Procesos Estocásticos para Ingenieros: Teoría y Aplicaciones

Francisco Montes Suay

Departament d'Estadística i Investigació Operativa
Universitat de València

Departament d'Estadística i Investigació Operativa
Universitat de València
46100-Burjassot
Spain

Índice general

Capítulo 1

Probabilidad. Variable aleatoria. Vector aleatorio

1.1. Probabilidad

1.1.1. Causalidad y aleatoriedad

A cualquiera que preguntemos cuanto tiempo tardaríamos en recorrer los 350 kilómetros que separan Valencia de Barcelona, si nos desplazamos con velocidad constante de 100 kms/hora, nos contestará sin dudar que 3 horas y media. Su actitud será muy distinta si, previamente a su lanzamiento, le preguntamos por la cara que nos mostrará un dado. Se trata de dos fenómenos de naturaleza bien distinta,

- *el primero* pertenece a los que podemos denominar **deterministas**, aquellos en los que la relación causa-efecto aparece perfectamente determinada. En nuestro caso concreto, la conocida ecuación $e = v \cdot t$, describe dicha relación,
- *el segundo* pertenece a la categoría de los que denominamos **aleatorios**, que se caracterizan porque aun repitiendo en las mismas condiciones el experimento que lo produce, el resultado variará de una repetición a otra dentro de un conjunto de posibles resultados.

La Teoría de la Probabilidad pretende emular el trabajo que los físicos y, en general, los científicos experimentales han llevado a cabo. Para entender esta afirmación observemos que la ecuación anterior, $e = v \cdot t$, es un resultado experimental que debemos ver como *un modelo matemático* que, haciendo abstracción del móvil concreto y del medio en el que se desplaza, describe la relación existente entre el espacio, el tiempo y la velocidad. La Teoría de la Probabilidad nos permitirá la obtención de *modelos aleatorios o estocásticos* mediante los cuales podremos conocer, en términos de probabilidad, el comportamiento de los fenómenos aleatorios.

1.1.2. Experimento, resultado, espacio muestral y suceso

Nuestro interlocutor sí que será capaz de responder que el dado mostrará una de sus caras. Al igual que sabemos que la extracción al azar de una carta de una baraja española pertenecerá a uno de los cuatro palos: oros, copas, espadas o bastos. Es decir, el *experimento* asociado a nuestro fenómeno aleatorio[1] da lugar a un *resultado*, ω, de entre un conjunto de posibles resultados.

[1]Una pequeña disquisición surge en este punto. La aleatoriedad puede ser inherente al fenómeno, lanzar un dado, o venir inducida por el experimento, extracción al azar de una carta. Aunque conviene señalarlo, no es

Este conjunto de posibles resultados recibe el nombre de *espacio muestral*, Ω. Subconjuntos de resultados con una característica común reciben el nombre de *sucesos aleatorios* o, simplemente, sucesos. Cuando el resultado del experimento pertenece al suceso A, decimos que **ha ocurrido** o **se ha realizado** A.

A continuación mostramos ejemplos de experimentos aleatorios, los espacios muestrales asociados y algunos sucesos relacionados.

Lanzamiento de dos monedas.- Al lanzar dos monedas el espacio muestral viene definido por Ω ={CC,C+,+C,++}. Dos ejemplos de sucesos en este espacio pueden ser:

A ={Ha salido una cara}={C+,+C},

B ={Ha salido más de una cruz}={++}.

Elegir un punto al azar en el círculo unidad.- Su espacio muestral es Ω ={Los puntos del círculo}. Ejemplos de sucesos:

$A = \{\omega;\ d(\omega, centro) < 0{,}5\}$,

$B = \{\omega;\ 0,3 < d(\omega, centro) < 0{,}75\}$.

Sucesos, conjuntos y σ-álgebra de sucesos

Puesto que los sucesos no son más que subconjuntos de Ω, podemos operar con ellos de acuerdo con las reglas de la teoría de conjuntos. Todas las operaciones entre conjuntos serán aplicables a los sucesos y el resultado de las mismas dará lugar a nuevos sucesos cuyo significado debemos conocer. Existen, por otra parte, sucesos cuya peculiaridad e importancia nos lleva a asignarles nombre propio. De estos y de aquellas nos ocupamos a continuación:

Suceso cierto o seguro: cuando llevamos a cabo cualquier experimento aletorio es seguro que el resultado pertenecerá al espacio muestral, por lo que Ω, en tanto que suceso, ocurre siempre y recibe el nombre de *suceso cierto* o *seguro*.

Suceso imposible: en el extremo opuesto aparece aquel suceso que no contiene ningún resultado que designamos mediante $\emptyset$ y que, lógicamente, no ocurre nunca, razón por la cual se le denomina *suceso imposible*.

Sucesos complementarios: la ocurrencia de un suceso, A, supone la no ocurrencia del suceso que contiene a los resultados que no están en A, es decir, A^c. Ambos sucesos reciben el nombre de *complementarios*.

Unión de sucesos: la unión de dos sucesos, $A \cup B$, da lugar a un nuevo suceso que no es más que el conjunto resultante de dicha unión. En consecuencia, $A \cup B$ *ocurre* cuando el resultado del experimento pertenece a A, a B o ambos a la vez.

Intersección de sucesos: la intersección de dos sucesos, $A \cap B$, es un nuevo suceso cuya *realización* tiene lugar si el resultado pertenece a ambos a la vez, lo que supone que ambos ocurren simultáneamente.

Sucesos incompatibles: Existen sucesos que al no compartir ningún resultado su intersección es el suceso imposible, $A \cap B = \emptyset$. Se les denomina, por ello, *sucesos incompatibles*. Un suceso A y su complementario A^c, son un buen ejemplo de sucesos incompatibles.

---

este lugar para profundizar en la cuestión

Siguiendo con el desarrollo emprendido parece lógico concluir que todo subconjunto de Ω será un suceso. Antes de admitir esta conclusión conviene una pequeña reflexión: la noción de suceso es un concepto que surge con naturalidad en el contexto de la experimentación aleatoria pero, aunque no hubiera sido así, la necesidad del concepto nos hubiera obligado a *inventarlo*. De la misma forma, es necesario que los sucesos poseean una mínima *estructura* que garantice la estabilidad de las operaciones *naturales* que con ellos realicemos, entendiendo por naturales la *complementación*, la *unión* y la *intersección*. Esta dos últimas merecen comentario aparte para precisar que no se trata de uniones e intersecciones en número cualquiera, puesto que más allá de la numerabilidad nos movemos con dificultad. Bastará pues que se nos garantice que uniones e intersecciones numerables de sucesos son estables y dan lugar a otro suceso. Existe una estructura algebraica que verifica las condiciones de estabilidad que acabamos de enumerar.

Definición 1.1 (σ-álgebra de conjuntos) *Una familia de conjuntos $\mathcal{A}$ definida sobre Ω decimos que es una σ-álgebra si:*

1. $\Omega \in \mathcal{A}$.
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.
3. $\{A_n\}_{n\geq 1} \subset \mathcal{A} \Rightarrow \bigcup_{n\geq 1} A_n \in \mathcal{A}$.

La familia de las partes de Ω, $\mathcal{P}(\Omega)$, cumple con la definición y es por tanto una σ-álgebra de sucesos, de hecho la más grande de las existentes. En muchas ocasiones excesivamente grande para nuestras necesidades, que vienen determinadas por el núcleo inicial de sucesos objeto de interés. El siguiente ejemplo permite comprender mejor este último comentario.

Ejemplo 1.1 *Si suponemos que nuestro experimento consiste en elegir al azar un número en el intervalo [0,1], nuestro interés se centrará en conocer si la elección pertenece a cualquiera de los posibles subintervalos de [0,1]. La σ-álgebra de sucesos generada a partir de ellos, que es la menor que los contiene, se la conoce con el nombre de* σ-álgebra de Borel en [0,1]*, $\beta_{[0,1]}$, y es estrictamente menor que $\mathcal{P}([0,1])$.*

En resumen, el espacio muestral vendrá acompañado de la correspondiente σ-álgebra de sucesos, la más conveniente al experimento. La pareja que ambos constituyen, $(\Omega, \mathcal{A})$, recibe el nombre de *espacio probabilizable*.

Señalemos por último que en ocasiones no es posible economizar esfuerzos y $\mathcal{A}$ coincide con $\mathcal{P}(\Omega)$. Por ejemplo cuando el espacio muestral es numerable.

1.1.3. Probabilidad y sus propiedades

Ya sabemos que la naturaleza aleatoria del experimento impide *predecir* de antemano el resultado que obtendremos al llevarlo a cabo. Queremos conocer si cada suceso de la σ-álgebra se realiza o no. Responder de una forma categórica a nuestro deseo es demasiado ambicioso. Es imposible predecir en cada realización del experimento si el resultado va a estar o no en cada suceso. En Probabilidad la pregunta se formula del siguiente modo: ¿qué posibilidad hay de que tenga lugar cada uno de los sucesos? La respuesta exige un tercer elemento que nos proporcione esa información: Una función de conjunto P, es decir, una función definida sobre la σ-álgebra de sucesos, que a cada uno de ellos le asocie un valor numérico que exprese la mayor o menor probabilidad o posibilidad de producirse cuando se realiza el experimento. Esta función de conjunto se conoce como *medida de probabilidad* o simplemente *probabilidad*. Hagamos un breve incursión histórica antes de definirla formalmente.

El concepto de probabilidad aparece ligado en sus orígenes a los juegos de azar, razón por la cual se tiene constancia del mismo desde tiempos remotos. A lo largo de la historia se han

hecho muchos y muy diversos intentos para formalizarlo, dando lugar a otras tantas *definiciones* de probabilidad que adolecían todas ellas de haber sido confeccionadas *ad hoc*, careciendo por tanto de la generalidad suficiente que permitiera utilizarlas en cualquier contexto. No por ello el interés de estas definiciones es menor, puesto que supusieron sucesivos avances que permitieron a Kolmogorov enunciar su conocida y definitiva axiomática en 1933. De entre las distintas aproximaciones, dos son las más relevantes:

Método frecuencialista.- Cuando el experimento es susceptible de ser repetido en las mismas condiciones una infinidad de veces, la probabilidad de un suceso A, $P(A)$, se define como el límite[2] al que tiende la *frecuencia relativa de ocurrencias* del suceso A.

Método clásico (Fórmula de Laplace).- Si el experimento conduce a un espacio muestral finito con n resultados posibles, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, todos ellos igualmente probables, la probabilidad de un suceso A que contiene m de estos resultados se obtiene mediante la fórmula

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

conocida como *fórmula de Laplace*, que la propuso a finales del siglo XVIII. La fórmula se enuncia como *el cociente entre el número de casos favorables y el número de casos posibles*. Obsérvese la incorreción formal de esta aproximación en la medida que exige equiprobabilidad en los resultados para poder definir, precisamente, la probabilidad, lo cual implica un conocimiento previo de aquello que se quiere definir.

Las anteriores definiciones son aplicables cuando las condiciones exigidas al experimento son satisfechas y dejan un gran número de fenómenos aleatorios fuera de su alcance. Estos problemas se soslayan con la definición axiomática propuesta por A.N.Kolmogorov en 1933:

Definición 1.2 (Probabilidad) *Una función de conjunto, P, definida sobre la σ-álgebra $\mathcal{A}$ es una probabilidad si:*

1. $P(A) \geq 0$ *para todo* $A \in \mathcal{A}$.
2. $P(\Omega) = 1$.
3. P *es numerablemente aditiva, es decir, si* $\{A_n\}_{n\geq 1}$ *es una sucesión de sucesos disjuntos de* $\mathcal{A}$*, entonces*

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

A la terna $(\Omega, \mathcal{A}, P)$ la denominaremos *espacio de probabilidad*.

Señalemos, antes de continuar con algunos ejemplos y con las propiedades que se derivan de esta definición, que los axiomas propuestos por Kolmogorov no son más que la generalización de las propiedades que posee la frecuencia relativa. La definición se apoya en las aproximaciones previas existentes y al mismo tiempo las incluye como situaciones particulares que son.

Ejemplo 1.2 (Espacio de probabilidad discreto. La fórmula de Laplace) *Supongamos un espacio muestral, Ω, numerable y como σ-álgebra la familia formada por todos los posibles subconjuntos de Ω. Sea p una función no negativa definida sobre Ω verificando: $\sum_{\omega\in\Omega} p(\omega) = 1$. Si definimos $P(A) = \sum_{\omega\in A} p(\omega)$, podemos comprobar con facilidad que P es una probabilidad.*

[2]Debemos advertir que no se trata aquí de un límite puntual en el sentido habitual del Análisis. Más adelante se introducirá el tipo de convergencia al que nos estamos refiriendo

Hay un caso particular de especial interés, el llamado espacio de probabilidad discreto uniforme*, en el que Ω es finito, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, y $p(\omega_i) = \frac{1}{n}$, $\forall \omega_i \in \Omega$. Entonces, para $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega i_m\}$ se tiene*

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

que es la fórmula de Laplace, obtenida ahora con rigor. El nombre de uniforme *se justifica porque la masa de probabilidad está uniformemente repartida al ser constante en cada punto.*

Un ejemplo de espacio de probabilidad discreto uniforme es el que resulta de lanzar dos dados. El espacio muestral, Ω =\{(1,1),(1,2),...,(6,5),(6,6)\}, está formado por las 6×6 posibles parejas de caras. Si los dados son correctos, cualquiera de estos resultados tiene la misma probabilidad, 1/36. Sea ahora A =\{ambas caras son pares\}, el número de puntos que contiene A son 9, por lo que aplicando la fórmula de Laplace, $P(A) = 9/36 = 1/4$.

Propiedades de la probabilidad

De la definición de probabilidad se deducen algunas propiedades muy útiles.

La probabilidad del vacío es cero.- $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \ldots$ y por la aditividad numerable, $P(\Omega) = P(\Omega) + \sum_{k \geq 1} P(\emptyset)$, de modo que $P(\emptyset) = 0$.

Aditividad finita.- Si $A_1, \ldots, A_n$ son elementos disjuntos de $\mathcal{A}$, aplicando la σ-aditividad, la propiedad anterior y haciendo $A_i = \emptyset,\ i > n$ tendremos

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

Se deduce de aquí fácilmente que $\forall A \in \mathcal{A}$, $P(A^c) = 1 - P(A)$.

Monotonía.- Si $A, B \in \mathcal{A},\ A \subset B$, entonces de $P(B) = P(A) + P(B - A)$ se deduce que $P(A) \leq P(B)$.

