas. En éstas el concepto equivale a decir que el valor de su integral, que siempre está definida, es finito.

IV. UNA IMPORTANTE PROPIEDAD DE LAS MEDIDAS ABSOLUTAMENTE CONTINUAS

Volvamos nuevamente a las condiciones iniciales del apartado II.a) y supongamos además que las dos medidas, λ y μ, son σ-finitas. Existe un importantísimo teorema que relaciona una medida absolutamente continua respecto a otra con ésta a través de una integral.

TEOREMA DE RADON-NIKODYM.- Sea μ una medida absolutamente continua respecto a λ y sean, además, ambas σ-finitas. Existe entonces una función medible finita, no negativa, f, de manera que para cualquier conjunto medible, $A\in\mathcal{B}$, tenemos

$$\mu(A) = \int_A f\,d\lambda.$$

A f se la conoce con el nombre de derivada de Radon-Nikodym de μ respecto de λ, y se la designa a veces mediante $\left[\frac{d\mu}{d\lambda}\right]$.

El teorema de Radon-Nikodym para medidas de probabilidad. El teorema adquiere especial relevancia para nosotros cuando μ es una medida de probabilidad, P, y λ es la medida de Lebesgue. Sabemos que tanto una como otra son medidas σ-finitas con lo que el teorema tiene total validez en semejante situación. Tendremos pues, que $\exists f$, medible, finita, no negativa, de manera que

$$P(A) = \int_A f \, d\lambda \quad , \text{ con } A \in \beta$$

siendo $\int_A f\, d\lambda$ una integral de Lebesgue así llamada por ser λ la medida de Lebesgue. Observemos que para $A =]a,b]$, tenemos

$$P(]a,b]) = \int_{A=]a,b]} f \, d\lambda$$

pero si F es la función de distribución de probabilidad definida a partir de P, la anterior igualdad puede escribirse

$$F(b) - F(a) = \int_{]a,b]} f \, d\lambda$$

en particular para $a = -\infty$, llegaremos a

$$F(b) = \int_{]-\infty,b]} f \, d\lambda \quad \forall b \in R$$

que preferimos escribir de la forma

$$F(x) = \int_{]-\infty,x]} f \, d\lambda \quad , \forall x \in R.$$

Por otra parte $F(x)$ por ser ~~monótona~~ tiene derivada casi por todas partes y además es finita casi por todas partes, una importante propiedad, que no demostramos, nos permite afirmar que $F' = f$ c.p.p y la última igualdad puede también escribirse

$$F(x) = \int_{]-\infty,x]} F' \, d\lambda \quad , \forall x \in R$$

porque además de ser finita c.p.p., F' es medible e integrable y no negativa c.p.p. por serlo f.

V. DEFINICIÓN DE VARIABLE ALEATORIA CONTINUA

Sea X una v.a. definida de $(S, \mathcal{A}, P)$ en (R, β, P'), decimos que X es una v.a. del tipo continuo, o simplemente una v.a. continua si la medida de probabilidad P' es absolutamente continua respecto de la medida de Lebesgue, lo que equivale a decir que F, función de distribución de probabilidad definida a partir de P', es absolutamente continua.

Como consecuencia de esta definición existirá una función f, medible, finita y no negativa tal que

$$F(x) = \int_{-\infty}^{x} f \, d\lambda \quad \text{o también} \quad P'(A) = \int_A f \, d\lambda \quad , \forall x \in R \text{ y } \forall A \in \beta,$$

a la función f se la conoce con el nombre de función de densidad de probabilidad de la variable aleatoria X. Esta función tiene además una propiedad importante, a saber

$$\int_{-\infty}^{+\infty} f \, d\lambda = 1$$

puesto

$$\int_{-\infty}^{+\infty} f \, d\lambda = F(+\infty) = \lim_{x \to \infty} F(x) = 1.$$

Caracterización de una distribución de probabilidad mediante la función de densidad de probabilidad

Recordemos que en el caso de variables aleatorias discretas podíamos caracterizar la distribución de probabilidad, es decir, la medida de probabilidad inducida o la función de distribución correspondiente con sólo conocer la función de cuantía. Exactamente el mismo procedimiento puede ser utilizado en nuestra situación actual, así como la función de distribución (o la medida de probabilidad) de una v.a. continua nos define una función, la de densidad de probabilidad, a partir de una función con características similares a ésta podemos recorrer el camino inverso para llegar al anterior punto de partida. Ello quiere decir que una v.a. continua, más concretamente, su distribución de probabilidad puede ser caracterizada mediante la correspondiente función de densidad. Veámoslo

TEOREMA.- Sea f una función integrable Lebesgue, es decir, medible y con $\int_R f \, d\lambda < +\infty$ tal que

a) f es finita

b) f es no negativa

c) $\int_{-\infty}^{+\infty} f \, d\lambda = 1$

entonces f es la función de densidad de probabilidad de alguna variable aleatoria del tipo continuo, X.

Demostración.- Basta comprobar que la función $F(x)$ definida mediante

$$F(x) = \int_{-\infty}^{x} f \, d\lambda$$

es una función de distribución de probabilidad, por cuanto, $F(+\infty) = \int_{-\infty}^{+\infty} f \, d\lambda = 1$.
A partir de F podemos definir una medida de probabilidad, P sobre los Borelianos de R que puede considerarse inducida por una variable aleatoria definida de $(R, \beta) \xrightarrow{en} (R, \beta)$, concretamente la identidad.

En resumen, cuando haya que caracterizar una v.a. continua lo haremos indistintamente con P', F, función de distribución (f.d.) o f, función de densidad de probabilidad (f.d.p.).

<u>Propiedad</u>.- Recordemos que sea cual sea la variable aleatoria X, P' la medida de probabilidad inducida (da

$$P'(\{a\}) = P(X = a) = F(a) - F(a-0)$$

Si X es del tipo discreto y $a \in S_a$ sabemos que $F(a) \neq F(a-0)$ y $P(X=a) = P'(\{a\}) \neq 0$.

Una variable aleatoria continua se caracteriza por poseer una f.d., F, que es absolutamente continua y por tanto continua, ello quiere decir que $F(x) = F(x-0)$, $\forall x \in R$, y de aquí

$$P'(\{a\}) = P(\cancel{X = a}) = 0 \quad , \forall a \in R. \quad \text{Si } X \text{ es una v.a. continua.}$$

<u>Relación entre la integral de Lebesgue y la integral de Riemann</u>. Antes de continuar veamos una interesante relación entre ambas integrales que nos será útil para estudiar las v.a. continuas.

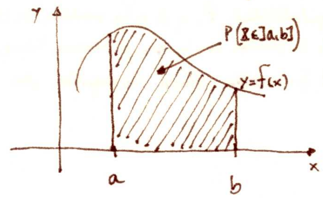
Se puede demostrar, <u>Teorema de Vitali</u>, que toda función integrable Riemann es integrable Lebesgue y además sus integrales son iguales. El recíproco no es cierto.

La importancia de este hecho estriba en que normalmente las variables aleatorias continuas que utilizaremos dan lugar a f.d.p., f, que son continuas y por tanto integrables Riemann (Una función es integrable Riemann si y sólo si los puntos de discontinuidad de la función forman un conjunto de medida de Lebesgue nula). Ello supone que todas las integrales que aparecen en la definición de v.a. continua pueden ser consideradas como integrales de Riemann, con las ventajas que ello comporta. [A señalar que las f.d.p.'s que utilizaremos serán tales que las integrales impropias de Riemann utilizadas son convergentes].

<u>Significado físico de la probabilidad</u>. Bajo las condiciones descritas en el párrafo precedente la probabilidad adquiere un significado físico del que carecía hasta ahora. En efecto la

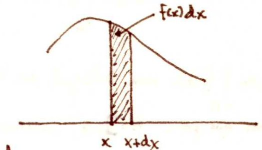
$$P(X \in]a,b]) = F(b) - F(a) = \int_a^b f(x) dx = \int_{]a,b]} f \, d\lambda$$

pero la integral de Riemann tiene el sentido de una superficie, más concretamente, la probabilidad descrita es el área comprendida bajo la curva $y = f(x)$ en el intervalo $]a,b]$.



Obsérvese que como F es absolutamente continua y $P(X=a) = 0$ la $P(X \in]a,b]) = P(X \in [a,b]) = P(X \in [a,b[) = P(X \in]a,b[)$

Por otra parte el hecho de que $\int_{-\infty}^{+\infty} f(x) dx = 1$, supone que el área encerrada bajo la curva $y = f(x)$ es igual a la unidad.

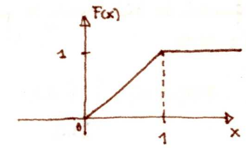


Observemos finalmente que, véase la figura, $f(x).dx$ es un elemento de área, o dicho de otra forma, es un elemento de probabilidad. $f(x)$ tendría así el sentido de densidad de probabilidad por unidad de longitud, de donde su nombre de función de densidad de probabilidad.

<u>EJEMPLOS</u>

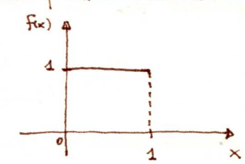
I) Sea X una v.a. con f.d. F dada por

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & 1 < x \end{cases}$$



La f.d.p., f, vendrá dada por la derivada de F en los puntos de continuidad de f, así

$$f(x) = F'(x) = \begin{cases} 0, & x < 0, \ x > 1 \\ 1, & 0 < x < 1 \end{cases}$$



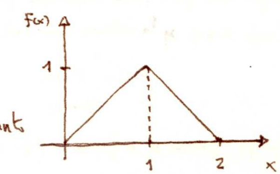
f no es continua en $x=0$ y $x=1$. Los valores $f(0)$ y $f(1)$ pueden definirse de cualquier manera, por ejemplo haciendo $f(0) = f(1) = 0$, tenemos

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

Una variable de este tipo se la conoce con el nombre de <u>v.a. uniforme</u> sobre el intervalo $[0,1]$.

II) Sea X una v.a. que tiene una f.d.p <u>triangular</u>, es decir

$$f(x) = \begin{cases} x, & 0 < x \leq 1 \\ 2-x, & 1 < x \leq 2 \\ 0, & \text{en el resto} \end{cases}$$



Observemos que $f(x)$ es, en efecto, una f.d.p por cuanto

a) $f(x)$ es no negativa a valores en R

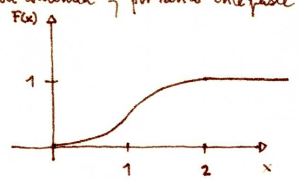
b) el área encerrada bajo su gráfica es 1. Es decir

$$\int_{-\infty}^{+\infty} f(x).dx = 1$$

La f.d. F, para la v.a. X vendrá dada por

$$F(x) = \int_{-\infty}^{x} f(t)\cdot dt$$

siendo una integral de Riemann, por cuanto f es una función continua y por tanto integrable Riemann. Así pues



$$F(x) = 0 \ , \quad \text{si } x \leq 0$$

$$F(x) = \int_0^x t\cdot dt = \frac{x^2}{2} \ , \quad \text{si } 0 < x \leq 1$$

$$F(x) = \int_0^1 t\cdot dt + \int_1^x (2-t)\cdot dt = 2x - \frac{x^2}{2} - 1 \ , \quad \text{si } 1 < x \leq 2$$

$$F(x) = 1 \ , \quad \text{si } x > 2$$

Si quisiéramos calcular la probabilidad del intervalo $].3, 1.5]$, tendremos

$$P(.3 < X \leq 1.5) = P(X \leq 1.5) - P(X \leq .3) = \left[2(1.5) - \frac{(1.5)^2}{2} - 1\right] - \left(\frac{0.3^2}{2}\right) = 0.83$$

III) Sea $k > 0$ una constante, y sea

$$f(x) = \begin{cases} kx(1-x) \ , & 0 < x < 1 \\ 0 \ , & \text{en el resto} \ . \end{cases}$$

Entonces

$$\int_0^1 f(x)\cdot dx = k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]\Big|_0^1 = k\cdot\left[\frac{1}{2} - \frac{1}{3}\right] = \frac{k}{6} \ .$$

Tendremos que $f(x)$ define una f.d.p. si $k = 6$. En este caso la f.d., F, viene dada por

$$F(x) = 0 \ , \quad \text{si } x \leq 0$$

$$F(x) = \int_0^x 6t(1-t)\, dt = 6\left(\frac{t^2}{2} - \frac{t^3}{3}\right) \ , \quad \text{si } 0 < x < 1$$

$$F(x) = 1 \ , \quad \text{si } x \geq 1$$

Así, $P(X > .3) = 1 - P(X \leq .3) = 1 - 6\left[\frac{0.09}{2} - \frac{0.027}{3}\right] = 1 - 6\,[0.045 - 0.009] = 1 - 6 \times 0.036 = 0.784$

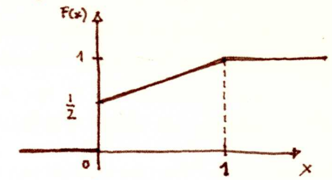
NOTA.- Debemos señalar que los dos tipos de variables aleatorias estudiadas, continuas y discretas, constituyen sólo una pequeña parte de la clase de todas las v.a.'s. Estos dos tipos, sin embargo, contienen prácticamente todas las variables aleatorias que puedan surgirnos en la práctica. Señalemos también aunque sin demostrarlo que cada función de distribución admite una descomposición de la forma

$$F(x) = a\cdot F_d(x) + (1-a)\cdot F_c(x)$$

donde F_d y F_c son ambas funciones de distribución. Concretamente F_d es la f.d. de una v.a. discreta, mientras que F_c es continua (no necesariamente absolutamente continua). De hecho se demuestra que a su vez, F_c, admite una descomposición posterior. Veamos un ejemplo que ilustre lo dicho y ponga de manifiesto una v.a., X, que no es ni discreta, ni continua.

IV) Sea X una v.a. con f.d., F, dada por

$$F(x) = \begin{cases} 0 \ , & x < 0 \\ \frac{1}{2} \ , & x = 0 \\ \frac{1}{2} + \frac{x}{2} \ , & 0 < x < 1 \\ 1 \ , & 1 \leq x \ . \end{cases}$$



Obsérvese que F tiene un salto en $x = 0$ y F es continua (de hecho, absolutamente continua) en el intervalo $(0,1)$. F es la f.d. de una v.a., X, que no es ni discreta, ni continua. De acuerdo con lo anterior F puede descomponerse como sigue,

$$F(x) = \frac{1}{2} F_d(x) + \frac{1}{2} F_c(x)$$

donde

$$F_d(x) = \begin{cases} 0 \ , & x < 0 \\ 1 \ , & x \geq 0 \end{cases} \qquad y \qquad F_c(x) = \begin{cases} 0 \ , & x \leq 0 \\ x \ , & 0 < x < 1 \\ 1 \ , & 1 \leq x \ . \end{cases}$$

F_d es la f.d. de una v.a. degenerada en el punto $x = 0$, llamada así por tener toda la probabilidad concentrada en un solo punto. Se trata de una variable aleatoria que toma un solo valor, es decir, es constante. F_c es la f.d. de una v.a. uniforme en el intervalo $(0,1)$.

ALGUNAS VARIABLES ALEATORIAS CONTINUAS

I) NORMAL (ó GAUSSIANA).

Se trata de una v.a. cuya f.d.p. viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad x\in\mathbb{R}.$$

Con estas características decimos que X se distribuye normal (μ,σ^2), y lo denotaremos mediante $N(\mu,\sigma^2)$, donde μ y σ^2 son llamados parámetros de la distribución, cuyo significado comprenderemos más adelante, en cualquier caso $\mu\in\mathbb{R}$ y $\sigma>0$. En el caso en que $\mu=0$ y $\sigma=1$, la distribución normal que se obtiene se la conoce con el nombre de distribución normal Standard o normal tipificada, y se la designa mediante $N(0,1)$. La función f es una f.d.p por cuanto:

a) Es continua y por tanto medible y además es finita

b) f es no negativa

c) $I=\int_{-\infty}^{+\infty} f(x)\,dx = 1.$

a) y b) son inmediatas de comprobar. Para demostrar c) lo haremos comprobando que $I^2=1$. En efecto

$$I^2 = \left[\int_{-\infty}^{+\infty} f(x)\,dx\right]^2 = \int_{-\infty}^{+\infty} f(x)dx \cdot \int_{-\infty}^{+\infty} f(y)dy = \frac{1}{2\pi\sigma}\int_{-\infty}^{+\infty}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx\,.$$

$$\cdot\frac{1}{\sigma}\int_{-\infty}^{+\infty}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]dy = \frac{1}{2\pi}\cdot\frac{1}{\sigma}\int_{-\infty}^{+\infty} e^{-z^2/2}\,\sigma\, dz\,.\,\frac{1}{\sigma}\int_{-\infty}^{+\infty} e^{-v^2/2}\,.\,\sigma\,.\,dv$$

para llegar a esta última igualdad hemos efectuado el cambio $\frac{x-\mu}{\sigma}=z$ y $\frac{y-\mu}{\sigma}=v$.

Por tanto

$$I^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(z^2+v^2)/2}\,dz\,dv = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} e^{-r^2/2}\, r\, dr\, d\theta$$

obtenida esta última igualdad mediante el cambio a coordenadas polares, a saber

$$r^2 = z^2+v^2 \qquad \Big\| \qquad z = r.\operatorname{sen}\theta$$
$$\theta = \operatorname{arctg}\frac{z}{v} \qquad \Big\| \qquad v = r.\cos\theta$$

siendo el Jacobiano de la transformación, r.

En consecuencia

$$I^2 = \frac{1}{2\pi}\int_0^{\infty} e^{-r^2/2}\, r\,dr\,.\int_0^{2\pi} d\theta = \int_0^{\infty} e^{-r^2/2}\, r\, dr = -e^{-r^2/2}\Big|_0^{\infty} = 1.$$

Entre otras propiedades de interés, la función $f(x)$ es simétrica respecto a μ, por cuanto para

$x_1 = \mu + k$

$x_2 = \mu - k$

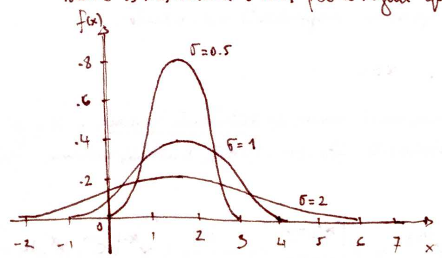
tenemos

$$\left.\begin{aligned} f(x_1) &= \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{k^2}{2\sigma^2}\right] \\ f(x_2) &= \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(-k)^2}{2\sigma^2}\right]\end{aligned}\right\} \quad f(x_1)=f(x_2)$$

Se comprueba fácilmente que la función alcanza un máximo para $x=\mu$ y que dicho máximo vale

$$f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma}$$

lo que supone que la forma de la función se altera con los distintos valores de σ, más concretamente es más estirada o más plana según que σ sea menor o mayor.



Gráficas de ~~las~~ f.d.p. de normales $N(1.5,\sigma^2)$ con $\sigma=0.5$, $\sigma=1$ y $\sigma=2$ respectivamente.

El nombre de normal le viene a esta variable aleatoria del hecho de ser la ~~que describe~~ la mayoría de los fenómenos aleatorios que se encuentran en el campo experimental. No obstante, su importancia radica en las propiedades límite de sucesiones de variables aleatorias cualesquiera, que en su momento estudiaremos.

La correspondiente función de distribución viene dada por

$$F(x) = \int_{-\infty}^{x} f(t).dt$$

pero hay que señalar que $f(t)$ carece de primitiva y por tanto los distintos valores de $F(x)$ se obtienen mediante unas tablas adecuadas, no obstante, las tablas son únicas, concretamente las de la normal tipificada, $N(0,1)$, y no una para cada par de valores (μ,σ^2), por cuanto resultaría imposible cubrir todo el conjunto de posibles valores de dicho par de parámetros. Esto no supone ningún serio problema a la hora de obtener valores de $F(x)$ para una $N(\mu,\sigma^2)$. En efecto, si nosotros queremos calcular, para $N(\mu,\sigma^2)$,

$$F(x)=\int_{-\infty}^{x} f(t)\cdot dt = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]dt$$

efectuando el cambio $z=\frac{t-\mu}{\sigma}$, tendremos que los límites de integración varían, convirtiéndose en $]-\infty,x']$, con $x'=\frac{x-\mu}{\sigma}$, quedando la integral convertida en

$$\phi(x')=\phi\left(\frac{x-\mu}{\sigma}\right)=\int_{-\infty}^{x'}\frac{1}{\sqrt{2\pi}}\exp\left[-z^2/2\right]dz$$

en donde $\phi(x)$ representa la función de distribución de la $N(0,1)$ cuyos valores aparecen en la tabla.

II) GAMMA.

Se trata de una variable aleatoria cuya función de densidad de probabilidad viene dada por

$$f(x)=\begin{cases}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\cdot x^{\alpha-1}e^{-x/\beta}, & x>0\\ 0, & x\le 0\end{cases}\qquad \alpha>0,\ \beta>0.$$

donde $\Gamma(\alpha)$ es el valor de la función Gamma correspondiente a α, es decir

$$\Gamma(\alpha)=\int_0^{\infty} y^{\alpha-1}\cdot e^{-y}\,dy,\quad \alpha>0$$

La distribución de la v.a. X se la conoce con el nombre de distribución Gamma y α y β son los parámetros de la distribución. Obviamente $f(x)\ge 0$, $\forall x\in\mathbb{R}$, es además continua y finita y además

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\Gamma(\alpha)\cdot\beta^{\alpha}}\int_{-\infty}^{+\infty} x^{\alpha-1}\cdot e^{-x/\beta}dx = \begin{bmatrix} x/\beta=y\\ x=\beta y\\ dx=\beta dy\end{bmatrix} = \frac{1}{\Gamma(\alpha)\cdot\beta^{\alpha}}\int_0^{+\infty} y^{\alpha-1}\cdot e^{-y}\cdot\beta^{\alpha}\cdot dy =$$

$$=\frac{1}{\Gamma(\alpha)}\cdot\Gamma(\alpha)=1.$$

Los distintos valores ~~valores~~ de la f.d., F, aparecen también tabulados, presentándose las tablas para diferentes valores de los parámetros α,β. Ello da idea de la dificultad que puede suponer obtener valores de F para cualesquiera α y β.

Respecto de la función Gamma debemos añadir que sus valores son sencillos de obtener

si α es entero o de la forma $n+\frac{1}{2}$, n natural. En efecto, mediante integraciones sucesivas por partes, puede verse que

$$\Gamma(\alpha)=(\alpha-1)\cdot\Gamma(\alpha-1)=(\alpha-1)(\alpha-2)\cdot\Gamma(\alpha-2)$$

de manera que para α entero,

$$\Gamma(\alpha)=(\alpha-1)(\alpha-2)\cdot\ldots\cdot 2\cdot\Gamma(1)$$

pero

$$\Gamma(1)=\int_0^{\infty}e^{-y}dy=1$$

y por tanto

$$\Gamma(\alpha)=(\alpha-1)!.$$

Para el caso en que $\alpha=n+\frac{1}{2}$, obtengamos $\Gamma(\frac{1}{2})$.

$$\Gamma\left(\tfrac{1}{2}\right)=\int_0^{+\infty}e^{-y}\cdot y^{-\frac{1}{2}}dy$$

haciendo $y^{\frac{1}{2}}=\frac{t}{\sqrt{2}}$, tendremos $y=\frac{t^2}{2}$; $dy=t\,dt$, $t\in(0,+\infty)$, y de aquí

$$\Gamma\left(\tfrac{1}{2}\right)=\int_0^{+\infty}\frac{\sqrt{2}}{t}e^{-t^2/2}\,t\,dt=\sqrt{2}\int_0^{+\infty}e^{-t^2/2}dt$$

pero la integral de la última parte de la igualdad dividida por $\sqrt{2\pi}$ es la mitad del área que cubre la f.d.p. de la $N(0,1)$ y dicha área vale $\frac{1}{2}$; así pues

$$\Gamma\left(\tfrac{1}{2}\right)=\frac{\sqrt{2}\cdot\sqrt{2\pi}}{2}=\sqrt{\pi}$$

lo que suele expresarse, en ocasiones, recordando que $\Gamma(\alpha)=(\alpha-1)!$ para α entero y haciendo extensiva esta notación para cualquier α, de la forma

$$\Gamma\left(\tfrac{1}{2}\right)=\left(-\tfrac{1}{2}\right)!=\sqrt{\pi}.$$

Es muy fácil, a partir de aquí, obtener

$$\Gamma\left(\tfrac{3}{2}\right)=\tfrac{1}{2}\Gamma\left(\tfrac{1}{2}\right)=\frac{\sqrt{\pi}}{2},\ \ldots\ .$$

Hay dos clases de distribuciones Gammas, que por su importancia y uso han adquirido nombre propio y son:

II.a) JI-CUADRADO

Se trata de un caso particular de la Gamma en la que $\alpha = r/2$, r entero no negativo, $\beta = 2$. La f.d.p. tiene la forma

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\frac{1}{2}r)\, 2^{r/2}} \cdot e^{-x/2} \cdot x^{(r/2)-1} & x>0 \\ 0 & x \le 0 \end{cases} \qquad r \geq 0 \text{, entero}$$

a esta f.d.p. se la designa mediante χ^2_r, y al parámetro r es conocido como el <u>número de grados de libertad</u> de la distribución, cuyo significado veremos más tarde.

II.b) EXPONENCIAL NEGATIVA

Se obtiene de la Gamma para $\alpha = 1$, $\beta = 1/\lambda$ y su f.d.p. es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x>0 \\ 0 & , x \le 0 \end{cases} \qquad \lambda > 0.$$

La distribución Gamma y sus derivadas aparecen en estadística cuando se estudian, entre otros, problemas de tiempo de espera. En particular la exponencial negativa está ligada a la distribución de Poisson, para el caso de desintegraciones de material radiactivo, por cuanto es la distribución de probabilidad de la v.a. que estudia el tiempo transcurrido entre dos impactos consecutivos (Verlo como ejercicio).

III) BETA

Es una v.a. cuya f.d.p. es de la forma

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} & 0<x<1 \\ 0 \text{ , en el resto} & \alpha>0,\ \beta>0. \end{cases}$$

f es una función no negativa, continua, excepto en ~~los~~ punto 1, finita y además normada, pues

$$\Gamma(\alpha)\cdot\Gamma(\beta) = \left(\int_0^\infty x^{\alpha-1} \cdot e^{-x}\, dx\right) \cdot \left(\int_0^\infty y^{\beta-1} \cdot e^{-y}\, dy\right) = \int_0^\infty \int_0^\infty x^{\alpha-1} y^{\beta-1} \cdot e^{-(x+y)}\, dx\, dy$$

haciendo el cambio $u = x/(x+y)$, tendremos

$$x = \frac{uy}{1-u}, \quad dx = \frac{y\,du}{1-u^2}, \quad u \in (0,1) \quad y \quad x+y = \frac{y}{1-u}$$

lo que nos lleva a

$$\Gamma(\alpha)\cdot\Gamma(\beta) = \int_0^\infty \int_0^1 \frac{u^{\alpha-1}}{(1-u)^{\alpha-1}} \cdot y^{\alpha-1} \cdot y^{\beta-1} \cdot e^{-y/(1-u)}\ y\ \frac{du}{(1-u)^2}\, dy =$$

$$= \int_0^\infty \int_0^1 \frac{u^{\alpha-1}}{(1-u)^{\alpha+1}}\, y^{\alpha+\beta-1} \cdot e^{-y/(1-u)}\, du\, dy$$

haciendo ahora $y/(1-u) = v$, tendremos $y = v(1-u)$, $dy = (1-u)dv$, $v \in (0,+\infty)$. Entonces la integral es

$$\Gamma(\alpha)\cdot\Gamma(\beta) = \int_0^\infty \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} v^{\alpha+\beta-1} \cdot e^{-v}\, du\, dv =$$

$$= \int_0^\infty v^{\alpha+\beta-1} \cdot e^{-v}\, dv \cdot \int_0^1 u^{\alpha-1} (1-u)^{\beta-1}\, du = \Gamma(\alpha+\beta) \int_0^1 u^{\alpha-1} (1-u)^{\beta-1}\, du$$

es decir que

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} = 1.$$

La distribución de probabilidad de X se conoce con el nombre de <u>distribución beta</u>, y deriva del hecho de que en Análisis la integral $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$, $\alpha,\beta>0$ es conocida como la función beta. A α y β se les conoce como los <u>parámetros de distribución</u>. Para la obtención de los valores de $F(x)$ vale todo lo dicho en el correspondiente párrafo de la distribución Gamma.

<u>OBSERVACIÓN</u>. Para $\alpha = \beta = 1$, teniendo en cuenta que $\Gamma(1) = \Gamma(2) = 1$, la f.d.p. adquiere la forma

$$f(x) = \begin{cases} 1 & , 0<x<1 \\ 0 & , \text{en el resto} \end{cases}$$

que es la f.d.p. de la v.a. uniforme en el intervalo $[0,1]$.

IV) CAUCHY

La f.d.p. viene dada por

$$f(x) = \frac{\sigma}{\pi} \cdot \frac{1}{\sigma^2 + (x-\mu)^2}, \quad x \in R,\ \mu \in R,\ \sigma > 0$$

Se trata, efectivamente, de una f.d.p. por cumplir todas las propiedades exigibles para ello. En particular (7)

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{\sigma}{\pi}\int_{-\infty}^{+\infty} \frac{1}{\sigma^2+(x-\mu)^2}\,dx = \frac{1}{\sigma\pi}\int_{-\infty}^{+\infty}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$$

haciendo el cambio $y = \frac{x-\mu}{\sigma}$, $dx = \sigma\,dy$, y tendremos

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{dy}{1+y^2} = \frac{1}{\pi}\cdot\left[\operatorname{arctg} y\right]_{-\infty}^{\infty} = 1$$

Una distribución de probabilidad de estas características se la conoce con el nombre de distribución de Cauchy y μ y σ son los parámetros de la distribución.

La correspondiente gráfica de la f.d.p. se asemeja mucho a la de la normal con iguales valores de μ y σ, con la diferencia que aquella tiene las colas mas elevadas.

V) LOGNORMAL

Se trata de una variable aleatoria con f.d.p. dada por

$$f(x) = \begin{cases} \dfrac{1}{x\beta\sqrt{2\pi}}\exp\left[-\dfrac{(\log x-\log\alpha)^2}{2\beta^2}\right], & x>0 \\ 0, & x\le 0 \end{cases}, \quad \text{con } \alpha,\beta>0.$$

La función $f(x)$ es continua, no negativa, finita y además

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{1}{\beta\sqrt{2\pi}}\int_{0}^{\infty}\frac{1}{x}\exp\left[-\frac{(\log x-\log\alpha)^2}{2\beta^2}\right]dx$$

haciendo el cambio $x = e^y$, $\log x = y$, $dx = e^y dy$, $y\in(-\infty,+\infty)$, tendremos

$$\int_{-\infty}^{+\infty} f(x)\,dx = \frac{1}{\beta\sqrt{2\pi}}\int_{-\infty}^{+\infty}\frac{1}{e^y}\exp\left[-\frac{(y-\log\alpha)^2}{2\beta^2}\right]e^y dy =$$

$$= \frac{1}{\beta}\left(\sqrt{2\pi}\right)^{-1}\int_{-\infty}^{+\infty}\exp\left[-\frac{(y-\log\alpha)^2}{2\beta^2}\right]dy$$

pero observemos que se trata de la integral de la f.d.p. de una normal, $N(\log\alpha,\beta^2)$ y por tanto dicha integral debe de ser la unidad. Precisamente el nombre de lognormal, con que se conoce a esta distribución, deriva del hecho expuesto, pues decir que X se distribuye como una lognormal es lo mismo que decir que $Y=\log X$ lo hace como una normal. Las constantes α y β se las conoce como los parámetros de la distribución.

EJERCICIOS (8)

I). Si X es una v.a. distribuida $N(3, 0.25)$, obtener, mediante las tablas, las siguientes probabilidades

a) $P(X<-1)$, b) $P(X>2.5)$, c) $P(-0.5<X<1.3)$.

Las tablas son de una normal tipificada y debemos por tanto efectuar el correspondiente cambio para poder utilizarlas. Tendremos pues:

a) $P(X<-1) = P\left(\frac{X-3}{0.5} < \frac{-1-3}{0.5}\right) = P(Z<-8) = \Phi(-8) \simeq 0.$

b) $P(X>2.5) = P\left(\frac{X-3}{0.5} > \frac{2.5-3}{0.5}\right) = P(Z>-1) = 1-P(Z\le -1) = 1-\Phi(-1) = 0.841345$

c) $P(-0.5<X<1.3) = P\left(\frac{-0.5-3}{0.5} < \frac{X-3}{0.5} < \frac{1.3-3}{0.5}\right) = P(-7<Z<-3.4) = \Phi(-3.4)-\Phi(-7) =$
$= 0.000337.$

II) Si una v.a. X se distribuye $N(\mu,\sigma^2)$ encontrar, en términos de μ y de σ, el valor de c tal que $P(X<c) = 2-9.P(X>c)$.

$$P(X<c) = 2-9.P(X>c) = 2-9[1-P(X<c)] = 2-9+9.P(X<c)$$

$$9.P(X<c)-P(X<c) = 7 \rightarrow 8.P(X<c)=7 \rightarrow P(X<c) = 0.875$$

luego

$$P\left(\frac{X-\mu}{\sigma} < \frac{c-\mu}{\sigma}\right) = 0.875 \quad \text{es decir} \quad \Phi\left(\frac{c-\mu}{\sigma}\right) = 0.875$$

Buscando en las tablas, encontramos

$$\Phi(1.15) = 0.875 \rightarrow \frac{c-\mu}{\sigma} = 1.15 \rightarrow \boxed{c = 1.15\sigma+\mu}$$

III) Un cierto proceso de manufactura produce bombillas, cuya duración, en horas, es una v.a. X distribuida como $N(2000, 200)$. Se supone que una bombilla es defectuosa si su duración es menor de 1800 horas. Si se compruetan 25 bombillas, ¿cual es la probabilidad de que al menos 15 de ellas sean defectuosas?

En el supuesto de que las distintas comprobaciones son independientes, cosa que cabe esperar así sea, el número de bombillas defectuosas es una distribución binomial, $B(25,p)$, donde p se obtiene a partir de la $N(2000,200)$ mediante

$$p = P(X<1.800) = P\left(\frac{X-2000}{\sqrt{200}} < \frac{1800-2000}{\sqrt{200}}\right) = P(Z<-\sqrt{200}) = \Phi(-\sqrt{200}) \simeq 0$$

y por tanto la probabilidad pedida será cero.