Probabilidad de una unión cualquiera de sucesos (fórmula de inclusión-exclusión).- Si $A_1, \ldots, A_n \in \mathcal{A}$, entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n). \tag{1.1}$$

Subaditividad.- Dados los sucesos $A_1, \ldots, A_n$, la relación existente entre la probabilidad de la unión de los A_i y la probabilidad de cada uno de ellos es la siguiente:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

Continuidad de la probabilidad.- Sea $\{A_n\}_{n \geq 1}$ una sucesión monótona de sucesos y sea A su límite. Se demuestra fácilmente que

$$P(A) = P(\lim_n A_n) = \lim_{n \to +\infty} P(A_n).$$

1.1.4. Probabilidad condicionada. Teorema de Bayes

Si compramos un número para una rifa que se celebra anualmente durante las fiestas de verano en nuestro pueblo y que está compuesta por 100 boletos numerados del 1 al 100, sabemos que nuestra probabilidad ganar el premio, suceso que designaremos por A, vale

$$P(A) = \frac{1}{100}$$

Supongamos que a la mañana siguiente de celebrase el sorteo alguien nos informa que el boleto premiado termina en 5. Con esta información, ¿continuaremos pensando que nuestra probabilidad de ganar vale 10^{-2}? Desde luego sería absurdo continuar pensándolo si nuestro número termina en 7, porque evidentemente la *nueva* probabilidad valdría $P'(A) = 0$, pero aunque terminara en 5 también nuestra probabilidad de ganar habría cambiado, porque los números que terminan en 5 entre los 100 son 10 y entonces

$$P'(A) = \frac{1}{10},$$

10 veces mayor que la inicial.

Supongamos que nuestro número es el 35 y repasemos los elementos que han intervenido en la nueva situación. De una parte, un suceso original $A = \{$*ganar el premio con el número 35*$\}$, de otra, un suceso $B\{$*el boleto premiado termina en 5*$\}$ de cuya ocurrencia se nos informa *a priori*. Observemos que $A \cap B = \{$*el número 35*$\}$ y que la nueva probabilidad encontrada verifica,

$$P'(A) = \frac{1}{10} = \frac{1/100}{10/100} = \frac{P(A \cap B)}{P(B)},$$

poniendo en evidencia algo que cabía esperar, que la *nueva* probabilidad a depende de $P(B)$. Estas propiedades observadas justifican la definición que damos a continuación.

Definición 1.3 (Probabilidad condicionada) *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sean A y B dos sucesos, con $P(B) > 0$, se define la* probabilidad de A condicionada a B *mediante la expresión,*

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

Teorema de factorización

A partir de la definición de probabilidad condicionada, la probabilidad de la intersección de dos sucesos puede expresarse de la forma $P(A \cap B) = P(A|B)P(B)$. El teorema de factorización extiende este resultado para cualquier intersección finita de sucesos.

Consideremos los sucesos $A_1, A_2, \ldots, A_n$, tales que $P(\cap_{i=1}^{n} A_i) > 0$, por inducción se comprueba fácilmente que

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(A_n | \cap_{i=1}^{n-1} A_i\right) P\left(A_{n-1} | \cap_{i=1}^{n-2} A_i\right) \ldots P(A_2|A_1)P(A_1). \tag{1.2}$$

Ejemplo 1.3 *En una urna que contiene 5 bolas blancas y 4 negras, llevamos a cabo 3 extracciones consecutivas sin reemplazamiento. ¿Cuál es la probabilidad de que las dos primeras sean blancas y la tercera negra?*

Cada extracción altera la composición de la urna y el total de bolas que contiene. De acuerdo con ello tendremos (la notación es obvia)

$$\begin{aligned} &P(B_1 \cap B_2 \cap N_3) = \\ &\quad = P(N_3|B_1 \cap B_2)P(B_2|B_1)P(B_1) = \tfrac{4}{7} \cdot \tfrac{4}{8} \cdot \tfrac{5}{9} \end{aligned}$$

Teorema de la probabilidad total

Si los sucesos $A_1, A_2, \ldots, A_n$ constituyen un partición del Ω, tal que $P(A_i) > 0$, $\forall i$, tendremos que cualquier suceso B podrá particionarse de la forma, $B = \cup_{i=1}^{n} B \cap A_i$ y tratándose de una unión disjunta podremos escribir

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \tag{1.3}$$

Este resultado se conoce con el nombre de *teorema de la probabilidad total*.

Teorema de Bayes

Puede tener interés, y de hecho así ocurre en muchas ocasiones, conocer la probabilidad asociada a cada elemento de la partición dado que ha ocurrido B, es decir, $P(A_i|B)$. Para ello, recordemos la definición de probabilidad condicionada y apliquemos el resultado anterior.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}.$$

Este resultado, conocido como el *teorema de Bayes*, permite conocer el cambio que experimenta la probabilidad de A_i como consecuencia de haber ocurrido B. En el lenguaje habitual del Cálculo de Probabilidades a $P(A_i)$ se la denomina probabilidad *a priori* y a $P(A_i|B)$ probabilidad *a posteriori*, siendo la ocurrencia de B la que establece la frontera entre el antes y el después. ¿Cuál es, a efectos prácticos, el interés de este resultado? Veámoslo con un ejemplo.

Ejemplo 1.4 *Tres urnas contienen bolas blancas y negras. La composición de cada una de ellas es la siguiente:* $U_1 = \{3B, 1N\}$, $U_2 = \{2B, 2N\}$, $U_3 = \{1B, 3N\}$. *Se elige al azar una de las urnas, se extrae de ella una bola al azar y resulta ser blanca. ¿Cuál es la urna con mayor probabilidad de haber sido elegida?*

Mediante U_1, U_2 *y* U_3, *representaremos también la urna elegida. Estos sucesos constituyen una partición de* Ω *y se verifica, puesto que la elección de la urna es al azar,*

$$P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}.$$

Si B=\{la bola extraída es blanca\}, *tendremos*

$$P(B|U_1) = \frac{3}{4}, \quad P(B|U_2) = \frac{2}{4}, \quad P(B|U_3) = \frac{1}{4}.$$

Lo que nos piden es obtener $P(U_i|B)$ *para conocer cuál de las urnas ha originado, más probablemente, la extracción de la bola blanca. Aplicando el teorema de Bayes a la primera de las urnas,*

$$P(U_1|B) = \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{1}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{2}{4} + \frac{1}{3} \cdot \frac{1}{4}} = \frac{3}{6},$$

y para las otras dos, $P(U_2|B) = 2/6$ *y* $P(U_3|B) = 1/6$. *Luego la primera de las urnas es la que con mayor probabilidad dió lugar a una extracción de bola blanca.*

El teorema de Bayes es uno de aquellos resultados que inducen a pensar que *la cosa no era para tanto*. Se tiene ante él la sensación que produce lo trivial, hasta el punto de atrevernos a pensar que lo hubiéramos podido deducir nosotros mismos de haberlo necesitado, aunque afortunadamente el Reverendo Thomas Bayes se ocupó de ello en un trabajo titulado *An Essay towards solving a Problem in the Doctrine of Chances*, publicado en 1763. Conviene precisar que Bayes no planteó el teorema en su forma actual, que es debida a Laplace.

1.1.5. Independencia

La información previa que se nos proporcionó sobre el resultado del experimento modificó la probabilidad inicial del suceso. ¿Ocurre esto siempre? Veámoslo.

Supongamos que en lugar de comprar un único boleto, el que lleva el número 35, hubiéramos comprado todos aquellos que terminan en 5. Ahora $P(A) = 1/10$ puesto que hemos comprado 10 boletos, pero al calcular la probabilidad condicionada a la información que se nos ha facilitado, $B = \{$*el boleto premiado termina en 5*$\}$, observemos que $P(A \cap B) = 1/100$ porque al intersección de ambos sucesos es justamente el boleto que está premiado, en definitiva

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/100}{10/100} = \frac{1}{10},$$

la misma que originalmente tenía A. Parecen existir situaciones en las que la información previa no modifica la probabilidad inicial del suceso. Observemos que este resultado tiene una consecuencia inmediata,

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

Esta es una situación de gran importancia en probabilidad que recibe el nombre de *independencia* de sucesos y que generalizamos mediante la siguiente definición.

Definición 1.4 (Sucesos independientes) *Sean A y B dos sucesos. Decimos que A y B son* independientes *si $P(A \cap B) = P(A)P(B)$.*

De esta definición se obtiene como propiedad,

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

y su simétrica $P(B|A) = P(B)$.

En ocasiones se define la independencia de dos sucesos a partir de este resultado, obteniéndose entonces como propiedad la que nosotros hemos dado como definición. Existe equivalencia entre ambas definiciones, aunque a fuerza de ser rigurosos, hay que matizar que definir el concepto a partir de la probabilidad condicional exige añadir la condición de que el suceso condicionante tenga probabilidad distinta de cero. Hay además otra ventaja a favor de la definición basada en la factorización de $P(A \cap B)$, pone de inmediato en evidencia la *simetría* del concepto.

El concepto de independencia puede extenderse a una familia finita de sucesos de la siguiente forma.

Definición 1.5 (Independencia mutua) *Se dice que los sucesos de la familia $\{A_1, \ldots, A_n\}$ son* mútuamente independientes *cuando*

$$P(A_{k_1} \cap \ldots \cap A_{k_m}) = \prod_{i=1}^{m} P(A_{k_i}) \tag{1.4}$$

siendo $\{k_1, \ldots, k_m\} \subset \{1, \ldots, n\}$ y los k_i distintos.

Conviene señalar que la independencia mutua de los n sucesos supone que han de verificarse $\binom{n}{n} + \binom{n}{n-1} + \ldots \binom{n}{2} = 2^n - n - 1$ ecuaciones del tipo dado en (1.4).

Si solamente se verificasen aquellas igualdades que implican a dos elementos diríamos que los sucesos son *independientes dos a dos*, que es un tipo de independencia menos restrictivo que el anterior como pone de manifiesto el siguiente ejemplo. Solo cuando $n = 2$ ambos conceptos son equivalentes.

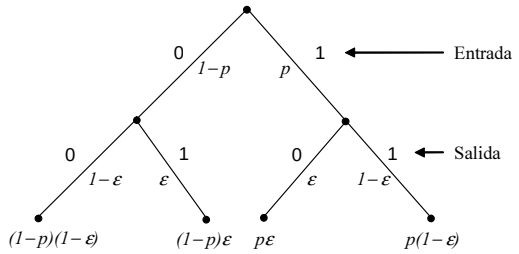
La definición puede generalizarse a familias o clases de sucesos.

Definición 1.6 (Clases independientes de sucesos) *Las clases de sucesos $\mathcal{A}_1, \ldots, \mathcal{A}_n \subset \mathcal{A}$ se dicen* independientes*, si al tomar A_i en cada $\mathcal{A}_i$, $i = 1, \ldots, n$, los sucesos de la familia $\{A_1, \ldots, A_n\}$ son independientes.*

Notemos que en la definición no se exige que los elementos de cada clase $\mathcal{A}_i$ sean independientes entre sí. De hecho A y A^c sólo lo son si $P(A) = 0$ ó $P(A) = 1$.

Para una *colección infinita de clases de sucesos* la anterior definición se extiende con facilidad. Diremos que $\{\mathcal{A}_n\}_{n\geq 1} \subset \mathcal{A}$ son *independientes* si cualquier subcolección finita lo es.

Ejemplo 1.5 (Teorema de Bayes y sistemas de comunicación) *Muchos sistemas de comunicación pueden modelizarse de la siguiente forma. El usuario entra un 0 o un 1 en el sistema y la correspondiente señal es transmitida. El receptor toma una decisión acerca del cuál fue la entrada del sistema en función de la señal recibida. Supongamos que el usuario envia un 0 con probabilidad* 1-p *y un 1 con probabilidad* p*, mientras que el receptor toma un decisión errónea con probabilidad ε. Para $i = 0, 1$, sea A_i=*{el emisor envía i}* y B_i=*{el receptor decidió i}*.*



De acuerdo con el esquema de la figura, las probabilidades del tipo $P(A_i \cap B_i)$, supuesta independencia entre las acciones del emisor y del receptor, valen

$$\begin{array}{ll} P(A_0 \cap B_0) = (1-p)(1-\varepsilon), & P(A_0 \cap B_1) = (1-p)\varepsilon \\ P(A_1 \cap B_0) = p\varepsilon, & P(A_1 \cap B_1) = p(1-\varepsilon). \end{array}$$

Es interesante, en este contexto, conocer las probabilidades del tipo $P(A_i|B_j)$, con $i, j = 0, 1$. Es decir, la probabilidad de que habiendo el receptor interpretado la señal como j *la que realmente se transmitió fuera* i*. De particular interés son, obviamente, $P(A_0|B_0)$ y $P(A_1|B_1)$.*

Para obtener estas probabilidades podemos hacer uso del Teorema de Bayes,

$$P(A_i|B_j) == \frac{P(B_j|A_i)P(A_i)}{\sum_{i=1}^{n} P(B_j|A_i)P(A_i)}. \tag{1.5}$$

El denominador no es más que $P(B_j)$, que se calcula fácilmente a partir de las anteriores probabilidades. Así, para j=0

$$P(B_0) = P(B_0|A_0)P(A_0) + P(B_0|A_1)P(A_1) = (1-\varepsilon)(1-p) + \varepsilon p,$$

y para j=1

$$P(B_1) = P(B_1|A_0)P(A_0) + P(B_1|A_1)P(A_1) = \varepsilon(1-p) + (1-\varepsilon)p.$$

En la tabla se muestran las cuatro probabilidades que se derivan de (1.5).

	$\boldsymbol{B_0}$	$\boldsymbol{B_1}$
$\boldsymbol{A_0}$	$\dfrac{(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)+\varepsilon p}$	$\dfrac{\varepsilon(1-p)}{(1-\varepsilon)(1-p)+\varepsilon p}$
$\boldsymbol{A_1}$	$\dfrac{\varepsilon p}{\varepsilon(1-p)+(1-\varepsilon)p}$	$\dfrac{(1-\varepsilon)p}{\varepsilon(1-p)+(1-\varepsilon)p}$

1.2. Variable aleatoria

1.2.1. Definición

Nuestro interés al examinar el resultado de un experimento aleatorio no es tanto el espacio de probabilidad resultante, como la o las características numéricas asociadas, lo que supone cambiar nuestro objetivo de Ω a $\mathcal{R}$ o $\mathcal{R}^k$. Hay dos razones que justifican este cambio:

1. el espacio de probabilidad es un espacio abstracto, mientras que $\mathcal{R}$ o $\mathcal{R}^k$ son espacios bien conocidos en los que resulta mucho más cómodo trabajar,

2. fijar nuestra atención en las características numéricas asociadas a cada resultado implica un proceso de abstracción que, al extraer los rasgos esenciales del espacio muestral, permite construir un modelo probabilístico aplicable a todos los espacios muestrales que comparten dichos rasgos.

Puesto que se trata de características numéricas ligadas a un experimento aleatorio, son ellas mismas *cantidades aleatorias*. Esto supone que para su estudio y conocimiento no bastará con saber que valores toman, habrá que conocer además la probabilidad con que lo hacen. Todo ello exige trasladar la información desde el espacio de probabilidad a $\mathcal{R}$ o $\mathcal{R}^k$ y la única forma que conocemos de trasladar información de un espacio a otro es mediante una aplicación. En nuestro caso la aplicación habrá de trasladar el concepto de suceso, lo que exige una mínima infraestructura en el espacio receptor de la información semejante a la σ-álgebra que contiene a los sucesos. Como nos vamos a ocupar ahora del caso unidimensional, una sola característica numérica asociada a los puntos del espacio muestral, nuestro espacio imagen es $\mathcal{R}$. En $\mathcal{R}$, los intervalos son el lugar habitual de trabajo, por lo que más conveniente será exigir a esta infraestructura que los contenga. Existe en $\mathcal{R}$ la llamada σ-álgebra de Borel, β, que tiene la propiedad de ser la menor de las que contienen a los intervalos, lo que la hace la más adecuada para convertir a $\mathcal{R}$ en espacio probabilizable: $(\mathcal{R}, \beta)$. Estamos ahora en condiciones de definir la variable aleatoria.

Definición 1.7 (Variable aleatoria) *Consideremos los dos espacios probabilizables* $(\Omega, \mathcal{A})$ *y* $(\mathcal{R}, \beta)$. *Una* variable aleatoria *es un aplicación,* $X : \Omega \longrightarrow \mathcal{R}$, *que verifica*

$$X^{-1}(B) \in \mathcal{A}, \quad \forall B \in \beta. \tag{1.6}$$

Cuando hacemos intervenir una variable aleatoria en nuestro proceso es porque ya estamos en presencia de un espacio de probabilidad, $(\Omega, \mathcal{A}, P)$. La variable aleatoria traslada la información probabilística relevante de Ω a $\mathcal{R}$ mediante una *probabilidad inducida* que se conoce como *ley de probabilidad de* X o *distribución de probabilidad de* X.

El concepto de σ-álgebra inducida

Dada la definición de variable aleatoria, es muy sencillo comprobar el siguiente resultado.

Lema 1.1 *La familia de sucesos*

$$\sigma(X) = \{X^{-1}(B),\ \forall B \in \beta\} = \{X^{-1}(\beta)\},$$

es una σ*-álgebra, denominada* σ-álgebra inducida por X, *que verifica* $\sigma(X) \subset \mathcal{A}$.

A los efectos de conocer el comportamiento probabilísticos de una variable aleatoria X, tres funciones nos proporcionan la información necesaria:

- la *probabilidad inducida*, P_X,
- la función de distribución de probabilidad, F_X, y
- la *función de cuantía o probabilidad*, si la variable es discreta, o la *función de densidad de probabilidad*, si la variables es continua, en ambos casos denotada por f_X.

Sus definiciones y propiedades se describen a continuación. Se puede demostrar, aunque está fuera del objetivo y alcance de estas notas, la equivalencia entre las tres funciones,

$$P_X \Longleftrightarrow F_X \Longleftrightarrow f_X.$$

Es decir, el conocimiento de una cualquiera de ellas permite obtener las otras dos.

1.2.2. Probabilidad inducida

X induce sobre $(\mathcal{R}, \beta)$ una probabilidad, P_X, de la siguiente forma,

$$P_X(A) = P(X^{-1}(A)),\ \forall A \in \beta.$$

Es fácil comprobar que P_X es una probabilidad sobre la σ-álgebra de Borel, de manera que $(\mathcal{R}, \beta, P_X)$ es un espacio de probabilidad al que podemos aplicar todo cuanto se dijo en el capítulo anterior. Observemos que P_X hereda las características que tenía P, pero lo hace a través de X. ¿Qué quiere esto decir? Un ejemplo nos ayudará a comprender este matiz.

Ejemplo 1.6 *Sobre el espacio de probabilidad resultante de lanzar dos dados, definimos las variables aletorias,* X=suma de las caras *e* Y=valor absoluto de la diferencia de las caras*. Aun cuando el espacio de probabilidad sobre el que ambas están definidas es el mismo,* P_X *y* P_Y *son distintas porque viene inducidas por variables distintas. En efecto,*

$$P_X(\{0\}) = P(X^{-1}(\{0\}) = P(\emptyset) = 0,$$

sin embargo,

$$P_Y(\{0\}) = P(Y^{-1}(\{0\}) = P(\{1,1\},\{2,2\},\{3,3\},\{4,4\},\{5,5\},\{6,6\}) = \frac{1}{6}.$$

La distribución de probabilidad de X, P_X, nos proporciona cuanta información necesitamos para conocer el comportamiento probabilístico de X, pero se trata de un objeto matemático complejo de incómodo manejo, al que no es ajena su condición de función de conjunto. Esta es la razón por la que recurrimos a funciones de punto para describir la aleatoriedad de X.

1.2.3. Función de distribución de probabilidad

A partir de la probabilidad inducida podemos definir sobre $\mathcal{R}$ la siguiente función,

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}\{(-\infty, x]\}) = P(X \leq x),\ \forall x \in \mathcal{R}. \tag{1.7}$$

Así definida esta función tiene las siguientes propiedades:

PF1) No negatividad.- Consecuencia inmediata de su definición.

PF2) Monotonía.- De la monotonía de la probabilidad se deduce fácilmente que $F_X(x_1) \leq F_X(x_2)$ si $x_1 \leq x_2$.

PF3) Continuidad por la derecha.- Consideremos una sucesión decreciente de números reales $x_n \downarrow x$. La correspondiente sucesión de intervalos verifica $\cap_n(-\infty, x_n] = (-\infty, x]$, y por la continuidad desde arriba de la probabilidad respecto del paso al límite tendremos $\lim_{x_n \downarrow x} F_X(x_n) = F_X(x)$.

Observemos por otra parte que $(-\infty, x] = \{x\} \cup \lim_{n\to+\infty}(-\infty, x - \frac{1}{n}]$, lo que al tomar probabilidades conduce a

$$F_X(x) = P(X = x) + \lim_{n\to+\infty} F_X\left(x - \frac{1}{n}\right) = P(X = x) + F(x-), \tag{1.8}$$

A partir de 1.8 se sigue que $F_X(x)$ *es continua en* x sí y solo sí $P(X = x) = 0$.

PF4) Valores límites.- Si $x_n \uparrow +\infty$ o $x_n \downarrow -\infty$ entonces $(-\infty, x_n] \uparrow \mathcal{R}$ y $(-\infty, x_n] \downarrow \emptyset$ y por tanto

$$F(+\infty) = \lim_{x_n\uparrow+\infty} F(x_n) = 1, \quad F(-\infty) = \lim_{x_n\downarrow-\infty} F(x_n) = 0.$$

A la función F_X se la conoce como *función de distribución de probabilidad de* X (en adelante simplemente función de distribución). En ocasiones se la denomina función de distribución acumulada, porque tal y como ha sido definida nos informa de la probabilidad acumulada por la variable X hasta el punto x. Nos permite obtener probabilidades del tipo $P(a < X \leq b)$ a partir de la expresión

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

1.2.4. Función de cuantía o probabilidad: variable aleatoria discreta

Existe una segunda función de punto que permite describir el comportamiento de X, pero para introducirla hemos de referirnos primero a las características del *soporte* de X, entendiendo por tal un conjunto $D_X \in \beta$ que verifica, $P_X(D_X) = P(X \in D_X) = 1$.

Cuando D_X es numerable, P_X es discreta y decimos que X es una variable aleatoria *discreta*. Como ya vimos en un ejemplo del capítulo anterior, $P_X(\{x_i\}) = P(X = x_i) > 0,\ \forall x_i \in D_X$, y siendo además $P(X \in D_X) = 1$, se deduce $P(X = x) = 0,\ \forall x \in D_X^c$. En este caso es fácil comprobar que la F_X asociada viene dada por

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i). \tag{1.9}$$

De acuerdo con esto, si $x_{(i)}$ y $x_{(i+1)}$ son dos puntos consecutivos del soporte tendremos que $\forall x \in [x_{(i)}, x_{(i+1)}[$, $F_X(x) = F_X(x_{(i)})$. Como además $P_X(x) = 0,\ \forall x \in D_X^c$, la función será también continua. Por otra parte $P(X = x_i) > 0$, para $x_i \in D_X$, con lo que los únicos puntos de discontinuidad serán lo del soporte, discontinuidad de salto finito cuyo valor es $F_X(x_{(i)}) - F_X(x_{(i-1)}) = P(X = x_i)$. Se trata por tanto de una función escalonada, cuyos saltos se producen en los puntos de D_X.

A la variable aleatoria discreta podemos asociarle una nueva función puntual que nos será de gran utilidad. La definimos para cada $x \in \mathcal{R}$ mediante $f_X(x) = P_X(\{x\}) = P(X = x)$, lo que supone que

$$f_X(x) = \begin{cases} P(X = x), & si\ x \in D_X \\ 0, & en\ el\ resto. \end{cases}$$

Esta función es conocida como *función de cuantía* o de *probabilidad* de X y posee las dos propiedades siguientes:

Pfc1) Al tratarse de una probabilidad, $f_X(x) \geq 0,\ \forall x \in \mathcal{R}$,

Pfc2) Como $P(X \in D_X) = 1$,

$$\sum_{x_i \in D_X} f_X(x_i) = 1.$$

La relación entre f_X y F_X viene recogida en las dos expresiones que siguen, cuya obtención es evidente a partir de (1.8) y (1.9). La primera de ellas permite obtener F_X a partir de f_X,

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i).$$

La segunda proporciona f_X en función de F_X,

$$f_X(x) = F_X(x) - F_X(x-).$$

1.2.5. Algunos ejemplos de variables aleatorias discretas

Variable aleatoria Poisson

La distribución de *Poisson* de parámetro λ es una de las distribuciones de probabilidad discretas más conocida. Una variable con esta distribución se caracteriza porque su soporte es $D_X = \{0, 1, 2, 3, \ldots\}$ y su función de cuantía viene dada por

$$f_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & si\ x \in D_X \\ \\ 0, & en\ el\ resto, \end{cases}$$

que cumple las propiedades Pfc1) y Pfc2). La función de distribución tiene por expresión

$$F_X(x) = \sum_{n \leq x} \frac{e^{-\lambda}\lambda^n}{n!}.$$

Diremos que X es una *variable Poisson de parámetro* λ y lo denotaremos $X \sim Po(\lambda)$. Esta variable aparece ligada a experimentos en los que nos interesa la ocurrencia de un determinado suceso a lo largo de un intervalo finito de tiempo[3], verificándose las siguientes condiciones:

1. la probabilidad de que el suceso ocurra en un intervalo pequeño de tiempo es proporcional a la longitud del intervalo, siendo λ el factor de proporcionalidad,

2. la probabilidad de que el suceso ocurra en dos o más ocasiones en un intervalo pequeño de tiempo es prácticamente nula.

Fenómenos como el número de partículas que llegan a un contador Geiger procedentes de una fuente radiactiva, el número de llamadas que llegan a una centralita telefónica durante un intervalo de tiempo, las bombas caídas sobre la región de Londres durante la Segunda Guerra mundial y las bacterias que crecen en la superficie de un cultivo, entre otros, pueden ser descritos mediante una variable aleatoria Poisson.

[3]En un planteamiento más general, el intervalo finito de tiempo puede ser sustituido por un subconjunto acotado de $\mathcal{R}^k$

Variable aleatoria Binomial

Decimos que X es una *variable Binomial* de parámetros n y p ($X \sim B(n,p)$) si $D_X = \{0,1,2,\ldots,n\}$ y

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & si\ x \in D_X \\ \\ 0, & en\ el\ resto, \end{cases}$$

que se comprueba fácilmente que verifica Pfc1) y Pfc2).

Cuando llevamos a cabo un experimento aleatorio cuyos rasgos esenciales son:

1. se llevan a cabo n repeticiones independientes de una misma prueba en las mismas condiciones,
2. en cada repetición observamos la ocurrencia (*éxito*) o no (*fracaso*) de un mismo suceso, A, y
3. la probabilidad de éxito es la misma en cada repetición, $P(A) = p$,

la variable que describe el número de éxitos alcanzado en las n repeticiones, es una Binomial de parámetros n y p.

Fenómenos aleatorios aparentemente tan diferentes como el número de hijos varones de un matrimonio con n hijos o el número de caras obtenidas al lanzar n veces una moneda correcta, son bien descritos mediante un variable Binomial. Este hecho, o el análogo que señalábamos en el ejemplo anterior, ponen de manifiesto el papel de modelo aleatorio que juega una variable aleatoria, al que aludíamos en la introducción. Esta es la razón por la que en muchas ocasiones se habla del *modelo Binomial* o del *modelo Poisson*.

Hagamos por último hincapié en un caso particular de variable aleatoria Binomial. Cuando $n = 1$ la variable $X \sim B(1,p)$ recibe el nombre de *variable Bernoulli* y se trata de una variable que solo toma los valores 0 y 1 con probabilidad distinta de cero. Es por tanto una variable dicotómica asociada a experimentos aleatorios en los que, realizada una sola prueba, nos interesamos en la ocurrencia de un suceso o su complementario. Este tipo de experimentos reciben el nombre de *pruebas Bernoulli*.

La distribución de Poisson como límite de la Binomial.- Consideremos la sucesión de variables aleatorias $X_n \sim B(n,p_n)$ en la que a medida que n aumenta, p_n disminuye de forma tal que $np_n \approx \lambda$. Más concretamente, $np_n \to \lambda$. Tendremos para la función de cuantía,

$$f_{X_n}(x) = \binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{n!}{x!(n-x)!} p_n^x (1-p_n)^{n-x},$$

y para n suficientemente grande,

$$\begin{aligned} f_{X_n}(x) &\approx \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!}\frac{n(n-1)\cdots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Al pasar al límite,

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} \to 1, \quad \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \quad \left(1-\frac{\lambda}{n}\right)^{-x} \to 1,$$

y tendremos

$$\lim_{n\to+\infty} f_{X_n}(x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

La utilidad de este resultado reside en permitir la aproximación de la función de cuantía de una $B(n,p)$ mediante la función de cuantía de una $Po(\lambda = np)$ cuando n es grande y p pequeño.