IV) Un proceso de fabricación produce bolas de rodamiento de $\frac{1}{2}$ pulgada que se suponen buenas si su diametro cae en el intervalo 0.5 ± 0.0006 y defectuosas en cualquier otro caso. La producción de un día se examina y se encuentra que la distribución de los diametros es aproximadamente normal con media $\mu = 0.5007$ pulgadas y $\sigma = 0.0005$ pulgadas. a) Calcular la proporción de bolas defectuosas. b) Si ajustes menores del proceso permiten cambiar el diametro de las bolas, pero no su desviación típica, indicar cuales serían los cambios aconsejables y calcular entonces la fracción de bolas defectuosas.

a) Las bolas defectuosas son aquellas cuyos diametros caen fuera del intervalo 0.5 ± 0.0006, entonces la proporción buscada, p, vendrá dada por (X representa el diametro)

$$p = P(X < 0.4994) + P(X > 0.5006) = P\left(\frac{X - 0.5007}{0.0005} < \frac{0.4994 - 0.5007}{0.0005}\right) +$$

$$+ P\left(\frac{X - 0.5007}{0.0005} > \frac{0.5006 - 0.5007}{0.0005}\right) = P\left(Z < -\frac{13}{5}\right) + P\left(Z > -\frac{1}{5}\right) = \phi(-2.6) + \phi(-0.2) =$$

$$= 0.004661 + 0.579260 = 0.583921$$

b) Disminuyendo el diametro medio hasta 0.5 pulgadas, tendremos

$$p = P\left(Z < -\frac{6}{5}\right) + P\left(Z > \frac{6}{5}\right) = 2 . P(Z < -1.2) = 2 . (0.116070) = 0.232140 .$$

Cualquier otro diametro medio, puede comprobarse que aumenta la proporción de piezas defectuosas.

V) Supongamos que el consumo medio de agua/mes de los residentes de cierta comunidad sigue una distribución lognormal con $\alpha = 10^4$ pies cúbicos y $\beta = 10^3$ pies cubicos por mes. Calcular la proporción de residentes que consumen más de 15×10^3 pies cúbicos mensualmente.

$$P(X > 15.10^3) = 1 - P(X \leq 15.10^3)$$

Sabemos que si X es lognormal con parámetros α y β, entonces $Y = \log X$ es una distribución $N(\log\alpha, \beta^2)$. Podemos replantear el problema en terminos de una normal calculando la proporción de residentes que consumen más de $\log(15 \times 10^3)$ pies cúbicos de agua mensualmente. Así pues

$$P(Y > \log(15 \times 10^3)) = 1 - P(Y \leq \log(15 \times 10^3)) = 1 - P\left(\frac{Y - \log 10^4}{10^3} \leq \frac{\log(15 \times 10^3) - \log(10^4)}{10^3}\right)$$

$$= 1 - P\left(Z \leq \frac{\log 1.5}{10^3}\right) \simeq 1 - \phi(0) = 1 - 0.5 = \underline{0.5}$$

VI) <u>Relación entre las distribuciones de POISSON y EXPONENCIAL NEGATIVA.</u>

Consideramos una variable aleatoria X, que denota el número de impactos, procedente de la desintegración de cierto material radiactivo, que alcanzan un contador Geiger en un intervalo de tiempo t. Ya sabemos que se trata de una v.a. con distribución de Poisson con parámetro λt, y por tanto

$$P(X = x) = \frac{(\lambda t)^x}{x!} . e^{-\lambda t}$$

Sea ahora T, una variable que denota el tiempo transcurrido entre dos impactos consecutivos. El suceso $T > t$ representa el hecho de que no se hayan producido impactos durante el intervalo de tiempo t, y por tanto

$$P(T > t) = P(X = 0) = e^{-\lambda t}$$

pero por otra parte, si designamos mediante $f(t)$, la f.d.p. de la v.a. T, tendremos

$$P(T > t) = 1 - P(T \leq t) = 1 - \int_0^t f(x) . dx = 1 - \int_{-\infty}^t f(x) dx = e^{-\lambda t}$$

derivando en ambos miembros de la igualdad, llegamos a

$$f(t) = \lambda e^{-\lambda t} \quad , t \geq 0$$

que es la f.d.p. de una exponencial negativa con parámetro λ.

VII) Sea X una v.a. distribuida como $U(-\alpha, \alpha)$. Determinar el valor del parámetro α para que sea cierto lo siguiente:

a) $P(-1 < X < 2) = 0.75$, b) $P(|x| < 1) = P(|x| > 2)$

Recordemos que la f.d.p. de una $U(-\alpha, \alpha)$ viene dada por

$$f(x) = \begin{cases} \frac{1}{2\alpha} & , -\alpha \leq x \leq \alpha \\ 0 & , \text{en el resto.} \end{cases}$$

Por tanto para a)

$$P(-1 < X < 2) = P(X < 2) - P(X \leq -1) = \int_{-\alpha}^2 f(x) dx - \int_{-\alpha}^{-1} f(x) dx = \int_{-1}^2 f(x) dx = \frac{1}{2\alpha} \int_{-1}^2 dx$$

$$= \frac{2 - (-1)}{2\alpha} = \frac{3}{2\alpha} = 0.75 \rightarrow \underline{\alpha = 2}$$

para b) tenemos

$$P(|X|<1)=P(-1<X<1)=\int_{-1}^{1} f(x)\,dx = 1/\alpha$$

$$P(|X|>2)=P(-\alpha<X<-2)+P(2<X<\alpha)=\frac{1}{2\alpha}\int_{-\alpha}^{-2} dx+\frac{1}{2\alpha}\int_{2}^{\alpha} dx=\frac{\alpha-2}{2\alpha}+\frac{\alpha-2}{2\alpha}=$$

$$=\frac{\alpha-2}{\alpha}$$

entonces

$$\frac{\alpha-2}{\alpha}=\frac{1}{\alpha} \longrightarrow \alpha-2=1 \longrightarrow \underline{\alpha=3}$$

Capítulo 6

Vectores aleatorios

VECTORES ALEATORIOS

Hagamos en primer lugar algunas consideraciones respecto de β^k, tribu de Borel en R^k. Recordemos que la tribu de Borel de un espacio topológico viene engendrada por el sistema de abiertos (ó cerrados) del espacio en cuestión. No obstante, hay ocasiones en que, si el espacio lo permite, podemos fijar nuestra atención en subfamilias que engendran igualmente dicha tribu. Así por ejemplo, recordemos que en el caso de R, β podía venir engendrada por intervalos de la forma $]-\infty, a]$, $a \in R$ o bien por cualquier otra familia de intervalos. Igual ocurre en R^k, de manera que podemos engendrar β^k a partir de cualquier familia de intervalos, por ejemplo $]-\infty, \vec{x}]$ con $\vec{x} = (x_1, \ldots, x_k)$. Pasemos ahora a la definición de vector aleatorio.

<u>Definición</u>. Sea $(S, \mathcal{A}, P)$ un espacio de probabilidad y sea el espacio medible (R^k, β^k).

Definimos entre ambos una aplicación medible $\vec{X}$, a la que denominaremos vector aleatorio.

El concepto de vector aleatorio no es más que una generalización inmediata del concepto de variable aleatoria que ya hemos estudiado. De la definición que acabamos de dar se deriva de inmediato la siguiente propiedad

<u>Propiedad</u>.- Una aplicación $\vec{X}$ de $(S, \mathcal{A}, P)$ en (R^k, β^k) es un vector aleatorio si y sólo si cada una de sus componentes X_i, $i = 1, \ldots, k$ es una variable aleatoria.

Demostración.- Sea $\vec{X}$ un vector aleatorio y definamos sobre R^k la función g_i de la forma

$$g_i(x_1 \ldots x_k) = x_i$$

g_i es la proyección y como tal es continua y por tanto medible, en consecuencia

$$X_i = g_i(\vec{X})$$

es una variable aleatoria por cuanto resulta de la composición de dos aplicaciones medibles.

Sean ahora las X_i, $i = 1, \ldots, k$, variables aleatorias. Para demostrar que $\vec{X}$ es un vector aleatorio hemos de comprobar que

$$\vec{X}^{-1}(B) \in \mathcal{A} \quad , \quad B \in \beta^k$$

pero teniendo en cuenta las consideraciones iniciales basta con tomar B en una de las familias que engendran β^k, por ejemplo $B =]-\infty, \vec{x}]$, entonces

$$\vec{X}^{-1}(B) = \{s \in S ; \vec{X}(s) \in B\} = \{s \in S ; X_i(s) \in]-\infty, x_i] , i = 1, \ldots, k\} =$$
$$= \bigcap_{i=1}^{k} \{s \in S ; X_i(s) \in]-\infty, x_i]\} = \bigcap_{i=1}^{k} \{X_i^{-1}(]-\infty, x_i])\} \in \mathcal{A}$$

y este último conjunto es de $\mathcal{A}$ por serlo cada $X_i^{-1}(]-\infty, x_i])$, ya que las X_i son variables aleatorias.

<u>Probabilidad inducida</u>. A partir de P y mediante $\vec{X}$ podemos definir sobre el espacio medible (R^k, β^k) una medida de probabilidad. En efecto, para cada $B \in \beta^k$

$$P'(B) = P(\vec{X}^{-1}(B)).$$

Análogamente a como lo hacíamos en el caso unidimensional podemos ahora definir una <u>Función de distribución de probabilidad</u> a partir de esta probabilidad inducida tomando una clase especial de conjuntos de Borel. En efecto, para $B =]-\infty, \vec{x}]$, tenemos

$$F(\vec{x}) = F(x_1, \ldots, x_k) = P'(]-\infty, \vec{x}]) = P(\vec{X}^{-1}(]-\infty, \vec{x}])) = P(\vec{X} \leq \vec{x}).$$

De esta definición se derivan, de inmediato, las <u>siguientes propiedades</u>:

1) La función $F(\vec{x})$ es continua por la derecha y no decreciente en cada una de las componentes de $\vec{x}$.
La comprobación es sencilla y similar a la presentada en el caso unidimensional.

2) Sean $a = (a_1, \ldots, a_k)$ y $b = (b_1, \ldots, b_k)$, tales que $a \leq b$, es decir $a_i \leq b_i$, $i = 1, \ldots, k$. El subconjunto de R^k, $\{a \leq b\}$ representa un rectángulo k-dimensional con extremos superior e inferior, b y a respectivamente.

Definamos ahora el siguiente operador $\Delta_{b_i - a_i}$ que actuará sobre cualquier función definida en R^k de la siguiente forma

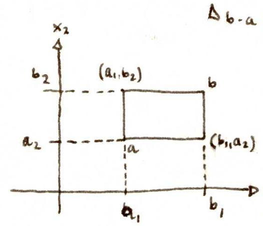
$$\Delta_{b_i - a_i}\{F(a_1, \ldots, a_i, \ldots a_k)\} = F(a_1, \ldots, b_i, \ldots a_k) - F(a_1, \ldots, a_i, \ldots, a_k).$$

Introducimos también el operador Δ_{b-a}, de la siguiente forma:

$$\Delta_{b-a}\{F(a)\} = \Delta_{b_1 - a_1} \Delta_{b_2 - a_2} \cdots \Delta_{b_k - a_k}\{F(a)\}$$

Así, por ejemplo, para $k = 2$, tendríamos

$$\Delta_{b-a}\{F(a)\} = \Delta_{b_1-a_1} \cdot \Delta_{b_2-a_2}\{F(a_1, a_2)\} = \Delta_{b_1-a_1}\{F(a_1, b_2) - F(a_1, a_2)\}$$
$$= \Delta_{b_1-a_1}\{F(a_1, b_2)\} - \Delta_{b_1-a_1}\{F(a_1, a_2)\} = F(b_1, b_2) - F(a_1, b_2)$$
$$- (F(b_1, a_2) - F(a_1, a_2)) = [F(b_1, b_2) - F(b_1, a_2)] - [F(a_1, b_2) -$$
$$- F(a_1, a_2)]$$



y observando la figura y teniendo en cuenta que $F(\vec{x}) = P(\vec{X} \leq \vec{x})$ llegamos a que $\Delta_{b-a}\{F(a)\} = P(\vec{X} \in]a, b]) \geq 0$

Esta es precisamente la segunda propiedad de $F(\vec{x})$, a saber:

Si $a \leq b$, entonces $\Delta_{b-a}\{F(a)\} \geq 0$.

3) $\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_k) = 0$, para cada x_i.

Se comprueba fácilmente observando que los intervalos de la forma $\{\vec{t} ; \vec{t} \leq \vec{x}\}$ tienden al vacío a medida que $x_i \to -\infty$.

4) $\lim\limits_{(x_1 \ldots x_k) \to (+\infty, \ldots, +\infty)} F(x_1, \ldots, x_k) = 1$. Por cuanto en el límite se trata de obtener la probabilidad $P(\vec{X} \in R^k)$ que obviamente es la unidad.

De la misma forma que para las variables aleatorias, una función de distribución de probabilidad podía utilizarse para inducir una probabilidad sobre los borelianos de R, y llevar así la equivalencia entre ambos conceptos (medida de probabilidad inducida y función de distribución de probabilidad), también aquí podemos actuar de forma semejante pero, eso sí, tomando algunas precauciones impuestas por los cambios que se han producido en la situación. En otras palabras, la generalización no puede hacerse directamente sino que hay que generalizar adecuadamente el concepto de función de distribución de probabilidad. Veámoslo:

<u>Definición</u>: Sea $F(\vec{x})$, $\vec{x} = (x_1 \ldots x_k)$ una función definida sobre R^k. Diremos que $F(\vec{x})$ es una <u>función de distribución de probabilidad</u> si verifica las siguientes condiciones:

c.1) $\Delta_{b-a}\{F(a)\} \geq 0$, si $a \leq b$

c.2) F es creciente y continua por la derecha en cada una de las componentes de $\vec{x}$.

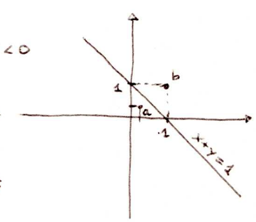
c.3) $\lim\limits_{x_i \to -\infty} F(x_1 \ldots x_k) = 0$ para cada x_i.

c.4) $\lim\limits_{(x_1 \ldots x_k) \to (+\infty, \ldots, +\infty)} F(x_1 \ldots x_k) = 1$

Obsérvese que esta definición coincide con la dada para una función de distribución de probabilidad. Sin embargo ahora hemos debido especificar la condición c.1) por cuanto la monotonía y la continuidad no bastan para que $F(\vec{x})$ pueda utilizarse en la definición de una probabilidad sobre β^k de R^k. Consideremos el ejemplo que sigue para mejor comprender lo dicho.

<u>Ejemplo 1</u>.- Sea $F(x,y)$ definida como sigue

$$F(x,y) = \begin{cases} 0, & \text{si } x<0, \text{ o } x+y<1, \text{ o } y<0 \\ 1, & \text{en el resto} \end{cases}$$



Esta función satisface las condiciones C2, C3 y C4 de la anterior definición. Sean $a = (\frac{1}{3}, \frac{1}{3})$, $b = (1,1)$, entonces

$$\Delta_{b-a}\{F(a)\} = F(1,1) + F(\tfrac{1}{3}, \tfrac{1}{3}) - F(1, \tfrac{1}{3}) - F(\tfrac{1}{3}, 1) =$$

$$= 1 + 0 - 1 - 1 = -1 < 0$$

por tanto no podemos utilizar $F(\vec{x})$ para la definición de una medida de probabilidad sobre β^k de R^k sobre la base de hacerlo mediante

$$P'(]a,b]) = \Delta_{b-a}\{F(a)\},$$

por cuanto nos encontraríamos con que $P'(]a,b]) = -1 < 0$, para $a = (\frac{1}{3}, \frac{1}{3})$, $b = (1,1)$.

A partir de una función de distribución de probabilidad, como la definida, podemos definir una medida de probabilidad sobre β^k de R^k, haciéndolo en primera instancia sobre la familia de los intervalos de la forma $\{\vec{x} \,;\, a < \vec{x} \leq b\} =]a,b]$, mediante

$$P(]a,b]) = \Delta_{b-a}\{F(a)\}.$$

A continuación, mediante el teorema de extensión, podemos hacer que la medida P trabaje sobre β^k. En definitiva esto viene a poner de manifiesto la equivalencia entre ambos conceptos a la hora de conocer las características de la distribución de probabilidad del vector aleatorio $\vec{X}$.

Existen, al igual que ocurría con las variables aleatorias, diversos tipos de vectores aleatorios. Nosotros tan sólo nos ocuparemos de los dos más interesantes y que son los equivalentes a las variables aleatorias estudiadas anteriormente.

<u>Tipos de vectores aleatorios</u>. Como hacíamos en el caso unidimensional podemos ahora introducir una nueva función, a saber

$$f(\vec{x}) = f(x_1, \ldots x_k) = P(\vec{X} = \vec{x}) = P(X_1 = x_1, \ldots, X_k = x_k).$$

Sea $A = \{\vec{x} \in R^k;\ f(\vec{x}) > 0\}$, si A es tal que

$$\sum_{\vec{x} \in A} f(\vec{x}) = 1$$

decimos entonces que $\vec{X}$ es un <u>vector aleatorio discreto</u>. Obsérvese que tal y como ocurría en el caso unidimensional A es a lo sumo numerable. A la función f se le denomina <u>función de cuantía</u> y guarda con F la siguiente relación:

$$F(\vec{x}) = \sum_{\vec{t} \in A \cap]-\infty, \vec{x}]} f(\vec{t})$$

lo que supone que la $F(\vec{x})$ es el equivalente k-dimensional a una función en escalera en R.

Un razonamiento similar al utilizado en el caso unidimensional, que no reiteraremos aquí, nos lleva a introducir el concepto de <u>vector aleatorio continuo</u>, como aquel para el que existe una función f, tal que, siendo no negativa,

$$F(\vec{x}) = \int_{-\infty}^{\vec{x}} f(\vec{t})\, d\vec{t} = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(t_1, \ldots t_k)\, dt_1 \ldots dt_k.$$

a la función f se la designa como <u>función de densidad de probabilidad</u>. Todas las consideraciones de tipo geométrico que en su momento hicimos para las variables aleatorias son aplicables ahora, sustituyendo hipervolumen k-dimensional, donde decíamos área, etc...

La función de densidad de probabilidad tiene, como consecuencia de la definición, las propiedades siguientes:

a) Si f es continua en $(x_1, x_2, \dots, x_k)$, entonces

$$f(x_1, \dots, x_k) = \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k}$$

b)

$$\lim_{(x_1 \dots x_k) \to (+\infty, \dots, +\infty)} F(x_1 \dots x_k) = 1 = \lim_{(x_1 \dots x_k) \to (+\infty, \dots, +\infty)} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_k} f(t_1, \dots t_k)\, dt_1 \dots dt_k =$$

$$= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1 \dots x_k)\, dx_1 \dots dx_k .$$

De acuerdo con todo lo expuesto, y dado que la distribución de probabilidad de un vector aleatorio $\vec{X}$ podemos conocerla a través de su función de distribución, en última instancia también podemos conocer, y por tanto caracterizar, dicha distribución a partir de la función de densidad de probabilidad.

Ejemplo 2. - Sea X una variable aleatoria cuyo valor designa la cara que nos muestra un dado al ser lanzado. Sea Y una variable aleatoria que toma los valores 0 ó 1, según que una moneda, previamente lanzada, nos muestre su cruz o su cara. Tanto el dado como la moneda son perfectos y los lanzamientos independientes. Definimos un vector aleatorio mediante (X, Y). Dicho vector será del tipo discreto con

$$A = \{(1,0), (2,0), \dots, (6,0), (1,1), \dots, (6,1)\}$$

y $f(x_1, x_2) = \frac{1}{12}$ para $(x_1, x_2) \in A$.

La función de distribución viene dada por

$$F(x,y) = \begin{cases} 0 & x<1, -\infty<y<+\infty ; -\infty<x<+\infty, y<0 \\ 1/12 & 1 \le x < 2, 0 \le y < 1 \\ 1/6 & 2 \le x < 3, 0 \le y < 1 ; 1 \le x < 2, 1 \le y \\ 1/4 & 3 \le x < 4, 0 \le y < 1 \\ 1/3 & 4 \le x < 5, 0 \le y < 1 ; 2 \le x < 3, 1 \le y \\ 5/12 & 5 \le x < 6, 0 \le y < 1 \end{cases}$$

$$F(x,y) = \begin{cases} 1/2 & 6 \le x, 0 \le y < 1 ; 3 \le x < 4, 1 \le y \\ 2/3 & 4 \le x < 5, 1 \le y \\ 5/6 & 5 \le x < 6, 1 \le y \\ 1 & 6 \le x, 1 \le y \end{cases}$$

Ejemplo 3. - Sea (X, Y) un vector aleatorio del tipo continuo con función de densidad de probabilidad dada por

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < +\infty ,\ 0 < y < +\infty \\ 0 & \text{en el resto.} \end{cases}$$

Efectivamente $f(x,y)$ es una f. d. p. por cuanto

a) $f(x,y) \ge 0 \quad \forall (x,y)$

b) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = \int_0^{+\infty}\int_0^{+\infty} e^{-(x+y)}\,dx\,dy = \left[(1-e^{-x})\right]_0^{+\infty} \int_0^{+\infty} e^{-y}\,dy =$$

$$= [1-e^{-y}]_0^{+\infty} = 1$$

Así pues la función de distribución del vector aleatorio (X, Y) será

$$F(x,y) = \int_0^{x} \int_0^{y} e^{-(t+u)}\,dt\,du = (1-e^{-x})(1-e^{-y}) \qquad \begin{matrix} 0<x<+\infty \\ 0<y<+\infty \end{matrix}$$

$F(x,y) = 0$ en el resto.

Funciones de distribución y de densidad marginales

Sea $\vec{X}$ un vector aleatorio y sea $F(\vec{x})$ su función de distribución. Consideremos los dos subvectores $\vec{X}_1$ y $\vec{X}_2$ de dimensiones m y $k-m$ respectivamente, $m \ge 1$, de manera que $\vec{X} = (\vec{X}_1, \vec{X}_2)$ ya sabemos que como tales vectores también serán aleatorios. Consideremos la función de distribución en un pto. genérico $\vec{x} = (\vec{x}_1, \vec{x}_2)$ y supongamos que $\vec{x}_2 \to +\infty$ en todas sus componentes, tendremos

$$F(\vec{x}_1, +\infty, \dots, +\infty) = \lim_{\vec{x}_2 \to (+\infty, \dots, +\infty)} F(\vec{x}_1, \vec{x}_2) = \lim_{\vec{x}_2 \to (+\infty, \dots, +\infty)} P(\vec{X}_1 \le \vec{x}_1, \vec{X}_2 \le \vec{x}_2) = P(\vec{X}_1 \le \vec{x}_1) =$$

$$= F_{\vec{X}_1}(\vec{x}_1)$$

es decir, se trata de la función de distribución del subvector $\vec{X}_1$. Actuando de esta manera podremos obtener la función de distribución de cualquier subvector de dimensión $m<k$. En particular para $m=1$, tendríamos las funciones de distribución de las variables aleatorias que componen el vector $\vec{X}$, a saber

$$F_i(x_i) = F(+\infty, \dots, x_i, \dots, +\infty).$$

Obsérvese que para cada dimensión, m, son exactamente $\binom{k}{m}$ el número de subvectores de dimensión m que pueden obtenerse. A todas estas funciones de distribución, así obtenidas, se las conoce con el nombre de <u>funciones de distribución marginales</u> (del subvector correspondiente $\vec{X}_1$). Por contraposición con esta nomenclatura a $F(\vec{x}) = F(x_1 \ldots x_k)$ se la conoce con el nombre de <u>función de distribución conjunta</u>.

Si $\vec{X}$ es un vector aleatorio discreto, recordemos que, si A es tal que $P(\vec{X} \in A) = 1$, teníamos

$$F(\vec{x}) = \sum_{\vec{t} \in A \cap]-\infty, \vec{x}]} f(\vec{t}) \quad , \text{ con } f(\vec{t}) = P(\vec{X} = \vec{t})$$

en términos de los subvectores $\vec{X}_1$ y $\vec{X}_2$ la anterior expresión podría escribirse

$$F(\vec{x}_1, \vec{x}_2) = \sum_{(\vec{t}_1, \vec{t}_2) \in A \cap]-\infty, \vec{x}]} f(\vec{t}_1, \vec{t}_2) = \sum_{\vec{t}_1 \in A_1 \cap]-\infty, \vec{x}_1]} \left[\sum_{\vec{t}_2 \in A_2 \cap]-\infty, \vec{x}_2]} f(\vec{t}_1, \vec{t}_2) \right]$$

donde A_i es la proyección de A en el subespacio correspondiente, al igual que $]-\infty, \vec{x}_i]$, $i = 1, 2$.

En esta situación

$$F_{\vec{X}_1}(\vec{x}_1) = \lim_{\vec{x}_2 \to +\vec{\infty}} F(\vec{x}_1, \vec{x}_2) = \lim_{\vec{t}_2 \to +\vec{\infty}} \left\{ \sum_{\vec{t}_1 \in A_1 \cap]-\infty, \vec{x}_1]} \left[\sum_{\vec{t}_2 \in A_2 \cap]-\infty, \vec{x}_2]} f(\vec{t}_1, \vec{t}_2) \right] \right\} =$$

$$= \sum_{\vec{t}_1 \in A_1 \cap]-\infty, \vec{x}_1]} \left[\sum_{\vec{t}_2 \in A_2} f(\vec{t}_1, \vec{t}_2) \right] = \sum_{\vec{t}_1 \in A_1 \cap]-\infty, \vec{x}_1]} f_{\vec{X}_1}(\vec{t}_1)$$

donde $f_{\vec{X}_1}(\vec{x}_1) = \sum_{\vec{t}_2 \in A_2} f(\vec{x}_1, \vec{t}_2)$ es la función de cuantía del vector aleatorio $\vec{X}_1$, como fácilmente se desprende de las últimas igualdades escritas. Al igual que a su correspondiente función de distribución también a ésta se la conoce como la ~~marginal con~~ <u>función de cuantía marginal</u>, mientras que a la $f(\vec{x}) = f(x_1 \ldots x_k)$ se la denomina <u>función de cuantía conjunta</u>.

Supongamos que el vector $\vec{X}$ es continuo, es decir, existe f tal que

$$F(\vec{x}) = F(x_1 \ldots x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(t_1 \ldots t_k)\, dt_1 \ldots dt_k$$

o sea, adoptando una nomenclatura coherente con los subvectores $\vec{X}_1$ y $\vec{X}_2$,

$$F(\vec{x}_1, \vec{x}_2) = \int_{\vec{t}_1 \le \vec{x}_1} \int_{\vec{t}_2 \le \vec{x}_2} f(\vec{t}_1, \vec{t}_2)\, d\vec{t}_2 \cdot d\vec{t}_1 .$$

Para la obtención de la función de distribución marginal de $\vec{X}_1$, tendremos

$$F_{\vec{X}_1}(\vec{x}_1) = \lim_{\vec{x}_2 \to +\vec{\infty}} F(\vec{x}_1, \vec{x}_2) = \int_{\vec{t}_1 \le \vec{x}_1} \left[\int_{\vec{t}_2 \in R^{k-m}} f(\vec{t}_1, \vec{t}_2)\, d\vec{t}_2 \right] d\vec{t}_1$$

De acuerdo con esto lo que figura dentro del corchete debe de ser la <u>función de densidad marginal</u> del subvector aleatorio $\vec{X}_1$, es decir

$$f_{\vec{X}_1}(\vec{x}_1) = \int_{\vec{x}_2 \in R^{k-m}} f(\vec{x}_1, \vec{x}_2)\, d\vec{x}_2 = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_m, x_{m+1}, \ldots, x_k)\, dx_{m+1} \cdots dx_k .$$

Veamos un par de ejemplos.

<u>Ejemplo 4</u>. Lanzamos tres veces ~~una~~ moneda. Sea X la variable que representa el número de caras en los tres lanzamientos e Y la diferencia, en valor absoluto, entre el número de caras y el número de cruces. Obtener las funciones de cuantía conjunta y marginales y las de distribución.

Para mejor trabajar dispondremos los valores en la siguiente tabla:

$Y \backslash X$	0	1	2	3	$P(Y=y)$
1	0	3/8	3/8	0	6/8
3	1/8	0	0	1/8	2/8
$P(X=x)$	1/8	3/8	3/8	1/8	1

La función de cuantía conjunta son los valores que aparecen en el centro de la tabla. Las marginales las que aparecen en los márgenes de la misma.

Las funciones de distribución se obtendrán sumando adecuadamente. Así por ejemplo, para la conjunta

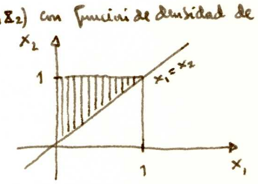
$$F(2.5, 1.5) = P(X \le 2.5, Y \le 1.5) = \sum_{\substack{x \le 2.5 \\ y \le 1.5}} f(x,y) = f(0,1) + f(1,1) + f(2,1) = 6/8$$

Para la marginal de la X

$$F_X(2.5) = P(X \le 2.5) = P(X=0) + P(X=1) + P(X=2) = 7/8 .$$

<u>Ejemplo 5</u>. Supongamos un vector aleatorio bidimensional (X_1, X_2) con función de densidad de probabilidad conjunta

$$f(x_1, x_2) = \begin{cases} 2 , & 0 < x_1 < x_2 < 1 \\ 0 , & \text{en el resto} \end{cases}$$



Las funciones de densidad marginales vienen dadas por

$$f_1(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 = \int_{x_1}^{1} 2\, dx_2 = 2(1 - x_1) \qquad 0 < x_1 < 1$$

$$f_1(x_1) = 0 \quad \text{en el resto}$$

$$f_2(x_2) = \int_0^{x_2} 2\,dx_1 = 2x_2 \quad , \quad 0<x_2<1$$

$$f_2(x_2) = 0 \quad \text{en el resto.}$$

La función de distribución conjunta viene dada por

$$F(x_1,x_2) = 0 \quad \text{si} \quad x_1 \le 0 \ \text{ó} \ x_2 \le 0$$

$$F(x_1,x_2) = \int_0^{x_1}\int_{t_1}^{x_2} f(t_1,t_2)\,dt_2\,dt_1 = \int_0^{x_1}\int_{t_1}^{x_2} 2\,dt_2\,dt_1 = 2\int_0^{x_1}(x_2-t_1)\,dt_1 = 2\left(x_2x_1 - \frac{x_1^2}{2}\right) =$$

$$= x_1(2x_2-x_1) \qquad 0<x_1<x_2<1$$

$$F(x_1,x_2) = 2x_1 - x_1^2 \quad , \quad 0<x_1<1, \ x_2 \ge 1$$

$$F(x_1,x_2) = x_2^2 \quad , \quad 0<x_2<1, \ x_1 \ge x_2$$

$$F(x_1,x_2) = 1 \quad , \quad x_1 \ge 1, \ x_2 \ge 1.$$

Las marginales son, para X_1

$$F_{X_1}(x_1) = \int_0^{x_1}(2-2t_1)\,dt_1 = 2x_1 - x_1^2 \quad , \quad 0<x_1<1$$

$$F_{X_1}(x_1) = 1 \quad , \quad x_1 \ge 1$$

$$F_{X_1}(x_1) = 0 \qquad x_1 \le 0$$

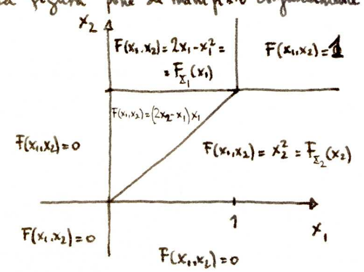
para X_2

$$F_{X_2}(x_2) = 0 \quad , \quad x_2 \le 0$$

$$F_{X_2}(x_2) = \int_0^{x_2} 2t_2\,dt_2 = x_2^2 \quad , \quad 0<x_2<1$$

$$F_{X_2}(x_2) = 1 \quad , \quad x_2 \ge 1$$

La figura pone de manifiesto conjuntamente, todos estos resultados



Funciones de densidad y de distribución condicionadas.

Caso discreto. Sea $\vec{X}$ un vector aleatorio discreto y consideremoslo nuevamente fraccionado en dos subvectores $(\vec{X}_1, \vec{X}_2) = \vec{X}$. Sea además $\vec{X}_1 = \vec{x}_1$ de manera que $f_{\vec{X}_1}(\vec{x}_1) > 0$, definamos una nueva función, del subvector $\vec{X}_2$, mediante

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1) = \frac{f(\vec{x}_1,\vec{x}_2)}{f_{\vec{X}_1}(\vec{x}_1)}$$

Así definida esta función tiene las siguientes propiedades:

a) $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1) \ge 0 \quad , \quad \forall \vec{x}_2$

b) $$\sum_{\vec{x}_2} f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1) = \sum_{\vec{x}_2} \frac{f(\vec{x}_1,\vec{x}_2)}{f_{\vec{X}_1}(\vec{x}_1)} = \frac{1}{f_{\vec{X}_1}(\vec{x}_1)}\left[\sum_{\vec{x}_2} f(\vec{x}_1,\vec{x}_2)\right] =$$

$$= \frac{1}{f_{\vec{X}_1}(\vec{x}_1)} \cdot f_{\vec{X}_1}(\vec{x}_1) = 1$$

pero estas son las condiciones que definen a las funciones de densidad de probabilidad, es decir $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)$ es una f.d.p., pero ¿qué significado tiene? Veámoslo.

Recordemos que $f(\vec{x}_1,\vec{x}_2) = P(\vec{X}_1 = \vec{x}_1, \vec{X}_2 = \vec{x}_2)$ y $f_{\vec{X}_1}(\vec{x}_1) = P(\vec{X}_1 = \vec{x}_1)$, así pues

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1) = \frac{P(\vec{X}_1 = \vec{x}_1, \vec{X}_2 = \vec{x}_2)}{P(\vec{X}_1 = \vec{x}_1)} = P(\text{[redacted]}, \vec{X}_2 = \vec{x}_2 / \vec{X}_1 = \vec{x}_1)$$

La función representa pues la probabilidad condicional de que el subvector $\vec{X}_2$ tome el valor $\vec{x}_2$ si el subvector $\vec{X}_1$ tomó el valor $\vec{x}_1$. Obsérvese además que

$$P(\vec{X}_2 \in B / \vec{X}_1 = \vec{x}_1) = \sum_{\vec{x}_2 \in B} P(\vec{X}_2 = \vec{x}_2 / \vec{X}_1 = \vec{x}_1) = \sum_{\vec{x}_2 \in B} f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1).$$

En definitiva, $f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)$ es la función de cuantía condicionada de $\vec{X}_2$, dado que $\vec{X}_1 = \vec{x}_1$.

Análogamente se define $f_{\vec{X}_1/\vec{X}_2}(\vec{x}_1/\vec{x}_2)$ siempre que $\vec{x}_2$ sea tal que $f_{\vec{X}_2}(\vec{x}_2) > 0$.