Variable aleatoria Hipergeométrica. Relación con el modelo Binomial

Si tenemos una urna con N bolas, de las cuales r son rojas y el resto, $N-r$, son blancas y extraemos n de ellas *con reemplazamiento*, el número X de bolas rojas extraídas será una $B(n,p)$ con $p = r/N$.

¿Qué ocurre si llevamos a cabo las extracciones *sin reemplazamiento*? La variable X sigue ahora una *distribución Hipergeométrica* ($X \sim H(n,N,r)$) con soporte $D_X = \{0,1,2,\ldots,\text{mín}(n,r)\}$ y cuya función de cuantía se obtiene fácilmente a partir de la fórmula de Laplace

$$f_X(x) = \begin{cases} \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, & si\ x \in D_X \\ \\ 0, & en\ el\ resto, \end{cases}$$

que cumple de inmediato la condición Pfc1). Para comprobar Pfc2) debemos hacemos uso de una conocida propiedad de los números combinatorios,

$$\sum_{i=0}^{n} \binom{a}{i}\binom{b}{n-i} = \binom{a+b}{n}.$$

La diferencia entre los modelos Binomial e Hipergeométrico estriba en el tipo de extracción. Cuando ésta se lleva a cabo con reemplazamiento las sucesivas extracciones son independientes y la probabilidad de *éxito* se mantiene constante e igual a r/N, el modelo es Binomial. No ocurre así si las extracciones son sin reemplazamiento. No obstante, si n es muy pequeño respecto a N y r, la composición de la urna variará poco de extracción a extracción y existirá lo que podríamos denominar una *quasi-independencia* y la distribución Hipergeométrica deberá comportarse como una Binomial. En efecto,

$$\begin{aligned} f_X(x) &= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \\ &= \frac{r!}{x!(r-x)!} \times \frac{(N-r)!}{(n-x)!(N-r-n+x)!} \times \frac{n!(N-n)!}{N!} \\ &= \binom{n}{x}\frac{r}{N} \times \frac{r-1}{N-1} \times \cdots \times \frac{r-x+1}{N-x+1} \times \frac{N-r}{N-x} \times \frac{N-r-1}{N-x-1} \times \\ &\quad \cdots \times \frac{N-r-n+x+1}{N-n+1} \\ &\approx \binom{n}{x} p^x (1-p)^{n-x}, \end{aligned}$$

con $p = r/N$.

Variable aleatoria Binomial Negativa

Consideremos n pruebas Bernoulli independientes con la misma probabilidad de éxito, p, en cada una de ellas. Nos interesamos en la ocurrencia del r-ésimo éxito. La variable que describe el mínimo número de pruebas adicionales necesarias para alcanzar los r éxitos, es una *variable aleatoria Binomial Negativa*, $X \sim BN(r,p)$, con soporte numerable $D_X = \{0,1,2,\ldots\}$ y con función de cuantía

$$f_X(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x, & si\ x \geq 0 \\ \\ 0, & en\ el\ resto. \end{cases}$$

El nombre de Binomial negativa se justifica a partir de la expresión alternativa que admite la función de cuantía,

$$f_X(x) = \begin{cases} \binom{-r}{x} p^r (-(1-p))^x, & si\ x \geq 0 \\ \\ 0, & en\ el\ resto, \end{cases}$$

obtenida al tener en cuenta que

$$\begin{aligned} \binom{-r}{x} &= \frac{(-r)(-r-1)\cdots(-r-x+1)}{x!} \\ &= \frac{(-1)^x r(r+1)\cdots(r+x-1)}{x!} \\ &= \frac{(-1)^x (r-1)! r(r+1)\cdots(r+x-1)}{(r-1)!x!} \\ &= (-1)^x \binom{r+x-1}{x}. \end{aligned}$$

La condición Pfc1) se cumple de inmediato, en cuanto a Pfc2) recordemos el desarrollo en serie de potencias de la función $f(x) = (1-x)^{-n}$,

$$\frac{1}{(1-x)^n} = \sum_{i \geq 0} \binom{n+i-1}{i} x^i, \quad |x| < 1.$$

En nuestro caso

$$f_X(x) = \sum_{x \geq 0} \binom{r+x-1}{x} p^r (1-p)^x = p^r \sum_{x \geq 0} \binom{r+x-1}{x} (1-p)^x = p^r \frac{1}{(1-(1-p))^r} = 1.$$

Un caso especial con nombre propio es el de $r = 1$. La variable aleatoria $X \sim BN(1,p)$ recibe el nombre de *variable aleatoria Geométrica* y su función de cuantía se reduce a

$$f_X(x) = \begin{cases} p(1-p)^x, & si\ x \geq 0 \\ \\ 0, & en\ el\ resto. \end{cases}$$

1.2.6. Función de densidad de probabilidad: variable aleatoria continua.

Una variable aleatoria X es *continua* si su función de distribución, F_X, es absolutamente continua. Lo que a efectos prácticos ello supone que existe una función $f_X(x)$, conocida como *función de densidad de probabilidad (fdp)* de X, tal que

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt.$$

En particular, dado cualquier intervalo $]a, b]$, se verifica

$$P(X \in]a, b]) = P(a < X \leq b) = \int_a^b f(x)\, dx.$$

Se derivan para f_X dos interesantes propiedades que la caracterizan:

Pfdp1) $f_X(x)$ es no negativa, y

Pfdp2) como $P(X \in \mathcal{R}) = 1$,

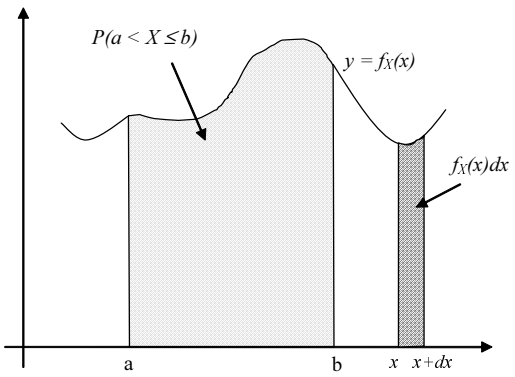
$$\int_{-\infty}^{+\infty} f(x)\, dx = 1.$$

Por otra parte, si x es un punto de continuidad de f_X, por una propiedad de la integral de Riemann se tiene

$$f_X(x) = F_X^{'}(x).$$

Significado físico de la fdp

La continuidad de F_X implica, recordemos (1.8), que en las variables aleatorias continuas $P(X = x) = 0,\ \forall x \in \mathcal{R}$. Este es un resultado que siempre sorprende y para cuya comprensión es conveniente interpretar físicamente la función de densidad de probabilidad.



La fdp es sencillamente eso, una densidad lineal de probabilidad que nos indica la cantidad de probabilidad por elemento infinitesimal de longitud. Es decir, $f_X(x)\, dx \approx P(X \in]x, x + dx])$. Ello explica que, para elementos con longitud cero, sea nula la correspondiente probabilidad. En este contexto, la probabilidad obtenida a través de la integral de Riemann pertinente se asimila a *un área*, la encerrada por f_X entre los límites de integración.

1.2.7. Algunos ejemplos de variables aleatorias continuas

Variable aleatoria Uniforme

La variable X diremos que tiene una *distribución uniforme en el intervalo* [a,b], $X \sim U(a,b)$, si su fdp es de la forma

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & si\ x \in [a,b] \\ \\ 0, & en\ el\ resto. \end{cases}$$

La función de distribución que obtendremos integrando f_X vale

$$F_X(x) = \begin{cases} 0, & si\ x \leq a \\ \\ \dfrac{x-a}{b-a}, & si\ a < x \leq b \\ \\ 1, & si\ x > b. \end{cases}$$

Surge esta variable cuando elegimos al azar un punto en el intervalo [a,b] y describimos con X su abscisa.

Variable aleatoria Normal

Diremos que X es una *variable aleatoria Normal* de parámetros μ y σ^2, $X \sim N(\mu, \sigma^2)$, si tiene por densidad,

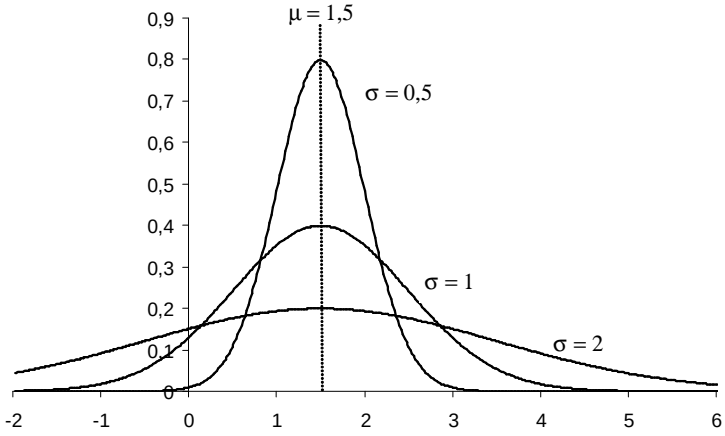
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty. \tag{1.10}$$

En tanto que densidad, f_X debe satisfacer las propiedades Pfdp1) y Pfdp2). La primera se deriva de inmediato de (1.10). Para comprobar la segunda,

$$\begin{aligned} I^2 &= \left[\int_{-\infty}^{+\infty} f_X(x)dx\right]^2 \\ &= \int_{-\infty}^{+\infty} f_X(x)dx \cdot \int_{-\infty}^{+\infty} f_X(y)dy \\ &= \frac{1}{2\pi} \cdot \frac{1}{\sigma} \int_{-\infty}^{+\infty} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx \cdot \frac{1}{\sigma} \int_{-\infty}^{+\infty} e^{-\dfrac{(y-\mu)^2}{2\sigma^2}} dy \qquad (1.11) \\ &= \frac{1}{2\pi} \cdot \frac{1}{\sigma} \int_{-\infty}^{+\infty} e^{-\dfrac{z^2}{2}} \sigma dz \cdot \frac{1}{\sigma} \int_{-\infty}^{+\infty} e^{-\dfrac{v^2}{2}} \sigma dv \qquad (1.12) \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\dfrac{z^2+v^2}{2}} dzdv \qquad (1.13) \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left[\int_0^{+\infty} e^{-\dfrac{r^2}{2}} rdr\right] d\theta = 1, \qquad (1.14) \end{aligned}$$

donde el paso de (1.11) a (1.12) se lleva a cabo mediante los cambios $z = (x-\mu)/\sigma$ y $v = (v-\mu)/\sigma$, y el de (1.13) a (1.14) mediante el cambio a polares $z = r\sin\theta$ y $v = r\cos\theta$.

La gráfica de f_X tiene forma de campana y es conocida como la campana de Gauss por ser Gauss quien la introdujo cuando estudiaba los errores en el cálculo de las órbitas de los planetas. En honor suyo, la distribución Normal es también conocida como distribución Gaussiana. Del significado de los parámetros nos ocuparemos más adelante, pero de (1.10) deducimos que $\mu \in \mathcal{R}$ y $\sigma > 0$. Además, el eje de simetría de f_X es la recta $x = \mu$ y el vértice de la campana (máximo de f_x) está en el punto de coordenadas $(\mu, 1/\sigma\sqrt{2\pi})$.



La figura ilustra el papel que los parámetros juegan en la forma y posición de la gráfica de la función de densidad de una $N(\mu, \sigma^2)$. A medida que σ disminuye se produce un mayor apuntamiento en la campana porque el máximo aumenta y porque, recordemos, el área encerrada bajo la curva es siempre la unidad.

Una característica de la densidad de la Normal es que carece de primitiva, por lo que su función de distribución no tiene expresión explícita y sus valores están tabulados o se calculan por integración numérica. Esto representa un serio inconveniente si recordamos que $P(a < X \leq b) = F_X(b) - F_X(a)$, puesto que nos obliga a disponer de una tabla distinta para cada par de valores de los parámetros μ y σ.

En realidad ello no es necesario, además de que sería imposible dada la variabilidad de ambos parámetros, porque podemos recurrir a la que se conoce como variable aleatoria *Normal tipificada*, $Z \sim N(0,1)$, cuya densidad es

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < +\infty.$$

En efecto, si para $X \sim N(\mu, \sigma^2)$ queremos calcular

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

efectuando el cambio $z = (t-\mu)/\sigma$ tendremos

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{z^2}{2}} dz = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

donde $\Phi(z)$ es la función de distribución de la $N(0,1)$.

Hay que señalar que el mayor interés de la distribución Normal estriba en el hecho de servir de modelo probabilístico para una gran parte de los fenómenos aleatorios que surgen en el campo de las ciencias experimentales y naturales.

El lema que sigue nos asegura que cualquier transformación lineal de un variable Normal es otra Normal.

Lema 1.2 *Sea $X \sim N(\mu, \sigma^2)$ y definamos $Y = aX + b$, entonces $Y \sim N(a\mu + b, a^2\sigma^2)$.*

Demostración.- Supongamos $a > 0$, la función de distribución de Y viene dada por

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

Su función de densidad es

$$f_Y(y) = F'_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2\right\}.$$

Si $a < 0$ entonces

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),$$

y la densidad será

$$f_Y(y) = F'_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\sigma}\right)^2\right\}.$$

En ambos casos se deduce que $Y \sim N(a\mu + b, a^2\sigma^2)$.

Variable aleatoria Exponencial

Diremos que la variable aleatoria X tiene una *distribución Exponencial de parámetro* λ, $X \sim Exp(\lambda)$, si su función de densidad es de la forma

$$f_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \lambda e^{-\lambda x}, & si\ x > 0,\ \lambda > 0. \end{cases}$$

La función de distribución de X vendrá dada por

$$F_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}, & si\ x > 0. \end{cases}$$

La distribución exponencial surge en problemas relacionados con tiempos de espera y está relacionada con la distribución de Poisson de igual parámetro. En efecto, si consideramos un proceso de desintegración radiactiva con λ desintegraciones por unidad de tiempo, el número de desintegraciones que se producen en el intervalo [0,t] es $N_t \sim Po(\lambda t)$, y el tiempo que transcurre ente dos desintegraciones consecutivas es $X \sim Exp(\lambda)$.

La falta de memoria de la variable aleatoria Exponencial.- La variable aleatoria Exponencial tiene una curiosa e interesante propiedad conocida como *falta de memoria*. Consiste en la siguiente igualdad,

$$P(X > x + t | X > t) = P(X > x), \quad \forall x, t \geq 0.$$

En efecto,

$$\begin{aligned} &P(X > x + t | X > t) = \\ &\quad = \frac{P\left(\{X > x + t\} \cap \{X > t\}\right)}{P(X > t)} = \frac{P(X > x + t)}{P(X > t)} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P(X > x). \end{aligned} \tag{1.15}$$

Variable aleatoria Gamma

Diremos que la variable aleatoria X tiene una *distribución Gamma de parámetros α y β*, $X \sim Ga(\alpha, \beta)$, si su función de densidad es de la forma

$$f_X(x) = \begin{cases} 0, & si\ x \leq 0 \\ \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & si\ x > 0,\ \alpha > 0,\ \beta > 0, \end{cases}$$

donde $\Gamma(\alpha)$ es el valor de la función Gamma en α, es decir

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \quad \alpha > 0.$$

Para comprobar que Pfdp2) se satisface, la Pfdp1) es de comprobación inmediata, bastará hacer el cambio $y = x/\beta$ en la correspondiente integral

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} \beta^\alpha dy = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

Los valores de la función de distribución $F_X(x)$ aparecen tabulados, con tablas para las diferentes combinaciones de los parámetros α y β.

Obsérvese que la distribución Exponencial de parámetro λ es un caso particular de la Gamma. En concreto $Exp(\lambda) = Gamma(1, 1/\lambda)$.

Observación 1.1 *Nos será de utilidad más tarde recordar alguna característica adicional de la función Gamma. En particular la obtención de sus valores cuando $\alpha = n$ o $\alpha = n + \frac{1}{2}$, n natural. Es fácil comprobar, mediante sucesivas integraciones por partes, que*

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) = (\alpha - 1)(\alpha - 2)\Gamma(\alpha - 2),$$

lo que para $\alpha = n$ da lugar a

$$\Gamma(n) = (n - 1)(n - 2) \ldots 2\Gamma(1).$$

Pero

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1 \quad y \quad \Gamma(n) = (n - 1)!.$$

Para el caso en que $\alpha = n + \frac{1}{2}$ *deberemos calcular* $\Gamma(\frac{1}{2})$*,*

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-x} x^{-1/2} dx = \left[y = \frac{t^2}{2}\right] = \sqrt{2} \int_0^\infty e^{-t^2/2} dt = \frac{\sqrt{2}\sqrt{2\pi}}{2} = \sqrt{\pi}. \tag{1.16}$$

La última integral en (1.16), dividida por $\sqrt{2\pi}$*, es la mitad del área que cubre la* fdp *de la* $N(0, 1)$*.*

1.3. Vector aleatorio

Cuando estudiamos simultáneamente k características numéricas ligadas al resultado del experimento, por ejemplo la *altura* y el *peso* de las personas, nos movemos en $\mathcal{R}^k$ como espacio imágen de nuestra aplicación. La σ-álgebra de sucesos con la que dotamos a $\mathcal{R}^k$ para hacerlo probabilizable es la correspondiente σ-álgebra de Borel β^k, que tiene la propiedad de ser la menor que contiene a los *rectángulos* $(a,b] = \prod_{i=1}^{k}(a_i, b_i]$, con $a = (a_1, \ldots, a_k)$, $b = (b_1, \ldots, b_k)$ y $-\infty \leq a_i \leq b_i < +\infty$. De entre ellos merecen especial mención aquellos que tiene el extremo inferior en $-\infty$, $S_x = \prod_{i=1}^{k}(-\infty, x_i]$, a los que denominaremos *región suroeste de* x, por que sus puntos están situados al suroeste de x.

Definición 1.8 (Vector aleatorio) *Un* vector aleatorio*,* $X = (X_1, X_2, \ldots, X_k)$*, es una aplicación de* Ω *en* $\mathcal{R}^k$*, que verifica*

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

La presencia de una probabilidad sobre el espacio $(\Omega, \mathcal{A})$ permite al vector inducir una probabilidad sobre $(\mathcal{R}^k, \beta^k)$.

1.3.1. Probabilidad inducida

X induce sobre $(\mathcal{R}^k, \beta^k)$ una probabilidad, P_X, de la siguiente forma,

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

Es sencillo comprobar que verifica los tres axiomas que definen una probabilidad, por lo que la terna $(\mathcal{R}^k, \beta^k, P_X)$ constituye un espacio de probabilidad con las características de $(\Omega, \mathcal{A}, P)$ heredadas a través de X.

1.3.2. Funciones de distribución conjunta y marginales

La función de distribución asociada a P_X se define para cada punto $x = (x_1, \ldots, x_k)$ de $\mathcal{R}^k$ mediante

$$F_X(x) = F_X(x_1, \ldots, x_k) = P_X(S_x) = P(X \in S_x) = P\left(\bigcap_{i=1}^{k}\{X_i \leq x_i\}\right). \tag{1.17}$$

De la definición se derivan las siguientes propiedades:

PFC1) No negatividad.- Consecuencia inmediata de ser una probabilidad.

PFC2) Monotonía en cada componente.- Si $x \leq y$, es decir, $x_i \leq y_i, \ i = 1, \ldots, k$, $S_x \subset S_y$ y $F_X(x) \leq F_X(y)$.

PFC3) Continuidad conjunta por la derecha.- Si $x^{(n)} \downarrow x$, entonces $F_X(x^{(n)}) \downarrow F_X(x)$,

PFC4) Valores límites.- Al tender a $\pm\infty$ las componentes del punto, se tiene

$$\lim_{\forall x_i \uparrow +\infty} F_X(x_1, \ldots, x_k) = 1,$$

o bien,

$$\lim_{\exists x_i \downarrow -\infty} F(x_1, \ldots, x_k) = 0.$$

que no son más que la versión multidimensional de las propiedades ya conocidas para el caso unidimensional. Existe ahora una quinta propiedad sin la cual no sería posible recorrer el camino inverso, obtener P_X a partir de F_X, y establecer la deseada y conveniente equivalencia entre ambos conceptos.

PFC5) Supongamos que $k = 2$ y consideremos el rectángulo $(a, b] = (a_1, b_1] \times (a_2, b_2]$ tal como lo muestra la figura. Indudablemente $P_X(]a, b]) \geq 0$, pero teniendo en cuenta (1.17) podemos escribir,

$$P_X(]a, b]) = F_X(b_1, b_2) - F_X(b_1, a_2) - F_X(a_1, b_2) + F_X(a_1, a_2) \geq 0.$$

Funciones de distribución marginales

Si el vector es aleatorio cabe pensar que sus componentes también lo serán. En efecto, la siguiente proposición, de sencilla demostración, establece una primera relación entre el vector y sus componentes.

Proposición 1.1 $X = (X_1, \ldots, X_k)$ *es un vector aleatorio* si y solo si *cada una de sus componentes es una variable aleatoria.*

Si las componentes del vector son variables aleatorias tendrán asociadas sus correspondientes probabilidades inducidas y funciones de distribución. La nomenclatura hasta ahora utilizada necesita ser adaptada, lo que haremos añadiendo los adjetivos *conjunta* y *marginal*, respectivamente. Puesto que P_X y F_X describen el comportamiento conjunto de las componentes de X, nos referiremos a ellas como *distribución conjunta* y *función de distribución conjunta* del vector X, respectivamente. Cuando, en el mismo contexto, necesitemos referirnos a la distribución de alguna componente lo haremos aludiendo a la *distribución marginal* o a la *función de distribución marginal* de X_i.

La pregunta que surge de inmediato es, *¿qué relación existe entre la distribución conjunta y las marginales?* Estamos en condiciones de dar respuesta en una dirección: cómo obtener la distribución marginal de cada componente a partir de la conjunta. Para ello, basta tener en cuenta que

$$\lim_{x_j \overset{j \neq i}{\to} \infty} \bigcap_{j=1}^{k} \{X_j \leq x_j\} = \{X_i \leq x_i\},$$

y al tomar probabilidades obtendremos

$$F_{X_i}(x_i) = \lim_{x_j \overset{j \neq i}{\to} \infty} F_X(x_1, \ldots, x_k). \tag{1.18}$$

El concepto de *marginalidad* puede aplicarse a cualquier subvector del vector original. Así, para $l \leq k$, si $X^l = (X_{i_1}, \ldots, X_{i_l})$ es un subvector de X, podemos hablar de la *distribución conjunta marginal de* X^l, para cuya obtención a partir de la conjunta procederemos de forma análoga a como acabamos de hacer para una componente. Si en la relación

$$\{X \in S_x\} = \bigcap_{i=1}^{k} \{X_i \leq x_i\},$$

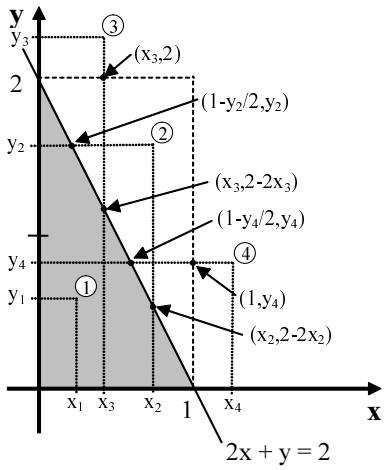
fijamos $x_{i_1}, \ldots, x_{i_l}$ y hacemos tender a ∞ el resto de las componentes, obtendremos

$$\{X^l \in S_{x^l}\} = \lim_{x_i \overset{i \neq i_1, \ldots, i_l}{\longrightarrow} \infty} \{X \in S_x\}.$$

Relación que nos permite obtener la función de distribución marginal conjunta de $X^l = (X_{i_1}, \ldots, X_{i_l})$ sin más que tomar probabilidades,

$$F_{X^l}(x_{i_1}, \ldots, x_{i_l}) = \lim_{x_i \overset{i \neq i_1, \ldots, i_l}{\longrightarrow} \infty} F_X(x_1, \ldots, x_k).$$

Ejemplo 1.7 *Elegimos un punto al azar sobre el triángulo T de vértices* (0,0), (1,0), (0,2).



Para encontrar la función de distribución conjunta del vector de sus componentes, (X, Y), observemos la figura y las distintas posiciones del punto. Como la masa de probabilidad está uniformemente repartida sobre el triángulo puesto que la elección del punto es al azar, tendremos que

$$P((X, Y) \in A) = \frac{\|A \cap T\|}{\|T\|},$$

donde $\|B\|$ es el área de B. Aplicado a la función de distribución dará lugar a

$$F_{XY}(x, y) = P((x, y) \in S_{xy}) = \|S_{xy} \cap T\|, \quad (1.19)$$

puesto que el área del triángulo vale 1.

Aplicando (1.19) obtenemos

$$F_{XY}(x, y) = \begin{cases} 0, & si\ x \leq 0\ o\ y \leq 0; \\ xy, & si\ (x, y)\ es\ del\ tipo\ \textbf{1}\ ; \\ xy - (x + y/2 - 1)^2, & si\ (x, y)\ es\ del\ tipo\ \textbf{2}\ ; \\ 2x - x^2, & si\ (x, y)\ es\ del\ tipo\ \textbf{3}\ ; \\ y - y^2/4, & si\ (x, y)\ es\ del\ tipo\ \textbf{4}\ ; \\ 1, & si\ x \geq 1\ e\ y \geq 2; \end{cases}$$

*Observemos que las expresiones de $F_{XY}(x, y)$ correspondientes a puntos del tipo **3** y **4** dependen solamente de x e y, respectivamente. Si recordamos la obtención de la función de distribución marginal veremos que se corresponden con $F_X(x)$ y $F_Y(y)$, respectivamente.*

1.3.3. Función de cuantía o probabilidad conjunta: vector aleatorio discreto

Si el soporte, D_X, del vector es numerable, lo que supone que también lo son los de cada una de sus componentes, diremos que X es un *vector aleatorio discreto*. Como en el caso unidimensional, una tercera función puede asociarse al vector y nos permite conocer su comportamiento aleatorio. Se trata de la *función de cuantía o probabilidad conjunta* y su valor, en cada punto de $\mathcal{R}^k$, viene dado por

$$f_X(x_1, \ldots, x_k) = \begin{cases} P(X_i = x_i,\ i = 1, \ldots, k), & si\ x = (x_1, \ldots, x_k) \in D_X \\ 0, & en\ el\ resto. \end{cases}$$

La función de cuantía conjunta posee las siguientes propiedades:

Pfcc1) Al tratarse de una probabilidad, $f_X(x) \geq 0,\ \forall x \in \mathcal{R}^k$,

Pfcc2) Como $P(X \in D_X) = 1$,

$$\sum_{x_1} \cdots \sum_{x_k} f_X(x_1, x_2, \ldots, x_k) = 1,\ (x_1, x_2, \ldots, x_k) \in D_X.$$

Entre F_X y f_X se establecen relaciones similares a las del caso unidimensional:

$$F_X(x) = \sum_{y \leq x,\ y \in D_X} f_X(y_1, y_2, \ldots, y_k),$$

y

$$f_X(x_1, x_2, \ldots, x_k) = F_X(x_1, x_2, \ldots, x_k) - F_X(x_1-, x_2-, \ldots, x_k-).$$

De ambas expresiones se deduce la equivalencia entre ambas funciones y también la de éstas con P_X,

$$P_X \Longleftrightarrow F_X \Longleftrightarrow f_X.$$

Funciones de cuantía marginales

Si el vector aleatorio X es discreto también lo serán cada una de sus componentes. Si por D_i designamos el soporte de $X_i,\ i = 1, \ldots, k$, se verifica,

$$\{X_i = x_i\} = \bigcup_{x_j \in D_j,\ j \neq i} \left(\bigcap_{j=1}^{k} \{X_j = x_j\} \right),$$

siendo disjuntos los elementos que intervienen en la unión. Al tomar probabilidades tendremos

$$f_{X_i}(x_i) = \sum_{x_j \in D_j,\ j \neq i} f_X(x_1, \ldots, x_k),$$

que permite obtener la función de cuantía marginal de la X_i a partir de la conjunta. La marginal conjunta de cualquier subvector aleatorio se obtendría de manera análoga, extendiendo la suma sobre todas las componentes del subvector complementario,

$$f_{X^l}(x_{i_1}, \ldots, x_{i_l}) = \sum_{x_j \in D_j,\ j \neq i_1, \ldots, i_l} f_X(x_1, \ldots, x_k).$$

Ejemplo 1.8 *Supongamos un experimento consistente en lanzar 4 veces una moneda correcta. Sea X el numero de caras en los 3 primeros lanzamientos y sea Y el número de cruces en los 3 últimos lanzamientos. Se trata de un vector discreto puesto que cada componente lo es. En concreto $D_X = \{0, 1, 2, 3\}$ y $D_Y = \{0, 1, 2, 3\}$.*

La función de cuantía conjunta se recoge en la tabla siguiente, para cualquier otro valor no recogido en la tabla $f_{XY}(x, y) = 0$.