Ejemplo 6. Recordemos el vector aleatorio bidimensional del ejemplo 4. Algunas de las condicionadas que allí pueden definirse serían, por ejemplo

$$f_{X/Y=1}(x/1) = 0 \quad , \text{ para } x=0,\ x=3$$

$$f_{X/Y=1}(x/1) = \tfrac{1}{2} \quad , \text{ para } x=1,\ x=2$$

$$f_{X/Y=3}(x/3) = \tfrac{1}{2} \quad , \text{ para } x=0,\ x=3$$

$$f_{X/Y=3}(x/3) = 0 \quad , \text{ para } x=1,\ x=2$$

$$f_{Y/X=0}(y/0) = 0 \,, \quad \text{ para } y=1$$

$$f_{Y/X=0}(y/0) = 1 \,, \quad \text{ para } y=3$$

El vector aleatorio condicionado $\vec{X}_2/\vec{X}_1=\vec{x}_1$ tiene como función de distribución condicionada a $F_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1)$, cuyo valor, en función de la función de cuantía condicionada es

$$F_{\vec{X}_2/\vec{X}_1}(\vec{x}_2/\vec{x}_1) = \sum_{\vec{t}_2 \leq \vec{x}_2} f_{\vec{X}_2/\vec{X}_1}(\vec{t}_2/\vec{x}_1).$$

Así por ejemplo, en el caso anterior tendríamos

$$F_{X/Y=1}(2/1) = \sum_{t\leq 2} f_{X/Y=1}(t/1) = f_{X/Y=1}(0/1) + f_{X/Y=1}(1/1) + f_{X/Y=1}(2/1) = 0 + \tfrac{1}{2} + \tfrac{1}{2} = 1.$$

Caso continuo.- Supongamos ahora que $\vec{X}$ es un vector aleatorio del tipo continuo y para facilitar el desarrollo que vendrá a continuación vamos a situarnos en el caso bidimensional, es decir, $\vec{X}=(X,Y)$. Recordemos que en el caso continuo $P(X=x)=0$ y $P(Y=y)=0$ y por tanto ahora, probabilidades del tipo $P(X\leq x/Y=y)$ no están definidas. Sea entonces $\varepsilon>0$ y supongamos que $P(y-\varepsilon<Y\leq y+\varepsilon)>0$, definiremos entonces la siguiente probabilidad condicional,

$$P(X\leq x/y-\varepsilon<Y\leq y+\varepsilon) = \frac{P(X\leq x,\ y-\varepsilon<Y\leq y+\varepsilon)}{P(y-\varepsilon<Y\leq y+\varepsilon)}.$$

Se trata, para todo intervalo fijo $]y-\varepsilon, y+\varepsilon]$, de la probabilidad condicional de que $X\leq x$, dado que $Y\in]y-\varepsilon,y+\varepsilon]$, es decir, de la función de distribución condicional de X, dado que $Y\in]y-\varepsilon,y+\varepsilon]$.

Consideremos ahora el siguiente límite

$$\lim_{\varepsilon\to 0^+} P(X\leq x/Y\in]y-\varepsilon,y+\varepsilon]).$$

Cuando dicho límite existe lo designaremos mediante $F_{X/Y}(x/y)$ y lo definiremos como la función de distribución condicional de X, dado $Y=y$. Bajo las condiciones ya conocidas de continuidad absoluta de la función de distribución considerada, sabemos que existe una función no negativa que verifica

$$F_{X/Y}(x/y) = \int_{-\infty}^{x} g(t)\,dt \quad , \forall x\in\mathbb{R}.$$

A esta función g la designaremos como función de densidad de probabilidad condicional de X, dado $Y=y$, y la denotaremos mediante $f_{X/Y}(x/y)$. Es más que evidente que la función g verifica las condiciones que definen a una f.d.p.

Llegados a este punto cabe preguntarnos qué relación existe entre la f.d.p. condicionada y las f.d.p. conjuntas y marginales del vector (X,Y). Lo veremos mediante el siguiente

TEOREMA.- Sea f la f.d.p. conjunta del vector aleatorio continuo (X,Y), y sea f_2 la marginal correspondiente de Y. Para cada punto de continuidad de f en el que además $f_2(y)>0$ y es continua, se verifica

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)}.$$

Demostración.- Recordemos que

$$F_{X/Y}(x/y) = \lim_{\varepsilon\to 0^+} \frac{P(X\leq x,\ y-\varepsilon<Y\leq y+\varepsilon)}{P(y-\varepsilon<Y\leq y+\varepsilon)} = \lim_{\varepsilon\to 0^+} \frac{\int_{-\infty}^{x}\left\{\int_{y-\varepsilon}^{y+\varepsilon} f(u,v)\,du\,dv\right\}}{\int_{y-\varepsilon}^{y+\varepsilon} f_2(v)\,dv}$$

dividiendo numerador y denominador por 2ε, pasando al límite y teniendo en cuenta la relación que liga a las f.d.p. con las funciones de distribución en los puntos de continuidad de aquellas, tendremos:

$$F_{X/Y}(x/y) = \frac{\int_{-\infty}^{x} f(u,y)\,du}{f_2(y)} = \int_{-\infty}^{x}\left\{\frac{f(u,y)}{f_2(y)}\right\}du$$

si derivamos ahora, respecto de x, la función de distribución condicional, tendremos,

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)}$$

que es el resultado buscado.

Observación.- De la igualdad obtenida en el teorema

$$F_{X/Y}(x/y) = \frac{\int_{-\infty}^{x} f(u,y)\,du}{f_2(y)}$$

tenemos

$$f_2(y)\cdot F_{X/Y}(x/y) = \int_{-\infty}^{x} f(u,y)\,du$$

Si integramos para y en toda la recta real, tendríamos

$$\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{x} f(u,y)\,du\right]dy = \int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)\,dy\right]du = \int_{-\infty}^{x} f_1(u)\,du = F_1(x)$$

que es la función de distribución marginal de X, y en la otra parte del igual quedaría

$$F_1(x) = \int_{-\infty}^{+\infty} f_2(y)\cdot F_{X/Y}(x/y)\,dy.$$

Estas definiciones pueden trasladarse de forma inmediata al caso de un espacio m-dimensional. En efecto, para $\vec{X} = (\vec{X}_1, \vec{X}_2)$, tendríamos

$$F_{\vec{X}_1/\vec{X}_2^0}(\vec{x}_1/\vec{x}_2) = \frac{\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k} f(x_1,\dots,x_m)\,dx_1\cdots dx_k}{f_{\vec{X}_2}(x_{k+1},\dots,x_m)}$$

y para la función de densidad de probabilidad condicional

$$f_{\vec{X}_1/\vec{X}_2^0}(\vec{x}_1/\vec{x}_2) = \frac{f(\vec{x})}{f_{\vec{X}_2}(\vec{x}_2)}, \quad \text{con } \vec{x} = (\vec{x}_1, \vec{x}_2).$$

Ejemplo 7.- Utilizando el vector bidimensional del ejemplo 5 (pag. 8), tendríamos

$$f_{X_1/X_2}(x_1/x_2) = \frac{f(x_1,x_2)}{f_2(x_2)} = \frac{2}{2x_2} = \frac{1}{x_2}, \quad 0<x_1<x_2$$

es decir, se trata de una distribución uniforme en el intervalo $]0,x_2[$.

Análogamente

$$f_{X_2/X_1}(x_2/x_1) = \frac{f(x_1,x_2)}{f_1(x_1)} = \frac{2}{2(1-x_1)} = \frac{1}{1-x_1}, \quad x_1<x_2<1$$

que es, nuevamente, una distribución uniforme en el intervalo $]x_1,1[$. Para las funciones de distribución correspondiente

$$F_{X_1/X_2=x_2}(x_1/x_2) = \int_0^{x_1} f_{X_1/X_2}(t/x_2)\,dt = \int_0^{x_1}\frac{1}{x_2}\,dt = \frac{x_1}{x_2}, \quad 0<x_1<x_2$$

para $x_2 = \frac{2}{3}$, $x_1 = \frac{1}{3}$, tendríamos

$$F_{X_1/X_2=\frac{2}{3}}(1/3|2/3) = P\left(X_1 \le \tfrac{1}{3} / X_2 = \tfrac{2}{3}\right) = \frac{1/3}{2/3} = \frac{1}{2}.$$

Por otra parte

$$F_1(x_1) = \int_{-\infty}^{+\infty} f_2(x_2)\cdot F_{X_1/X_2=x_2}(x_1/x_2)\,dx_2 = \int_{x_1}^{1} 2x_2\left[\frac{x_1}{x_2}\right]dx_2 = x_1\int_{x_1}^{1} 2\,dx_2 = x_1\left[2x_2\right]_{x_1}^{1} =$$

$$= 2x_1 - 2x_1^2, \quad 0<x_1<1.$$

$$F_1(x_1) = 0, \quad x_1 \le 0$$

$$F_1(x_1) = 1, \quad x_1 \ge 1.$$

como ya obtuvimos.

Truncamiento de una variable aleatoria

Hemos visto que si X es una variable aleatoria continua, mediante el cambio $Y=g(X)$ con $g(x)=x$, para $|x|<b$, $g(x)=-b$, $x\le -b$ y $g(x)=b$, $x\ge b$, podíamos llevar a cabo un truncamiento de la variable X, pero existía el inconveniente de obtener una nueva variable, Y, que no era ni discreta, ni continua, por cuanto su función de distribución no era continua, ni tampoco en escalera. Haciendo uso de las probabilidades condicionadas podemos llevar a cabo un truncamiento de la variable que resuelve este inconveniente. Veamos como.

Definición.- Si X es una variable aleatoria [redacted] sobre el espacio de probabilidad $(S, \mathcal{A}, P)$ y si T es un Boreliano de la recta real, tal que $0<P(X\in T)<1$, entonces a la distribución condicional $P(X\le x/X\in T)$, definida para cualquier x, se le llama distribución truncada de X.

Obsérvese que la definición es válida para cualquier tipo de variable, así, para el caso de una variable discreta, las cosas quedan de la siguiente manera:

$$P(X=x_i/X\in T) = \frac{P(X=x_i, X\in T)}{P(X\in T)} = \begin{cases} \dfrac{f(x_i)}{\sum_{x_j\in T} f(x_j)}, & \text{si } x_i\in T \\ 0, & \text{en el resto} \end{cases}$$

El caso más interesante, aquel en que X es continua, da lugar a

$$P(X \le x / X \in T) = \frac{P(X \le x, X \in T)}{P(X \in T)} = \frac{\int_{]-\infty,x] \cap T} f(t).dt}{\int_T f(t).dt}$$

La función de densidad de probabilidad para la variable truncada es

$$h(x) = \begin{cases} \dfrac{f(x)}{\int_T f(t).dt} & , x \in T \\ 0 & , x \in T^c . \end{cases}$$

El efecto del truncamiento es acumular sobre T toda la masa de probabilidad, por cuanto

$$P(X \in T / X \in T) = 1.$$

Este es un método muy utilizado en teoría de la probabilidad y muy útil en particular cuando se estudian teoremas límite. Otra interesante aplicación del método de truncamiento de una variable sobreviene cuando la variable original no tiene esperanza finita, pero de ello nos ocuparemos en el próximo capítulo.

Ejemplo 8.- Sea X una v.a. $N(0,1)$ y sea $T =]-\infty, 0]$. Entonces $P(X \in T) = 1/2$ debido a la simetría de X. Para la variable truncada tendremos pues

$$h(x) = \begin{cases} 2 f(x) = \dfrac{\sqrt{2}}{\sqrt{\pi}} \cdot e^{-\frac{1}{2}x^2} & , x \le 0 \\ 0 & , \text{en el resto.} \end{cases}$$

Algunos vectores aleatorios usuales.

$\boxed{\text{Discretos}}$

1. Distribución multinomial.- Se trata de una generalización de la variable aleatoria binomial que en su momento estudiamos. Surge en situaciones como la siguiente:

Supongamos que llevamos a cabo n pruebas, independientes unas de otras, de tal manera que en cada una de ellas podemos tener como resultado uno de los r sucesos $E_1, E_2, \ldots, E_r$, cada uno de los cuales teniendo una probabilidad p_i, $i=1,\ldots,r$ de ocurrir, verificándose además $p_1 + p_2 + \ldots + p_r = 1$. En estas condiciones la probabilidad de obtener, al final de las n pruebas, k_1 veces E_1, k_2 veces E_2, ..., k_r veces E_r, con $k_1+k_2+\ldots+k_r = n$ vendrá dada por:

$$\binom{n}{k_1}\binom{n-k_1}{k_2} \cdots \binom{n-(k_1+k_2+\ldots+k_{r-1})}{k_r} p_1^{k_1} \cdot p_2^{k_2} \cdots p_r^{k_r}$$

que también puede escribirse

$$\frac{n!}{k_1!\, k_2! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r} .$$

Definamos ahora un vector aleatorio $\vec{X} = (X_1, \ldots, X_r)$ donde X_i designa el número de veces que ocurre el [redacted] suceso E_i. Para $\vec{x} \in A$, siendo $A = \{\vec{x} \in R^r ;\ x_i \ge 0,\ i=1,\ldots,r \wedge \sum_{i=1}^{r} x_i = n\}$, tendremos que la función de cuantía vale

$$f(\vec{x}) = P(X_1 = x_1, \ldots, X_r = x_r) = \frac{n!}{x_1! \cdots x_r!} p_1^{x_1} \cdots p_r^{x_r}$$

mientras que si $\vec{x} \in A^c$, entonces

$$f(\vec{x}) = 0$$

La función así definida es efectivamente una función de cuantía, por cuanto es no negativa en R^r y además, al sumarla para todos los elementos de A, y teniendo en cuenta el significado de los coeficientes multinomiales $\left(\frac{n!}{x_1! \ldots x_r!}\right)$, tendremos

$$\sum_{\vec{x} \in A} f(\vec{x}) = \sum_{\vec{x} \in A} \frac{n!}{x_1! \ldots x_r!} p_1^{x_1} \ldots p_r^{x_r} = (p_1 + p_2 + \ldots + p_r)^n = 1^n = 1.$$

Obtención de algunas marginales.- Si queremos obtener la marginal de X_1, procederemos así

$$f_1(x_1) = \sum_{\vec{X}_2 \in A_2} f(\vec{x}) \quad \text{con } \vec{X}_2 = (x_2, \ldots, x_r) \text{ y } A_2 = \{\vec{x}_2 ;\ x_i \ge 0,\ i=2,\ldots,r,\ \sum_{i=2}^{r} x_i = n - x_1\}$$

entonces

$$f_1(x_1) = \sum_{\vec{x}_2 \in A_2} f(x_1, \vec{x}_2) = \sum_{\vec{x}_2 \in A_2} \binom{n}{x_1}\binom{n-x_1}{x_2} \cdots \binom{x_r}{x_r} p_1^{x_1} \cdots p_r^{x_r} =$$

$$= \binom{n}{x_1} p_1^{x_1} \sum_{\vec{x}_2 \in A_2} \binom{n-x_1}{x_2}\binom{n-(x_1+x_2)}{x_3} \cdots \binom{x_r}{x_r} p_2^{x_2} \cdots p_r^{x_r} =$$

$$= \binom{n}{x_1} p_1^{x_1} (p_2 + \cdots + p_r)^{n-x_1} = \binom{n}{x_1} p_1^{x_1} (1-p_1)^{n-x_1}$$

se trata pues de una binomial con parámetros n, p_1. Obsérvese que razonando intuitivamente sobre el significado de la marginal se espera el resultado encontrado.

Para marginales que cualquier otra dimensión se obtienen nuevamente multinomiales de dimensión coincidente con la del vector marginal considerado.

Para las distribuciones condicionadas llegaríamos a un resultado análogo.

2. Distribución binomial negativa en el caso bivariante. Supongamos que E_1, E_2, E_3 constituyen una partición del espacio muestral, de manera que $P(E_1) = p_1$, $P(E_2) = p_2$, $P(E_3) = p_3 = 1-p_1-p_2$. Realizamos un experimento cuyo resultado es necesariamente E_1, E_2 ó E_3. Vamos a repetir este experimento, de manera independiente en cada ocasión, hasta que el suceso E_3 se haya realizado k veces, ello significa que E_1 habrá ocurrido x veces y E_2 habrá ocurrido y veces y todo ello antes de la k-ésima ocurrencia del suceso E_3. La probabilidad de que las cosas hayan ocurrido así, vendrá dada por

$$\binom{x+y+k-1}{x} \cdot \binom{y+k-1}{y} \cdot p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k = \frac{(x+y+k-1)!}{x!\, y!\, (k-1)!} p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k$$

Definamos ahora un vector aleatorio bidimensional (X, Y) donde X e Y designan el número de ocurrencias de los sucesos E_1 y E_2 antes de la k-ésima ocurrencia del suceso E_3. La función de cuantía de este vector aleatorio viene dada por

$$f(x,y) = P(X = x, Y = y) = \frac{(x+y+k-1)!}{x!\, y!\, (k-1)!} p_1^x \cdot p_2^y \cdot (1-p_1-p_2)^k \quad , \ (x,y) \in A$$

donde $A = \{(x,y) \in R^2 ; \ x = 0,1,2,\ldots \ \ y = 0,1,2,\ldots\}$, y

$$f(x,y) = 0, \quad (x,y) \in A^c.$$

Se trata de una función de cuantía por cuanto, $f(x,y) \geq 0$, $\forall (x,y) \in R^2$, además, teniendo en cuenta que

$$(1-p_1-p_2)^{-k} = \sum_{x,y \in A} \binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x \cdot p_2^y$$

entonces

$$\sum_{(x,y) \in A} f(x,y) = (1-p_1-p_2)^k \cdot (1-p_1-p_2)^{-k} = 1.$$

Distribuciones marginales. La marginal de Y, por ejemplo, la obtendríamos como sigue:

$$f_2(y) = \sum_{x \in A_1} f(x,y) \qquad \text{donde} \quad A_1 = \{x ; \ x = 0,1,2,\ldots\}$$

entonces

$$f_2(y) = \sum_{x \geq 0} \binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x p_2^y (1-p_1-p_2)^k =$$

$$= \binom{y+k-1}{y} p_2^y (1-p_1-p_2)^k \sum_{x \geq 0} \binom{x+y+k-1}{x} p_1^x$$

pero recordemos que $(1-x)^{-n} = \sum_{j \geq 0} \binom{n+j-1}{j} x^j$, por tanto

$$f_2(y) = \binom{y+k-1}{y} p_2^y (1-p_1-p_2)^k (1-p_1)^{-(y+k)} = \binom{y+k-1}{y} \left(\frac{p_2}{1-p_1}\right)^y \cdot \left(\frac{1-p_1-p_2}{1-p_1}\right)^k =$$

$$= \binom{y+k-1}{y} q_2^y (1-q_2)^k \qquad \text{con} \quad q_2 = \frac{p_2}{1-p_1} \ , \ \text{siendo} \ 0 < q_2 < 1 \ , \ y = 0,1,\ldots$$

se trata pues de una binomial negativa con parámetros k, y $q_2 = \frac{p_2}{1-p_1}$. Razonando de forma análoga llegaríamos a

$$f_1(x) = \binom{x+k-1}{x} q_1^x (1-q_1)^k \ . \ \text{con} \quad q_1 = \frac{p_1}{1-p_2} \ , \quad x = 0,1,\ldots\ldots$$

Distribuciones condicionadas. Para $Y = y$, la variable aleatoria condicionada X/y, tiene como función de cuantía

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)} = \frac{\binom{x+y+k-1}{x}\binom{y+k-1}{y} p_1^x p_2^y (1-p_1-p_2)^k}{\binom{y+k-1}{y}\left(\frac{p_2}{1-p_1}\right)^y \cdot \left(\frac{1-p_1-p_2}{1-p_1}\right)^k} =$$

$$= \binom{x+y+k-1}{x} p_1^x (1-p_1)^{y+k} \qquad . \ x = 0,1,2,\ldots \quad y = 0,1,\ldots$$

que es una binomial negativa con parámetros $y+k$ y p_1.

Capítulo 7

Esperanza matemática o valor esperado

ESPERANZA MATEMATICA o VALOR ESPERADO

INTRODUCCION

El concepto de speranza aparece inicialmente, en Cálculo de Probabilidades, ligado a los juegos de azar y a las "stadísticas" de población. En el primer caso hacía referencia a la cantidad de dinero que un jugador speraba obtener de su participación en determinado juego. En el otro caso el concepto "speranza de vida", significaba la vida media de un individuo a la vista de los datos considerados en la "stadística" en cuestión.

Vayamos al caso del jugador y supongamos que éste participa en el siguiente juego: Se lanza un dado, si la cara que presenta el dado s un número par el jugador recibe tantas pesetas como puntos indica la cara, si por el contrario s impar el número de puntos el jugador deberá pagar dicha cantidad en pesetas. ¿Qué ganancia spera obtener el jugador en cada lanzamiento?

En el supuesto de un dado correcto si el jugador lleva a cabo un elevado número de lanzamientos, N, las distintas caras del dado deben aparecer, muy aproximadamente, el mismo número de veces, a saber, $N/6$. Así pues, en ~~todas~~ estas jugadas el jugador habrá ganado (o perdido) lo siguiente:

$$(-1).\frac{N}{6} + 2.\frac{N}{6} + (-3).\frac{N}{6} + 4.\frac{N}{6} + (-5).\frac{N}{6} + 6.\frac{N}{6}$$

por tanto en cada jugada sperará obtener una cantidad E, que se obtendrá de dividir por N la anterior, así pues

$$E = (-1).\frac{1}{6} + 2.\frac{1}{6} + (-3).\frac{1}{6} + 4.\frac{1}{6} + (-5)\frac{1}{6} + 6.\frac{1}{6} = \frac{1}{6}\left[2+4+6-(1+3+5)\right] =$$

$$= \frac{1}{6}(12-9) = \frac{3}{6} = 0.5$$

s decir, spera ganar 50 céntimos en cada jugada (de haber obtenido una cantidad negativa habría que interpretarlo como una pérdida).

Replanteemos ahora el problema de otra forma. Definamos una variable aleatoria X cuyos valores van a ser los pagos o las ganancias del jugador en cada jugada, de acuerdo con sto X tomará los siguientes valores (positivos ganancias, negativos pagos)

$$X = \{-1, 2, -3, 4, -5, 6\}$$

además, teniendo en cuenta la naturaleza del experimento probabilístico asociado al juego, tendremos:

$$P(X = x_i) = \frac{1}{6} \quad , \quad x_i = -1, 2, -3, 4, -5, 6$$

de acuerdo con sto, E podría también obtenerse mediante la siguiente expresión

$$E = \sum_{i=1}^{6} x_i . P(X = x_i) = \frac{1}{6}\left[\sum_{i=1}^{6} x_i\right] = 0.5$$

a la expresión $\sum_{i=1}^{6} x_i . P(X=x_i)$ se la designa mediante $E(X)$ y se la denomina "<u>esperanza de la variable aleatoria</u> X". Todavía cabe un nuevo replanteamiento de la cuestión. X, como variable aleatoria que s, supone una aplicación del spacio de probabilidad $(\Omega, \mathcal{A}, P)$ asociado al experimento de lanzar un dado y el spacio medible (R, β), s decir, la recta real con su tribu de Borel, como tal aplicación, X puede ser descrita de la siguiente manera:

$$X = \sum_{i=1}^{6} x_i . \varphi_{A_i} \qquad A_i = \{w \in \Omega ; X(\omega) = x_i\} \qquad x_i = -1, 2, -3, 4, -5, 6$$

s decir, se trata de una función simple. Recordemos que la integral de una función simple, siempre que fuera finita (o no negativa la función), se escribía

$$\int_{\Omega} s \, d\mu = \sum_{i=1}^{n} a_i . \mu(A_i) \quad , \text{si} \quad s = \sum_{i=1}^{n} a_i \varphi_{A_i}$$

aplicado a X, tendremos

$$\int_{\Omega} X \, dP = \sum_{i=1}^{6} x_i . P(A_i)$$

pero cada $A_i = \{w_i\}$, siendo w_i el suceso "han salido i puntos en la cara del dado" y por tanto $P(A_i) = \frac{1}{6}$, $i = 1, \ldots, 6$, y en definitiva, teniendo en cuenta la definición de speranza de la variable X que hemos dado, tendremos

$$E(X) = \int_{\Omega} X \, dP .$$

Esta última igualdad s la que se utiliza como definición del concepto de speranza matemática de una variable aleatoria del tipo que sea. Definición que hemos visto su significado en el caso ~~de~~ discreto y que se adopta como tal para el caso continuo, pero haciendo algunas consideraciones de las que nos vamos a ocupar a continuación.

DEFINICION.- Sea X una variable aleatoria definida del spacio de probabilidad $(\Omega, \mathcal{A}, P)$ en (R, β, P') (P' probabilidad inducida) se define la <u>speranza matemática de X</u>, o el <u>valor sperado de X</u>, $E(X)$, mediante

$$E(X) = \int_{\Omega} X . dP \quad , \text{siempre que sea finita.}$$

Observación.- A la vista de la definición de esperanza que hemos dado tiene sentido el hablar de la existencia o no de ésta para una variable aleatoria, puesto que la definición viene condicionada por la finitud de la integral que en ella aparece.

Recordemos que si X es no negativa la integral

$$\int_\Omega X\, dP$$

está siempre definida, aunque puede no ser finita. Para el caso de X cualquiera se dice que X es integrable, si es finita, cosa que ocurría cuando X^+ y X^- tienen integral finita, en cuyo caso

$$\int_\Omega X\, dP = \int_\Omega X^+ dP - \int_\Omega X^- dP.$$

Si por otra parte recordamos la relación que existe entre las integrales, sobre un mismo conjunto, de X y $|X|$, podemos afirmar que:

"La $E(X)$ existe si y sólo si $E(|X|)$, que siempre está definida, es finita, es decir, existe".

La esperanza para v.a. discretas. Una v.a. discreta, X, se caracterizaba por la existencia de un conjunto $A \subset R$, tal que $P(X \in A) = 1$, siendo a lo sumo, A, numerable. Ello nos permite escribir X de la siguiente forma

$$X = \sum_{a_i \in A} a_i \, \varphi_{A_i} \quad . \ A_i = \{\omega \in \Omega;\ X(\omega) = a_i\} \quad . \ \cup A_i = \Omega \quad , \ A_i \cap A_j = \phi \ . \ i \neq j$$

así pues, si existe

$$E(X) = \int_\Omega X\, dP = \int_\Omega \Big(\sum_{a_i \in A} a_i \varphi_{A_i}\Big) dP = \int_{\cup A_i} \Big(\sum_{a_i \in A} a_i \varphi_{A_i}\Big) dP = \sum_{a_i \in A} \int_{A_i} a_i\, dP =$$

$$= \sum_{a_i \in A} a_i \int_{A_i} dP = \sum_{a_i \in A} a_i . P(A_i) = \sum_{a_i \in A} a_i . P(X = a_i).$$

La esperanza para v.a. continuas. En primer lugar enunciaremos, sin demostrar, un importante teorema que se aplica a los cambios de variable en una integral. Supongamos un espacio de medida $(\Omega, \mathcal{B}, \mu)$ y un espacio medible $(\Omega', \mathcal{B}')$ y entre ellos una aplicación T. Sea además f una función medible de $(\Omega', \mathcal{B}')$ en (R, β). Sobre $(\Omega', \mathcal{B}')$ inducimos una medida, μ', mediante la aplicación medible T, como sigue

$$\mu'(A') = \mu(T^{-1}(A')).$$

¿Cómo podemos trasladar la integral de la función medible $f \circ T$ sobre Ω, a una integral sobre Ω'? El teorema que enunciamos nos da la solución

TEOREMA I.- Con los elementos anteriores y siempre que f sea no negativa o bien integrable, tendremos

$$\int_\Omega f \circ T\, d\mu = \int_{\Omega'} f\, d\mu'.$$

Este teorema tiene aplicación inmediata a nuestra situación. En efecto, hagamos $T \equiv X$, variable aleatoria del tipo continuo con la que vamos a trabajar, en esta situación $(\Omega', \mathcal{B}') \equiv (R, \beta)$, $\mu \equiv P$ y $\mu' \equiv P'$, probabilidad inducida. Finalmente sea f la identidad. Tendremos pues que $f \circ T \equiv X$ y podemos escribir, si $E(X)$ existe

$$E(X) = \int_\Omega X\, dP = \int_R x\, dP'$$

Por otra parte recordemos que la v.a. continua se caracteriza por la existencia de una función, la f.d.p. que permite escribir

$$F(x) = \int_{-\infty}^{x} f(t).dt$$

siendo F la función de distribución, y también

$$P'(A) = \int_A f\, d\lambda$$

siendo $A \in \beta$ y λ la medida de Lebesgue. A las medidas que gozan de esta propiedad (poseer una densidad respecto de otra medida; en nuestro caso f es la densidad de P' respecto a la medida de Lebesgue) es de aplicación el siguiente teorema, que también enunciamos y no demostramos

TEOREMA II.- Si la medida ν tiene una densidad g respecto de la medida μ, entonces para toda función h, medible, no negativa o integrable, tenemos

$$\int_\Omega h\, d\nu = \int_\Omega h.g.d\mu.$$

Este teorema aplicado a nuestra situación concreta con $h \equiv x$, medible e integrable, $\nu \equiv P'$ y $\mu \equiv \lambda$, medida de Lebesgue y la densidad $g \equiv f$, da lugar a:

$$E(X) = \int_{\Omega} X \, dP \underset{\text{por Teorema I}}{=} \int_{R} x \, dP' \underset{\text{por Teorema II}}{=} \int_{R} x \, f(x) \, d\lambda$$

Finalmente si la función $x.f(x)$ es integrable Riemann, como es el caso normalmente, el valor de su integral coincidirá con la integral de Lebesgue existe y obtendremos finalmente para la $E(X)$, siempre que exista

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x \, f(x) \, dx}$$

Esperanza de una función de una variable aleatoria. Si transformamos una variable aleatoria X mediante una aplicación g medible, la nueva variable aleatoria $g(X)$ tiene una esperanza que se define como sigue:

Definición.- Sea g una aplicación medible y X una variable aleatoria, si $E[|g(X)|] < +\infty$ definimos la esperanza de $g(X)$ mediante

$$E(g(X)) = \int_{\Omega} g(X) \, dP$$

siendo $(\Omega, \mathcal{A}, P)$ el espacio de probabilidad sobre el que está definida X.

Un razonamiento análogo al utilizado en los párrafos anteriores nos lleva a las siguientes expresiones de la $E(g(X))$, según que X sea discreta o continua.

X discreta:

$$E(g(X)) = \sum_{a_i \in A} g(a_i) . P(X = a_i)$$

X continua:

$$E(g(X)) = \int_{R} g(x) . f(x) \, d\lambda$$

y si $g(x).f(x)$ es tal que resulta ser integrable Riemann, entonces

$$E(g(X)) = \int_{R} g(x) \, f(x) \, dx .$$

Momentos de una variable aleatoria

a) Momentos absolutos respecto del origen.

Si hacemos $g(X) = |X|^k$, entonces $E(g(X)) = E(|X|^k)$ que se denomina momento absoluto de orden k respecto del origen. Obsérvese que si el momento absoluto de orden k respecto del origen existe para una v.a. X, entonces para $j \leq k$, tendremos

$|x|^j \leq 1 + |x|^k$ y de aquí

$$\int_{R} |X|^j \, dP' \leq \int_{R} (1 + |x|^k) \, dP' = 1 + E(|X|^k) < +\infty$$

y por tanto $E(|X|^j)$ también existe.

b) Momentos respecto del origen.

Si hacemos $g(X) = (X)^k$, entonces $E(g(X)) = E(X^k)$ que se denomina momento de orden k respecto del origen.

Obsérvese que para $k=1$, se obtiene la esperanza de la variable aleatoria como momento de orden uno respecto del origen.

c) Momentos respecto de una constante c.

Haciendo $g(X) = |X-c|^k$ o bien $g(X) = (X-c)^k$ se obtienen los momentos de orden k, absolutos o no, respecto de una constante c. Hay un caso especial de este tipo de momentos que es aquel para el cual $c = E(X)$, entonces a los momentos de esta manera obtenidos se les denomina momentos centrales de orden k (absolutos o no). De entre los momentos centrales tiene especial interés el momento central de segundo orden, a saber

$$E(X - E(X))^2$$

al que se le denomina varianza de la variable aleatoria.

A continuación nos vamos a ocupar de estudiar con detalle la esperanza y la varianza de una variable aleatoria a través de sus propiedades más relevantes.

Propiedades de la Esperanza de una variable aleatoria

Hay una serie de propiedades que se derivan de inmediato de la definición de esperanza, son:

1) $E(c) = c$, donde c es una constante

2) $E[c\,g(X)] = c\,E[g(X)]$, en particular $E[cX] = c\,E[X]$.

3) $E[g(X) + d] = E[g(X)] + d$, d constante, en particular $E[X+d] = E(X) + d$

de la combinación de las dos últimas, llegamos a

4) $E[ag(X)+b] = a\,E[g(X)] + b$, con a y b constantes.

5) $E\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i . E[g_i(X)]$, c_i constantes $i=1,\dots,n$.

6) Si $X \geq 0$, $E(X) \geq 0$.

7) De 6) se deriva $E(X) \geq E(Y)$ si $X \geq Y$, siendo ambas variables aleatorias.

Otras propiedades requieren una demostración explícita. Por ejemplo

Propiedad 8.- Sea X una v.a. no negativa con función de distribución F. Entonces

$$E(X) = \int_0^\infty (1-F(x))\,dx,$$

en el sentido de que si alguno de los miembros de la igualdad existe también lo hace el otro y sus valores coinciden.

Demostración.- a) X es una v.a. continua.

Sabemos que la $E(X)$ cuando existe viene dada por

$$E(X) = \int_0^\infty x f(x)\,dx = \lim_{n\to\infty} \int_0^n x f(x)\,dx$$

Integrando por partes obtendremos

$$\int_0^n x f(x)\,dx = [xF(x)]_0^n - \int_0^n F(x)\,dx = n.F(n) - \int_0^n F(x)\,dx$$

sumando y restando n, tendremos

$$\int_0^n x f(x)\,dx = -n + nF(n) + n - \int_0^n F(x)\,dx = -n[1-F(n)] + \int_0^n [1-F(x)]\,dx.$$

Pero

$$n[1-F(n)] = n\int_n^\infty f(x)\,dx < \int_n^\infty x f(x)\,dx$$

y puesto que $E(|X|) < \infty$, entonces

$$\lim_{n\to\infty} \int_n^\infty x f(x)\,dx = 0 \longrightarrow \lim_{n\to\infty} n[1-F(n)] = 0,$$

tenemos pues finalmente

$$E(X) = \lim_{n\to\infty} \int_0^n x f(x)\,dx = \lim_{n\to\infty} \int_0^n [1-F(x)]\,dx = \int_0^\infty [1-F(x)]\,dx.$$

Por otra parte si partimos de $\int_0^\infty [1-F(x)]\,dx < +\infty$, entonces

$$\int_0^n x f(x)\,dx \leq \int_0^n [1-F(x)]\,dx \leq \int_0^\infty [1-F(x)]\,dx, \ \forall n$$ y de aquí $E(|X|) < +\infty$.

b) X es una v.a. discreta

Sea $\{x_1, \dots, x_n, \dots\}$ el conjunto A tal que $P(X \in A) = 1$ y hagamos $P(X = x_j) = p_j$, entonces

$$E(X) = \sum_{j\geq 1} x_j . p_j .$$

Sea $I = \int_0^\infty [1-F(x)]\,dx$. Entonces

$$I = \sum_{k\geq 1} \int_{(k-1)/n}^{k/n} P\{X > x\}\,dx$$

y puesto que $P(X>x)$ es una función no creciente de x, tendremos

$$P\left(X > \tfrac{k}{n}\right) \leq P(X>x) \leq P\left(X > \tfrac{k-1}{n}\right), \quad \tfrac{k}{n} \leq x \leq \tfrac{k-1}{n}, \ \forall n$$

y de aquí, integrando y sumando sobre k,

$$\frac{1}{n}\sum_{k\geq 1} P\left(X > \tfrac{k}{n}\right) \leq I \leq \frac{1}{n}\sum_{k\geq 1} P\left(X > \tfrac{k-1}{n}\right), \ \forall n.$$

Hagamos

$$U_n = \frac{1}{n}\sum_{k\geq 1} P\left(X > \tfrac{k-1}{n}\right) \quad y \quad L_n = \frac{1}{n}\sum_{k\geq 1} P\left(X > \tfrac{k}{n}\right).$$

Tenemos

$$L_n = \frac{1}{n}\sum_{k\geq 1}\sum_{j\geq k} P\left\{\tfrac{j}{n} < X \leq \tfrac{j+1}{n}\right\} = \frac{1}{n}\sum_{k\geq 2}(k-1)P\left\{\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right\}$$

reordenando la serie. Así

$$L_n = \sum_{k\geq 2} \frac{k}{n} P\left\{\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right\} - \sum_{k\geq 2} \frac{1}{n} P\left\{\tfrac{k-1}{n} < X \leq \tfrac{k}{n}\right\} =$$

$$= \sum_{k\geq 2} \frac{k}{n}\left[\sum_{(k-1)/n < x_j \leq k/n} p_j\right] - \frac{1}{n}P\left\{X > \tfrac{1}{n}\right\} \geq \sum_{k\geq 1} \sum_{(k-1)/n < x_j \leq k/n} x_j p_j - \frac{1}{n} =$$

$$= E(X) - \frac{1}{n}, \ \forall n$$

Análogamente podemos demostrar que para cada n, tenemos

$$U_n \leq E(X) + \frac{1}{n}.$$

Tenemos pues

$$E(X) - \frac{1}{n} \leq L_n \leq I \leq U_n \leq E(X) + \frac{1}{n}, \ \forall n$$ y se sigue el teorema.

Una consecuencia de la propiedad 8 es esta nueva propiedad.

Propiedad 9.- Para una v.a. cualquiera X, $E(|X|)<+\infty$ si y sólo si las integrales

$$\int_{-\infty}^{0} P\{X\le x\}\,dx \quad y \quad \int_{0}^{\infty} P\{X>x\}\,dx$$

son finitas y además en este caso

$$E(X)=\int_{0}^{\infty} P\{X>x\}\,dx-\int_{-\infty}^{0} P\{X\le x\}\,dx .$$

Demostración.- Si $E(|X|)<+\infty$, existe la esperanza de X y sabemos además que

$$E(X)=E(X^+)-E(X^-)$$

siendo

$$X^+=\begin{cases}0, & X<0\\ X, & X\ge 0\end{cases} \quad y \quad X^-=\begin{cases}0, & X>0\\ -X, & X\le 0\end{cases}$$

Se trata en ambos casos de variables aleatorias no negativas y podemos aplicarles la propiedad anterior. Tendremos pues

$$E(X^+)=\int_0^\infty (1-G(y))\,dy=\int_0^\infty P\{X^+>y\}\,dy$$, donde G es la f.d. de X^+

pero teniendo en cuenta la definición de X^+, $P(X^+>y)=P(X>y)$, $y\ge 0$ y por tanto

$$E(X^+)=\int_0^\infty P(X>x)\,dx .$$

Por otra parte

$$E(X^-)=\int_0^\infty (1-H(y))\,dy=\int_0^\infty P\{X^->y\}\,dy$$, donde H es la f.d. de X^-

y teniendo en cuenta la definición de X^- tenemos, $P(X^->y)=P(X<-y)=$ $=P(X\le -y)-P(X=-y)$ y por tanto

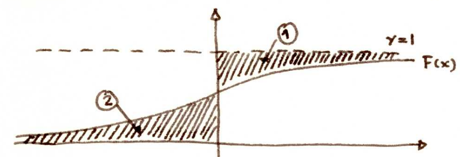
$$E(X^-)=\int_0^\infty P(X\le -x)\,dx-\int_0^\infty P(X=-x)\,dx=\int_{-\infty}^0 P(X\le x)\,dx-\int_{-\infty}^0 P(X=x)\,dx$$

pero sea cual sea el tipo de v.a. de la que nos ocupemos la función $P(X=x)$ es distinta de cero a lo sumo en un conjunto numerable y por tanto su integral de Lebesgue es nula. En definitiva

$$E(X)=\int_0^\infty P(X>x)\,dx-\int_{-\infty}^0 P(X\le x)\,dx$$

La implicación contraria se deriva del hecho de ser $|X|=X^+ + X^-$ y por tanto, si $E(X^+)<+\infty$ y $E(X^-)<+\infty$, entonces $E(|X|)=E(X^+)+E(X^-)$ y además es finita.

Comentario.- Obsérvese que la anterior expresión para la $E(X)$ tiene una significación geométrica en función de las áreas que cubren las funciones $F(x)$ y $1-F(x)$ en los semiejes negativo y positivo respectivamente. Concretamente la diferencia entre las áreas ① y ② de la figura.



Propiedades de la varianza

1) $\sigma^2(c)=0$, si c es una constante.

2) $\sigma^2(c\,g(X))=c^2\,\sigma^2[g(X)]$. En particular $\sigma^2(cX)=c^2\sigma^2(X)$

3) $\sigma^2(g(X)+d)=E\{[g(X)+d]-E[g(X)+d]\}^2=E\{g(X)-E[g(X)]+d-E(d)\}^2=$
$=E\{[g(X)-E(g(X))]+d-d\}^2=E\{g(X)-E(g(X))\}^2=\sigma^2(g(X))$.

Combinando 2 y 3 tenemos

4) $\sigma^2(c\,g(X)+d)=c^2\,\sigma^2(g(X))$

5) $\sigma^2[g(X)]=E[g(X)-E(g(X))]^2=E\{[g(X)]^2-2g(X).E(g(X))+[E(g(X))]^2\}=$
$=E[g(X)]^2-2[E(g(X))]^2+[E(g(X))]^2=E[g(X)]^2-[E(g(X))]^2$

lo que para una variable aleatoria se escribe

$$\sigma^2(X)=E[X^2]-[E(X)]^2.$$

TEOREMA.- Para una v.a. X, con $E|X|^2<+\infty$, la $E[X-c]^2$ se minimiza para $c=E(X)$, es decir, para la varianza.

Demostración.-

$$\sigma^2(X)=E[X-E(X)]^2=E\{X-c-[E(X)-c]\}^2=E\{[X-c]^2-2(X-c)(E(X)-c)+[E(X)-c]^2\}=E[X-c]^2-2(E(X)-c)^2+[E(X)-c]^2=E(X-c)^2-[E(X)-c]^2$$

y siendo $[E(X)-c]^2\ge 0$, entonces $\sigma^2(X)\le E[X-c]^2$, alcanzándose la igualdad para $c=E(X)$.

Tipificación de una v.a.

Una v.a. X se dice que está tipificada cuando $E(X)=0$ y $\sigma^2(X)=1$. Las propiedades anteriormente expuestas de la esperanza y de la varianza permiten obtener a partir de cualquier v.a. con $E(X)=\mu$ y $\sigma^2(X)=\sigma^2$, una nueva variable Y que está tipificada. En efecto, definamos

$$Y = \frac{X - E(X)}{\sqrt{\sigma^2(X)}} = \frac{X-\mu}{\sigma}$$

entonces

$$E(Y) = \frac{1}{\sigma}\left(E(X)-\mu\right) = 0$$

$$\sigma^2(Y) = \frac{1}{\sigma^2}\,\sigma^2(X-\mu) = \frac{1}{\sigma^2}\,\sigma^2(X) = 1.$$

Relación entre los momentos centrales y los momentos respecto del origen.

Los momentos respecto del origen se definen mediante

$$m_k = E(X^k)$$

mientras que los momentos centrales, son

$$\mu_k = E(X-\mu)^k \quad , \text{donde } \mu = E(X).$$

Recordando el desarrollo del binomio de Newton (k entero positivo), tendremos

$$\mu_k = E\left[(X-\mu)^k\right] = E\left[X^k - \binom{k}{1}X^{k-1}\cdot\mu + \binom{k}{2}X^{k-2}\cdot\mu^2 + \cdots + (-1)^k\mu^k\right] =$$

$$= m_k - \binom{k}{1}\mu\cdot m_{k-1} + \binom{k}{2}\mu^2\cdot m_{k-2} + \cdots + (-1)^k\cdot\mu^k$$

Por otra parte

$$m_k = E(X^k) = E\left[\left((X-\mu)+\mu\right)^k\right] = E\left[(X-\mu)^k + \binom{k}{1}\mu\cdot(X-\mu)^{k-1} + \cdots + \mu^k\right] =$$

$$= \mu_k + \binom{k}{1}\mu\cdot\mu_{k-1} + \binom{k}{2}\mu^2\cdot\mu_{k-2} + \cdots + \mu^k.$$

Estas relaciones nos permiten obtener unos de otros momentos conociendo tan sólo unos de ellos.

ALGUNAS DESIGUALDADES DE INTERES

Se trata en este apartado de obtener algunas desigualdades para los momentos de una v.a. El principal resultado es el teorema que demostraremos a continuación, que proporciona una cota superior para las colas de probabilidad en términos de algún momento de la v.a.

TEOREMA.- Sea $g(X)$ una función medible, no negativa de una v.a. X. Si existe $E[g(X)]$ entonces, para todo $\varepsilon>0$, tenemos

$$P\{g(X) \geq \varepsilon\} \leq \frac{E[g(X)]}{\varepsilon}.$$

Demostración.- Recordemos que

$$E[g(X)] = \int_\Omega g(X)\,dP$$

recordemos también que toda integral de una función medible no negativa es una medida sobre la familia $\mathcal{A}$ de los conjuntos medibles. Sea

$$A = \{\omega\in\Omega;\ g(X(\omega)) < \varepsilon\}$$

y

$$A^c = \{\omega\in\Omega;\ g(X(\omega)) \geq \varepsilon\}.$$

Tendremos, $\Omega = A\cup A^c$, y de aquí

$$E[g(X)] = \int_\Omega g(X)\,dP = \int_{A\cup A^c} g(X)\,dP = \int_A g(X)\,dP + \int_{A^c} g(X)\,dP$$

pero siendo $g(X)\geq 0$, también $\int_A g(X)\,dP \geq 0$ y por tanto

$$E[g(X)] \geq \int_{A^c} g(X)\,dP \geq \varepsilon\int_{A^c} dP = \varepsilon\cdot P(A^c)$$

luego

$$P(A^c) = P\{g(X)\geq\varepsilon\} \leq \frac{E[g(X)]}{\varepsilon}.$$

COROLARIO I (Desigualdad de MARKOV).- Hagamos $g(X) = |X|^r$ y $\varepsilon = k^r$, con $r>0, k>0$.

Entonces

$$P\{|X|^r \geq k^r\} = P\{|X|\geq k\} \leq \frac{E[|X|^r]}{k^r}.$$

COROLARIO 2 (Desigualdad de Tchebycher).- Apoyándonos en la anterior desigualdad para $g(X) = |X - E(X)|^2 = |X-\mu|^2$, con $\mu = E(X)$ y para $\varepsilon = k^2 \cdot \sigma^2(X) = k^2 \cdot \sigma^2$, con $\sigma^2 = \sigma^2(X)$, llegamos a la conocida desigualdad de Tchebycher

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

que en ocasiones, si tomamos $\varepsilon = k^2$, también se nos presenta de la forma

$$P\left(|X-\mu| \geq k\right) \leq \frac{\sigma^2}{k^2}.$$

Observación.- Si X es una v.a. con $E(X) = \mu$ y $\sigma^2(X) = 0$, la desigualdad de Tchebycher nos conduce a

$$P(|X-\mu| \geq k) = 0 \quad \forall k > 0.$$

Sea $A_n = \{\omega : |X(\omega)-\mu| \geq \frac{1}{n}\}$, tendremos $P(A_n) = 0$. Además, por construcción $A_n \subset A_{n+1}$ por lo que $A_n^c \supset A_{n+1}^c$, siendo $P(A_n^c) = 1$. Si pasamos al límite tendremos

$$P(\cap A_n^c) = \lim_n P(A_n^c) = 1$$

pero

$$\cap A_n^c = \{\omega : |X(\omega)-\mu| < \tfrac{1}{n}, \forall n\} = \{\omega : |X(\omega)-\mu| = 0\} = \{\omega : X(\omega) = \mu\}$$

es decir

$$P(X = \mu) = 1.$$

La v.a. es pues constante con probabilidad 1.

En otro orden de cosas es posible mejorar, en ocasiones, la desigualdad de Tchebycher si suponemos la existencia de momentos de orden superior. Vamos a comprobarlo, pero previamente necesitaremos el siguiente Lema

LEMA.- Sea X una v.a. con $E(X) = 0$ y $\sigma^2(X) = \sigma^2$. Entonces

$$P(X \geq x) \leq \frac{\sigma^2}{\sigma^2 + x^2}, \text{ si } x > 0$$

y

$$P(X > x) \geq \frac{x^2}{\sigma^2 + x^2}, \text{ si } x < 0.$$

Demostración.- Sea $h(t) = (t+c)^2$, $c > 0$. Se trata de una función medible no negativa. Entonces, para $0 < x < t$, tenemos $h(t) \geq (x+c)^2$. De aquí se sigue

$$P(X > x) \leq P\{h(X) \geq (x+c)^2\} \leq \frac{E(X+c)^2}{(x+c)^2}, \quad c > 0,\ x > 0.$$

Ahora bien, como $E(X) = 0$, entonces $E(X^2) = \sigma^2$ y por tanto

$$P(X > x) \leq \frac{\sigma^2 + c^2}{(x+c)^2}, \quad c > 0,\ x > 0$$

pero $f(c) = \frac{\sigma^2 + c^2}{(x+c)^2}$ tiene un mínimo para $c = \sigma^2/x$ y por tanto

$$P(X > x) \leq \frac{\sigma^2 + \sigma^4/x^2}{\left(x + \frac{\sigma^2}{x}\right)^2} = \frac{\sigma^2(x^2+\sigma^2)}{(x^2+\sigma^2)^2} = \frac{\sigma^2}{\sigma^2 + x^2}, \quad x > 0.$$

La otra desigualdad se demuestra de forma similar. Además, se puede demostrar que estas dos desigualdades no pueden mejorarse.

Con la ayuda del lema podemos ahora obtener una nueva desigualdad que mejora, en ocasiones, la desigualdad de Tchebycher.

TEOREMA.- Sea X una v.a. tal que $E(|X|^4) < +\infty$, $E(X) = 0$ y $E(X^2) = \sigma^2$. Entonces

$$P\{|X| \geq k\sigma\} \geq \frac{\mu_4 - \sigma^4}{\mu_4 + \sigma^4 k^4 - 2k^2\sigma^4}, \text{ para } k > 1, \text{ donde } \mu_4 = E(X^4).$$

Demostración.- En la primera de las desigualdades del lema anterior hagamos $X = (X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)$ y $x = 1$. Entonces

$$P\{X^2 - \sigma^2 \geq k^2\sigma^2 - \sigma^2\} \geq \frac{\sigma^2\{(X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)\}}{1 + \sigma^2\{(X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)\}} =$$

$$= \frac{E\{[(X^2-\sigma^2) - E(X^2-\sigma^2)]^2\}}{\sigma^4(k^2-1)^2 + E\{[(X^2-\sigma^2) - E(X^2-\sigma^2)]^2\}} = \frac{E\{(X^2-\sigma^2)^2\}}{\sigma^4(k^2-1)^2 + E\{(X^2-\sigma^2)^2\}} =$$

$$= \frac{\mu_4 - \sigma^4}{\sigma^4(k^2-1)^2 + \mu_4 - \sigma^4} = \frac{\mu_4 - \sigma^4}{\mu_4 + \sigma^4 k^4 - 2k^2\sigma^4}, \quad k > 1.$$

Puede demostrarse que esta última cota mejora la de Tchebycher si $k^2 \geq \mu_4/\sigma^4$ mientras que si $1 \leq k^2 < \mu_4/\sigma^4$ es peor que aquella.

Una propiedad que asegura la existencia de los momentos

Ya hemos visto que la existencia de la esperanza de una determinada función aleatoria $g(X)$ está supeditada a la existencia de la correspondiente integral, es decir, a la integrabilidad de la función. El teorema que demostramos a continuación nos asegura la existencia de todos los momentos de una variable si ésta verifica ciertas condiciones:

TEOREMA.- Sea una v.a. X satisfaciendo

$$\frac{P\{|X|>\alpha k\}}{P\{|X|>k\}} \to 0 , \text{ cuando } k \to \infty , \ \forall \alpha > 1 .$$

Entonces X posee todos los momentos. (Obsérvese que si $\alpha = 1$, el límite es 1, mientras que si $\alpha < 1$, $P\{|X|>\alpha k\} \geq P\{|X|>k\}$ y el límite no puede ser 0).

Demostración.- Sea $\varepsilon > 0$ (cuyo valor fijaremos más tarde), elegimos k_0 de manera que

$$\frac{P\{|X|>\alpha k\}}{P\{|X|>k\}} < \varepsilon . \quad \forall k \geq k_0 .$$

Así mismo elegimos k_1 de manera que $P\{|X|>k\} < \varepsilon$, $\forall k \geq k_1$.

Sea $N = \max(k_1, k_0)$. Para cada entero r, fijo, tenemos

$$\frac{P\{|X|>\alpha^r k\}}{P\{|X|>k\}} = \prod_{p=1}^{r} \frac{P\{|X|>\alpha^p k\}}{P\{|X|>\alpha^{p-1} k\}} , \quad k \geq N$$

Como $k' = \alpha^{p-1} k > k \geq k_0$, tendremos que

$$\frac{P\{|X|>\alpha^p k\}}{P\{|X|>\alpha^{p-1} k\}} = \frac{P\{|X|>\alpha k'\}}{P\{|X|>k'\}} < \varepsilon \quad , p = 1, \dots, r$$

y por tanto

$$\frac{P\{|X|>\alpha^r k\}}{P\{|X|>k\}} < \varepsilon^r \longrightarrow P\{|X|>\alpha^r k\} < \varepsilon^{r+1} , \quad k \geq N .$$

Por otra parte, como la variable aleatoria $|X|^n$ es no negativa, utilizando una propiedad antes demostrada tenemos

$$E[|X|^n] = \int_0^\infty [1 - F_{|X|^n}(y)] \, dy$$

que en términos de la f.d. de la v.a. $|X|$, y tras hacer el pertinente cambio de variable, adopta la forma:

$$E[|X|^n] = n \int_0^\infty x^{n-1} . P\{|X|>x\} \, dx =$$

$$= n \int_0^N x^{n-1} . P\{|X|>x\} \, dx + n \int_N^\infty x^{n-1} . P\{|X|>x\} \, dx$$

puesto que la primera integral es finita, necesitamos demostrar la finitud de la segunda para comprobar la existencia de la $E(X^n)$, entonces

$$\int_N^\infty x^{n-1} . P\{|X|>x\} \, dx = \sum_{r=1}^\infty \int_{\alpha^{r-1}N}^{\alpha^r N} x^{n-1} . P\{|X|>x\} \, dx \leq$$

$$\leq \sum_{r=1}^\infty (\alpha^r N)^{n-1} . P\{|X|>\alpha^{r-1} N\} [\alpha^r N - \alpha^{r-1} N] \leq$$

$$\leq \sum_{r=1}^\infty (\alpha^r N)^{n-1} . \varepsilon^r . 2\alpha^r N = 2N^n \sum_{r=1}^\infty (\alpha^n \varepsilon)^r =$$

$$= 2N^n . \frac{\alpha^n \varepsilon}{1 - \alpha^n \varepsilon} < \infty$$

siempre que ε haya sido adecuadamente elegido $(\varepsilon \alpha^n < 1)$. Tenemos pues que $E(|X|^n) < \infty$, $\forall n$, y la función posee todos sus momentos.

MOMENTOS DE ALGUNAS VARIABLES ALEATORIAS

Caso discreto

1.- **Bernouilli**. Recordemos que se trata de una v.a. discreta que toma los valores 0 y 1 con probabilidad p y q respectivamente, de manera que $p + q = 1$.

$$E(X) = 0.p + 1.q = q$$

$$E(X^2) = 0^2 . p + 1^2 . q = q \qquad \text{var}(X) = E(X^2) - [E(X)]^2 = q - q^2 = q(1-q) = q.p.$$

2.- **Binomial**. Sea X una v.a. Binomial con parámetros $n.p$. Veamos cual es su esperanza.

$$E(X) = \sum_{x=0}^{n} x.P(X=x) = \sum_{x=0}^{n} x.\binom{n}{x} p^x . q^{n-x} = \sum_{x=0}^{n} x \frac{n(n-1)\dots(n-x+1)}{x!} p^x . q^{n-x} =$$

$$= np \sum_{x=1}^{n} \frac{(n-1)\dots(n-x+1)}{(x-1)!} p^{x-1} . q^{n-x} = n.p \sum_{x-1=0}^{n-1} \binom{n-1}{x-1} p^{x-1} . q^{n-x} = np$$

Para obtener la $\text{var}(X)$, observemos que

$$E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X)$$

y de aquí

$$var(X) = E(X(X-1)) + E(X) - [E(X)]^2.$$

ahora bien

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x . q^{n-x} = \sum_{x=0}^{n} x(x-1) \frac{n(n-1)\dots(n-x+1)}{x!} p^x . q^{n-x} =$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)\dots(n-x+1)}{(x-2)!} p^{x-2} . q^{n-x} = n(n-1)p^2 \sum_{x-2=0}^{n-2} \binom{n-2}{x-2} p^{x-2} q^{n-x} =$$

$$= n(n-1)p^2.$$

Por tanto

$$var(X) = n(n-1)p^2 + np - n^2p^2 = np[(n-1)p + 1 - np] = np(1-p) = n.p.q$$

Nota. Los momentos del tipo $E[X(X-1)\dots(X-n+1)]$ reciben el nombre de momentos factoriales de n-ésimo orden. Su utilidad ha quedado de manifiesto en la obtención de la varianza anterior.

3.- Poisson. Si X es una v.a. Poisson con parámetro λ

$$E(X) = \sum_{x=0}^{\infty} x . e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda . e^{-\lambda} \sum_{x-1=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda . e^{-\lambda} . e^{\lambda} = \lambda.$$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda} \sum_{x-2=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2$$

De aquí

$$var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Obsérvese que para una v.a. $P(\lambda)$, los momentos factoriales de orden n, valen λ^n. En efecto

$$E(X(X-1)\dots(X-n+1)) = \sum_{x=0}^{\infty} x(x-1)\dots(x-n+1) e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^n e^{-\lambda} \sum_{x-n=0}^{\infty} \frac{\lambda^{x-n}}{(x-n)!} = \lambda^n.$$

Caso Continuo

1.- Uniforme. Sea X una v.a. uniforme en el intervalo $[a,b]$, entonces

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_a^b x . \frac{1}{b-a}\, dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \left[\frac{x^2}{2}\right]_a^b = \frac{1}{2}(a+b)$$

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{b-a} \left[\frac{x^3}{3}\right]_a^b = \frac{1}{3} \frac{b^3-a^3}{b-a}$$

$$var(X) = \frac{b^3-a^3}{3(b-a)} - \left\{\frac{1}{2}(b+a)\right\}^2 = \frac{1}{3} . \frac{b^3-a^3}{b-a} - \frac{1}{4} . (b+a)^2 = \frac{4(b^3-a^3) - 3(b^2+a^2+2ab)(b-a)}{12(b-a)} =$$

$$= \frac{b^3 - 3b^2a + 3ba^2 - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

2.- Normal tipificada. Si X es una v.a. $N(0,1)$, entonces $E(X^{2n+1}) = 0$ y $E(X^{2n}) = \frac{(2n)!}{2^n (n!)}$, $n \geq 0$.

En efecto:

$$E(X^{2n+1}) = \int_{-\infty}^{+\infty} x^{2n+1} f(x)\, dx = \int_{-\infty}^{+\infty} x^{2n+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x^{2n+1} e^{-x^2/2}\, dx +$$

$$+ \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx = -\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx = 0$$

Por otra parte

$$E(X^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} . 2 \int_0^{\infty} x^{2n} e^{-x^2/2}\, dx$$

ahora bien

$$\int_0^{\infty} x^{2n} . e^{-x^2/2}\, dx = \int_0^{\infty} -x^{2n-1}(-x)\, e^{-x^2/2}\, dx = \int_0^{\infty} -x^{2n-1}\, d e^{-x^2/2} = \left[-x^{2n-1} . e^{-x^2/2}\right]_0^{\infty} +$$

$$+ (2n-1) \int_0^{\infty} x^{2n-2} . e^{-x^2/2}\, dx = (2n-1) \int_0^{\infty} x^{2n-2} . e^{-x^2/2}\, dx$$

Si designamos $E(X^{2n})$ por m_{2n}, tendremos

$m_{2n} = (2n-1)\, m_{2n-2}$

$m_{2n-2} = (2n-3)\, m_{2n-4}$

- - - - - - - - -

$m_2 = 1 . m_0$

$m_0 = 1$, pues $m_0 = E(X^0) = E(1) = 1$

y de aquí

$$var(X) = E(X(X-1)) + E(X) - [E(X)]^2.$$

ahora bien

$$E(X(X-1)) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x . q^{n-x} = \sum_{x=0}^{n} x(x-1)\frac{n(n-1)\cdots(n-x+1)}{x!} p^x . q^{n-x} =$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)\cdots(n-x+1)}{(x-2)!} p^{x-2} . q^{n-x} = n(n-1)p^2 \sum_{x-2=0}^{n-2} \binom{n-2}{x-2} p^{x-2} q^{n-x} =$$

$$= n(n-1)p^2.$$

Por tanto

$$var(X) = n(n-1)p^2 + np - n^2p^2 = np[(n-1)p + 1 - np] = np(1-p) = n.p.q$$

Nota. Los momentos del tipo $E[X(X-1)\cdots(X-n+1)]$ reciben el nombre de momentos factoriales de n-simo orden. Su utilidad ha quedado de manifiesto en la obtención de la varianza anterior.

3.- Poisson. Si X es una v.a. Poisson con parámetro λ

$$E(X) = \sum_{x=0}^{\infty} x . e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda . e^{-\lambda} \sum_{x-1=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda . e^{-\lambda} . e^{\lambda} = \lambda.$$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda} \sum_{x-2=0}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = \lambda^2$$

De aquí

$$var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Obsérvese que para una v.a. $P(\lambda)$, los momentos factoriales de orden n, valen λ^n. En efecto

$$E(X(X-1)\cdots(X-n+1)) = \sum_{x=0}^{\infty} x(x-1)\cdots(x-n+1)\, e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^n e^{-\lambda} \sum_{x-n=0} \frac{\lambda^{x-n}}{(x-n)!} = \lambda^n.$$

Caso Continuo

1.- Uniforme. Sea X una v.a. uniforme en el intervalo $[a,b]$, entonces

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_a^b x . \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{2}(a+b)$$

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}\,\frac{b^3-a^3}{b-a}$$

$$var(X) = \frac{b^3-a^3}{3(b-a)} - \left\{\frac{1}{2}(b+a)\right\}^2 = \frac{1}{3}\cdot\frac{b^3-a^3}{b-a} - \frac{1}{4}.(b+a)^2 = \frac{4(b^3-a^3) - 3(b^2+a^2+2ab)(b-a)}{12(b-a)} =$$

$$= \frac{b^3 - 3b^2a + 3ba^2 - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

2.- Normal tipificada. Si X es una v.a. $N(0,1)$, entonces $E(X^{2n+1}) = 0$ y $E(X^{2n}) = \frac{(2n)!}{2^n (n!)}$, $n \geq 0$.

En efecto:

$$E(X^{2n+1}) = \int_{-\infty}^{+\infty} x^{2n+1} f(x)\, dx = \int_{-\infty}^{+\infty} x^{2n+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x^{2n+1} e^{-x^2/2}\, dx +$$

$$+ \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx = -\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\, dx = 0$$

Por otra parte

$$E(X^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} . 2\int_0^{\infty} x^{2n} e^{-x^2/2}\, dx$$

ahora bien

$$\int_0^{\infty} x^{2n} . e^{-x^2/2}\, dx = \int_0^{\infty} -x^{2n-1}(-x)\, e^{-x^2/2}\, dx = \int_0^{\infty} -x^{2n-1}\, de^{-x^2/2} = \left[-x^{2n-1} . e^{-x^2/2}\right]_0^{\infty} +$$

$$+ (2n-1)\int_0^{\infty} x^{2n-2} . e^{-x^2/2}\, dx = (2n-1)\int_0^{\infty} x^{2n-2} . e^{-x^2/2}\, dx$$

Si designamos $E(X^{2n})$ por m_{2n}, tendremos

$m_{2n} = (2n-1)\, m_{2n-2}$

$m_{2n-2} = (2n-3)\, m_{2n-4}$

- - - - - - - - -

$m_2 = 1 . m_0$

$m_0 = 1$, pues $m_0 = E(X^0) = E(1) = 1$

multiplicando ambos lados de las igualdades tenemos:

$$m_{2n} = (2n-1)(2n-3)\cdots 1 = \frac{2n(2n-1)(2n-2)(2n-3)\cdots 2.1}{2n(2n-2)\cdots 2} =$$

$$= \frac{(2n)!}{2(n).2(n-1)\cdots 2.1} = \frac{(2n)!}{2^n.(n!)}.$$

Entonces para la $E(X)$ y para la $var(X)$, tendremos

$$E(X)=0$$

$$var(X) = E(X^2) = \frac{2!}{2^1\, 1!} = \frac{2}{2} = 1.$$

Observación.- Si queremos obtener la Esperanza y la Varianza de una v.a. $N(\mu,\sigma^2)$, tendremos que al definir una nueva variable, a partir de ésta, mediante

$$Z = \frac{X-\mu}{\sigma}$$

la variable Z resultará ser $N(0,1)$, por tanto

$$E(Z) = 0$$

pero

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\left[E(X)-\mu\right] = 0 \rightarrow E(X)=\mu$$

además

$$var(Z) = 1 = var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} var(X) \rightarrow var(X) = \sigma^2$$

Así pues, la esperanza y la varianza de una $N(\mu,\sigma^2)$ coinciden con los parámetros de la distribución.

3.- Gamma. Sea X una v.a. Gamma con parámetros α y β.

$$E(X) = \int_0^\infty x\, f(x)\, dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x.x^{\alpha-1} e^{-x/\beta}\, dx = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha . \frac{e^{-x/\beta}}{-\beta}\, dx =$$

$$= \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha\, d e^{-x/\beta} = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\left\{\left[x^\alpha . e^{-x/\beta}\right]_0^\infty - \alpha\int_0^\infty x^{\alpha-1}.e^{-x/\beta}\, dx\right\} =$$

$$= \frac{\alpha\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^{\alpha-1}.e^{-x/\beta}\, dx = \alpha\beta, \quad \text{pues} \int_0^\infty x^{\alpha-1}.e^{-x/\beta}\, dx = \Gamma(\alpha).\beta^\alpha$$

Análogamente

$$E(X^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+1}.e^{-x/\beta}\, dx = \beta^2\alpha(\alpha+1).$$

Para la varianza

$$var(X) = E(X^2) - [E(X)]^2 = \beta^2\alpha(\alpha+1) - \beta^2\alpha^2 = \beta^2\alpha.$$

Casos particulares:

a) Si X es una χ^2_r, tendremos que entonces $\alpha = r/2$, $\beta = 2$, con r entero, por tanto

$$E(X) = \frac{r}{2}.2 = r \qquad , \quad var(X) = \frac{r}{2}.4 = 2r$$

b) Si X es una exponencial negativa con parámetro λ, entonces $\alpha = 1$, $\beta = 1/\lambda$ y por tanto

$$E(X) = 1.1/\lambda = \frac{1}{\lambda} \qquad , \quad var(X) = 1.\frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

4.- Cauchy. Esta situación es interesante por cuanto se trata de una variable que carece de momentos. En efecto

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \frac{\sigma}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{\sigma^2 + (x-\mu)^2}$$

podemos suponer $\mu = 0$, $\sigma = 1$, entonces

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{1+x^2} = \frac{1}{2\pi}\left[\log(1+x^2)\right]_{-\infty}^{+\infty}$$ que es una indeterminación.

De cualquier forma obsérvese que

$$E(|X|) = \frac{1}{\pi}\int_{-\infty}^{+\infty}|x|.\frac{dx}{1+x^2} = \frac{2}{\pi}\int_0^\infty \frac{x\,dx}{1+x^2} = \frac{1}{\pi}\left[\log(1+x^2)\right]_0^{+\infty} = +\infty$$

y por tanto no existe la esperanza.