	X				
Y	**0**	**1**	**2**	**3**	$\boldsymbol{f_Y(y)}$
0	*0*	*0*	*1/16*	*1/16*	*2/16*
1	*0*	*2/16*	*3/16*	*1/16*	*6/16*
2	*1/16*	*3/16*	*2/16*	*0*	*6/16*
3	*1/16*	*1/16*	*0*	*0*	*2/16*
$\boldsymbol{f_X(x)}$	*2/16*	*6/16*	*6/16*	*2/16*	*1*

En los márgenes de la tabla parecen las funciones de cuantía marginales de X e Y, obtenidas al sumar a lo largo de la correspondiente fila o columna.

1.3.4. Algunos ejemplos de vectores aleatorios discretos

Vector aleatorio Multinomial

La versión k-dimensional de la distribución Binomial es el llamado vector aleatorio *Multinomial*. Surge este vector en el contexto de un experimento aleatorio en el que nos interesamos en la ocurrencia de alguno de los k sucesos, $A_1, A_2, \ldots, A_k$, que constituyen una partición de Ω. Si $P(A_i) = p_i$ y repetimos n veces el experimento de manera que las repeticiones son independientes, el vector $X = (X_1, \ldots, X_k)$, con $X_i =$*número de ocurrencias de* A_i, decimos que tiene una *distribución Multinomial*, $X \sim M(n; p_1, p_2, \ldots, p_k)$. La función de cuantía conjunta viene dada por,

$$f_X(n_1, \ldots, n_k) = \begin{cases} \dfrac{n!}{n_1! n_2! \ldots n_k!} \displaystyle\prod_{i=1}^{k} p_i^{n_i}, & si\ 0 \leq n_i \leq n,\ i = 1, \ldots, k,\ \sum n_i = n \\ \\ 0, & en\ el\ resto, \end{cases}$$

que verifica Pfcc1) y Pfcc2), porque es no negativa y al sumarla para todos los posibles $n_1, n_2, \ldots, n_k$ obtenemos el desarrollo del polinomio $(p_1 + p_2 + \ldots + p_k)^n$, de suma 1 porque los A_i constituían una partición de Ω.

Para obtener la marginal de X_i observemos que

$$\frac{n!}{n_1! n_2! \ldots n_k!} \prod_{i=1}^{k} p_i^{n_i} = \binom{n}{n_i} p_i^{n_i} \frac{(n - n_i)!}{n_1! \ldots n_{i-1}! n_{i+1}! \ldots n_k!} \prod_{j \neq i} p_j^{n_j},$$

y al sumar para el resto de componentes,

$$\begin{aligned} f_{X_i}(n_i) &= \binom{n}{n_i} p_i^{n_i} \sum \frac{(n - n_i)!}{n_1! \ldots n_{i-1}! n_{i+1}! \ldots n_k!} \prod_{j \neq i} p_j^{n_j}, \\ &= \binom{n}{n_i} p_i^{n_i} (p_1 + \ldots + p_{i-1} + p_{i+1} + \ldots + p_k)^{n - n_i} \\ &= \binom{n}{n_i} p_i^{n_i} (1 - p_i)^{n - n_i}, \end{aligned}$$

llegamos a la conclusión que $X_i \sim B(n, p_i)$, como era de esperar, pues al fijar X_i sólo nos interesamos por la ocurrencia de A_i y el experimento que llevamos a cabo puede ser descrito mediante un modelo Binomial.

1.3.5. Función de densidad de probabilidad conjunta: vector aleatorio continuo

Decimos que X es un *vector aleatorio continuo* si existe entonces un función, f_X sobre $\mathcal{R}^k$, conocida como *función de densidad de probabilidad conjunta* de X, tal que

$$P(X \in (a, b]) = P(a_i < X_i \leq b_i,\ i = 1, \ldots, k) = \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} f_X(x_1, \ldots, x_k)\, dx_1 \ldots dx_k.$$

Al igual que ocurría en el caso unidimensional, esta función tiene dos propiedades que la caracterizan,

Pfdpc1) $f_X(x)$ es no negativa, y

Pfdcp2) como $P(X \in \mathcal{R}^k) = 1$,

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(x_1, \dots, x_k) dx_1 \dots dx_k = 1.$$

Como consecuencia de esta definición, entre la función de distribución conjunta y la de densidad de probabilidad conjunta se establecen las siguientes relaciones:

$$F_X(x_1, \dots, x_k) = P_X(S_x) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f(t_1, \dots, t_k)\, dt_1 \dots dt_k, \tag{1.20}$$

y si $x \in \mathcal{R}^k$ es un punto de continuidad de f_X,

$$f_X(x_1, \dots, x_k) = \frac{\partial^k F_X(x_1, \dots, x_k)}{\partial x_1, \dots, \partial x_k}. \tag{1.21}$$

Funciones de densidad marginales

Para obtener la densidad marginal de X_i a partir de la función de la densidad conjunta tengamos en cuenta (1.18) y (1.20) y que integración y paso al límite pueden permutarse por ser la densidad conjunta integrable,

$$\begin{aligned} F_{X_i}(x_i) &= \lim_{x_j \overset{j \neq i}{\to} \infty} F_X(x_1, \dots, x_k) \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{x_i} \dots \int_{-\infty}^{+\infty} f(t_1, \dots, t_i, \dots, t_k)\, dt_1 \dots dt_i \dots dt_k. \end{aligned}$$

Pero la derivada de F_{X_i} es una de las densidades de X_i y como las condiciones de la densidad conjunta permiten también intercambiar derivación e integración, tendremos finalmente

$$\begin{aligned} &f_{X_i}(x_i) = \\ &\quad = \int_{\mathcal{R}^{k-1}} f(t_1, \dots, x_i, \dots, t_k)\, dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_k. \end{aligned} \tag{1.22}$$

Para el caso de un subvector, X^l, la densidad conjunta se obtiene de forma análoga,

$$\begin{aligned} &f_{X^l}(x_{i_1}, \dots, x_{i_l}) = \\ &\quad = \int_{\mathcal{R}^{k-l}} f_X(t_1, \dots, t_k) \prod_{j \neq i_1, \dots, i_l} dt_j. \end{aligned}$$

Ejemplo 1.9 *La función de densidad conjunta del vector aleatorio bidimensional (X, Y) viene dada por*

$$f_{XY}(x, y) = \begin{cases} 8xy, & si\ 0 \leq y \leq x \leq 1 \\ 0, & en\ el\ resto. \end{cases}$$

Si queremos obtener las marginales de cada componente, tendremos para X

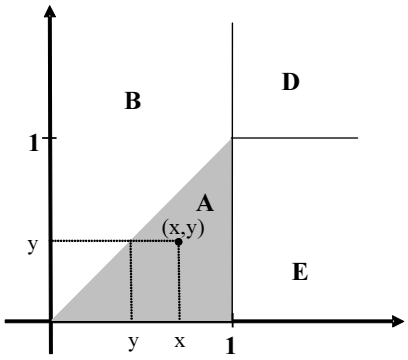
$$f_X(x) = \int_0^x f_{XY}(x, y) dy = \int_0^x 8xy dy = 4x^3, \quad 0 \leq x \leq 1,$$

y cero en el resto. Para Y,

$$f_Y(y) = \int_y^1 f_{XY}(x, y) dx = \int_y^1 8xy dx = 4y(1 - y^2), \quad 0 \leq y \leq 1,$$

y cero en el resto.

Obtengamos ahora la función de distribución conjunta, $F_{XY}(x,y)$. Observemos para ello el gráfico, la función de densidad es distinta de 0 en la región A por lo que $F_{XY}(x,y) = 0$ si $x \leq 0$ o $y \leq 0$.



Si $(x,y) \in A$,

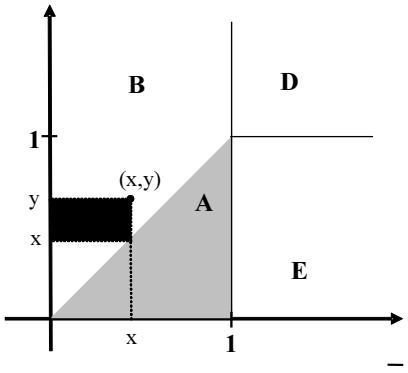
$$F_{XY}(x,y) = \int\int_{S_{xy} \cap A} f_{XY}(u,v) du dv,$$

pero $S_{xy} \cap A = \{(u,v); 0 \leq v \leq u \leq y\} \cup \{(u,v); y \leq u \leq x,\ 0 \leq v \leq y\}$, por tanto

$$F_{XY}(x,y) = \int_0^y \left[\int_0^u 8uvdv\right] du + \int_y^x \left[\int_0^y 8uvdv\right] du = y^2(2x^2 - y^2). \tag{1.23}$$

Si $(x,y) \in B$, como $f_{XY}(x,y) = 0$ en B, el rectángulo superior de la figura (en negro) no acumula probabilidad y por tanto

$$F_{XY}(x,y) = P\left(S_{xy} \cap A\right) = P\left(S_{xx} \cap A\right).$$



Así pues,

$$F_{XY}(x,y) = \int_0^x \left[\int_0^u 8uvdv\right] du = x^4. \tag{1.24}$$

Observemos que (1.24) puede obtenerse a partir de (1.23) haciendo $y = x$. En efecto, de acuerdo con (1.18), (1.24) no es más que $F_X(x)$, la función de distribución marginal de X.

Si $(x, y) \in E$, un razonamiento similar conduce a

$$F_{XY}(x, y) = y^2(2 - y^2),$$

que no es más que la función de distribución marginal de Y, que podríamos haber obtenido haciendo $x = 1$ en (1.23).

Por último, si $(x, y) \in D$, $F_{XY}(x, y) = 1$. Para su obtención basta hacer $x = 1$ e $y = 1$ en (1.23).

Resumiendo

$$F_{XY}(x,y) = \begin{cases} 0, & si\ x \leq 0\ o\ y \leq 0; \\ y^2(2x^2 - y^2), & si\ (x,y) \in A; \\ x^4, & si\ (x,y) \in B; \\ y^2(2-y^2), & si\ (x,y) \in E; \\ 1, & si\ (x,y) \in D. \end{cases}$$

Ejemplo 1.10 *La función de densidad conjunta del vector (X_1, X_2, X_3) es*

$$f(x_1,x_2,x_3) = \begin{cases} \dfrac{48x_1x_2x_3}{(1+x_1^2+x_2^2+x_3^2)^4}, & si\ x_1,x_2,x_3 \geq 0 \\ 0, & en\ el\ resto. \end{cases}$$

Obtengamos en primer lugar las densidades marginales de cada componente. Dada la simetría de la densidad conjunta bastará con obtener una cualquiera de ellas.

$$\begin{aligned} f_1(x_1) &= \int_0^\infty \int_0^\infty \frac{48x_1x_2x_3}{(1+x_1^2+x_2^2+x_3^2)^4} dx_2 dx_3 \\ &= 48x_1 \int_0^\infty x_2 \left[\int_0^\infty \frac{x_3}{(1+x_1^2+x_2^2+x_3^2)^4} dx_3 \right] dx_2 \\ &= \int_0^\infty \int_0^\infty \frac{8x_1x_2}{(1+x_1^2+x_2^2)^3} dx_2 \qquad (1.25) \\ &= \frac{2x_1}{(1+x_1^2)^2}. \end{aligned}$$

Luego

$$f_i(x_i) = \begin{cases} \dfrac{2x_i}{(1+x_i^2)^2}, & si\ x_i \geq 0 \\ 0, & en\ el\ resto. \end{cases} \qquad (1.26)$$

Por otra parte, en el transcurso de la obtención de $f_i(x_i)$ el integrando de (1.25) es la densidad marginal conjunta de (X_1, X_2), que por la simetría antes mencionada es la misma para cualquier pareja de componentes. Es decir, para $i, j = 1, 2, 3$

$$f_{ij}(x_i,x_j) = \begin{cases} \dfrac{8x_ix_j}{(1+x_i^2+x_j^2)^3}, & si\ x_i,x_j \geq 0 \\ 0, & en\ el\ resto. \end{cases}$$

Para obtener la función de distribución conjunta recordemos que

$$F(x_1,x_2,x_3) = P(X_i \leq x_i, i = 1,2,3) = P\left(\bigcap_{i=1}^3 \{X_i \leq x_i\} \right) = \int_0^{x_1} \int_0^{x_2} \int_0^{x_3} f(u,v,z) du dv dz.$$

pero en este caso será más sencillo recurrir a esta otra expresión,

$$F(x_1,x_2,x_3) = 1 - P\left(\left[\bigcap_{i=1}^{3}\{X_i \le x_i\}\right]^c\right) = 1 - P\left(\bigcup_{i=1}^{3} A_i\right),$$

con $A_i = \{X_i > x_i\}$. *Si aplicamos la fórmula de inclusión-exclusión (1.1),*

$$F(x_1,x_2,x_3) = 1 - \left[\sum_{i=1}^{3} P(A_i) - \sum_{1\le i<j\le 3} P(A_i\cap A_j) + P(A_1\cap A_2\cap A_3)\right]. \tag{1.27}$$

La obtención de las probabilidades que aparecen en (1.27) involucran a las densidades antes calculadas. Así, para $P(A_i)$

$$P(A_i) = P(X_i > x_i) = \int_{x_i}^{\infty} \frac{2u}{(1+u^2)^2}du = \frac{1}{1+x_i^2}.$$

Para $P(A_i\cap A_j)$,

$$\begin{aligned} P(A_i\cap A_j) &= P(X_i > x_i, X_j > x_j) \\ &= \int_{x_i}^{\infty}\int_{x_j}^{\infty} \frac{8uv}{(1+u^2+v^2)^3}dudv \\ &= \frac{1}{1+x_i^2+x_j^2}. \end{aligned}$$

Finalmente

$$P(A_1\cap A_2\cap A_3) = \int_{x_1}^{\infty}\int_{x_2}^{\infty}\int_{x_3}^{\infty} ds\frac{48uvz}{(1+u^2+v^2+z^2)^4}dudvdz = \frac{1}{1+x_1^2+x_2^2+x_3^2}.$$

Sustituyendo en (1.27) tendremos,

$$F(x_1,x_2,x_3) = 1 - \sum_{i=1}^{3}\frac{1}{1+x_i^2} + \sum_{1\le i<j\le 3}\frac{1}{1+x_i^2+x_j^2} - \frac{1}{1+x_1^2+x_2^2+x_3^2}.$$

1.3.6. Algunos ejemplos de vectores aleatorios continuos

Vector aleatorio Uniforme en el círculo unidad

Al elegir un punto al azar en, C_1, círculo unidad, podemos definir sobre el correspondiente espacio de probabilidad un vector aleatorio de las coordenadas del punto, (X,Y). La elección al azar implica una probabilidad uniforme sobre C_1, lo que se traduce en un densidad conjunta constante sobre todo el círculo, pero como por otra parte $\int\int_{C_1} f(x,y)\,dxdy = 1$, la densidad conjunta vendrá dada por

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{\pi}, & si\ (x,y)\in C_1 \\ 0, & en\ el\ resto. \end{cases}$$

Para obtener la densidad marginal de X,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy = \frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1-x^2},$$

por lo que

$$f_X(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x^2}, & si\ |x| \leq 1 \\ \\ 0, & en\ el\ resto, \end{cases}$$

La marginal de Y, por simetría, tiene la misma expresión.