Observación.- El valor principal de Cauchy para una integral impropia se define mediante

$$\lim_{b\to\infty}\int_{-b}^{b} g(x)\, dx$$

Obsérvese que en nuestro caso, el V.P.C de la integral anterior es finito, concretamente 0, aunque la integral no es convergente (desde luego en el caso de convergencia ambos valores coinciden).

multiplicando ambos lados de las igualdades tenemos:

$$m_{2n} = (2n-1)(2n-3) \ldots 1 = \frac{2n(2n-1)(2n-2)(2n-3)\ldots 2.1}{2n(2n-2)\ldots 2} =$$

$$= \frac{(2n)!}{2(n).2(n-1)\ldots 2.1} = \frac{(2n)!}{2^n.(n!)}.$$

Entonces para la $E(X)$ y para la $var(X)$, tendremos

$$E(X)=0$$

$$var(X) = E(X^2) = \frac{2!}{2!\,1!} = \frac{2}{2} = 1.$$

Observación.- Si queremos obtener la Esperanza y la varianza de una v.a. $N(\mu,\sigma^2)$, tendremos que al definir una nueva variable, a partir de ésta, mediante

$$Z = \frac{X-\mu}{\sigma}$$

la variable Z resultaba ser $N(0,1)$, por tanto

$$E(Z) = 0$$

pero

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\left[E(X)-\mu\right] = 0 \rightarrow E(X)=\mu$$

además

$$var(Z) = 1 = var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}var(X) \rightarrow var(X) = \sigma^2$$

Así pues, la esperanza y la varianza de una $N(\mu,\sigma^2)$ coinciden con los parámetros de la distribución.

3.- Gamma. Sea X una v.a. Gamma con parámetros α y β.

$$E(X) = \int_0^\infty x\,f(x)\,dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x.x^{\alpha-1}e^{-x/\beta}dx = \frac{-\beta}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^\alpha.\frac{e^{-x/\beta}}{-\beta}dx =$$

$$= \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha\, d e^{-x/\beta} = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\left\{\left[x^\alpha.e^{-x/\beta}\right]_0^\infty - \alpha\int_0^\infty x^{\alpha-1}.e^{-x/\beta}dx\right\} =$$

$$= \frac{\alpha\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^{\alpha-1}.e^{-x/\beta}dx = \alpha\beta \,, \text{ pues } \int_0^\infty x^{\alpha-1}.e^{-x/\beta}dx = \Gamma(\alpha).\beta^\alpha$$

Análogamente

$$E(X^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+1}.e^{-x/\beta}dx = \beta^2\alpha(\alpha+1).$$

Para la varianza

$$var(X) = E(X^2) - [E(X)]^2 = \beta^2\alpha(\alpha+1) - \beta^2\alpha^2 = \beta^2\alpha.$$

Casos particulares:

a) Si X es una χ^2_r, tendremos que entonces $\alpha = r/2$, $\beta=2$, con r entero, por tanto

$$E(X) = \frac{r}{2}.2 = r \qquad , \quad var(X) = \frac{r}{2}.4 = 2r$$

b) Si X es una exponencial negativa con parámetro λ, entonces $\alpha=1$, $\beta=1/\lambda$ y por tanto

$$E(X) = 1.1/\lambda = \frac{1}{\lambda} \qquad , \quad var(X) = 1.\frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

4.- Cauchy. Esta situación es interesante por cuanto se trata de una variable que carece de momentos. En efecto

$$E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \frac{\sigma}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{\sigma^2+(x-\mu)^2}$$

podemos suponer $\mu=0$, $\sigma=1$, entonces

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{1+x^2} = \frac{1}{2\pi}\left[\log(1+x^2)\right]_{-\infty}^{+\infty} \text{ que es una indeterminación.}$$

De cualquier forma obsérvese que

$$E(|X|) = \frac{1}{\pi}\int_{-\infty}^{+\infty}|x|.\frac{dx}{1+x^2} = \frac{2}{\pi}\int_0^\infty\frac{x\,dx}{1+x^2} = \frac{1}{\pi}\left[\log(1+x^2)\right]_0^{+\infty} = +\infty$$

y por tanto no existe la esperanza.

Observación.- El valor principal de Cauchy para una integral impropia se define mediante

$$\lim_{b\to\infty}\int_{-b}^{b} g(x)\,dx$$

Obsérvese que en nuestro caso, el V.P.C de la integral anterior es finito, concretamente 0, aunque la integral no es convergente (desde luego en el caso de convergencia ambos valores coinciden).

OTRAS CARACTERISTICAS NUMERICAS DE UNA DISTRIBUCION DE PROBABILIDAD

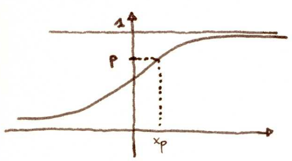
El concepto de esperanza nos ha permitido obtener los momentos de diversos tipos de una distribución de probabilidad (correctamente de la v.a. correspondiente), que, en definitiva, no son más que características numéricas que nos permiten un conocimiento, sino exhaustivo, bastante aproximado y, en cualquier caso, de forma rápida de la distribución de probabilidad asociada a la v.a. Pero no son estas las únicas características numéricas de interés, a continuación introducimos otras, conocidas genéricamente como cuantiles.

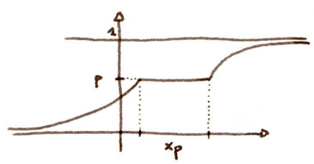
Definición.- Sea X una v.a. con f.d. F y sea p, tal que $0<p<1$. Toda cantidad x_p que satisface simultáneamente

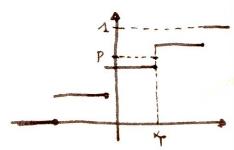
$$P(X \leq x_p) = F(x_p) \geq p \quad y \quad P(X \geq x_p) \geq 1-p$$

se la denomina cuantil de orden p de la v.a. X (o de la distribución de probabilidad correspondiente).

Dadas la definición y las propiedades de la f.d. un cuantil de orden p existe siempre, pero no tiene porque ser único. En particular todas las soluciones de la ecuaciones antes planteadas son cuantiles de orden p, pudiendo su número ser no numerable; por ejemplo, si la recta $y=p$ y la f.d. $F(x)$ tienen un intervalo común, todo el intervalo está constituido por cuantiles del orden deseado. Por contra si la función de distribución es estrictamente creciente la solución es única y único el cuantil. Veanse en las figuras algunas de las situaciones que pueden somir.







Algunos cuantiles tienen nombre propio. Si $p=0.1, 0.2, \ldots, 0.9$ a los x_p correspondientes se les denomina deciles, si $p=0.25, 0.5, 0.75$ a los correspondientes x_p se les conoce como cuartiles de la distribución y de ellos tiene especial importancia la mediana, que es el segundo cuartil, es decir el correspondiente a $p=1/2$.

Como ocurre con todos los cuantiles la mediana no tiene por que ser única, pero si la distribución de probabilidad es simétrica, el centro de simetría de la misma coincide, desde luego, con la mediana. El papel de la mediana es importante, como medida de centralización, especialmente en aquellas v.a. que carecen de media, puesto que, como ya hemos advertido, nunca carece v.a. de mediana. Esta posee además otra importante propiedad:

Propiedad de la mediana.- Sea X una v.a. continua. El momento absoluto $E(|X-c|)$ alcanza un mínimo si c se elige igual a la mediana de la distribución.

Demostración.- Observemos en primer lugar que si X es una v.a. continua aún cuando la mediana no sea única, verifica las igualdades de las condiciones que la definen. Es decir, si m es una mediana, tenemos

$$F(m) = \frac{1}{2} \quad y \quad 1-F(m) = \frac{1}{2}.$$

Hecha esta observación, tenemos

$$E(|X-c|) = \int_{-\infty}^{+\infty} |x-c|\, f(x)dx = \int_{-\infty}^{c} (c-x) f(x)dx + \int_{c}^{+\infty} (x-c) f(x)dx =$$

$$= \int_{-\infty}^{m} (c-m+m-x) f(x)dx + \int_{m}^{c} (c-x) f(x)\, dx + \int_{c}^{+\infty} (x-m+m-c) f(x)dx =$$

$$= \int_{-\infty}^{m} (m-x) f(x)\, dx + \int_{c}^{+\infty} (x-m) f(x)\, dx + \int_{-\infty}^{m} (c-m) f(x)dx + \int_{m}^{c} (c-x) f(x)\, dx +$$

$$+ \int_{c}^{+\infty} (m-c) f(x)\, dx = E(|X-m|) - \int_{m}^{c} (x-m) f(x)dx + \int_{m}^{c} (c-x) f(x)\, dx +$$

$$+ \int_{-\infty}^{m} (c-m) f(x)\, dx + \int_{m}^{c} (m-c) f(x)dx - \int_{m}^{c} (m-c) f(x)\, dx + \int_{c}^{+\infty} (m-c) f(x)\, dx =$$

$$= E(|X-m|) + \int_{-\infty}^{m} (c-m) f(x)\, dx + \int_{m}^{+\infty} (m-c) f(x)dx + \int_{m}^{c} (c-x-x+m-m+c) f(x)dx =$$

$$= E(|X-m|) + 2\int_{m}^{c} (c-x) f(x)\, dx + (c-m)\, F(m) + (m-c)\left[1-F(m)\right] =$$

$$= E(|X-m|) + 2\int_{m}^{c} (c-x) f(x)\, dx \qquad c>m$$

Analogamente, para $m>c$, $E(|X-c|) = E(|X-m|) + 2\int_{c}^{m} (x-c) f(x)dx$. En ambos casos el segundo sumando del segundo miembro de la igualdad es positivo para $c \neq m$, y por tanto es cierta la propiedad enunciada.

Otra característica de interés es la MODA cuya definición damos a continuación.

Definición.- Sea X una v.a. con función de densidad de probabilidad o de cuantía, f, según sea continua o discreta. Una moda de f, si existe, es cualquier número que maximiza $f(x)$.

De acuerdo con la definición, si f es diferenciable dos veces, la moda puede obtenerse mediante diferenciación de la función, todo se reduce a un problema clásico de máximos de una función. Sin embargo esta actuación no es posible en el caso discreto. Para el caso de la Binomial y de la distribución de Poisson los dos teoremas que vienen a continuación nos dan la solución correspondiente.

TEOREMA (moda de la Binomial).- Sea X una v.a. $B(n,p)$. Consideremos el número $(n+1)p$ y sea $[(n+1)p]=m$, la parte entera de $(n+1)p$. Entonces si $(n+1)p \neq [(n+1)p]$, $f(x)$ tiene una única moda para $x=m$. Si $(n+1)p=[(n+1)p]$, entonces las modas son dos, $x_1=m$ y $x_2=m-1$.

Demostración.- Tenemos

$$\frac{f(x)}{f(x-1)} = \frac{\binom{n}{x} p^x q^{n-x}}{\binom{n}{x-1} p^{x-1} q^{n-x+1}} = \frac{\frac{n!}{x!(n-x)!} p^x \cdot q^{n-x}}{\frac{n!}{(x-1)!(n-x+1)!} p^{x-1} \cdot q^{n-x+1}} = \frac{n-x+1}{x} \cdot \frac{p}{q}$$

por tanto $f(x) \geqslant f(x-1)$, cociente, si y sólo si

$(n-x+1)p \geqslant x(1-p)$, o bien $np-xp+p \geqslant x-xp$, o bien $(n+1)p \geqslant x$.

Así pues, si $(n+1)p$ no es entero, $f(x)$ aumenta para $x \leq [(n+1)p]$ y decrece a continuación. La moda será pues $m=[(n+1)p]$.

Si $(n+1)p$ es entero, entonces de los anteriores cálculos se deduce que para $x=(n+1)p$ $f(x)=f(x-1)$, luego en $x=(n+1)p-1$ hay también una moda.

TEOREMA (moda de una v.a. Poisson).- Sea X una v.a. $P(\lambda)$. Entonces si λ no es entero, $f(x)$ tiene una única moda en $x=[\lambda]$. Si λ es entero $f(x)$ tiene dos modas en $x_1=\lambda$, $x_2=\lambda-1$

Demostración.- Tenemos

$$\frac{f(x)}{f(x-1)} = \frac{e^{-\lambda} \cdot \lambda^x / x!}{e^{-\lambda} \cdot \lambda^{x-1}/(x-1)!} = \frac{\lambda}{x}$$

por tanto $f(x) \geqslant f(x-1)$ si y sólo si $\lambda \geqslant x$. Así si λ no es entero la función aumenta mientras $x \leq [\lambda]$ y después disminuye. El máximo se alcanza en $x=[\lambda]$. Si λ es entero, entonces para $x=\lambda$, $f(x)=f(x-1)$, luego el maximo se alcanza en $x=\lambda$ y en $x=\lambda-1$.

Capítulo 8

Esperanza y momentos de un vector aleatorio

ESPERANZA Y MOMENTOS DE UN VECTOR ALEATORIO

DEFINICION.- Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $\vec{X}$ una aplicación medible de $(\Omega, \mathcal{A})$ en (R^m, β^m), es decir, un vector aleatorio. Si $\vec{X}$ es una aplicación medible y g, definida de R^m en R, también lo es, la composición $g(\vec{X})$ es una variable aleatoria, cuya esperanza se define de la forma consabida. A saber, si $g(\vec{X})$ es integrable, entonces

$$E(g(\vec{X})) = \int_{\Omega} g(\vec{x})\, dP.$$

Ya sabemos que la anterior expresión tiene un significado u otro en función del tipo de variable aleatoria que $g(\vec{X})$ es. Así, para el caso discreto, sabemos que

$$E(g(\vec{X})) = \sum_{\vec{x} \in A} g(\vec{x}).P(\vec{X}=\vec{x}) \qquad \text{donde } A \text{ es tal que } P(\vec{X}\in A)=1.$$

Si nos situamos en el caso continuo

$$E(g(\vec{X})) = \int_{R^m} g(\vec{x})\, f(\vec{x})\, d\vec{x} = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1 \ldots x_m)\, f(x_1 \ldots x_m)\, dx_1 \ldots dx_m$$

donde $f(\vec{x})$ es la función de densidad de probabilidad del vector aleatorio $\vec{X}$.

Los conceptos de momentos centrales respecto del origen así como los momentos absolutos y los momentos sin más tienen una definición paralela a la establecida para el caso de una variable aleatoria estudiada anteriormente con detalle. Surgen sin embargo, en esta versión, situaciones nuevas de interés. En efecto:

a) Si $g(\vec{X}) = X_1^{n_1} \cdots X_m^{n_m}$, con $n_i \geq 0$, enteros para $i=1,\ldots,m$ a la correspondiente esperanza

$$E(g(\vec{X})) = E(X_1^{n_1} \cdots X_m^{n_m})$$

si existe, se le conoce como momento conjunto de orden $(n_1,\ldots,n_m)$. Se suele designar mediante $m_{n_1 \ldots n_m}$. Obsérvese que para $n_i=1$, $n_j=0$, $j\neq i$, tenemos el momento $m_{0\ldots1\ldots0}$ que coincide con la esperanza de la variable aleatoria X_i (no olvidemos que las componentes del vector aleatorio son todas variables aleatorias).

b) Análogamente, si $g(\vec{X}) = (X_1 - E(X_1))^{n_1} \cdots (X_m - E(X_m))^{n_m}$, el momento resultante al obtener la esperanza es conocido como el momento central conjunto de orden $(n_1 \ldots n_m)$. Los correspondientes momentos absolutos surgirían de manera similar. Por ser especialmente interesante vamos a estudiar el caso siguiente.

VECTOR ALEATORIO BIDIMENSIONAL.- Para un vector aleatorio bidimensional (X,Y), el momento central de orden $(1,1)$ es conocido con el nombre de covarianza y su definición, cuando existe la correspondiente esperanza es

$$cov(X,Y) = E((X-E(X))(Y-E(Y))) = E((X-m_{10})(Y-m_{01})) = E(XY) - E(X).E(Y) = m_{11} - m_{10}.m_{01}$$

Veamos a continuación una interesante propiedad

DESIGUALDAD DE CAUCHY-SCHWARZ.- Sean X e Y dos variables aleatorias con varianzas finitas. Entonces la $cov(X,Y)$ existe. Además

$$[E(XY)]^2 \leq E(X^2).E(Y^2)$$

verificándose la igualdad si y sólo si existe un número real α tal que $P\{(\alpha X + Y) = 0\} = 1$.

Demostración.- Para cualquier par de números reales a y b tenemos

$$|ab| \leq \frac{a^2+b^2}{2}$$

lo que significa que la $E(XY) < +\infty$ si $E(Y^2) < +\infty$ y la $E(X^2) < +\infty$. Por otra parte, para cualquier número real α, tenemos

$$E[(\alpha X + Y)^2] = \alpha^2 E(X^2) + 2\alpha E(XY) + E(Y^2)$$

tratándose de una función no negativa, $E((\alpha X+Y)^2) \geq 0$, $\forall \alpha$, lo que supone que la ecuación de segundo grado de la derecha de la igualdad tiene a lo sumo una raíz y por tanto su discriminante será no positivo, es decir

$$[E(XY)]^2 \leq E(X^2).E(Y^2).$$

Obsérvese que si $[E(XY)]^2 = E(X^2).E(Y^2)$ la ecuación tiene raíz doble, a saber $\alpha_0 = -\frac{E(XY)}{E(X^2)}$, lo que sustituido en la ecuación supone que $E[(\alpha_0 X + Y)^2] = 0$ y dadas las características de la función $(\alpha_0 X + Y)^2$ y las propiedades de la integral de una función no negativa tendremos $P\{(\alpha_0 X + Y) = 0\} = 1$.

Introducido el concepto de covarianza podemos ahora conocer el valor de la varianza de una suma finita de variables aleatorias.

VARIANZA DE UNA SUMA.- Sean $X_1,\ldots,X_n$ variables aleatorias cuyas varianzas existen. Sean $a_1,\ldots,a_n$ números reales. Definimos

$$S = \sum_{i=1}^{n} a_i X_i.$$

Entonces la varianza de S existe, y viene dada por

$$var(S) = \sum_{i=1}^{n} a_i^2 . var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i\neq j}}^{n} a_i a_j \, cov(X_i X_j).$$

En particular si las variables aleatorias son tales que todas las posibles covarianzas son nulas, entonces

$$var(S) = \sum_{i=1}^{n} a_i^2 . var(X_i).$$

Demostración.- El resultado se obtiene de desarrollar el cuadrado

$$\left[\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i)\right]^2$$

y después tomar esperanzas.

ALGUNAS DESIGUALDADES DE INTERES

Lema.- Para a y b reales cualesquiera, tenemos

$$|a+b|^r \leq C_r\left(|a|^r + |b|^r\right), \text{ donde } C_r = 1, \text{ si } 0 \leq r \leq 1 \text{ y } C_r = 2^{r-1} \text{ si } r > 1.$$

Demostración.- Para $r=0$ y $r=1$ la desigualdad es trivialmente cierta. Observemos por otra parte que bastará comprobar la desigualdad para $0 < a \leq b$. Así pues, $0 < a \leq b$, y hagamos $x = a/b$. Entonces

$$\frac{(a+b)^r}{a^r + b^r} = \frac{(1+x)^r}{1+x^r}$$

Sea $f(x) = (1+x)^r/(1+x^r)$, su derivada vale

$$f'(x) = \frac{r(1+x)^{r-1}}{(1+x^r)^2}\left(1 - x^{r-1}\right), \quad 0 < x \leq 1,\ r > 0.$$

Se sigue que $f'(x) > 0$, si $r > 1$, para $r = 1$, entonces $f'(x) = 0$ y para $r < 1$, entonces $f'(x) \leq 0$. En consecuencia

$$\max_{0 \leq x \leq 1} f(x) = f(0) = 1, \quad \text{si } r \leq 1$$

$$\max_{0 \leq x \leq 1} f(x) = f(1) = 2^{r-1}, \quad \text{si } r \geq 1.$$

Con lo que queda demostrado el lema.

Obsérvese que hubiéramos podido obtener, de inmediato, una cota del tipo

$$|a+b|^r \leq 2^r\left(|a|^r + |b|^r\right)$$

puesto que

$$|a+b| \leq \max\left(2|a|, 2|b|\right)$$

y de aquí al elevar a la r-ésima potencia obtenemos la desigualdad anterior.

Como una aplicación de esta desigualdad obtenida en el lema tenemos la siguiente propiedad.

Teorema.- Sean X e Y variables aleatorias y $r > 0$ un número fijo. Si $E(|X|^r)$ y $E(|Y|^r)$ son ambas finitas, entonces también lo es $E(|X+Y|^r)$.

Demostración.- Basta tomar $a = X$, $b = Y$ y aplicar la desigualdad anterior, luego tomamos esperanzas y llegamos al resultado enunciado en el teorema.

DESIGUALDAD DE HÖLDER.- Para x e y reales cualesquiera y para p y q, dos números reales positivos tales que $p > 1$ y $1/p + 1/q = 1$, tenemos

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$$

Demostración.- Observemos que para $x > 0$, la función $w = \log x$ es cóncava, por tanto para $0 < t < 1$ y para $x_1 > 0$, $x_2 > 0$, tenemos

$$\log\left[t x_1 + (1-t) x_2\right] \geq t \log x_1 + (1-t)\log x_2.$$

tomando antilogaritmos, tenemos

$$x_1^t \cdot x_2^{(1-t)} \leq t x_1 + (1-t) x_2$$

Elegimos ahora $x_1 = |x|^p$, $x_2 = |y|^q$, $t = 1/p$, $1-t = 1/q$, con $p > 1$ y $\frac{1}{p} + \frac{1}{q} = 1$, llegando entonces a

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$$

desigualdad conocida como la desigualdad de Hölder.

Aplicando la desigualdad de Hölder obtenemos otra conocida desigualdad en términos de

~~DESIGUALDAD DE MINKOWSKI.- Para $p \geq 1$,~~

$$\cancel{\left\{E\left[|X+Y|^p\right]\right\}^{1/p} \leq \left\{E\left(|X|^p\right)\right\}^{1/p} + \left\{E\left(|Y|^p\right)\right\}^{1/p}}$$

~~Demostración.- El caso $p=1$ es consecuencia inmediata del hecho $|X+Y| \leq |X| + |Y|$.~~
~~Sea $p > 1$, tenemos~~

$$\cancel{|X+Y|^p = |X+Y| \cdot |X+Y|^{p-1} \leq |X| \cdot |X+Y|^{p-1} + |Y| \cdot |X+Y|^{p-1}}$$

~~tomando esperanzas y utilizando la desigualdad de~~

esperanzas que es, a menudo, conocida también como la desigualdad de Hölder.

DESIGUALDAD DE HÖLDER PARA ESPERANZAS.- Sean $p > 1$, $q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$. Entonces

$$E(|XY|) \leq \left\{E\left(|X|^p\right)\right\}^{1/p}\left\{E\left(|Y|^q\right)\right\}^{1/q}.$$

Demostración.- Utilizando la desigualdad de Hölder, con $x = X\left\{E\left(|X|^p\right)\right\}^{-1/p}$ e $y = Y\left\{E\left(|Y|^q\right)\right\}^{-1/q}$ llegamos a

$$|XY| \leq p^{-1}|X|^p\left\{E\left(|X|^p\right)\right\}^{1/p - 1}\left\{E\left(|Y|^q\right)\right\}^{1/q} +$$

$$+ q^{-1}|Y|^q\left\{E\left(|Y|^q\right)\right\}^{1/q - 1}\left\{E\left(|X|^p\right)\right\}^{1/p}$$

tomando esperanzas llegamos a la desigualdad buscada.

Un corolario importante de esta última desigualdad nos conduce a la ya conocida desigualdad de Cauchy-Schwarz. En efecto, para $p = q = 2$, tenemos

$$E(|XY|) \le \{E(X^2)\}^{1/2} \cdot \{E(Y^2)\}^{1/2},$$

que puede también escribirse bajo la forma más conocida

$$\{E(|XY|)\}^2 \le E(X^2) \cdot E(Y^2).$$

Finalicemos con una desigualdad debida a Minkowski

DESIGUALDAD DE MINKOWSKI.- Para $p \ge 1$,

$$\{E(|X+Y|^p)\}^{1/p} \le \{E(|X|^p)\}^{1/p} + \{E(|Y|^p)\}^{1/p}.$$

Demostración.- Para $p = 1$, la desigualdad es consecuencia inmediata de $|X+Y| \le |X| + |Y|$.

Sea $p > 1$, entonces

$$|X+Y|^p = |X+Y| \cdot |X+Y|^{p-1} \le |X| \cdot |X+Y|^{p-1} + |Y| \cdot |X+Y|^{p-1}$$

tomando esperanzas y utilizando la correspondiente desigualdad de Hölder, en la que hacemos $Y = |X+Y|^{p-1}$ $(p > 1)$, tenemos

$$E(|X+Y|^p) \le \{E(|X|^p)\}^{1/p} \cdot \{E(|X+Y|^{(p-1)q})\}^{1/q} + \{E(|Y|^p)\}^{1/p} \cdot \{E(|X+Y|^{(p-1)q})\}^{1/q} =$$

$$= \left[\{E(|X|^p)\}^{1/p} + \{E(|Y|^p)\}^{1/p}\right] \cdot \{E(|X+Y|^{(p-1)q})\}^{1/q}.$$

Excluyendo el caso trivial en que $E(|X+Y|^p) = 0$, y teniendo en cuenta que $(p-1)q = p$, reagrupando adecuadamente en la anterior desigualdad llegamos al resultado buscado.

ESPERANZA CONDICIONADA

Recordemos que si $\vec{X}$ es un vector aleatorio al que fraccionamos en dos subvectores $\vec{X}_1, \vec{X}_2$, para $\vec{X}_1 = \vec{x}_1$ tal que $f_{\vec{X}_1}(\vec{x}_1) > 0$, la función

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2) = \frac{f(\vec{x})}{f_{\vec{X}_1}(\vec{x}_1)}, \quad \text{con } \vec{x} = (\vec{x}_1, \vec{x}_2) \text{ y } f \text{ la f.d.p. conjunta}$$

define una f.d.p.; más concretamente define una distribución de probabilidad para el subvector $\vec{X}_2$ condicionada a que el otro subvector $\vec{X}_1$ tome un valor fijo $\vec{x}_1$.

Tendrá pues sentido estudiar la esperanza de este subvector condicionado, utilizando para ello la correspondiente f.d.p. y la definición usual ya establecida para el concepto de esperanza. Por razones de simplicidad vamos a ceñirnos al caso de un vector bidimensional en el que los subvectores serán cada una de las variables que lo componen, la generalización a dimensiones mayores es inmediata.

DEFINICIÓN.- Sea $\vec{X} = (X, Y)$ un vector aleatorio definido sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y sea h una función medible definida de R en R. Supongamos que existe la $E(h(X))$. Entonces la esperanza condicionada de $h(X)$, dado Y, que se escribe $E(h(X)/Y)$, es una variable aleatoria que toma el valor $E(h(X)/y)$, definido mediante

$$E(h(X)/y) = \int_\Omega h(X) \, dP_{Y=y}$$

donde $P_{Y=y}$ quiere significar la distribución de probabilidad condicionada al hecho de que Y toma el valor y. Según el tipo de vector aleatorio que sea $\vec{X} = (X, Y)$ aquella expresión adopta la forma:

1) $$E(h(X)/y) = \sum_{x \in A} h(x) \cdot P(X = x / Y = y) = \sum_{x \in A} h(x) \cdot \frac{f(x,y)}{f_2(y)}$$

donde $A = B_y$, la sección de B mediante y, siendo B tal que $P(\vec{X} \in B) = 1$. Esto para el caso en que $\vec{X}$ sea un vector aleatorio del tipo discreto

2) En el caso continuo

$$E(h(X)/y) = \int_R h(x) \cdot f_{X/Y=y}(x) \, dx = \int_R h(x) \frac{f(x,y)}{f_2(y)} \, dx.$$

Obsérvese que en ambos casos $E(h(X)/y)$ es, efectivamente, una función del valor y.

Obviamente, una definición similar puede darse para $E(h(Y)/X)$, siempre que exista $E(h(Y))$.

Dada la definición establecida es evidente que la esperanza condicionada goza de todas las propiedades inherentes al concepto de esperanza y que ya hemos estudiado. Recordemos a título de ejemplo:

PROPIEDADES DE LA ESPERANZA CONDICIONADA

a) $E(c/Y) = c$, donde c es constante.

b) $E([aX+b]/Y) = a\,E(X/Y) + b$, a y b constantes.

c) Si h_1 y h_2 son funciones medibles, de manera que sus esperanzas, $E(h_i(X))$ existen, tenemos

$$E\{[a_1 h_1(X) + a_2 h_2(X)]/Y\} = a_1 E(h_1(X)/Y) + a_2 E(h_2(X)/Y).$$

d) Si $X \geq 0$, entonces $E(X/Y) \geq 0$.

e) Si $X_1 \geq X_2$, entonces $E(X_1/Y) \geq E(X_2/Y)$.

Momentos de todo tipo de una distribución de probabilidad condicional se definen de forma análoga a como se establecieron en un principio y gozan también, claro está, de las mismas propiedades. Sin embargo, las características especiales de este tipo de distribuciones dan lugar a algunas propiedades que les son características. Veámoslas.

PROPIEDAD 1.- Si $E(h(X))$ existe, entonces

$$E[E(h(X)/Y)] = E(h(X)).$$

Demostración.- Si $\vec{X}$ es un vector aleatorio continuo, entonces la v.a. $E(h(X)/y)$ viene definida mediante

$$g(y) = E(h(X)/y) = \int_R h(x)\,\frac{f(x,y)}{f_2(y)}\,dx$$

como función medible de la variable aleatoria Y que es, su esperanza vendrá dada por

$$E[g(Y)] = \int_R g(y) \cdot f_2(y)\,dy = \int_R \left[\int_R h(x) \cdot \frac{f(x,y)}{f_2(y)}\,dx\right] f_2(y)\,dy =$$

$$= \int_R \left[\int_R h(x) \cdot f(x,y)\,dx\right] dy = \int_R h(x) \left[\int_R f(x,y)\,dy\right] dx =$$

$$= \int_R h(x)\,f_1(x)\,dx = E(h(X)).$$

Para el caso discreto la demostración es análoga.

PROPIEDAD 2.- Si $E(X^2) < +\infty$, entonces

$$\operatorname{var}(X) = \operatorname{var}(E(X/Y)) + E(\operatorname{var}(X/Y)).$$

Demostración.- Haciendo uso de la propiedad anterior sabemos que

$$\operatorname{var}(X) = E((X-E(X))^2) = E\left[E((X-E(X))^2/Y)\right] = E\left[E\{(X^2 + (E(X))^2 - 2X E(X))/Y\}\right] =$$

$$= E\left[E(X^2/Y) + (E(X))^2 - 2E(X) \cdot E(X/Y)\right] = E\left[E(X^2/Y) - (E(X/Y))^2 + (E(X/Y))^2 +\right.$$

$$\left. + (E(X))^2 - 2E(X) \cdot E(X/Y)\right] = E\left[\operatorname{var}(X/Y) + (E(X/Y) - E(X))^2\right] =$$

$$= E(\operatorname{var}(X/Y)) + E\left[(E(X/Y) - E(X))^2\right] = E(\operatorname{var}(X/Y)) + \operatorname{var}(E(X/Y)).$$

COROLARIO.- Si $E(X^2) < +\infty$, entonces de la propiedad 2 se sigue de inmediato que

$$\operatorname{var}(X) \geq \operatorname{var}(E(X/Y)).$$

Obsérvese por otra parte que cuando se verifica la igualdad $\operatorname{var}(X) = \operatorname{var}(E(X/Y))$, ello supone que

$$E(\operatorname{var}(X/Y)) = 0$$

es decir

$$E((X - E(X/Y))^2/Y) = 0$$

y de acuerdo con las propiedades de las integrales de funciones medibles no negativas, ello supone que la función $(X - E(X/Y))^2$ es nula casi por todas partes, o sea

$$P(X = E(X/Y)) = 1$$

lo que supone que X es, casi por todas partes, una función de Y, por cuanto $E(X/Y)$ lo es.

EL PRINCIPIO DE LOS MINIMOS CUADRADOS

Sean X e Y dos variables aleatorias de manera que existe entre ellas una relación de dependencia que deseamos determinar. Supongamos ~~que la relación funcional entre ambas es y = h(x) y además~~ que $E(Y^2)$ y $E[(h(X))^2]$ son finitas. Vamos a tratar de encontrar una función $h(x)$ **que aproxime el valor de la relación funcional entre X e Y.**

El principio de los mínimos cuadrados consiste en la elección de $h(x)$ de tal suerte que la cantidad

$$E\{(Y-h(X))^2\}$$

sea mínima. ¿Cuál ha de ser entonces $h(X)$?

Observemos que, si (X,Y) es un vector aleatorio continuo, tenemos

$$E\{(Y-h(X))^2\} = \int_{R^2} [y-h(x)]^2 f(x,y)\,dx\,dy = \int_{R^2} [y-h(x)]^2 f_{Y/X}(y) \cdot f_X(x)\,dy\,dx =$$

$$= \int_R f_X(x) \left\{ \int_R [y-h(x)]^2 f_{Y/X}(y)\,dy \right\} dx,$$

que sea mínima esta expresión supone que se minimize lo que figura entre llaves en el integrando y de acuerdo con las propiedades estudiadas para la varianza, dicha cantidad se minimiza sí cuando $h(x) = E(Y/x)$.

Un resultado análogo se verifica para el caso en que (X,Y) es discreto.

DEFINICION.- La relación $y = E(Y/x)$ se conoce como la regresión de Y sobre X, análogamente $x = E(X/y)$ es la regresión de X sobre Y.

LINEA DE REGRESION.- En ocasiones, por motivos de sencillez fundamentalmente, se está interesado en aproximar la relación entre X e Y mediante una linea recta, $y = ax+b$. En esta situación y haciendo uso del principio de los mínimos cuadrados antes citado, a y b deberán ser elegidos de manera que

$$L = E\{(Y-aX-b)^2\}$$

sea mínimo. Siempre bajo el supuesto de la existencia de $E(Y^2)$ y $E(X^2)$, la obtención de a, b se reduce a un problema de máximos y mínimos, bastando con derivar parcialmente L respecto a ambos parámetros e igualar, posteriormente, a cero dichas derivadas. Se obtienen así para a y b, las expresiones

$$a = \frac{cov(X,Y)}{var(X)}, \quad b = E(Y) - E(X)\cdot\frac{cov(X,Y)}{var(X)}$$

siendo la ecuación de la recta, conocida como línea de regresión de Y sobre X,

$$y - E(Y) = \frac{cov(X,Y)}{var(X)}(x - E(X)). \quad (1)$$

Una expresión análoga se obtiene para la línea de regresión de X sobre Y, a saber

$$x - E(X) = \frac{cov(X,Y)}{var(Y)}(y - E(Y)). \quad (2)$$

Obsérvese que la relación (2) no se obtiene a partir de la (1) despejando en aquella $x - E(X)$, puesto que en ambas relaciones los papeles de x e y aparecen cambiados. La relación (1) se obtiene considerando a x como variable casual, mientras que en la (2) este papel viene desempeñado por y.

Un procedimiento análogo permitiría aproximar la relación entre X e Y mediante un polinomio de grado n, por ejemplo, $y = a+bx+cx^2+dx^3$, para $n=3$.