Si en lugar de llevar a cabo la elección en C_1, elegimos el punto en el cuadrado unidad, $Q = [0,1] \times [0,1]$, la densidad conjunta es constante e igual a 1 en el cuadrado y las marginales son ambas $U(0,1)$.

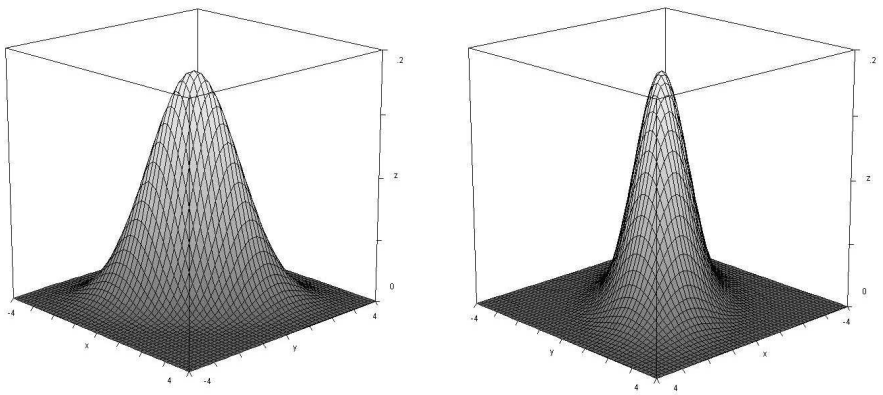
Vector aleatorio Normal bivariante

El vector aleatorio bidimensional (X,Y) tiene una distribución Normal bivariante de parámetros $\mu_X \in \mathcal{R}$, $\mu_y \in \mathcal{R}$, $\sigma_x > 0$, $\sigma_y > 0$ y ρ, $|\rho| < 1$, si su función de densidad conjunta es de la forma,

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{q(x,y)}{2}}, \quad (x,y) \in \mathcal{R}^2,$$

donde

$$q(x,y) = \frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}.$$



La figura nos muestras sendas gráficas de la normal bivariante con parámetros $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$ y $\rho = 0,5$. La gráfica está centrada en (μ_x, μ_y) (parámetros de posición) y su forma depende de σ_x, σ_y y ρ (parámetros de forma). Para ver el efecto de este último la gráfica de la derecha ha sido rotada $-90°$.

Para ver que $f_{XY}(x,y)$ es una densidad es inmediato comprobar que verifica la primera condición, en cuanto a la segunda, $\int_{\mathcal{R}^2} f_{XY}(x,y)dxdy = 1$, observemos que

$$\begin{aligned}(1-\rho^2)q(x,y) &= \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \\ &= \left[\left(\frac{x-\mu_x}{\sigma_x}\right) - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2, \qquad (1.28)\end{aligned}$$

pero el primer sumando de (1.28) puede escribirse

$$\begin{aligned}\left(\frac{x-\mu_x}{\sigma_x}\right) - \rho\left(\frac{y-\mu_y}{\sigma_y}\right) &= \left(\frac{x-\mu_x}{\sigma_x}\right) - \frac{\rho\sigma_x}{\sigma_x}\left(\frac{y-\mu_y}{\sigma_y}\right) \\ &= \frac{1}{\sigma_x}\left[x - \left(\mu_x + \rho\sigma_x\frac{y-\mu_y}{\sigma_y}\right)\right] \\ &= \frac{1}{\sigma_x}(x-b),\end{aligned}$$

con $b = \mu_x + \rho\sigma_x\frac{y-\mu_y}{\sigma_y}$. Sustituyendo en (1.28)

$$(1-\rho^2)q(x,y) = \left(\frac{x-b}{\sigma_x}\right)^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

y de aquí

$$\begin{aligned}&\int_{\mathcal{R}^2} f_{XY}(x,y)dxdy = \\ &\quad = \int_{-\infty}^{+\infty} \frac{1}{\sigma_y\sqrt{2\pi}} e^{\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)} \left[\int_{-\infty}^{+\infty} \frac{1}{\sigma_x\sqrt{2\pi(1-\rho^2)}} e^{\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x-b}{\sigma_x}\right)^2\right)} dx\right] dy = 1, \qquad (1.29)\end{aligned}$$

porque el integrando de la integral interior es la función de densidad de una $N(b, \sigma_x^2(1-\rho^2))$ e integra la unidad. La integral resultante vale también la unidad por tratarse de la densidad de una $N(\mu_y, \sigma_y^2)$, que es precisamente la densidad marginal de Y (basta recordar la expresión (1.22) que permite obtener la densidad marginal a partir de la conjunta). Por simetría $X \sim N(\mu_x, \sigma_x^2)$.

Esta distribución puede extenderse a n dimensiones, hablaremos entonces de *Normal multivariante*. La expresión de su densidad la daremos en el próximo capítulo y utilizaremos una notación matricial que la haga más sencilla y compacta.

1.4. Independencia de variables aleatorias

La independencia entre dos sucesos A y B supone que ninguno de ellos aporta información de interés acerca del otro. Pretendemos ahora trasladar el concepto a la relación entre variables aleatorias, pero siendo un concepto originalmente definido para sucesos, la traslación deberá hacerse por medio de sucesos ligados a las variables. Para ello necesitamos recurrir al concepto de σ-álgebra inducida por una variable aleatoria que definimos en la página 11.

Definición 1.9 (Variables aleatorias independientes) *Decimos que las variables* $X_i,\ i = 1,\ldots,k$ *son* independientes *si las* σ*-álgebras que inducen lo son.*

Si recordamos lo que significaba la independencia de familias de sucesos, la definición implica que al tomar un $A_i \in \sigma(X_i),\ i = 1,\ldots,k$,

$$P(A_{j_1} \cap \ldots \cap A_{j_n}) = \prod_{l=1}^{n} P(A_{j_l}),\quad \forall (j_1,\ldots,j_n) \subset (1,\ldots,k).$$

Teniendo en cuenta cómo han sido inducidas las $\sigma(X_i)$, admite una expresión alternativa en términos de las distintas variables,

$$P(A_{j_1} \cap \ldots \cap A_{j_n}) = P(X_{j_l} \in B_{j_l}, l = 1,\ldots,n) = \prod_{l=1}^{n} P(X_{j_l} \in B_{j_l}),\ B_{j_l} \in \beta, \tag{1.30}$$

donde $A_{j_1} = X^{-1}(B_{j_l})$.

Comprobar la independencia de variables aleatorias mediante (1.30) es prácticamente imposible. El teorema que sigue, cuya demostración omitimos, permite una caracterización más sencilla de la independencia.

Teorema 1.1 (Teorema de factorización) *Sea* $X = (X_1,\ldots,X_k)$ *un vector aleatorio cuyas funciones de distribución y densidad o cuantía conjuntas son, respectivamente,* $F_X(x_1,\ldots,x_k)$ *y* $f_X(x_1,\ldots,x_k)$*. Sean* $F_j(x_j)$ *y* $f_j(x_j)$*,* $j = 1,\ldots,k$*, las respectivas marginales. Las variables aleatorias* $X_1,\ldots,X_k$ *son independientes* sí y solo sí *se verifica alguna de las siguientes condiciones equivalentes:*

1. $F_X(x_1,\ldots,x_k) = \prod_{j=1}^{k} F_j(x_j),\ \forall (x_1,\ldots,x_k) \in \mathcal{R}^k$
2. $f_X(x_1,\ldots,x_k) = \prod_{j=1}^{k} f_j(x_j),\ \forall (x_1,\ldots,x_k) \in \mathcal{R}^k$

Observación 1.2 *Hemos visto anteriormente que a partir de la distribución conjunta del vector es posible conocer la distribución de cada una de sus componentes. El teorema de factorización implica que a partir de las marginales podemos reconstruir la distribución conjunta, si bien es cierto que no siempre, pues se exige la independencia de las variables. La recuperación en cualquier circunstancia requiere de la noción de distribución condicionada.*

Ejemplo 1.11 *En la sección 1.3.6 estudiábamos el vector aleatorio determinado por las coordenadas de un punto elegido al azar en el círculo unidad. La densidad conjunta venía dada por*

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{\pi}, & si\ (x,y) \in C_1 \\ \\ 0, & en\ el\ resto. \end{cases}$$

Por simetría, las marginales de X *e* Y *son idénticas y tienen la forma,*

$$f_X(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1-x^2}, & si\ |x| \leq 1 \\ \\ 0, & en\ el\ resto. \end{cases}$$

De inmediato se comprueba que $f_{XY}(x,y) \neq f_X(x) f_Y(y)$ *y ambas variables* no son independientes.

1.5. Distribuciones condicionadas

1.5.1. Caso discreto

Consideremos un vector aleatorio bidimensional (X, Y), con soportes para cada una de sus componentes D_x y D_y, respectivamente, y con función de cuantía conjunta $f_{XY}(x, y)$.

Definición 1.10 *La* función de cuantía condicionada *de Y dado $\{X = x\}$, $x \in D_x$, se define mediante,*

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{f_{XY}(x, y)}{f_X(x)}.$$

La función de distribución condicionada *de Y dado $\{X = x\}$, $x \in D_x$, se define mediante,*

$$F_{Y|X} = P(Y \leq y|X = x) = \frac{\sum_{v \leq y,\ v \in D_y} f_{XY}(x, v)}{f_X(x)} = \sum_{v \leq y,\ v \in D_y} f_{Y|X}(v|x).$$

La función $f_{Y|X}(y|x)$ es efectivamente una función de cuantía por cuanto cumple con las dos consabidas condiciones,

1. es no negativa por tratarse de una probabilidad condicionada, y
2. suma la unidad sobre D_y,

$$\sum_{y \in D_y} f_{Y|X}(y|x) = \frac{\sum_{y \in D_y} f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

El concepto de distribución condicional se extiende con facilidad al caso k-dimensional. Si $X = (X_1, \ldots, X_k)$ es un vector aleatorio k-dimensional y $X^l = (X_{i_1}, \ldots, X_{i_l})$, $l \leq k$ y $X^{k-l} = (X_{j_1}, \ldots, X_{j_{k-l}})$ son subvectores de dimensiones complementarias, con soportes D_{x^l} y $D_{x^{k-l}}$, respectivamente, la *función de cuantía condicionada* de X^l dado $X^{k-l} = (x_{j_1}, \ldots, x_{j_{k-l}})$, $(x_{j_1}, \ldots, x_{j_{k-l}}) \in D_{x^{k-l}}$, se define mediante,

$$f_{X^l|X^{k-l}}(x_{i_1}, \ldots, x_{i_l}|x_{j_1}, \ldots, x_{j_{k-l}}) = \frac{f_X(x_1, \ldots, x_k)}{f_{X^{k-l}}(x_{j_1}, \ldots, x_{j_{k-l}})},$$

donde el argumento del numerador, $(x_1, \ldots, x_k)$, está formado por las componentes $(x_{i_1}, \ldots, x_{i_l})$ y $(x_{j_1}, \ldots, x_{j_{k-l}})$ adecuadamente ordenadas.

Ejemplo 1.12 *Consideremos dos variables aleatorias independientes X e Y, con distribución de Poisson de parámetros μ y λ, respectivamente. Queremos encontrar al distribución de la variable condicionada $X|X + Y = r$.*

Recordemos que

$$f_{X|X+Y}(k|r) = P(X = k|X + Y = r) = \frac{P(X = k, Y = r - k)}{P(X + Y = r)} = \frac{f_{XY}(k, r - k)}{f_{X+Y}(r)}. \tag{1.31}$$

La distribución conjunta del vector (X, Y) es conocida por tratarse de variables independientes,

$$f_{XY}(k, r - k) = \frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{r-k}}{r - k!} e^{-\lambda}. \tag{1.32}$$

La distribución de la variable $X+Y$ se obtiene de la forma

$$\begin{aligned} f_{X+Y}(r) &= P\left(\bigcup_{k=0}^{r}\{X=k, Y=r-k\}\right) \\ &= \sum_{k=0}^{r} f_{XY}(k, r-k) \\ &= \sum_{k=0}^{r} \frac{\mu^k \lambda^{r-k}}{k!r-k!} e^{-(\mu+\lambda)} \\ &= \frac{e^{-(\mu+\lambda)}}{r!} \sum_{k=0}^{r} \frac{r!}{k!r-k!} \mu^k \lambda^{r-k} \\ &= \frac{(\mu+\lambda)^r}{r!} e^{-(\mu+\lambda)}. \end{aligned} \tag{1.33}$$

Lo que nos dice que $X+Y \sim Po(\mu+\lambda)$.

Sustituyendo (1.32) y (1.33) en (1.31),

$$\begin{aligned} f_{X|X+Y}(k|r) &= \frac{\frac{\mu^k}{k!} e^{-\mu} \frac{\lambda^{r-k}}{r-k!} e^{-\lambda}}{\frac{(\mu+\lambda)^r}{r!} e^{-(\mu+\lambda)}} \\ &= \frac{r!}{k!(r-k)!} \frac{\mu^k \lambda^{r-k}}{(\mu+\lambda)^r} \\ &= \binom{r}{k} \left(\frac{\mu}{\mu+\lambda}\right)^k \left(1-\frac{\mu}{\mu+\lambda}\right)^{r-k}, \end{aligned}$$

concluimos que $X|X+Y=r \sim B(r, \mu/(\mu+\lambda))$.