DEFINICION.- Si $E(X^2)$ y $E(Y^2)$ existen, a las cantidades

$$\gamma_{X/Y} = \frac{cov(X,Y)}{var(Y)} \quad y \quad \gamma_{Y/X} = \frac{cov(X,Y)}{var(X)}$$

se les designa coeficientes de regresión lineal de Y sobre X y X sobre Y respectivamente, pero al revés.

DEFINICION.- Si $E(X^2)$ y $E(Y^2)$ existen, se define el coeficiente de correlación lineal entre X e Y, mediante

$$\rho = \frac{cov(X,Y)}{(var(X)\cdot var(Y))^{1/2}}$$

a

$$\rho^2 = \gamma_{X/Y}\cdot\gamma_{Y/X}$$

se le denomina en ocasiones coeficiente de determinación. Obsérvese que el signo de ρ es el mismo que el de la $cov(X,Y)$.

DEFINICION.- Decimos que dos variables X e Y son incorreladas si y sólo si $\rho = 0$.

Obviamente si $cov(X,Y)=0$, entonces $\rho=0$ y las variables son incorreladas. Posteriormente, cuando introduzcamos el concepto de independencia entre variables, volveremos sobre este concepto.

A continuación estudiaremos algunas propiedades básicas del coeficiente de correlación.

PROPIEDAD 1.- a) El coeficiente de correlación lineal entre X e Y satisface la relación

$$|\rho| \le 1.$$

b) La igualdad $\rho = \pm 1$ se verifica si existen constantes a y b tales que

$$P\{Y = aX+b\} = 1.$$

Demostración.- a) Recordemos que de acuerdo con la desigualdad de Cauchy-Schwarz

$$[cov(X,Y)]^2 \le var(X)\cdot var(Y),$$

y de aquí la desigualdad del enunciado.

b) Además, recordemos que si $\rho = \pm 1$, entonces $[cov(X,Y)]^2 = var(X)\cdot var(Y)$

EL PRINCIPIO DE LOS MINIMOS CUADRADOS

Sean X e Y dos variables aleatorias de manera que existe entre ellas una relación de dependencia que deseamos determinar. Supongamos ~~que la relación funcional entre ambas es y = h(x) y además~~ que $E(Y^2)$ y $E[(h(X))^2]$ son finitas. Vamos a tratar de encontrar una función $h(x)$ **que aproxime el valor de la relación funcional entre X e Y.**

El principio de los mínimos cuadrados consiste en la elección de $h(x)$ de tal suerte que la cantidad

$$E\{(Y-h(X))^2\}$$

sea mínima. ¿Cuál ha de ser entonces $h(X)$?

Observemos que, si (X,Y) es un vector aleatorio continuo, tenemos

$$E\{(Y-h(X))^2\} = \int_{R^2} [y-h(x)]^2 f(x,y)\,dx\,dy = \int_{R^2} [y-h(x)]^2 f_{Y/X}(y)\cdot f_X(x)\,dy\,dx =$$

$$= \int_R f_X(x)\left\{\int_R [y-h(x)]^2 f_{Y/X}(y)\,dy\right\} dx,$$

que sea mínima esta expresión supone que se minimice lo que figura entre llaves en el integrando y de acuerdo con las propiedades estudiadas para la varianza, dicha cantidad se minimiza así cuando $h(x) = E(Y/x)$.

Un resultado análogo se verifica para el caso en que (X,Y) es discreto.

DEFINICION.- La relación $y = E(Y/x)$ se conoce como la regresión de Y sobre X, análogamente $x = E(X/y)$ es la regresión de X sobre Y.

LINEA DE REGRESION.- En ocasiones, por motivos de sencillez fundamentalmente, se está interesado en aproximar la relación entre X e Y mediante una línea recta, $y = ax+b$. En esta situación y haciendo uso del principio de los mínimos cuadrados antes citado, a y b deberán ser elegidos de manera que

$$L = E\{(Y-aX-b)^2\}$$

sea mínimo. Siempre bajo el supuesto de la existencia de $E(Y^2)$ y $E(X^2)$, la obtención de a y b se reduce a un problema de máximos y mínimos, bastando con derivar parcialmente L respecto a ambos parámetros e igualar, posteriormente, a cero dichas derivadas. Se obtienen así para a y b, las expresiones

$$a = \frac{cov(X,Y)}{var(X)}, \qquad b = E(Y) - E(X)\cdot\frac{cov(X,Y)}{var(X)}$$

siendo la ecuación de la recta, conocida como línea de regresión de Y sobre X,

$$y - E(Y) = \frac{cov(X,Y)}{var(X)}(x - E(X)). \qquad (1)$$

Una expresión análoga se obtiene para la línea de regresión de X sobre Y, a saber

$$x - E(X) = \frac{cov(X,Y)}{var(Y)}(y - E(Y)). \qquad (2)$$

Obsérvese que la relación (2) no se obtiene a partir de la (1) despejando en aquella $x - E(X)$, puesto que en ambas relaciones los papeles de x e y aparecen cambiados. La relación (1) se obtiene considerando a x como variable causal, mientras que en la (2) este papel viene desempeñado por y.

Un procedimiento análogo permitiría aproximar la relación entre X e Y mediante un polinomio de grado n, por ejemplo, $y = a + bx + cx^2 + dx^3$, para $n=3$.

DEFINICION.- Si $E(X^2)$ y $E(Y^2)$ existen, a las cantidades

$$\gamma_{X/Y} = \frac{cov(X,Y)}{var(Y)} \quad y \quad \gamma_{Y/X} = \frac{cov(X,Y)}{var(X)}$$

se las designa coeficientes de regresión lineal de Y sobre X y X sobre Y respectivamente, pero al revés.

DEFINICION.- Si $E(X^2)$ y $E(Y^2)$ existen, se define el coeficiente de correlación lineal entre X e Y, mediante

$$\rho = \frac{cov(X,Y)}{(var(X)\cdot var(Y))^{1/2}}$$

a

$$\rho^2 = \gamma_{X/Y}\cdot\gamma_{Y/X}$$

se le denomina en ocasiones coeficiente de determinación. Obsérvese que el signo de ρ es el mismo que el de la $cov(X,Y)$.

DEFINICION.- Decimos que dos variables X e Y son incorreladas si y sólo si $\rho = 0$.

Obviamente si $cov(X,Y) = 0$, entonces $\rho = 0$ y las variables son incorreladas. Posteriormente, cuando introduzcamos el concepto de independencia entre variables, volveremos sobre este concepto.

A continuación estudiaremos algunas propiedades básicas del coeficiente de correlación.

PROPIEDAD 1.- a) El coeficiente de correlación lineal entre X e Y satisface la relación

$$|\rho| \leq 1.$$

b) La igualdad $\rho = \pm 1$ se verifica si existen constantes a y b tales que

$$P\{Y = aX + b\} = 1.$$

Demostración.- a) Recordemos que de acuerdo con la desigualdad de Cauchy-Schwarz

$$[cov(X,Y)]^2 \leq var(X)\cdot var(Y),$$

y de aquí la desigualdad del enunciado.

b) Además, recordemos que si $\rho = \pm 1$, entonces $[cov(X,Y)]^2 = var(X)\cdot var(Y)$

y ello ocurrirá si y sólo si $P\{aX+Y=0\}=1$, que es de la forma $P\{Y=aX+b\}=1$, mediante una adecuada traslación de Y.

<u>COROLARIO</u>.- Las dos líneas de regresión (1) y (2) coinciden si y sólo si $\rho=1$ o $\rho=-1$.

<u>PROPIEDAD 2</u>.- Sean $E(X^2)<+\infty$, $E(Y^2)<+\infty$ y hagamos $U=aX+b$, $V=cY+d$, entonces

$$\rho_{X,Y}=\pm\rho_{U,V};$$

donde $\rho_{X,Y}$ y $\rho_{U,V}$ son, respectivamente, los coeficientes de correlación lineal de X e Y y de U y V.

Demostración.- La propiedad se deriva de las propiedades de los momentos ya estudiadas.

EJEMPLOS

① Sea (X,Y) un vector aleatorio con función de densidad conjunta

$$f(x,y)=\begin{cases} x+y & , 0<x<1, 0<y<1 \\ 0 & , \text{en el resto.}\end{cases}$$

Entonces

$$E(X^l Y^m)=\int_0^1\int_0^1 x^l y^m (x+y)\,dx\,dy=\int_0^1\int_0^1 x^{l+1}y^m\,dx\,dy+\int_0^1\int_0^1 x^l y^{m+1}\,dx\,dy$$

$$=\frac{1}{(l+2)(m+1)}+\frac{1}{(l+1)(m+2)},$$ con l y m enteros positivos.

A partir de este resultado, tenemos

$l=1, m=0$, $E(X)=\frac{7}{12}$, $l=0, m=1$, $E(Y)=\frac{7}{12}$.

$l=2, m=0$, $E(X^2)=\frac{5}{12}$, $l=0, m=2$, $E(Y^2)=\frac{5}{12}$

$$\mathrm{var}(X)=\frac{5}{12}-\frac{49}{144}=\frac{11}{144}=\mathrm{var}(Y)$$

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{3}-\frac{49}{144}=\frac{48-49}{144}=-\frac{1}{144}.$$

② Recordemos el ejemplo 5 (pag 8. de <u>vectores aleatorios</u>). El vector aleatorio (X,Y), con f.d.p. conjunta

$$f(x,y)=\begin{cases} 2 & , 0<x<y<1 \\ 0 & , \text{en resto}\end{cases}$$

tenía por densidades marginales $f_1(x)=\begin{cases} 2(1-x) & , 0<x<1 \\ 0 & , \text{en el resto}\end{cases}$

y $f_2(y)=\begin{cases} 2y & , 0<y<1 \\ 0 & , \text{en el resto.}\end{cases}$

Asimismo las funciones de densidad condicionadas vendrán dadas por (véase ejemplo 7, pag. 13 del mismo capítulo)

$$f_{X/y}(x)=\frac{1}{y}, \quad 0<x<y$$

$$f_{Y/x}(y)=\frac{1}{1-x}, \quad x<y<1.$$

En ambos casos se trata de distribuciones de probabilidad uniformes en el correspondiente intervalo. A partir de aquí

$$E(Y/x)=\int_x^1 y f_{Y/x}(y)\,dy=\int_x^1 y\cdot\frac{1}{1-x}\,dy=\frac{1}{1-x}\cdot\frac{1-x^2}{2}=\frac{1+x}{2}, \quad 0<x<1.$$

$$E(X/y)=\int_0^y x f_{X/y}(x)\,dx=\int_0^y x\cdot\frac{1}{y}\,dx=\frac{1}{y}\cdot\frac{y^2}{2}=\frac{y}{2}, \quad 0<y<1.$$

Análogamente

$$E(X^2/y)=\int_0^y x^2\cdot\frac{1}{y}\,dx=\frac{1}{y}\cdot\frac{y^3}{3}=\frac{y^2}{3}, \quad 0<y<1$$

y por tanto

$$\mathrm{var}(X/y)=E(X^2/y)-[E(X/y)]^2=\frac{y^2}{3}-\frac{y^2}{4}=\frac{y^2}{12}, \quad 0<y<1$$

Por otra parte

$$E(X)=\int_0^1 x f_1(x)\,dx=\int_0^1 x\,2(1-x)\,dx=\int_0^1(2x-2x^2)\,dx=\left[x^2-\frac{2x^3}{3}\right]_0^1=$$

$$=1-\frac{2}{3}=\frac{1}{3}$$

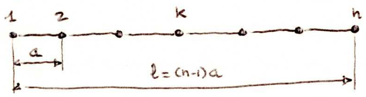
pero

$$E(E(X/Y))=\int_0^1 E(X/y)f_2(y)\,dy=\int_0^1\frac{y}{2}\cdot 2y\,dy=\int_0^1 y^2\,dy=\left[\frac{y^3}{3}\right]_0^1=\frac{1}{3},$$

que coincide, como ya sabíamos, con $E(X)$.

③ La igualdad $E(X) = E\left(E(X/Y)\right)$ es en ocasiones muy útil para obtener la primera de las esperanzas a partir de la segunda, cuando la obtención directa de aquella es complicada. Veamos el siguiente ejemplo:

"Un trabajador opera con determinado tipo de máquinas en un número n. Las máquinas están alineadas y separadas entre sí una distancia a. Supongamos que el operador se traslada de una máquina a otra según un orden de prioridad totalmente aleatorio. Encontrar la longitud media del camino que recorre el operador para atender dos máquinas consecutivamente".



Numeramos las máquinas de izquierda a derecha, de 1 a n, y designamos por B_k el suceso, "el operador está atendiendo la máquina k". Puesto que las máquinas son todas del mismo tipo, y el orden de prioridad es aleatorio, la probabilidad p_i^k, de que la próxima máquina en requerir los servicios del operador sea la i, es $1/n$ $(1 \le i \le n)$. La longitud del camino que deberá recorrer el operador para atender dicha máquina, vendrá dada por

$$\lambda_i^k = \begin{cases} (k-i)a, & k \ge i \\ (i-k)a, & k < i \end{cases}$$

designando por λ la v.a. correspondiente, tendremos

$$E(\lambda/B_k) = \sum_{i=1}^{n} \lambda_i^k \cdot p_i^k = \frac{1}{n}\left[\sum_{i=1}^{k}(k-i)a + \sum_{i=k+1}^{n}(i-k)a\right] =$$

$$= \frac{a}{n}\left[\frac{k(k-1)}{2} + \frac{(n-k)(n-k+1)}{2}\right] = \frac{a}{2n}\left[2k^2 - 2k(n+1) + n(n+1)\right]$$

Por otra parte la probabilidad de que el operador esté sirviendo a la máquina k, es la misma para todos los k, es decir $P(B_k) = \frac{1}{n}$, $1 \le k \le n$, de acuerdo con esto

$$E(\lambda) = E\left[E(\lambda/B_k)\right] = \sum_{k=1}^{n} E(\lambda/B_k) \cdot P(B_k) =$$

$$= \sum_{k=1}^{n} \frac{a}{2n^2}\left[2k^2 - 2k(n+1) + n(n+1)\right] = \frac{a}{2n^2}\left[2\sum_{k=1}^{n} k^2 - 2(n+1)\sum_{k=1}^{n} k + n^2(n+1)\right]$$

$$= \frac{a}{2n^2}\left[2 \cdot \frac{n(n+1)(2n+1)}{6} - 2(n+1) \cdot \frac{n(n+1)}{2} + n^2(n+1)\right] =$$

$$= \frac{a}{2n^2}\left[\frac{n(n+1)(2n+1)}{3} - n(n+1)\right] = \frac{a}{2n^2}\left[n(n+1)\left\{\frac{2n+1}{3} - 1\right\}\right] =$$

$$= \frac{a}{2n^2} \cdot \frac{n(n+1)}{3} \cdot (2n-2) = \frac{2an}{6n^2}(n^2-1) = \frac{a(n^2-1)}{3n} = \frac{\ell(n+1)}{3n} = \boxed{\frac{\ell}{3}\left(1+\frac{1}{n}\right)}$$

④ Sea (X,Y) un vector aleatorio con densidad conjunta dada por

$$f(x,y) = \begin{cases} e^{-y}, & 0<x<y<+\infty \\ 0, & \text{en el resto}. \end{cases}$$

Las correspondientes marginales vendrán dadas por

$$f_1(x) = \int_R f(x,y)\,dy = \int_x^{+\infty} e^{-y}\,dy = \left(-e^{-y}\right)_x^{+\infty} = e^{-x}, \quad 0<x<+\infty$$
$$= 0, \quad \text{en el resto}$$

$$f_2(y) = \int_R f(x,y)\,dx = \int_0^{y} e^{-y}\,dx = e^{-y} \cdot y, \quad 0<y<+\infty$$
$$= 0, \quad \text{en el resto}.$$

A partir de aquí las densidades condicionales se obtienen fácilmente,

$$f_{X/Y}(x) = \frac{f(x,y)}{f_2(y)} = \frac{e^{-y}}{e^{-y} \cdot y} = \frac{1}{y}, \quad 0<x<y$$
$$= 0, \quad \text{en el resto}.$$

$$f_{Y/X}(y) = \frac{f(x,y)}{f_1(x)} = \frac{e^{-y}}{e^{-x}} = e^{x-y}, \quad x<y<+\infty$$
$$= 0, \quad \text{en el resto}.$$

Entonces

$$E(X/y) = \int_R x\, f_{X/y}(x)\,dx = \int_0^y x \cdot \frac{1}{y}\,dx = \frac{1}{y} \cdot \frac{y^2}{2} = \frac{y}{2}, \quad 0<y<+\infty$$

$$E(Y/x) = \int_R y\, f_{Y/x}(y)\,dy = \int_x^{+\infty} y \cdot e^{x-y}\,dy = \int_0^{+\infty}(u+x)\,e^{-u}\,du = x+1, \quad 0<x<+\infty$$

En este caso la relación entre X e Y, por ser lineal en ambos casos, coincidirá con las líneas de regresión de X sobre Y e Y sobre X. De acuerdo con esto los coeficientes de regresión

Capítulo 9

Algunos teoremas de convergencia

ALGUNOS TEOREMAS DE CONVERGENCIA

Antes de comenzar el capítulo de funciones características enunciaremos algunos teoremas que proporcionan las condiciones mediante las cuales es lícito permutar la integración y el paso al límite cuando trabajamos con familias de funciones medibles. Muchos de estos teoremas de convergencia, pues así son conocidos, deben su origen a Lebesgue que fue quien se ocupó en primer lugar, al menos con cierto rigor y sistematización, de la teoría de la medida y de la integración abstractas, de aquí que muchos de estos teoremas lleven su nombre.

TEOREMA DE LA CONVERGENCIA MONOTONA DE LEBESGUE.- Sea $\{f_n\}$ una sucesión monotona creciente de funciones medibles no negativas tales que $\lim_n f_n = f$. Entonces

$$\lim_n \int_\Omega f_n \, d\mu = \int_\Omega \lim_n f_n \, d\mu = \int_\Omega f \, d\mu .$$

(El enunciado del teorema es también válido si la convergencia de las funciones es casi por todas partes).

COROLARIO.- Sea $f_n \geq 0$, $\forall n$ y sea $f = \sum_n f_n$, entonces la sucesión $g_n = \sum_{i=1}^{n} f_i$ verifica las condiciones del teorema anterior y podemos escribir

$$\int_\Omega \sum_n f_n \, d\mu = \int_\Omega \lim_n g_n \, d\mu = \lim_n \int_\Omega g_n \, d\mu = \lim_n \sum_{i=1}^{n} \int_\Omega f_i \, d\mu = \sum_n \int_\Omega f_n d\mu$$

Del teorema de la convergencia monotona de Lebesgue se deriva a su vez el conocido lema

LEMA DE FATOU.- Para funciones medibles no negativas, f_n, se verifica siempre

$$\int_\Omega \underline{\lim}_n f_n \, d\mu \leq \underline{\lim}_n \int_\Omega f_n \, d\mu .$$

donde $\underline{\lim}_n$ denota el límite inferior.

A su vez el lema de Fatou permite obtener el otro importante teorema de convergencia, válido ahora para cualesquiera funciones medibles

TEOREMA DE LA CONVERGENCIA DOMINADA DE LEBESGUE.- Sean f_n funciones medibles tales que $|f_n| \leq g$, (casi por todas partes), donde g es una función medible integrable. Además $f_n \to f$ (casi por todas partes), entonces f y f_n son integrables y además

$$\lim_n \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu .$$

Si en este teorema reemplazamos la función g por una constante positiva y exigimos además que la medida sea finita, entonces

$$\int_\Omega M \, d\mu = M.\mu(\Omega) < +\infty$$

lo que supone que M es una función integrable. Obtenemos entonces un nuevo teorema que se conoce como el

TEOREMA DE LA CONVERGENCIA ACOTADA.- Si $\mu(\Omega) < +\infty$ y las f_n están uniformemente acotadas, entonces $f_n \to f$ (casi por todas partes) implica que $\lim_n \int_\Omega f_n d\mu = \int_\Omega f \, d\mu$

Al igual que para el el teorema de la convergencia monotona, también para el de la convergencia dominada existe, como corolario, una versión para series, a saber

COROLARIO DEL T.C.D. - Si $\sum_n f_n$ converge (casi por todas partes) y $\left|\sum_{k=1}^{n} f_k\right| < g$ (casi por todas partes), donde g es una función integrable, entonces $\sum_n f_n$ es integrable y tenemos

$$\sum_n \int_\Omega f_n \, d\mu = \int_\Omega \sum_n f_n \, d\mu .$$

OBSERVACION.- Si bien todos estos teoremas han sido enunciados para sucesiones se demuestra que son también válidos para conjuntos de índices cualesquiera

Como consecuencia de esta observación enunciamos finalmente un teorema, también de utilidad en el estudio de las funciones características, y que resuelve el problema de derivar una integral mediante la derivación bajo el integrando.

TEOREMA.— Supongamos que $f(\omega,t)$ es una función medible para cada $t \in]a,b[$.

1) Supongamos que $f(\omega,t)$ es continua (casi por todas partes) para $t=t_0$ y supongamos además que para cada t perteneciente a un entorno de t_0, $|f(\omega,t)| \leq g(\omega)$ (casi por todas partes), siendo g integrable. Entonces $\int_\Omega f(\omega,t) d\mu$ es continua para $t=t_0$.

2) Supongamos que para $\omega \in A$ (A medible tal que $\mu(\Omega - A) = 0$), $f(\omega,t)$ tiene en $]a,b[$ derivada, $f'(\omega,t)$, respecto de t. Supongamos además que $|f'(\omega,t)| \leq g(\omega)$, $\omega \in A$, y $t \in]a,b[$, siendo g integrable. Entonces

$$\frac{d}{dt} \int_\Omega f(\omega,t) \, d\mu = \int_\Omega f'(\omega,t) \, d\mu , \quad t \in]a,b[.$$

Capítulo 10

Función característica

FUNCION CARACTERISTICA

Existen determinados problemas en teoría de la Probabilidad de difícil y complicada solución, concretamente todos aquellos relacionados con el estudio de las distribuciones límite y la distribución de probabilidad de sumas de variables aleatorias. Una solución relativamente sencilla a estos problemas se obtiene mediante el uso de las funciones características, funciones que se derivan de la aplicación de técnicas clásicas del análisis matemático, las transformadas de Fourier, a la teoría de la Probabilidad. Todo lo que vamos a hacer en este capítulo se ocupa de estas funciones características.

DEFINICION.- Sea X una variable aleatoria, se define la función característica de la v.a. X como $E(e^{itX})$. A dicha función se la denota mediante $\varphi_X(t)$, o bien $\varphi(t)$ si no existe confusión posible.

De acuerdo con la definición de esperanza y según las características de la v.a. X, la función característica de X se obtendrá mediante:

a) X continua

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx .$$

b) X discreta

$$\varphi(t) = \sum_{x\in A} e^{itx} . f(x) \qquad \text{, donde } A \text{ es tal que } P(X\in A) = 1.$$

Señalemos de inmediato que la función característica de cualquier v.a. existe siempre, por cuanto la correspondiente esperanza que la define siempre lo hace ya que la función e^{itx} está acotada, $|e^{itx}| \le 1$, y por tanto es integrable.

A continuación vamos a conocer las propiedades más importantes de las funciones características en forma de teoremas.

PROPIEDADES DE LAS FUNCIONES CARACTERISTICAS

TEOREMA 1.- Una función característica es uniformemente continua en R, y además verifica

$$\varphi(0) = 1 , \quad |\varphi(t)| \le 1, \quad t\in R.$$

Demostración.- En efecto

$$\varphi(0) = E\left(e^{i0X}\right) = E(1) = 1$$

por otra parte

$$|\varphi(t)| = \left|\int_\Omega e^{itX} dP\right| \le \int_\Omega |e^{itX}| . dP \le \int_\Omega dP = P(\Omega) = 1 .$$

Comprobemos la continuidad uniforme. Estudiemos para ello la diferencia $\varphi(t+h) - \varphi(t)$, tenemos

$$\varphi(t+h) - \varphi(t) = \int_\Omega e^{i(t+h)X} dP - \int_\Omega e^{itX} dP = \int_\Omega e^{itX}(e^{ihX} - 1)\, dP$$

entonces

$$|\varphi(t+h) - \varphi(t)| \le \int_\Omega |e^{itX}(e^{ihX}-1)|\, dP \le \int_\Omega |e^{ihX}-1|\, dP = E\left(|e^{ihX}-1|\right)$$

pero $|e^{ihX}-1| \le |e^{ihX}| + |1| = 2$

por tanto $E\left(|e^{ihX}-1|\right) \le E(2) = 2$, entonces es lícito permutar el paso al límite y la integración, de acuerdo con los teoremas de convergencia, tendremos pues

$$\lim_{h\to 0} |\varphi(t+h) - \varphi(t)| \le \lim_{h\to 0} E\left(|e^{ihX}-1|\right) = E\left(\lim_{h\to 0}|e^{ihX}-1|\right) = E(0) = 0$$

y por tanto se da la continuidad uniforme.

TEOREMA 2.- Si $Y = aX+b$, donde a y b son constantes, entonces

$$\varphi_Y(t) = e^{ibt} . \varphi_X(at).$$

Demostración.-

$$\varphi_Y(t) = E\left(e^{itY}\right) = E\left(e^{it(aX+b)}\right) = E\left(e^{itb} . e^{itaX}\right) = e^{itb} . E\left(e^{i(at)X}\right) =$$

$$= e^{itb} . \varphi_X(at).$$

TEOREMA 3.- Si una variable aleatoria X tiene un momento absoluto de orden n, entonces su función característica es diferenciable n veces y para $k\le n$, tenemos

$$\varphi^{(k)}(0) = i^k . E(X^k).$$

Demostración.- Derivando la función característica k veces, $k\le n$, tenemos

$$\varphi^{(k)}(t) = i^k \int_\Omega X^k . e^{itX} . dP$$

Pero

$$\left|\int_\Omega X^k . e^{itX} dP\right| \le \int_\Omega |X^k| . dP$$

y por tanto, de acuerdo con la hipótesis del teorema, $\varphi^{(k)}(t)$ existe, y la derivación está permitida. Haciendo $t=0$, tenemos

$$\varphi^{(k)}(0) = i^k \int_{\Omega} X^k \, dP = i^k . E(X^k).$$

Apoyándonos en este teorema y mediante el logaritmo de la función característica podemos obtener una sencilla y muy útil expresión para la media y la varianza de una variable aleatoria. En efecto, sea

$$\psi(t) = \log \varphi(t)$$

entonces

$$\psi'(t) = \frac{\varphi'(t)}{\varphi(t)}$$

y derivando nuevamente

$$\psi''(t) = \frac{\varphi''(t).\varphi(t) - [\varphi'(t)]^2}{[\varphi(t)]^2}$$

teniendo en cuenta que $\varphi(0) = 1$ y de acuerdo con el teorema anterior,

$$\psi'(0) = \varphi'(0) = i.E(X)$$

$$\psi''(0) = \varphi''(0) - [\varphi'(0)]^2 = i^2 E(X^2) - [i\,E(X)]^2 = -\operatorname{var}(X),$$

de aquí

$$E(X) = \frac{1}{i}\psi'(0)$$

$$\operatorname{var}(X) = -\psi''(0).$$

A la derivada k-ésima de $\psi(t)$ multiplicada por i^k se la conoce como el <u>cumulante</u> (o semi-invariante) de k-ésimo orden de la variable aleatoria. Este cumulante es una función de los k primeros momentos de la variable. Por ejemplo,

$$i^3.\psi'''(0) = -\left\{E(X^3) - 3E(X^2).E(X) + 2(E(X))^3\right\}.$$

FORMULA DE INVERSION Y TEOREMA DE UNICIDAD

Hemos visto que a partir de la distribución de probabilidad de una v.a. X es siempre posible determinar su función característica. Sería importante que la proposición inversa también fuese cierta. Eso es lo que tratamos de establecer con el siguiente teorema

FORMULA DE INVERSION DE LEVY.– Sean F y φ las funciones de distribución y característica, respectivamente, de una variable aleatoria X. Se verifica entonces:

$$F(b) - F(a) = \lim_{T\to\infty} \frac{i}{2\pi}\int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{t}\varphi(t).dt$$, o bien su equivalente

$$\underset{h>0}{F(x_0+h) - F(x_0-h)} = \lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\operatorname{sen} ht}{t}e^{-itx_0}.\varphi(t).dt .$$

siempre que a y b sean puntos de continuidad de F (ó x_0+h, x_0-h).

Demostración.– Sea

$$J = \frac{1}{\pi}\int_{-T}^{T}\frac{\operatorname{sen} ht}{t}.e^{-itx_0}\varphi(t)\,dt = \frac{1}{\pi}\int_{-T}^{T}\frac{\operatorname{sen} ht}{t}e^{-itx_0}\left[\int_R e^{itx}\,dP'\right]dt$$

donde $\int_R e^{itx}\,dP' = E(e^{itX})$, siendo P' la probabilidad inducida por P, mediante X en R. Podemos también escribir J de la siguiente forma

$$J = \frac{1}{\pi}\int_{-T}^{T}\left[\int_R \frac{\operatorname{sen} ht}{t}.e^{it(x-x_0)}\,dP'\right]dt$$

pero

$$\left|\frac{\operatorname{sen} ht}{t}.e^{it(x-x_0)}\right| \le \left|\frac{\operatorname{sen} ht}{t}\right| < h \quad \text{y} \quad \int_R h\,dP' = h$$

en definitiva la función $\frac{\operatorname{sen} ht}{t}.e^{it(x-x_0)}$ es integrable (T es finito) y por tanto podemos aplicar el teorema de Fubini para permutar el orden de integración, es decir

$$J = \frac{1}{\pi}\int_R dP'.\int_{-T}^{T}\frac{\operatorname{sen} ht}{t}.e^{it(x-x_0)}\,dt = \frac{1}{\pi}\int_R dP'\int_{-T}^{T}\frac{\operatorname{sen} ht}{t}\left(\cos t(x-x_0) + i\operatorname{sen} t(x-x_0)\right)dt =$$

$$= \frac{2}{\pi}\int_R dP'\int_0^T \frac{\operatorname{sen} ht}{t}.\cos t(x-x_0)\,dt$$

aplicando la fórmula $\operatorname{sen}\alpha.\cos\beta = \frac{1}{2}[\operatorname{sen}(\alpha+\beta) + \operatorname{sen}(\alpha-\beta)]$, con $\alpha = ht$, $\beta = t(x-x_0)$,

$$J = \frac{2}{\pi}\int_R dP'\int_0^T \frac{1}{2t}[\operatorname{sen} t(x-x_0+h) - \operatorname{sen}(x-x_0-h)]\,dt = \frac{1}{\pi}\int_R dP'\left[\int_0^T \frac{\operatorname{sen} t(x-x_0+h)}{t}dt - \right.$$

$$\left. - \int_0^T \frac{\operatorname{sen} t(x-x_0-h)}{t}dt\right] = \frac{1}{\pi}\int_R g(x,T)\,dP'$$

donde $g(x,T)$ es la función que figura entre corchetes en el anterior miembro de la igualdad.

Por el calculo integral sabemos que:

$$\lim_{T\to\infty}\int_0^T \frac{\operatorname{sen} x}{x}\,dx = \frac{\pi}{2} \quad \text{y está acotada para todo } T>0$$

Por otra parte

$$\int_0^T \frac{\operatorname{sen}\alpha x}{x}\,dx = \int_0^{\alpha T}\frac{\operatorname{sen} y}{y}\,dy$$

En consecuencia

$$\lim_{T\to\infty}\frac{2}{\pi}\int_0^T \frac{\operatorname{sen}\alpha x}{x}\,dx = \begin{cases} 1, & \alpha>0\\ 0, & \alpha=0\\ -1, & \alpha<0\end{cases}$$

aplicando este resultado a la función $g(x,T)$, tendremos:

$$\lim_{T\to\infty}\frac{1}{\pi}\,g(x,T) = \begin{cases} 0, & x<x_0-h\\ 1/2, & x=x_0-h\\ 1, & x_0-h<x<x_0+h\\ 1/2, & x=x_0+h\\ 0, & x>x_0+h\end{cases}$$

además $|g(x,T)|$ está acotado y podemos hacer uso de los teoremas de convergencia para permutar la integración y el paso al limite. Tendremos en definitiva:

$$\lim_{T\to\infty} J = \lim_{T\to\infty}\frac{1}{\pi}\int_R g(x,T)\,dP' = \frac{1}{\pi}\int_R \lim_{T\to\infty} g(x,T)\,dP' = \int_{x_0-h}^{x_0+h} dP' = F(x_0+h)-F(x_0-h).$$

No olvidemos que P' define F de la forma ya conocida y que x_0+h, y x_0-h son puntos de continuidad de F.

COROLARIO.- Si la función característica es absolutamente integrable en R, entonces la función de distribución es absolutamente continua, su derivada es continua uniformemente y tenemos

$$f(x) = F'(x) = \frac{1}{2\pi}\int_R e^{-ixt}\,\varphi(t)\,dt.$$

Demostración.- En efecto

$$\frac{F(x_0+h)-F(x_0-h)}{2h} = \frac{1}{2\pi}\int_R \frac{\operatorname{sen} ht}{ht}\cdot e^{-itx_0}\varphi(t)\,dt$$

tenemos que

$$\left|\frac{\operatorname{sen} ht}{ht}\cdot e^{-itx_0}\cdot\varphi(t)\right| \le |\varphi(t)|$$

y como $|\varphi(t)|$ es integrable, podemos permutar el paso al limite con la integración, obteniendo

$$\lim_{h\to 0}\frac{F(x_0+h)-F(x_0-h)}{2h} = f(x_0) = F'(x_0) = \frac{1}{2\pi}\int_R e^{-itx_0}\varphi(t)\,dt.$$

Existe pues la derivada de $F(x)$ para todos sus puntos de continuidad. Por otra parte la continuidad absoluta deriva del hecho de poder escribir

$$F(x_0+h)-F(x_0-h) \le 2h\cdot\frac{1}{2\pi}\int_R |\varphi(t)|\,dt$$

que podemos hacer tan pequeño como queramos eligiendo h adecuadamente.

Finalmente, para comprobar la continuidad de f, tendremos

$$|f(x+h)-f(x)| = \frac{1}{2\pi}\left|\int_R e^{-itx}(e^{-ith}-1)\,\varphi(t)\,dt\right| \le \frac{1}{2\pi}\int_R |e^{-ith}-1|\cdot|\varphi(t)|\,dt$$

pero

$$|e^{-ihx}-1| = |\cos hx - i\operatorname{sen} hx - 1| = |(\cos hx-1) - i\operatorname{sen} hx| = \sqrt{(\cos hx-1)^2+\operatorname{sen}^2 hx} =$$
$$= \sqrt{\cos^2 hx+\operatorname{sen}^2 hx+1-2\cos hx} = \sqrt{2(1-\cos hx)} = \sqrt{2\cdot 2\operatorname{sen}^2\frac{hx}{2}} = 2\left|\operatorname{sen}\frac{hx}{2}\right|$$

Luego

$$|f(x+h)-f(x)| \le \frac{1}{2\pi}\int_R 2\left|\operatorname{sen}\frac{th}{2}\right|\cdot|\varphi(t)|\,dt,$$

pero como la integral es finita, podemos elegir A tal que, dado $\varepsilon>0$,

$$\frac{1}{2\pi}\int_R 2\left|\operatorname{sen} th/2\right|\cdot|\varphi(t)|\,dt = \frac{1}{2\pi}\int_{|t|\le A} 2\cdot\left|\operatorname{sen}\frac{th}{2}\right|\cdot|\varphi(t)|\,dt + \frac{1}{2\pi}\int_{|t|>A} 2\cdot\left|\operatorname{sen}\frac{th}{2}\right|\cdot|\varphi(t)|\,dt$$

La segunda integral del segundo miembro es menor que $\varepsilon/2$ y la primera también sin más que elegir h adecuadamente. (Obsérvese que hemos comprobado la continuidad uniforme).

La formula de inversión nos permite demostrar uno de los resultados más importantes del capítulo, el llamado teorema de unicidad.

TEOREMA DE UNICIDAD.- Existe una correspondencia biunívoca entre la función de distribución y la función característica de una variable aleatoria.

Demostración.- Ya sabemos, por definición, que toda función de distribución da lugar a una función característica.

Si φ es la función característica de X, que tiene a F por función de distribución, entonces por el teorema de inversión sabemos que

$$F(x)-F(y) = \lim_{T\to\infty}\frac{i}{2\pi}\int_{-T}^{T}\frac{e^{-itx}-e^{-ity}}{t}\cdot\varphi(t)\,dt, \quad \text{para } x \text{ e } y \text{ pts. de continuidad de } F.$$

por tanto

$$F(x) = \lim_{y\to-\infty}\lim_{T\to\infty}\frac{i}{2\pi}\int_{-T}^{T}\frac{e^{-itx}-e^{-ity}}{t}\,\varphi(t)\,dt$$

donde el paso al límite en y se lleva a cabo a través de puntos de continuidad de F, cosa que siempre es posible. Tenemos pues a F determinada salvo en los puntos de discontinuidad, pero en ellos podemos actuar como sigue:

Sea x un pto de discontinuidad de F. Sea $\{x_n\}$ una sucesión de puntos de continuidad de F tal que $x_n \downarrow x$, entonces $F(x_n) \downarrow F(x)$ por ser F continua por la derecha. Siendo los $F(x_n)$ determinados unívocamente de acuerdo con lo anterior, también lo estará $F(x)$ a través del correspondiente límite.

La utilidad de este teorema de unicidad quedará puesta de manifiesto cuando nos ocupemos del problema de la independencia de variables y más concretamente de la suma de variables aleatorias independientes.

OTRA FORMA DEL TEOREMA DE INVERSION

Existe otra manera de presentar el teorema de inversión que es especialmente útil para determinar la función de cuantía en el caso de variables aleatorias discretas. Recordemos antes que P', la probabilidad inducida por P ~~a partir~~ mediante ξ sobre el espacio medible (R, β), nos servía a su vez para definir la función de distribución a través de

$$F(b)-F(a) = P'(]a,b]) \quad \text{, o sea} \quad F(x) = P'(]-\infty, x])$$

de acuerdo con esto

$$P'(\{a\}) = \lim_n P'(]a-\tfrac{1}{n}, a]) = \lim_n \left(F(a) - F(a-\tfrac{1}{n})\right) = F(a) - F(a-0)$$

Si F es continua en a, entonces $P'(\{a\}) = 0$, de lo contrario $P'(\{a\}) \neq 0$ y mide el salto que da la función en dicho punto, que recordemos, equivale al valor de la función de cuantía en dicho punto, por cuanto

$$f(a) = P(\xi = a) = P'(\{a\}).$$

Hecha esta introducción, que nos será útil posteriormente, enunciemos el teorema de inversión en su nueva forma

TEOREMA.- Sea ξ una v.a. y sean F y φ las funciones de distribución y característica, respectivamente. Entonces, para cualquier valor de x_0 tenemos

$$F(x_0) - F(x_0 - 0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx_0} \varphi(t)\,dt.$$

Demostración.-

$$\frac{1}{2T}\int_{-T}^{T} e^{-itx_0}\varphi(t)\,dt = \frac{1}{2T}\int_{-T}^{T} e^{-itx_0}\left[\int_R e^{itx}\,dP'\right]dt$$

dada la integrabilidad de la función podemos aplicar el teorema de Fubini y permutar el orden de integración, de manera que

$$\frac{1}{2T}\int_{-T}^{T} e^{-itx_0}\left[\int_R e^{itx}\,dP'\right]dt = \frac{1}{2T}\int_R\int_{-T}^{T} e^{it(x-x_0)}\,dt\,dP' = \frac{1}{2T}\int_R dP'\left[\int_{-T}^{T}\left(\cos t(x-x_0) + i \operatorname{sen} t(x-x_0)\right)dt\right]$$

$$= \frac{1}{2T}\int_R dP'\left[2\int_0^T \cos t(x-x_0)\,dt\right] = \frac{1}{T}\int_R dP'\left[\left[\frac{\operatorname{sen} t(x-x_0)}{x-x_0}\right]_0^T\right] = \frac{1}{T}\int_R dP'\,\frac{\operatorname{sen} T(x-x_0)}{x-x_0}$$

pero

$$\lim_{T\to\infty} \frac{1}{T}\cdot\frac{\operatorname{sen} T(x-x_0)}{x-x_0} = 0, \quad \text{si } x \neq x_0$$

mientras que

$$\lim_{T\to\infty} \frac{1}{T}\,\frac{\operatorname{sen} T(x-x_0)}{x-x_0} = 1, \quad \text{si } x = x_0$$

entonces, aplicando el teorema de convergencia que nos permite permutar la integración y el paso al límite tendremos

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itx_0}\varphi(t)\,dt = \int_R \lim_{T\to\infty}\frac{\operatorname{sen} T(x-x_0)}{T(x-x_0)}\,dP' = \int_{\{x_0\}} dP' = P'(\{x_0\})$$

y de acuerdo con lo dicho al principio del párrafo

$$P'(\{x_0\}) = F(x_0) - F(x_0 - 0).$$

Esta fórmula es especialmente útil para obtener la función de cuantía de una variable aleatoria discreta a partir de su función característica. Recordemos lo dicho en la introducción, aquellos puntos con probabilidad distinta de cero,

$$f(x) = P(\xi = x) = P'(\{x\}) = F(x) - F(x-0) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itx}\varphi(t)\,dt.$$

EL CASO DE LAS DISTRIBUCIONES EN ENREJADO (LATTICE DISTRIBUTIONS)

DEFINICION.- Se dice que una variable aleatoria discreta posee una distribución en enrejado si el enrejado $a + bn$, $n = 0, \pm 1, \ldots$ soporta la distribución. Es decir, si los únicos valores que toma la variable con probabilidad distinta de cero son de la forma $x = a + bn$, con $n = 0, \pm 1, \ldots$ siendo a y b enteros y b positivo.

Son muchos los casos de variables de este tipo. Por ejemplo, la Binomial $B(n_0, p)$ es un caso particular en la que $a = 0$, $b = 1$, $n = 0, 1, \ldots, n_0$. La Poisson y la Binomial negativa son también distribuciones en enrejado en las que $a=0$, $b=1$ y $n=0,1,\dots$.

Este tipo de distribuciones presentan algunas peculiaridades en cuanto a sus funciones características, por ejemplo, puede demostrarse que $\exists t\neq 0$; $|\varphi(t)|=1$, si X es una distribución en enrejado. El recíproco de esta propiedad es también cierto. Otra propiedad que las caracteriza la enunciamos y demostramos a continuación

TEOREMA.- Una variable aleatoria X posee una distribución en enrejado si y sólo si su función característica es periódica, con periodo 2π. Es decir $\varphi(t+2\pi)=\varphi(t)$, $t\in\mathbb{R}$.

Demostración.- Directo. Si X es una distribución en enrejado, entonces $x_n=a+bn$ será un número entero y tendremos

$$\varphi(t)=\sum_n e^{itx_n}\cdot P(X=x_n)$$

y

$$\varphi(t+2\pi)=\sum_n e^{i(t+2\pi)x_n}P(X=x_n)=\sum_n e^{itx_n}\cdot e^{ix_n 2\pi}\cdot P(X=x_n)$$

pero

$$e^{ix_n 2\pi}=\cos x_n 2\pi+i\,\mathrm{sen}\,x_n 2\pi=1$$

por tanto

$$\varphi(t+2\pi)=\varphi(t).$$

Inverso. Sea X una variable aleatoria discreta que toma los valores $\{x_n\}_{n\geq 1}$, de manera que su función característica verifica $\varphi(t)=\varphi(t+2\pi)$, entonces

$$\varphi(t)=\sum_n e^{itx_n}P(X=x_n)=\sum_n e^{itx_n}\cdot e^{ix_n 2\pi}P(X=x_n)=\varphi(t+2\pi)$$

por tanto

$$e^{ix_n 2\pi}=1,\ \forall n$$

y de aquí x_n deberá tomar valores en $\{0,\pm1,\pm2,\dots\}$ o en algún conjunto que pueda ponerse en biyección con este.

Para terminar con este tipo de distribuciones de probabilidad estudiaremos la formulación especial que el teorema de inversión adquiere en estas variables.

TEOREMA DE INVERSION PARA DISTRIBUCIONES EN ENREJADO

Si X es una v.a. con distribución en enrejado, entonces su función característica permite obtener la función de cuantía mediante la siguiente fórmula de inversión:

$$p_n=f(X=n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-int}\varphi(t)\,dt.$$

Hemos supuesto que X toma valores $n=0,\pm1,\pm2,\dots$ por cuanto haciendo el cambio $Y=\frac{X-a}{b}$ podemos convertirla en una variable de esas características.

Demostración.- Sabemos que

$$\varphi(t)=\sum_n e^{itn}p_n$$

tomando n' entero, $n'\neq n$,

$$e^{-in't}\cdot\varphi(t)=\sum_n e^{it(n-n')}p_n=\sum_{n\neq n'}e^{it(n-n')}p_n+p_{n'}$$

integrando en el intervalo $[-\pi,\pi]$,

$$\int_{-\pi}^{\pi}e^{-in't}\varphi(t)\,dt=\int_{-\pi}^{\pi}\Big[\sum_{n\neq n'}e^{it(n-n')}p_n+p_{n'}\Big]dt=\sum_{n\neq n'}\int_{-\pi}^{\pi}e^{it(n-n')}p_n\,dt+2\pi p_{n'}$$

siendo lícito permutar el sumatorio con la integral por el teorema de la convergencia acotada.

Pero

$$\int_{-\pi}^{\pi}e^{it(n-n')}dt=\left[\frac{e^{it(n-n')}}{i(n-n')}\right]_{-\pi}^{\pi}=\frac{\cos(n-n')\pi+i\,\mathrm{sen}(n-n')\pi-\cos(n-n')(-\pi)-i\,\mathrm{sen}(n-n')(-\pi)}{i(n-n')}=$$

$$=\frac{2\,\mathrm{sen}(n-n')\pi}{n-n'}=0$$

por tanto

$$\int_{-\pi}^{\pi}e^{-in't}\varphi(t)\,dt=2\pi\cdot p_{n'}.$$

DESARROLLO EN SERIE DE POTENCIAS DE LA FUNCION CARACTERISTICA

Ya hemos comprobado en otra parte de este capítulo que si $E(X^n)$ existe, entonces

$$\varphi^{(k)}(0)=i^k\cdot E(X^k)\quad,\ k\leq n.$$

Aplicando la fórmula de Taylor podemos escribir

$$\varphi(t)=\sum_{k=0}^{n-1}\frac{(it)^k\cdot E(X^k)}{k!}+\frac{\varphi^{(n)}(t_1)}{n!}t^n,\quad \text{con } t_1<t$$

Si X es una variable aleatoria que posee momentos de cualquier orden, entonces $\varphi(t)$ genera una serie de Taylor que converge a $\varphi(t)$ siempre que el resto tienda a cero a medida que aumenta n, es decir

$$\varphi(t) = \sum_{n\geq 0} \frac{(it)^k \cdot E(X^k)}{k!} \quad , \forall t \in A$$

donde $A = \{ t \in R;\ \lim_{n\to\infty} \frac{t_1^n \cdot \varphi^{(n)}(t_1)}{n!} = 0 \text{, con } |t_1| < |t| \}$

o bien, teniendo en cuenta el valor de $\varphi^{(n)}(t_1)$ podemos también caracterizar A, exigiendo a sus elementos que verifiquen $\lim_{n\to\infty} \frac{|t|^n \cdot E(|X|^n)}{n!} = 0$.

FUNCION GENERATRIZ DE MOMENTOS

La función generatriz de momentos de una variable aleatoria X se define mediante

$$M(s) = E(e^{sX}) = \int_\Omega e^{sX} dP = \int_R e^{sx} dP'$$

para todos aquellos valores de s para los que la correspondiente esperanza existe.

Como ya sabemos, si la v.a. X es del tipo discreto, entonces

$$M(s) = \sum_{x_n \in A} e^{sx_n} P(X = x_n) \quad , A \text{ tal que } P(X \in A) = 1.$$

mientras que para X v.a. continua, tendremos

$$M(s) = \int_R e^{sx} f(x)\, dx.$$

Esta función es conocida en análisis como la transformada de Laplace de P' (o si se prefiere de F).

Respecto de la existencia de $M(s)$ obsérvese que $\int_0^\infty e^{sx} dP'$ es finita para $s \leq 0$ y si es finita para algún s positivo, también lo será para todo s menor que aquel. Esto, junto con consideraciones análogas hechas respecto de la semirrecta real negativa, permite afirmar que $M(s)$ existe, o está definida, en algún intervalo que contiene a 0. Si X es no negativa este intervalo contiene a $]-\infty,0]$ y puede que parte de $]0,+\infty[$, mientras que si X es no positiva entonces el intervalo contiene a $[0,+\infty[$ y puede que parte de $]-\infty,0]$. Hay situaciones, no obstante, en las que el intervalo contiene sólo el punto 0, por ejemplo, una v.a. que toma los valores $\pm n$, $n=1,2,\ldots$ con probabilidad $P(X=n) = P(X=-n) = c/n^2$.

En cualquier caso la existencia de $M(s)$ en un intervalo simétrico da lugar a una serie de importantes propiedades que transcribimos a continuación por medio del siguiente teorema.

TEOREMA.- Si $M(s)$ está definida en el intervalo $[-s_0, s_0]$, $s_0 > 0$, entonces la variable aleatoria X posee momentos de cualquier orden y además $M^{(k)}(0) = E(X^k)$ (de aquí el nombre de función generatriz de momentos que recibe $M(s)$).

Demostración.- Si $M(s)$ está definida en $[-s_0, s_0]$, ello supone que e^{sX} es integrable para $|s| \leq s_0$. Por otra parte

$$e^{|sx|} \leq e^{sx} + e^{-sx}$$

y por tanto para $|s| \leq s_0$, $e^{|sX|}$ es integrable. Observemos además que, $\forall k$

$|x|^k$ es de menor orden que $e^{|sx|}$, cuando $x \to \pm\infty$, si s es fijo y >0,

ello supone que X posee momentos de cualquier orden.

Fijémonos ahora en la serie $e^{sX} = \sum_{k\geq 0} s^k X^k / k!$, sus sumas parciales están mayoradas por la serie

$$e^{|sX|} = \sum_{k \geq 0} |sX|^k / k! \quad \text{(en ambos casos } |s| \leq s_0)$$

y no olvidemos que $e^{|sX|}$ es una función integrable, podemos entonces aplicar el teorema de la convergencia dominada y escribir

$$E(e^{sX}) = M(s) = \sum_{k \geq 0} \frac{E(X^k)}{k!} s^k \quad , \quad |s| \leq s_0.$$

Resulta pues que $M(s)$ posee un desarrollo en serie de Taylor alrededor del punto 0, y de acuerdo con las propiedades de dicho desarrollo los coeficientes son las derivadas de la función en el pto 0, es decir

$$M^{(k)}(0) = E(X^k).$$

COROLARIO.- Si $M(s)$ está definida en $|s| \leq s_0$ y de alguna manera podemos obtener un desarrollo en serie de potencias de $M(s)$, de la forma

$$M(s) = \sum_{k \geq 0} a_k s^k \quad , \quad |s| \leq s_0.$$

de acuerdo con la unicidad del desarrollo en serie en el campo de convergencia de la serie, tendremos que

$$a_k = M^{(k)}(0) / k!$$

pero como $M^{(k)}(0) = E(X^k)$, ello nos permite obtener los momentos de X, mediante

$$E(X^k) = a_k \cdot k!.$$

Relación entre la función generatriz y la función característica.- Cuando la función generatriz de momentos existe, observese que

$$\varphi(t) = E(e^{itX})$$
$$M(t) = E(e^{tX})$$

y por tanto

$$\varphi(t) = M(it)$$

esta relación supone que muchas de las propiedades estudiadas para la función característica sean aplicables a la función generatriz, no obstante, una de las más importante, la existencia, siempre, de la función característica, ya hemos visto que no es ~~siempre~~ verificada por $M(s)$. Este motivo es más que suficiente para que la función generatriz quede en pura anecdota y sea la característica la mas utilizada en teoria de la probabilidad, sobretodo en todo lo referente a teoremas de inversión y de unicidad.

Observemos finalmente que

$$\varphi(t) = \sum_{k\geq 0} \frac{(it)^k}{k!} E(X^k)$$

está mayorada por

$$\sum_{k\geq 0} \frac{|t|^k}{k!} E(|X|^k) = E(e^{|tX|})$$

y hemos dicho que si $M(t)$ está definida en $|t| \leq t_0$, entonces $e^{|tX|}$ es integrable, por tanto $\varphi(t)$ es desarrollable en serie de Taylor si la correspondiente función generatriz de momentos existe.

FUNCION CARACTERISTICA DE UN VECTOR ALEATORIO

Sea $\vec{X}$ un vector aleatorio de componentes $(X_1, X_2, \ldots, X_n)$, se define la función característica del vector como la esperanza de la variable $e^{i(t_1X_1+t_2X_2+\cdots+t_nX_n)}$. Es decir

$$\varphi_{\vec{X}}(t_1, t_2, \ldots, t_n) = E\left(\exp\left[i\sum_{k=1}^{n} t_k X_k\right]\right), \quad t_k \in R.$$

De acuerdo con esta definición, generalización inmediata de la dada para variables aleatorias, la función característica del vector $\vec{X}$ existe siempre y goza además de las siguientes propiedades.

PROPIEDADES DE LA F.C. DE UN VECTOR ALEATORIO

1.- $\varphi_{\vec{X}}(0, \ldots, 0) = 1$

2.- $|\varphi_{\vec{X}}(t_1, \ldots, t_n)| \leq 1$

3.- $\varphi_{\vec{X}}(t_1, \ldots, t_n)$ es uniformemente continua en R^n.

4.- Si la función característica de $\vec{X}$ es $\varphi_{\vec{X}}(t_1, \ldots, t_n)$, la del vector $\vec{Y}$, cuyas componentes se definen $Y_k = a_k X_k + b_k$, es

$$\varphi_{\vec{Y}}(t_1, \ldots, t_n) = \exp\left(i\sum_{k=1}^{n} t_k b_k\right) \cdot \varphi_{\vec{X}}(a_1t_1, \ldots, a_nt_n).$$

5.- Si el momento conjunto de orden $-(n_1, \ldots, n_n)$ existe, entonces

$$\left.\frac{\partial^{n_1+\cdots+n_n}}{\partial t_1^{n_1}\cdots\partial t_n^{n_n}} \varphi_{\vec{X}}(t_1, \ldots, t_n)\right|_{t_1=\cdots=t_n=0} = i^{\sum_{k=1}^{n} n_k} \cdot E\left(X_1^{n_1}\cdots X_n^{n_n}\right)$$

En particular

$$\left.\frac{\partial^r}{\partial t_k^r} \varphi_{\vec{X}}(t_1, \ldots, t_n)\right|_{t_1=\cdots=t_n=0} = i^r \cdot E(X_k^r), \quad k = 1, \ldots, n$$

6.- La función característica de cualquier subvector de dimensión m, $\vec{X}_1$, se obtiene a partir de la de $\vec{X}$ sin más que igualar a cero las t_k del subvector complementario, $\vec{X}_2$, de dimensión $n-m$. Así

$$\varphi_{\vec{X}_1}(t_1, \ldots, t_m) = \varphi_{\vec{X}}(t_1, \ldots, t_m, 0, \ldots, 0)$$

En particular, la función característica de cada componente se obtiene mediante

$$\varphi_{X_k}(t_k) = \varphi_{\vec{X}}(0, \ldots, 0, t_k, 0, \ldots, 0).$$

Todas estas propiedades tienen una demostración análoga, si bien ligeramente más complicada, a la de sus correspondientes versiones unidimensionales.

La función característica del vector $\vec{X}$ permite obtener fácilmente la función característica de la suma de sus componentes. En efecto

TEOREMA.- Si el vector aleatorio $\vec{X}$ tiene por función característica $\varphi_{\vec{X}}(t_1, \ldots, t_n)$, entonces la variable aleatoria $X_1+X_2+\cdots+X_n$ tiene por función característica $\varphi(t, t, \ldots, t)$.

Demostración.- Sea $Y=\sum_{k=1}^{n} X_k$, de acuerdo con la definición de función característica, tendremos

$$\varphi_Y(t)=E\left(e^{itY}\right)=E\left(\exp\left[it\sum_k X_k\right]\right)=E\left(\exp\left[i\sum_k tX_k\right]\right)$$

pero

$$\varphi_{\vec{X}}(t_1, \ldots, t_k)=E\left(\exp\left[i\sum_k t_k X_k\right]\right)$$

que para $t_k=t$, $k=1,\ldots,n$ da lugar a $\varphi_Y(t)$.

Finalmente enunciaremos, sin demostración por cuanto nuevamente ésta es análoga a la correspondiente unidimensional, las versiones vectoriales de la fórmula de inversión y del teorema de unicidad.

FORMULA DE INVERSION.- Sea $\vec{X}$ un vector aleatorio con función característica $\varphi_{\vec{X}}(t_1, \ldots, t_n)$, entonces

$$P\{a_k < X_k \le b_k,\ k=1,2,\ldots,n\}=\lim_{T\to\infty}\frac{1}{(2\pi)^n}\int_{-T}^{T}\cdots\int_{-T}^{T}\prod_{k=1}^{n}\frac{e^{-it_k a_k}-e^{-it_k b_k}}{it_k}\varphi_{\vec{X}}(t_1, \ldots, t_n)\,dt_1\ldots dt_n$$

o bien su equivalente

$$P\{a_k - h < X_k \le a_k+h,\ k=1,2,\ldots,n\}=\lim_{T\to\infty}\frac{1}{(2\pi)^n}\int_{-T}^{T}\cdots\int_{-T}^{T}\prod_{k=1}^{n}\frac{\operatorname{sen} h t_k}{it_k}e^{-it_k a_k}\varphi_{\vec{X}}(t_1, \ldots, t_n)\,dt_1\ldots dt_n$$

donde a_k y b_k son números reales que satisfacen la siguiente condición:

la probabilidad de tomar valores en la superficie del paralelepípedo $\{a_k < X_k \le b_k,\ k=1,\ldots,n\}$ es igual a 0.

COROLARIO.- Si la función característica es absolutamente integrable, entonces el vector aleatorio es continuo y su función de densidad conjunta se puede obtener a partir de

$$f_{\vec{X}}(x_1, \ldots, x_n)=\frac{1}{(2\pi)^n}\int_R\cdots\int_R e^{-i\left(\sum_k t_k x_k\right)}\varphi_{\vec{X}}(t_1, \ldots, t_n)\,dt_1\ldots dt_n .$$

Existe, al igual que para las variables aleatorias, otra versión de la fórmula de inversión que permite obtener fácilmente las funciones de cuantía para vectores aleatorios del tipo discreto:

OTRA FORMA DE LA FORMULA DE INVERSION.- Sea $\varphi_{\vec{X}}(t_1, \ldots, t_n)$ la función característica de un vector aleatorio $\vec{X}$, entonces

$$P\left(X_1=x_1, \ldots, X_n=x_n\right)=\lim_{T\to\infty}\frac{1}{(2T)^n}\int_{-T}^{T}\cdots\int_{-T}^{T}e^{-i\left(\sum_k t_k x_k\right)}\varphi_{\vec{X}}(t_1, \ldots, t_n)\,dt_1\ldots dt_n .$$

Basado en la fórmula de inversión puede enunciarse, también ahora, el siguiente teorema de unicidad.

TEOREMA DE UNICIDAD.- Existe una correspondencia biunívoca entre las funciones característica y de distribución de un vector aleatorio.

Para finalizar el capítulo señalemos que también podría hablarse ahora de desarrollo en serie de potencias de la función característica de un vector aleatorio. Desarrollo que se basa en la fórmula de Taylor para variables n-dimensionales.

Igualmente existe el concepto de <u>función generatriz de momentos</u> para un vector aleatorio $\vec{X}$, cuya definición es análoga a la correspondiente noción unidimensional. Igualmente gozan de las propiedades ya descritas para tal concepto en el caso de variables aleatorias, modificadas adecuadamente para su versión vectorial.

EJEMPLOS

<u>Funciones características de algunas distribuciones de probabilidad.</u>

1) BINOMIAL.

Sea X una v.a. con distribución de probabilidad $B(n,p)$, entonces

$$\varphi(t)=\sum_{x=0}^{n}e^{itx}\binom{n}{x}p^x q^{n-x}=\sum_{x=0}^{n}\binom{n}{x}(pe^{it})^x q^{n-x}=(p\cdot e^{it}+q)^n$$

De acuerdo con la relación $M(it)=\varphi(t)$, entonces en aquellos t para los que exista,

$$M(t)=(p\cdot e^{t}+q)^n .$$

Se comprueba que $M(t)$ existe para todo t.

Por otra parte

$$\frac{d}{dt}\varphi(t)=n(p\cdot e^{it}+q)^{n-1}\cdot ipe^{it} \quad \text{y} \quad \left.\frac{d}{dt}\varphi(t)\right|_{t=0}=inp=i.E(X).$$

La función $\psi(t)=\log\varphi(t)=n\log(pe^{it}+q)$, al derivarla

$$\psi'(t)=\frac{inpe^{it}}{pe^{it}+q} \rightarrow \psi'(0)=inp=iE(X)$$

$$\psi''(t)=\frac{i^2npe^{it}(pe^{it}+q)-i^2np^2e^{2it}}{(pe^{it}+q)^2} \rightarrow \psi''(0)=\frac{-npq}{1}=-npq=-\mathrm{var}(X).$$

2) POISSON.

Para una v.a. Poisson, $P(\lambda)$, su función característica vale

$$\varphi(t) = \sum_{x\geq 0} e^{itx} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x\geq 0} \frac{(\lambda e^{it})^x}{x!} = e^{-\lambda} . e^{\lambda e^{it}} = e^{\lambda e^{it} - \lambda}$$

3) MULTINOMIAL.

Si el vector $\vec{X}$ posee una distribución multinomial con parámetros $n;\ p_1, p_2, \ldots, p_k$, su función característica vale

$$\varphi_{\vec{X}}(t_1 \ldots t_k) = \sum_{x_1 \ldots x_k} e^{i \sum_j t_j x_j} \frac{n}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} =$$

$$= \sum_{x_1 \ldots x_k} \frac{n}{x_1! \cdots x_k!} (p_1 e^{it_1})^{x_1} \cdots (p_k e^{it_k})^{x_k} = (p_1 e^{it_1} + \cdots + p_k e^{it_k})^n$$

$$\frac{\partial^k}{\partial t_1 \cdots \partial t_k} \varphi_{\vec{X}}(t_1 \ldots t_k) \Big|_{t_1 = t_2 = \cdots = t_k = 0} = n(n-1) \cdots (n-k+1)\, i^k p_1 \cdots p_k (p_1 e^{it_1} + \cdots + p_k e^{it_k})^{n-k} =$$

$$= n(n-1) \cdots (n-k+1)\, i^k p_1 \cdots p_k$$

con lo que

$$E(X_1 \cdots X_k) = n(n-1) \cdots (n-k+1)\, p_1 p_2 \cdots p_k$$

Obsérvese que la función característica de la componente X_j podemos obtenerla mediante

$$\varphi_{X_j}(t_j) = \varphi_{\vec{X}}(0, \ldots, 0, t_j, 0, \ldots, 0) = (p_1 + p_2 + \cdots + p_j e^{it_j} + \cdots + p_k)^n =$$

$$= \Big(p_j e^{it_j} + \sum_{i \neq j} p_i\Big)^n = (p_j e^{it_j} + (1-p_j))^n = (p_j e^{it_j} + q_j)^n$$

y por el teorema de unicidad X_j es una $B(n, p_j)$, como en su momento habíamos establecido estudiando las marginales de la multinomial.

4) NORMAL.

Supongamos que X es una variable aleatoria $N(0,1)$, su función característica valdrá

$$\varphi(t) = \int_R e^{itx} f(x)\, dx = \int_R e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_R e^{itx - \frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_R e^{(2itx - x^2 + t^2)/2} . e^{-t^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_R e^{-(x-it)^2/2} dx = e^{-t^2/2}$$

Para obtener la función característica de una v.a. X, $N(\mu, \sigma^2)$, podemos utilizar la propiedad

$$\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$$

puesto que si X es $N(\mu, \sigma^2)$, la variable $Z = \frac{X - \mu}{\sigma}$ es $N(0,1)$ y por tanto

$$\varphi_X(t) = \varphi_{\sigma Z + \mu}(t) = e^{it\mu} \varphi_Z(\sigma t) = e^{it\mu} . e^{-t^2\sigma^2/2} = e^{it\mu - t^2\sigma^2/2}.$$

Podríamos haber obtenido este resultado por otro camino. En efecto, sabemos que si X es $N(0,1)$, entonces

$$E(X^{2n+1}) = 0 \quad \text{y} \quad E(X^{2n}) = \frac{(2n)!}{2^n . n!}$$

sabemos además que $\varphi(t)$ admite un desarrollo en serie, puesto que existen todos los momentos, para todos aquellos t que verifican

$$\lim_{n\to\infty} \frac{|t|^n . E(|X|^n)}{n!} = 0$$

pero

$$\lim_{n\to\infty} \frac{|t|^{2n} . E(|X|^{2n})}{(2n)!} = \lim_{n\to\infty} \frac{t^{2n} . (2n)!}{(2n)!\, 2^n . n!} = \lim_{n\to\infty} \frac{t^{2n}}{2^n . n!} = 0, \quad \forall t$$

luego

$$\varphi(t) = \sum_{k\geq 0} \frac{(it)^{2k} . E(X^{2k})}{(2k)!} = \sum_{k\geq 0} \frac{(it)^{2k} . (2k)!}{(2k)!\, 2^k . k!} = \sum_{k\geq 0} \Big(-\frac{t^2}{2}\Big)^k . \frac{1}{k!} =$$

$$= e^{-t^2/2}.$$

5) GAMMA.

Si X se distribuye como una Gamma con parámetros α y β, entonces

$$\varphi(t) = \frac{1}{\Gamma(\alpha) . \beta^\alpha} \int_0^\infty e^{itx} x^{\alpha-1} . e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} . e^{-x(1-i\beta t)/\beta} dx$$

haciendo $x(1 - i\beta t) = y$, tenemos

$$x = y/(1-i\beta t), \quad dx = dy/(1-i\beta t), \quad y \in [0, +\infty[$$

$$\varphi(t) = (1-i\beta t)^{-\alpha} . \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty y^{\alpha-1} . e^{-y/\beta} dy = (1-i\beta t)^{-\alpha}.$$

Así, para una χ^2_r, en la que $\alpha = r/2$, $\beta = 2$, tendremos como función característica

$$\varphi_{\chi^2_r}(t) = (1-2it)^{-r/2}.$$

Para una <u>exponencial negativa</u>, $\alpha = 1$, $\beta = 1/\lambda$, tendremos

$$\varphi(t) = \left(1 - \frac{it}{\lambda}\right)^{-1} = \frac{\lambda}{\lambda - it}.$$

6) NORMAL BIVARIANTE

Supongamos que $\vec{X}$ es un vector normal bivariante de manera que sus componentes son N(0,1) cada una de ellas. Su función de densidad conjunta viene dada por

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[x^2 - 2\rho xy + y^2\right]\right\}$$

La función característica será

$$\varphi_{\vec{X}}(t_1,t_2) = \int_R \int_R e^{i(t_1x+t_2y)} f(x,y)\,dx\,dy = \int_{R^2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{(it_1x + it_2y - \frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} dx\,dy$$

Con un cambio de variable adecuado llegamos a

$$\varphi_{\vec{X}}(t_1,t_2) = \int_R e^{-\frac{1}{2}(t_1^2+2\rho t_1t_2+t_2^2)} \frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)}\,du\,dv = e^{-\frac{1}{2}(t_1^2+2\rho t_1t_2+t_2^2)}.$$

Si queremos ahora obtener la función característica de un vector $\vec{X}$ normal bivariante completamente general, es decir con función de densidad conjunta dada por

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

si hacemos los cambios $Z_1 = \frac{X-\mu_1}{\sigma_1}$ y $Z_2 = \frac{Y-\mu_2}{\sigma_2}$, las variables conjuntamente, (Z_1, Z_2) dan lugar a un vector aleatorio normal bivariante tipificado, es decir, con función de densidad conjunta como la anterior, ello supone, de acuerdo con una propiedad de las ~~[illegible]~~ funciones características que

$$\varphi_{\vec{X}}(t_1,t_2) = e^{i(\mu_1t_1+\mu_2t_2)} \cdot \varphi_{\vec{Z}}(\sigma_1t_1, \sigma_2t_2) =$$

$$= \exp\left\{i\mu_1t_1 + i\mu_2t_2 - \frac{1}{2}(\sigma_1^2t_1^2 + 2\rho\sigma_1\sigma_2t_1t_2 + \sigma_2^2t_2^2)\right\}$$

Si quisiéramos ahora conocer la función característica de la variable $U = X + Y$, podemos

utilizar una propiedad de las funciones características de los vectores aleatorios que dice

$$\varphi_{\sum_k X_k}(t) = \varphi_{\vec{X}}(t, t, \ldots, t)$$

así pues

$$\varphi_U(t) = \varphi_{\vec{X}}(t,t) = \exp\left\{(i\mu_1 + i\mu_2)t - \frac{1}{2}(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)t^2\right\} =$$

$$= \exp\left\{it(\mu_1+\mu_2) - \frac{1}{2}t^2(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)\right\}$$

pero esta es la función característica de una normal con esperanza $\mu = \mu_1 + \mu_2$ y con varianza $\sigma^2 = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2 = \sigma_1^2 + 2\operatorname{cov}(X_1, X_2) + \sigma_2^2$. Resultado este que pone de manifiesto la utilidad del teorema de unicidad.

Capítulo 11

Independencia

INDEPENDENCIA

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sean $\mathcal{A}_j$, $j=1,\dots,k$ familias de conjuntos tales que $\mathcal{A}_j \subset \mathcal{A}$, $j=1,\dots,k$.

<u>DEFINICION 1</u>.- Decimos que las familias $\mathcal{A}_j$ son <u>independientes</u>, si para cada $A_j \in \mathcal{A}_j$, $j=1,\dots,k$ los sucesos $A_1,\dots,A_k$ son independientes. Es decir

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \prod_{j=1}^{n} P(A_{i_j}) \qquad \text{donde } (i_1,\dots,i_n) \subset (1,\dots,k).$$

De esta definición se desprende de inmediato que las subfamilias de familias independientes son también independientes.

INDEPENDENCIA DE VARIABLES ALEATORIAS

Sea X una v.a. definida sobre $(\Omega, \mathcal{A}, P)$. Sea β la tribu de Borel, la familia de conjuntos de $\mathcal{A}$, $X^{-1}(\beta)$ es también una σ-algebra como fácilmente puede comprobarse. A esta σ-algebra se la conoce como la σ-algebra <u>inducida</u> por X. Consideremos ahora un conjunto de variables aleatorias X_j, $j=1,\dots,k$, definidas todas ellas sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y sean $\mathcal{A}_j$, las σ-algebras inducidas por dichas variables.

<u>DEFINICION 2</u>.- Decimos que las variables aleatorias X_j, $j=1,\dots,k$ son <u>independientes</u> si las σ-algebras que inducen lo son.

Teniendo en cuenta la definición de la σ-algebra inducida, $\forall A_j \in \mathcal{A}_j$, $\exists B_j \in \beta / A_j = X^{-1}(B_j)$ entonces

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \prod_{j=1}^{n} P(A_{i_j}) \qquad (i_1,\dots,i_n) \subset (1,\dots,k) \quad (1)$$

pero

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = P(X_{i_j} \in B_{i_j},\ j=1,\dots,n) = \prod_{j=1}^{n} P(X_{i_j} \in B_{i_j}).$$

Como los $A_j \in \mathcal{A}_j$, $j=1,\dots,k$ están definidos como antiimagenes de conjuntos de Borel, esto supone que de verificarse la última de las desigualdades escrita, se verificará la anterior (1). En definitiva, otra definición para la independencia de las variables aleatorias en términos de conjuntos de Borel, puede ser esta

<u>DEFINICION 3</u>.- Las variables aleatorias X_j, $j=1,\dots,k$ son <u>independientes</u> si

$$P(X_{i_j} \in B_{i_j},\ j=1,\dots,n) = \prod_{j=1}^{n} P(X_{i_j} \in B_{i_j}), \quad (i_1,\dots,i_n) \subset (1,\dots,k)$$

cualesquiera que sean los conjuntos de Borel B_{i_j}.

Las variables aleatorias que no son independientes se dice que son <u>dependientes</u>.

Antes de demostrar el teorema de factorización, el más importante de este capítulo, veamos algunos lemas previos que nos serán de utilidad.

<u>LEMA 1</u>.- Sean X_j, $j=1,\dots,k$ variables aleatorias independientes y sean $g_j : (R,\beta) \to (R,\beta)$ funciones medibles, de manera que $g_j(X_j)$ es una variable aleatoria. Entonces las variables aleatorias $g_j(X_j)$, $j=1,\dots,k$ son también independientes.

Demostración.- Sea X una v.a. y sea g una función medible. Designemos por $\mathcal{A}_X$ y $\mathcal{A}_{g(X)}$ las σ-algebras inducidas por las v.a. X y $g(X)$. Sea $A \in \mathcal{A}_{g(X)}$, entonces $\exists B \in \beta$, tal que $[g(X)]^{-1}(B) = A$. Pero

$$A = [g(X)]^{-1}(B) = X^{-1}[g^{-1}(B)] = X^{-1}(B')$$

donde $B' = g^{-1}(B)$ y como g es medible $B' \in \beta$ y por tanto $X^{-1}(B') = A \in \mathcal{A}_X$, es decir $\mathcal{A}_{g(X)} \subset \mathcal{A}_X$.

De acuerdo con esto, para cada $j=1,\dots,k$, tendremos

$$\mathcal{A}_{g_j(X_j)} \subset \mathcal{A}_{X_j}$$

y como la familia $\mathcal{A}_{X_j}$, $j=1,\dots,k$ es independiente lo es cualquier familia de subfamilias y $g_j(X_j)$, $j=1,\dots,k$ constituyen un conjunto de v.a. independientes.

<u>LEMA 2</u>.- Una condición necesaria y suficiente para que las variables aleatorias X_j, $j=1,\dots,k$ sean independientes es que las familias $\mathcal{C}_j$, $j=1,\dots,k$, con

$$\mathcal{C}_j = X_j^{-1}\{\,]-\infty, x],\ x \in R\}$$

sean independientes.

Demostración.- <u>Directo</u>.- Si las v.a. son independientes, las familias $\mathcal{A}_j = X_j^{-1}(B)$, $B \in \beta$ son independientes y por tanto también lo serán las subfamilias de estas y las $\mathcal{C}_j$ son subfamilias de las $\mathcal{A}_j$.

<u>Inverso</u>.- Para cada $\mathcal{C}_j$ definimos

$\mathcal{C}_j^* = \{$menor clase de conjuntos stable para las sumas numerables y diferencias propias, que contiene a $\mathcal{C}_j\}$.

Las $\mathcal{C}_j^*$, así definidas también son independientes. Veamoslo. Definamos

$$\mathcal{D}_i = \{D_i \in \mathcal{A} \ ;\ P(D_i \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{k} A_j) = P(D_i) \cdot \prod_{\substack{j=1 \\ j\neq i}}^{k} P(A_j),\ A_j \in \mathcal{C}_j\}$$

esta clase de conjuntos verifica las siguientes propiedades:

a) $\mathcal{C}_i \subset \mathcal{D}_i$

$$\forall A_i \in \mathcal{C}_i \quad . \; P\Big(A_i \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j\Big) = P(A_i) . \prod_{\substack{j=1\\ j\neq i}}^{k} P(A_j) \quad , \; A_j \in \mathcal{C}_j$$

pues los $\mathcal{C}_j$ son independientes.

b) Sean $\{D_{in}\}_{n\geq 1}$, $D_{in} \in \mathcal{D}_i . \forall n$ y $D_{im} \cap D_{in} = \phi$, $m \neq n$. Entonces

$$\bigcup_{n\geq 1} D_{in} \in \mathcal{D}_i .$$

$$P\Big(\big(\bigcup_{n\geq 1} D_{in}\big) \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j\Big) = \sum_{n\geq 1} P\Big(D_{in} \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j\Big) = \prod_{\substack{j=1\\ j\neq i}}^{k} P(A_j) \Big[\sum_{n\geq 1} P(D_{in})\Big] =$$

$$= P\big(\bigcup_{n\geq 1} D_{in}\big) . \prod_{\substack{j=1\\ j\neq i}}^{k} P(A_j) \quad , \; A_j \in \mathcal{C}_j .$$

c) $D_{i_1} , D_{i_2} \in \mathcal{D}_i$, con $D_{i_2} \subset D_{i_1}$, entonces $D_{i_1} - D_{i_2} \in \mathcal{D}_i$.

$$(D_{i_1} - D_{i_2}) \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j = D_{i_1} \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j - D_{i_2} \cap \bigcap_{\substack{j=1\\ j\neq i}}^{k} A_j \quad , \; A_j \in \mathcal{C}_j$$

y de aquí el resultado.

Dadas las propiedades que verifica $\mathcal{D}_j$ y del hecho de estar definida $\mathcal{C}_j^*$ como "la menor clase ...." , entonces $\mathcal{C}_j^* \subset \mathcal{D}_j$. Además por definición de $\mathcal{D}_i$, las familias

$$\{\mathcal{D}_i , \mathcal{C}_j , j=1,\ldots,k , j\neq i\}$$

son independientes , en particular como $\mathcal{C}_j^* \subset \mathcal{D}_j$, tendremos que las familias

$$\{\mathcal{C}_i^* , \mathcal{C}_j , j=1,\ldots,k , j\neq i\}$$

son independientes. Repitiendo el razonamiento y sustituyendo en cada ocasión una clase $\mathcal{C}_j$ por su correspondiente $\mathcal{D}_j$ llegaremos a la conclusión de que las familias

$$\{\mathcal{D}_j , j=1,\ldots,k\}$$

así como las familias

$$\{\mathcal{C}_j^* , j=1,\ldots,k\}$$

son independientes. Pero las $\mathcal{C}_j^*$ gozan de otra importante propiedad . En efecto , al ser las $\mathcal{C}_j$ estables para intersecciones finitas , tenemos que $\mathcal{C}_j^* = \mathcal{A}_j = X_j^{-1}(B)$, $B \in \beta$. Veámoslo:

Sea $\mathcal{C}_j' = \{ C' \in \mathcal{C}_j^* ; \; C' \cap B \in \mathcal{C}_j^* , \forall B \in \mathcal{C}_j \}$

1) $C_1' , C_2' \in \mathcal{C}_j'$, $C_2' \subset C_1' \rightarrow C_1' - C_2' \in \mathcal{C}_j'$

$\forall B \in \mathcal{C}_j$; $(C_1' - C_2') \cap B = C_1' \cap B - C_2' \cap B \in \mathcal{C}_j^*$

2) $\{C_n'\}_{n\geq 1} \subset \mathcal{C}_j'$, $C_n' \cap C_m' = \phi$, $n \neq m \rightarrow \bigcup_{n\geq 1} C_n' \in \mathcal{C}_j'$

$\forall B \in \mathcal{C}_j$, $\big(\bigcup_{n\geq 1} C_n'\big) \cap B = \bigcup_{n\geq 1} (C_n' \cap B) \in \mathcal{C}_j^*$

Resulta pues que $\mathcal{C}_j'$ es estable para las mismas operaciones que lo era $\mathcal{C}_j^*$, además $\forall C \in \mathcal{C}_j$, $C \cap B \in \mathcal{C}_j \subset \mathcal{C}_j^*$, $B \in \mathcal{C}_j$, tendremos pues que $\mathcal{C}_j \subset \mathcal{C}_j'$ y por ser $\mathcal{C}_j^*$ mínima , tendremos $\mathcal{C}_j' \supset \mathcal{C}_j^*$, pero por definición $\mathcal{C}_j^* \supset \mathcal{C}_j'$, en definitiva $\mathcal{C}_j^* \equiv \mathcal{C}_j'$.

Sea ahora $\mathcal{C}_j'' = \{ C \in \mathcal{C}_j^* ; \; C \cap B \in \mathcal{C}_j^* , \forall B \in \mathcal{C}_j^* \}$. De igual manera se comprueba que esta nueva clase resulta estable para las mismas operaciones que lo era $\mathcal{C}_j^*$. Además , $\forall C \in \mathcal{C}_j$, $C \cap B \in \mathcal{C}_j^*$. $\forall B \in \mathcal{C}_j^*$ ya que $\mathcal{C}_j^* \equiv \mathcal{C}_j'$, en definitiva $\mathcal{C}_j \subset \mathcal{C}_j''$ y razonando como antes llegamos a $\mathcal{C}_j'' \equiv \mathcal{C}_j^*$, es decir que $\mathcal{C}_j^*$ es estable para la intersección finita y siendo además estable para las diferencias propias y las sumas numerables , tendremos que es una σ-algebra que necesariamente debe de coincidir con $\mathcal{A}_j = X_j^{-1}(B)$, $B \in \beta$ (se comprueba por doble inclusión).

En definitiva las familias

$$\{\mathcal{C}_j^* , j=1,\ldots,k\} \equiv \{ \mathcal{A}_j = X_j^{-1}(B) , B \in \beta , j=1,\ldots,k\}$$

son independientes , lo que supone que las variables aleatorias X_j , $j=1,\ldots,k$, lo son.

<u>Observación</u>.- La demostración se ha llevado a cabo haciendo abstracción del contenido de las familias $\mathcal{C}_j$. Ello supone que tiene validez general y que en realidad hemos demostrado este teorema , de enunciado más general

<u>TEOREMA</u>.- Las σ-algebras engendradas por familias $\mathcal{C}_j$, independientes y estables para la $\cap$ finita , son también independientes.

Consideremos ahora las funciones de distribución y de densidad conjuntas del vector aleatorio que las k v.a. $X_1,\ldots,X_k$ constituyen . El teorema que a continuación demostramos establece un criterio de independencia para dichas variables basado en las citadas funciones . Constituye este teorema, sin duda , el resultado más interesante de este capítulo

<u>TEOREMA DE FACTORIZACION</u>.- Sean $F_{\vec{X}}(x_1,\ldots,x_k)$ y $f_{\vec{X}}(x_1,\ldots,x_k)$ las funciones de distribución y densidad o cuantía conjuntas del vector aleatorio $\vec{X} = (X_1,\ldots,X_k)$. Sean $F_j(x_j)$ y $f_j(x_j)$, $j=1,\ldots,k$ las correspondientes marginales. Las v.a. X_j , $j=1,\ldots,k$ son independientes si y sólo si se verifica alguna de las dos condiciones (equivalentes) siguientes:

1) $F_{\vec{X}}(x_1,\ldots,x_k) = \prod_{j=1}^{k} F_j(x_j)$, $x_j \in R$, $j=1,\ldots,k$

2) $f_{\vec{X}}(x_1 \ldots x_k) = \prod_{j=1}^{k} f_j(x_j)$, $x_j \in R$, $j = 1, \ldots, k$

Demostración.-

1) Si las v.a. son independientes, sabemos que

$$P(X_j \in B_j \,,\, j = 1, \ldots, k) = \prod_{j=1}^{k} P(X_j \in B_j) \quad , B_j \in \beta$$

tomando $B_j =]-\infty, x_j]$, tendremos

$$P(X_j \in]-\infty, x_j] \,,\, j = 1, \ldots, k) = F_{\vec{X}}(x_1 \ldots x_k) = \prod_{j=1}^{k} F_j(x_j) \quad , x_j \in R$$

El inverso se verifica sin más que utilizar el lema 2.

2) Distinguiremos el caso continuo del discreto a la hora de la demostración.

2.a) Caso discreto.

Sea $A \subset R^k$ tal que $P(\vec{X} \in A) = 1$. Sea $\vec{x} = (x_1 \ldots x_k) \in A$, entonces

$$f_{\vec{X}}(x_1 \ldots x_k) = P(X_j = x_j \,,\, j = 1 \ldots k)$$

y siendo independientes sabemos que

$$P(X_j = x_j \,,\, j = 1 \ldots k) = \prod_{j=1}^{k} P(X_j = x_j) = \prod_{j=1}^{k} f_j(x_j), \quad x_j \in A_j$$

Para cualquier $\vec{x} \in R^k - A$, la igualdad es trivialmente cierta por cuanto es igual a cero en ambos lados.

Para demostrar el inverso, sean $B_j =]-\infty, y_j]$, $y_j \in R$, $j = 1 \ldots k$ y sea $B = B_1 \times B_2 \times \ldots \times B_k$, tenemos

$$P(\vec{X} \in B) = P(X_j \in B_j \,,\, j = 1, \ldots, k) = F_{\vec{X}}(y_1, \ldots, y_k) =$$

$$= \sum_{\vec{x} \in B} f_{\vec{X}}(x_1, \ldots, x_k) = \sum_{\vec{x} \in B} f_1(x_1) \ldots f_k(x_k) = \sum_{(x_1 \ldots x_k) \in B_1 \times \ldots \times B_k} f_1(x_1) \ldots f_k(x_k) =$$

$$= \prod_{j=1}^{k} \left[\sum_{x_j \in B_j} f_j(x_j) \right] = \prod_{j=1}^{k} F_j(y_j)$$

y por 1), las variables aleatorias son independientes.

2b) Caso continuo

Si hay independencia entre las variables aleatorias, tendremos

$$F_{\vec{X}}(x_1 \ldots x_k) = \prod_{j=1}^{k} F_j(x_j)$$

derivando parcialmente en ambos miembros llegamos a

$$f_{\vec{X}}(x_1 \ldots x_k) = \prod_{j=1}^{k} f_j(x_j)$$

El punto $(x_1 \ldots x_k)$ debe constituir un punto de continuidad de la función de ~~[illegible]~~ densidad. En cualquier caso, sabemos que para los vectores aleatorios de tipo continuo el número de puntos de discontinuidad tiene medida nula y por tanto no afecta al resultado.

Para el inverso, sea

$$f_{\vec{X}}(x_1 \ldots x_k) = \prod_{j=1}^{k} f_j(x_j)$$

sea $G_j =]-\infty, y_j]$, $y_j \in R$, $j = 1 \ldots k$. Entonces

$$F_{\vec{X}}(y_1 \ldots y_k) = \int_{\prod_{j=1}^{k} G_j} f_{\vec{X}}(x_1 \ldots x_k)\, dx_1 \ldots dx_k = \int_{\prod_{j=1}^{k} G_j} \prod_{j=1}^{k} f_j(x_j)\, dx_1 \ldots dx_k =$$

$$= \prod_{j=1}^{k} F_j(x_j).$$

El teorema de factorización es también aplicable a las funciones características, pero necesitamos previamente estudiar la esperanza del producto de variables aleatorias cuando éstas son independientes.

LEMA 3.- Sean X_j, $j = 1 \ldots k$ v.a. y sean $g_j : (R, \beta) \to (R, \beta)$ transformaciones medibles, $j = 1 \ldots k$. Si las variables aleatorias X_j, $j = 1 \ldots k$ son independientes, entonces

$$E\left[\prod_{j=1}^{k} g_j(X_j)\right] = \prod_{j=1}^{k} \left[E(g_j(X_j))\right].$$

Demostración.- Supongamos que el vector $\vec{X} = (X_1 \ldots X_k)$ es continuo, entonces

$$E\left(\prod_{j=1}^{k} g_j(X_j)\right) = \int_{R^k} \prod_{j=1}^{k} g_j(X_j)\, f_{\vec{X}}(x_1 \ldots x_k)\, dx_1 \ldots dx_k = \int_{R^k} \prod_{j=1}^{k} g_j(X_j) f_j(x_j) dx_j = \prod_{j=1}^{k} E(g_j(X_j))$$

Análogo razonamiento se aplica para el caso discreto.

Observación.- El inverso del lema no es cierto como pondremos de manifiesto más tarde mediante un adecuado contraejemplo.

COROLARIO 1.- Si X_1 y X_2 son independientes, entonces son incorreladas, siempre que existan los correspondientes momentos.

Demostración.- Recordemos que la incorrelación de dos variables aleatorias suponía que su coeficiente de correlación era nulo. A su vez

$$\rho = \frac{cov(X_1, X_2)}{[var(X_1) . var(X_2)]^{1/2}}$$

pero

$$cov(X_1 . X_2) = E(X_1 X_2) - E(X_1) . E(X_2) = E(X_1) . E(X_2) - E(X_1) . E(X_2) = 0$$

y por tanto $\rho = 0$.

Hay que señalar que la implicación de signo contrario, incorrelación $\to$ independencia no es cierta en general salvo para el caso de distribuciones normales multivariantes que más tarde comentaremos.

COROLARIO 2.- Si el vector $\vec{X} = (X_1, \ldots X_k)$ está constituido por variables aleatorias independientes, entonces

$$Y = \sum_{j=1}^{k} a_j X_j$$

tiene como varianza

$$var(Y) = \sum_{j=1}^{k} a_j^2 \sigma_j^2 \quad , \text{donde } \sigma_j^2 = var(X_j).$$

Demostración.- Recordemos que

$$var(Y) = \sum_{j=1}^{k} a_j^2 \sigma_j^2 + 2 \sum_{j>i} a_i a_j cov(X_i, X_j)$$

y siendo independientes $cov(X_i, X_j) = 0$, $\forall i, j$ y de aquí el resultado.

Estamos ahora en condiciones de enunciar la versión del teorema de caracterización para funciones características.

TEOREMA (de factorización para funciones características).- Las variables aleatorias X_j, $j = 1, \ldots, k$ son independientes, si y sólo si

$$\varphi_{\vec{X}}(t_1, \ldots t_k) = \prod_{j=1}^{k} \varphi_{X_j}(t_j).$$

donde $\vec{X} = (X_1 \ldots X_k)$, y φ representa la función característica.

Demostración.- Si las X_j son independientes entonces $g_j(X_j) = e^{it_j X_j}$ también lo son, $j = 1 \ldots k$. Entonces

$$\varphi_{\vec{X}}(t_1 \ldots t_k) = E[e^{i \sum_{j=1}^{k} t_j X_j}] = E[\prod_{j=1}^{k} e^{it_j X_j}]$$

Aplicando el lema 3, tendremos

$$\varphi_{\vec{X}}(t_1 \ldots t_k) = \prod_{j=1}^{k} E[e^{it_j X_j}] = \prod_{j=1}^{k} \varphi_{X_j}(t_j).$$

Supongamos ahora que se verifica la anterior factorización. Sea $F_{\vec{X}}(x_1 \ldots x_k)$ la función de distribución del vector $\vec{X} = (X_1 \ldots X_k)$. Definamos

$$G(x_1 \ldots x_k) = \prod_{j=1}^{k} F_j(x_j)$$

donde $F_j(x_j)$ es la función de distribución de X_j. La función característica correspondiente a la distribución de probabilidad que $G(\vec{x})$ representa viene dada por $\prod_{j=1}^{k} \varphi_{X_j}(t_j)$, pero de acuerdo con la factorización citada dicha función será igual a la función característica $\varphi_{\vec{X}}(t_1 \ldots t_k)$ del vector $\vec{X}$, aplicando el teorema de unicidad $F_{\vec{X}}(x_1 \ldots x_k) \equiv G(x_1 \ldots x_k)$ y por tanto son independientes las variables aleatorias X_j, $j = 1 \ldots k$.

La independencia de las componentes de un vector aleatorio tiene especial incidencia a la hora de obtener las distribuciones condicionadas. En efecto, sea $\vec{X} = (X_1, \ldots, X_k)$ un vector cuyas componentes son independientes. Sean $\vec{X}_1 = (X_1, \ldots, X_m)$ y $\vec{X}_2 = (X_{m+1}, \ldots, X_k)$ dos subvectores que tienen también sus componentes independientes. Entonces

$$f_{\vec{X}_1/\vec{X}_2}(\vec{x}_1) = \frac{f_{\vec{X}}(\vec{x})}{f_{\vec{X}_2}(\vec{x}_2)} \quad , \text{con } \vec{x} = (\vec{x}_1, \vec{x}_2)$$

pero aplicando el teorema de factorización tenemos

$$f_{\vec{X}_1/\vec{X}_2}(\vec{x}_1) = \frac{\prod_{j=1}^{k} f_j(x_j)}{\prod_{j=m+1}^{k} f_j(x_j)} = \prod_{j=1}^{m} f_j(x_j) = f_{\vec{X}_1}(\vec{x}_1).$$

La independencia supone pues que el conocimiento previo de lo ocurrido con $\vec{X}_2$ no afecta para nada a lo que ocurrirá a $\vec{X}_1$.

Incorrelación e independencia en el caso normal.- Hemos visto que la independencia de dos variables aleatorias supone su incorrelación. Ya hemos señalado que el recíproco no es cierto como pondremos de manifiesto más tarde mediante un contraejemplo, salvo en un caso concreto del que nos ocupamos en el teorema siguiente:

TEOREMA.- Sea $\vec{X} = (X_1, X_2)$ un vector normal bivariante. Entonces X_1 y X_2 son independientes si y sólo si son incorrelados.

Demostración.- Ya sabemos que

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-q/2}$$

donde

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\,\frac{x_1-\mu_1}{\sigma_1}\cdot\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

Ya sabemos que las marginales correspondientes vienen dadas por

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right] \quad , \quad f_2(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right].$$

Observemos que si $\rho = 0$, entonces

$$f(x_1, x_2) = f_1(x_1)\cdot f_2(x_2)$$

con lo cual las variables X_1, X_2 son independientes.

INDEPENDENCIA DE VECTORES ALEATORIOS

El concepto de independencia se extiende de inmediato a familias de vectores aleatorios.

Definición.- Los vectores $\vec{X}_j$, $j=1,\dots,k$ son independientes si lo son las σ-álgebras que inducen.

Puede también enunciarse un teorema de factorización para los vectores y puede también afirmarse que cualquier transformación medible de vectores aleatorios da lugar a nuevos vectores aleatorios.

Obsérvese, no obstante, que la independencia entre vectores aleatorios no afirma nada acerca de los componentes de dichos vectores, de manera que puede ser perfectamente lícito que en cada vector no haya independencia entre sus componentes. Véase el siguiente ejemplo.

Ejemplo.- Sea (X, Y) un vector aleatorio con función de densidad conjunta

$$f(x,y) = \frac{1}{4}\left[1 + xy(x^2-y^2)\right] \quad , \quad |x|\le 1,\ |y|\le 1$$

y sea (U,V) un vector aleatorio con densidad conjunta

$$g(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\} \qquad u\in\mathbb{R},\ v\in\mathbb{R},\ |\rho|<1.$$

La densidad conjunta de (X, Y, U, V) viene dada por

$$h(x,y,u,v) = \frac{1}{8\pi\sqrt{1-\rho^2}}\left[1+xy(x^2-y^2)\right]\exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}$$

y por tanto (X,Y), (U,V) son independientes. Sin embargo X e Y, así como U y V, no son independientes.

(9) ALGUNAS APLICACIONES DE LAS FUNCIONES CARACTERISTICAS

Veremos en este apartado la utilidad de las funciones características para obtener la distribución de probabilidad de ciertas variables aleatorias definidas como sumas de variables aleatorias independientes.

TEOREMA (Binomial).- Sean X_j unas v.a. $B(n_j, p)$, $j=1,\dots,k$ e independientes. Entonces

$$X = \sum_{j=1}^{k} X_j \quad \text{es una } B(n,p) \text{ con } n = \sum_{j=1}^{k} n_j.$$

Demostración.- Sabemos que

$$\varphi_{X_j}(t) = (pe^{it}+q)^{n_j}$$

entonces

$$\varphi_X(t) = E\left[e^{itX}\right] = E\left[e^{it\sum_j X_j}\right] = E\left[\prod_{j=1}^{k} e^{itX_j}\right] = \prod_{j=1}^{k} E(e^{itX_j}) =$$

$$= \prod_{j=1}^{k}\varphi_{X_j}(t) = \prod_{j=1}^{k}(pe^{it}+q)^{n_j} = (pe^{it}+q)^{\sum n_j}.$$

En realidad esta propiedad aditiva de las binomiales con igual p las caracteriza, por cuanto puede enunciarse el recíproco del teorema que acabamos de demostrar, lo que supone que esto sólo ocurre para las binomiales de estas características.

TEOREMA (Poisson).- Sean X_j, $P(\lambda_j)$, $j=1,\dots,k$ e independientes. Entonces

$$X = \sum_{j=1}^{k} X_j \quad \text{es } P(\lambda), \text{ con } \lambda = \sum_{j=1}^{k}\lambda_j.$$

Demostración.-

$$\varphi_X(t) = \prod_{j=1}^{k}\varphi_{X_j}(t) = \prod_{j=1}^{k}\exp(\lambda_j e^{it}-\lambda_j) = \exp\left(e^{it}\sum\lambda_j - \sum\lambda_j\right) =$$

$$= \exp(e^{it}\lambda - \lambda) \quad , \text{ con } \lambda = \sum_{j=1}^{k}\lambda_j.$$

Al igual que antes, también esta propiedad caracteriza a las distribuciones de Poisson por cuanto también, en esta ocasión, es cierto el recíproco. Hay además una curiosa propiedad que se deriva del teorema que acabamos de demostrar.

COROLARIO.- Sean X e Y independientes con distribución $P(\lambda_1)$ y $P(\lambda_2)$ respectivamente. Entonces la distribución condicional de X, dada $X+Y$ es una binomial.

Demostración.- La distribución de $Z = X+Y$, de acuerdo con lo anterior, es $P(\lambda_1+\lambda_2)$. Entonces, para enteros m y n, $m<n$, tendremos

$$P(X=m / X+Y=n) = \frac{P(X=m,\ Y=n-m)}{P(X+Y=n)} = \frac{P(X=m)\cdot P(Y=n-m)}{P(X+Y=n)} =$$

$$= \frac{e^{-\lambda_1}\cdot \lambda_1^{m}/m! \cdot e^{-\lambda_2}\,\lambda_2^{n-m}/(n-m)!}{e^{-(\lambda_1+\lambda_2)}\cdot(\lambda_1+\lambda_2)^{n}/n!} =$$

$$= \frac{n!}{m!\,(n-m)!}\cdot\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{m}\cdot\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-m} = \binom{n}{m} p^{m}\cdot q^{n-m}$$

donde $p=\lambda_1/(\lambda_1+\lambda_2)$ y $q=1-p=\lambda_2/(\lambda_1+\lambda_2)$, y se trata de una Binomial con parámetros n y $p=\lambda_1/(\lambda_1+\lambda_2)$.

TEOREMA (Normal).- Sean X_j, $N(\mu_j,\sigma_j^2)$, $j=1,\dots,k$ e independientes. Entonces

$$X=\sum_{j=1}^{k} a_j X_j \text{ es } N(\mu,\sigma^2), \text{ con } \mu=\sum_{j=1}^{k} a_j\mu_j \text{ y } \sigma^2=\sum_{j=1}^{k} a_j^2\sigma_j^2.$$

Demostración.-

$$\varphi_X(t)=\prod_{j=1}^{k} E\left(e^{ita_jX_j}\right)=\prod_{j=1}^{k}\varphi_{X_j}(a_jt)=\prod_{j=1}^{k}\exp\left\{ita_j\mu_j-\frac{\sigma_j^2a_j^2t^2}{2}\right\}=$$

$$=\exp\left\{it\sum_{j=1}^{k}a_j\mu_j-\frac{t^2}{2}\sum_{j=1}^{k}a_j^2\sigma_j^2\right\}=\exp\left\{it\mu-\frac{t^2\sigma^2}{2}\right\}$$

con $\mu=\sum_{j=1}^{k}a_j\mu_j$ y $\sigma^2=\sum_{j=1}^{k}a_j^2\sigma_j^2$.

Una aplicación interesante de este teorema se produce en la siguiente situación:

Sean X_j $j=1,\dots,k$ variables aleatorias independientes todas ellas con igual media, μ e igual varianza σ^2. Definamos una nueva variable

$$\bar X=\sum_{j=1}^{k} X_j/k$$

por las propiedades de la media y la varianza y dada la independencia de las X_k, sabemos que

$$E(\bar X)=\mu \quad \text{y} \quad var(\bar X)=\sigma^2/k.$$

Si todas las X_j son, además, normales, entonces

COROLARIO.- Sean X_j, $N(\mu,\sigma^2)$, $j=1,\dots,k$ e independientes. Entonces $\bar X=\sum_{j=1}^{k}X_j/k$ es $N(\mu,\sigma^2/k)$

Demostración.- Basta aplicar el teorema con $a_j=1/k$, $\mu_j=\mu$ y $\sigma_j^2=\sigma^2$, $j=1,\dots,k$. Obsérvese que la variable $Z=\sqrt{k}(\bar X-\mu)/\sigma$ es $N(0,1)$.

 TEOREMA (χ^2).- Sean X_j, $\chi^2_{r_j}$, $j=1,\dots,k$ e independientes. Entonces

$$X=\sum_{j=1}^{k}X_j \text{ es } \chi^2_r \text{ con } r=\sum_{j=1}^{k}r_j.$$

Se demuestra como los anteriores. Una aplicación interesante de este teorema es la siguiente:

Sean X_j $N(\mu_j,\sigma_j^2)$, $j=1,\dots,k$ e independientes. Hemos visto en otra parte que las variables

$$Z_j=\frac{X_j-\mu_j}{\sigma_j} \text{ son } N(0,1) \text{ y desde luego independientes, } j=1,\dots,k.$$

También son independientes las variables Z_j^2 y además, como ya sabemos, son χ^2_1, entonces aplicando el teorema

$$\sum_{j=1}^{k}Z_j^2=\sum_{j=1}^{k}\left(\frac{X_j-\mu_j}{\sigma_j}\right)^2 \text{ es } \chi^2_k.$$

TEOREMA (Cauchy).- Sean X_j v.a. Cauchy con $\mu=0$, $\sigma=1$, $j=1,\dots,k$ e independientes. Entonces

$$X=\sum_{j=1}^{k}X_j$$

es una variable kY, donde Y es Cauchy con $\mu=0$, $\sigma=1$ y por tanto $\bar X=X/k$ es Cauchy con $\mu=0$, $\sigma=1$.

Demostración.-

$$\varphi_X(t)=\prod_{j=1}^{k}\varphi_{X_j}(t)=\prod_{j=1}^{k}e^{-|t|}=e^{-k|t|}$$

que es la función característica de una variable $Z=kY$, con Y Cauchy $\mu=0$, $\sigma=1$.

La segunda afirmación es inmediata a partir de este resultado.

⇒ Un ejemplo para demostrar que la incorrelación no supone independencia.

Sean X_1 y X_2 variables aleatorias que toman los valores $-1, 0, 1$ con arreglo a las siguientes probabilidades

X_2 \ X_1	-1	0	1	
-1	α	β	α	$2\alpha+\beta$
0	β	0	β	2β
1	α	β	α	$2\alpha+\beta$
	$2\alpha+\beta$	2β	$2\alpha+\beta$	$4\alpha+4\beta$

$\alpha, \beta > 0 \quad , \quad \alpha+\beta = \frac{1}{4}$

Entonces

$$Cov(X_1, X_2) = E(X_1 X_2) - E(X_1) \cdot E(X_2)$$

$$E(X_1 X_2) = (-1)(-1)\cdot\alpha + (-1)1\cdot\alpha + 1\cdot(-1)\alpha + 1\cdot 1\cdot\alpha = \alpha - \alpha - \alpha + \alpha = 0$$

$$E(X_1) = -(2\alpha+\beta) + 2\alpha+\beta = 0$$

$$E(X_2) = -(2\alpha+\beta) + 2\alpha+\beta = 0$$

En consecuencia

$$\rho = \frac{Cov(X_1, X_2)}{\sigma(X_1) \cdot \sigma(X_2)} = 0$$

pero es evidente que la función de cuantía conjunta no verifica la factorización

$$f(x_1, x_2) \neq f_1(x_1) \cdot f_2(x_2)$$ donde f_i es la marginal de X_i.