Distribuciones condicionadas en la Multinomial

Si el vector $X=(X_1,\ldots,X_k) \sim M(n; p_1,\ldots,p_k)$ sabemos que la marginal del subvector $X^l=(X_1,\ldots,X_l)$ es una $M(n; p_1,\ldots,p_l,(1-p^*))$, $p^*=p_1+\cdots+p_l$ (en definitiva la partición de sucesos que genera la multinomial queda reducida a $A_1,\ldots,A_l,A^*$, con $A^*=\left(\cup_{i=1}^{l} A_i\right)^c$). La distribución condicionada de $X^{k-l}=(X_{l+1},\ldots,X_k)$ dado $X^l=(n_1,\ldots,n_l)$, viene dada por

$$\begin{aligned} f_{X^{k-l}|X^l}(n_{l+1},\ldots,n_k|n_1,\ldots,n_l) &= \frac{\frac{n!}{n_1!n_2!\ldots n_k!}\prod_{i=1}^{k} p_i^{n_i}}{\frac{n!}{n_1!\ldots n_l!(n-n^*)!}(1-p^*)^{(n-n^*)}\prod_{i=1}^{l} p_i^{n_i}} \\ &= \frac{(n-n^*)!}{n_{l+1}!\ldots n_k!}\prod_{i=l+1}^{k}\left(\frac{p_i}{1-p^*}\right)^{n_i} \end{aligned}$$

con $n^*=n_1+\cdots+n_l$ y $\sum_{i=l+1}^{k} n_i = n-n^*$. Se trata, en definitiva, de una $M(n-n^*; \frac{p_{l+1}}{1-p^*},\ldots,\frac{p_k}{1-p^*})$.

1.5.2. Caso continuo

Si tratamos de trasladar al caso continuo el desarrollo anterior nos encontramos, de entrada, con una dificultad aparentemente insalvable. En efecto, si tenemos un vector bidimensional

(X, Y) en la expresión

$$P(Y \leq y | X = x) = \frac{P\left(\{Y \leq y\} \cap \{X = x\}\right)}{P(X = x)}$$

el denominador $P(X = x)$ es nulo. Puede parecer que el concepto de distribución condicionada carezca de sentido en el contexto de variables continuas.

Pensemos, no obstante, en la elección de un punto al azar en C_1, círculo unidad. Fijemos la abscisa del punto en $X = x$, $|x| < 1$, y consideremos cómo se distribuirá la ordenada Y sobre la correspondiente cuerda. Estamos hablando de la distribución condicionada de $Y|X = x$, que no sólo tiene sentido, si no que intuimos que será uniforme sobre la cuerda. ¿Cómo comprobar nuestra intuición? Si aceptamos como definición de función densidad condicionada la que hemos encontrado para el caso discreto,

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)},$$

y recordamos las expresiones de las densidades conjuntas y las marginales obtenidas en la sección 1.3.6 y el ejemplo 1.11, tendremos

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{1/\pi}{2\sqrt{1-x^2}/\pi}, & si\ |y| \leq \sqrt{1-x^2} \\ \\ 0, & en\ el\ resto, \end{cases}$$

que confirma nuestra intuición, $Y|X = x \sim U(-\sqrt{1-x^2}, +\sqrt{1-x^2})$. Parece lógico pensar que la densidad condicionada sea, efectivamente, la que hemos supuesto.

Una obtención rigurosa de las expresiones de $f_{Y|X}(y|x)$ y $F_{Y|X}(y|x)$ está fuera del alcance de esta introducción, pero una aproximación válida consiste en obtener $F_{Y|X}(y|x) = P(Y \leq y|X = x)$ como límite de $P(Y \leq y|x - \varepsilon < X \leq x + \varepsilon)$ cuando $\varepsilon \downarrow 0$ y siempre que $f_X(x) > 0$. Veámoslo.

$$\begin{aligned} F_{Y|X}(y|x) &= \lim_{\varepsilon \downarrow 0} P(Y \leq y | x - \varepsilon < X \leq x + \varepsilon) \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(Y \leq y, x - \varepsilon < X \leq x + \varepsilon)}{P(x - \varepsilon < X \leq x + \varepsilon)} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\int_{-\infty}^{y} \left[\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(u, v) du\right] dv}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(u) du}. \end{aligned}$$

Dividiendo numerador y denominador por 2ε, pasando al límite y teniendo en cuenta la relación (1.21) que liga a las funciones de densidad y de distribución en los puntos de continuidad de aquellas,

$$F_{Y|X}(y|x) = \frac{\int_{-\infty}^{y} f_{XY}(x, v) dv}{f_X(x)} = \int_{-\infty}^{y} \frac{f_{XY}(x, v)}{f_X(x)} dv, \quad f_X(x) > 0.$$

Al derivar en la expresión anterior respecto de v obtendremos una de las posibles densidades condicionadas,

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}, \quad f_X(x) > 0,$$

justamente la que hemos utilizado anteriormente.

Ambas expresiones se generalizan fácilmente para el caso de un vector X, k-dimensional, y subvectores X^l y X^{k-l} de dimensiones complementarias l y $k-l$, respectivamente.

$$F_{X^l|X^{k-l}}(x_{i_1},\ldots,x_{i_l}|x_{j_1},\ldots,x_{j_{k-l}}) = \frac{\int_{-\infty}^{x_{i_1}}\cdots\int_{-\infty}^{x_{i_l}} f_X(x_1,\ldots,x_k)dx_{i_1}\cdots dx_{i_l}}{f_{X^{k-l}}(x_{j_1},\ldots,x_{j_{k-l}})},$$

y

$$f_{X^l|X^{k-l}}(x_{i_1},\ldots,x_{i_l}|x_{j_1},\ldots,x_{j_{k-l}}) = \frac{f_X(x_1,\ldots,x_k)}{f_{X^{k-l}}(x_{j_1},\ldots,x_{j_{k-l}})},$$

con $f_{X^{k-l}}(x_{j_1},\ldots,x_{j_{k-l}}) > 0$, y donde el argumento de ambos numeradores, $(x_1,\ldots,x_k)$, está formado por las componentes $(x_{i_1},\ldots,x_{i_l})$ y $(x_{j_1},\ldots,x_{j_{k-l}})$ adecuadamente ordenadas.

Ejemplo 1.13 *Elegimos al azar X en [0,1] y a continuación Y, también al azar, en $[0,X^2]$. Es decir*

$$f_X(x) = \begin{cases} 1, & x\in[0,1]; \\ 0, & \textit{en el resto.} \end{cases} \quad \text{y} \quad f_{Y|X}(y|x) = \begin{cases} 1/x^2, & y\in[0,x^2]; \\ 0, & \textit{en el resto.} \end{cases}$$

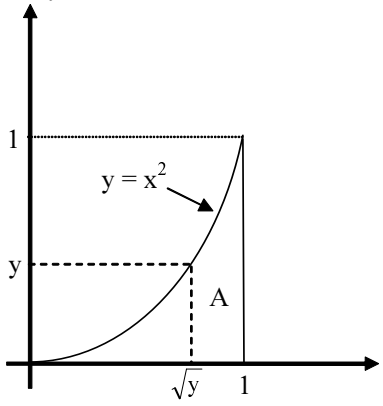
La densidad conjunta de (X,Y) vale

$$f_{XY}(x,y) = f_X(x)f_{Y|X}(y|x) = \begin{cases} \dfrac{1}{x^2}, & x\in[0,1],\ y\in[0,x^2]; \\ \\ 0, & \textit{en el resto.} \end{cases}$$

La densidad marginal de Y es

$$f_Y(y) = \int_{\sqrt{y}}^1 \frac{1}{x^2}dx = \frac{1}{\sqrt{y}} - 1, \quad y\in[0,1],$$

y vale 0 fuera del intervalo.



Cabe preguntarse si la elección de X e Y que hemos hecho se corresponde con la elección al azar de un punto en el recinto A de la figura, determinado por la parábola $y=x^2$ entre $x=0$ y $x=1$. La respuesta es negativa, puesto que la densidad conjunta vendría dada en este caso por

$$f^*_{XY}(x,y) = \begin{cases} \dfrac{1}{\textit{área de } A} = 3, & (x,y)\in A \\ \\ 0, & \textit{en el resto.} \end{cases}$$

y evidentemente, $f_{XY}(x,y) \neq f^*_{XY}(x,y)$.

Puede comprobarse que en este caso

$$f^*_X(x) = \begin{cases} 3x^2, & x\in[0,1]; \\ 0, & \textit{en el resto.} \end{cases} \quad \text{y} \quad f^*_{Y|X}(y|x) = \begin{cases} 1/x^2, & y\in[0,x^2]; \\ 0, & \textit{en el resto.} \end{cases}$$

Es decir, elegida la componente X la elección de Y continua siendo al azar en el intervalo $[0,X^2]$, pero a diferencia de cómo elegíamos X inicialmente, ahora ha de elegirse con la densidad $f^*_X(x)$.

Distribuciones condicionadas en la Normal bivariante

Si (X,Y) es un vector Normal bivariante,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}}}{\frac{1}{\sigma_x\sqrt{2\pi}}e^{\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)}} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y-\mu_y}{\sigma_y}\right)-\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right\}^2} \\ &= \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left\{y-\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right)\right\}^2}. \end{aligned}$$

Es decir, $Y|X=x \sim N\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma_y^2(1-\rho^2)\right)$.

1.6. Función de una o varias variables aleatorias

1.6.1. Caso univariante

Si g es una función medible, $Y=g(X)$ es una variable aleatoria porque,

$$Y:\Omega \xrightarrow{X} \mathcal{R} \xrightarrow{g} \mathcal{R},$$

e $Y^{-1}(B)=X^{-1}[g^{-1}(B)] \in \mathcal{A}$.

Tiene sentido hablar de la distribución de probabilidad asociada a Y, que como ya hemos visto podrá ser conocida mediante cualquiera de las tres funciones: P_Y, F_Y o f_Y. Lo inmediato es preguntarse por la relación entre las distribuciones de probabilidad de ambas variables. Es aparentemente sencillo, al menos en teoría, obtener F_Y en función de F_X. En efecto,

$$F_Y(y)=P(Y\leq y)=P(g(X)\leq y)=P(X\in g^{-1}\{]-\infty,y]\}). \tag{1.34}$$

Si la variable X es discreta se obtiene la siguiente relación entre las funciones de cuantía f_Y y f_X,

$$f_Y(y)=P(Y=y)=P(g(X)=y)=\sum_{\{g^{-1}(y)\cap D_X\}} f_X(x). \tag{1.35}$$

Pero la obtención de $g^{-1}\{]-\infty,y]\}$ o $g^{-1}(y)$ no siempre es sencilla. Veamos ejemplos en los que (1.34) puede ser utilizada directamente.

Ejemplo 1.14 *Sea $X\sim U(-1,1)$ y definamos $Y=X^2$. Para obtener F_Y, sea $y\in[0,1]$,*

$$F_Y(y)=P(Y\leq y)=P(X^2\leq y)=P(-\sqrt{y}\leq X\leq\sqrt{y})=F_X(\sqrt{y})-F_X(-\sqrt{y})=\sqrt{y}.$$

Entonces,

$$F_Y(y)=\begin{cases} 0, & si\ y<0; \\ \sqrt{y}, & si\ 0\leq y\leq 1; \\ 1, & si\ y>1. \end{cases}$$

Ejemplo 1.15 *Si X es una variable discreta con soporte D_X, definamos Y mediante*

$$Y = signo(X) = \begin{cases} \frac{X}{|X|}, & si\ X \neq 0 \\ 0, & si\ X = 0. \end{cases}$$

Con esta definición, $D_Y = \{-1, 0, 1\}$, y su función de cuantía viene dada por

$$f_Y(y) = \begin{cases} \sum_{x<0} f_X(x), & si\ y = -1 \\ f_X(0), & si\ y = 0 \\ \sum_{x>0} f_X(x), & si\ y = 1 \end{cases}$$

Cuando la variable aleatoria es discreta (1.34) y (1.35) son la únicas expresiones que tenemos para obtener la distribución de probabilidad de Y. El caso continuo ofrece, bajo ciertas condiciones, otra alternativa.

Teorema 1.2 *Sea X una variable aleatoria continua y sea g monótona, diferenciable con $g'(x) \neq 0,\ \forall x$. Entonces $Y = g(X)$ es una variable aleatoria con función de densidad,*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|, & si\ y \in g(\{D_X\}) \\ 0, & en\ el\ resto. \end{cases}$$

Demostración.- Como g es medible por ser continua, Y será una variable aleatoria. Supongamos ahora que g es monótona creciente. Tendremos, para $y \in g(\{D_X\})$,

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Derivando respecto de y obtendremos una función de densidad para Y,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dg^{-1}(y)} \cdot \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|.$$

En caso de monotonía decreciente para g,

$$F_Y(y) = P(Y \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

El resto se obtiene análogamente. ♠

Ejemplo 1.16 *Consideremos la variable aleatoria X cuya densidad viene dada por*

$$f_X(x) = \begin{cases} 0, & si\ x < 0, \\ \frac{1}{2}, & si\ 0 \leq x \leq 1, \\ \frac{1}{2x^2}, & si\ x > 1, \end{cases}$$

Definimos una nueva variable mediante la transformación $Y = 1/X$. La transformación cumple con las condiciones del teorema, $x = g^{-1}(y) = 1/y$ y $\frac{dg^{-1}(y)}{dy} = -\frac{1}{y^2}$, por tanto la densidad de Y vendrá dada por

$$f_Y(y) = \begin{cases} 0, & si\ y < 0, \\ \frac{1}{2} \cdot \frac{1}{y^2}, & si\ 1 \leq y < \infty, \\ \frac{1}{2(1/y)^2} \cdot \frac{1}{y^2}, & si\ 0 < y < 1, \end{cases}$$

que adecuadamente ordenado da lugar a la misma densidad que poseía X.

Hay dos transformaciones especialmente interesantes porque permiten obtener variables aleatorias con distribuciones preestablecidas. La primera conduce siempre a una $U(0,1)$ y la otra, conocida como *transformación integral de probabilidad*, proporciona la distribución que deseemos. Necesitamos previamente la siguiente definición.

Definición 1.11 (Inversa de una función de distribución) *Sea F una función en $\mathcal{R}$ que verifica las propiedades PF1) a PF4) de la página 12, es decir, se trata de una función de distribución de probabilidad. La* inversa de F *es la función definida mediante*

$$F^{-1}(x) = \text{ínf}\{t:\ F(t) \geq x\}.$$

Observemos que F^{-1} existe siempre, aun cuando F no sea continua ni estrictamente creciente. Como contrapartida, F^{-1} no es una inversa puntual de F, pero goza de algunas propiedades interesantes de fácil comprobación.

Proposición 1.2 *Sea F^{-1} la* inversa *de F. Entonces,*

a) *para cada x y t, $F^{-1}(x) \leq t \longleftrightarrow x \leq F(t)$,*

b) *F^{-1} es creciente y continua por la izquierda, y*

c) *si F es continua, entonces $F(F^{-1}(x)) = x,\ \forall x \in [0,1]$.*

Podemos ya definir las dos transformaciones antes mencionadas.

Proposición 1.3 (Transformada integral de probabilidad) *Sea $U \sim U(0,1)$, F una función de distribución de probabilidad y definimos $X = F^{-1}(U)$. Entonces, $F_X = F$.*

Demostración.- Como F^{-1} es monótona, X es una variable aleatoria. Por a) en la proposición anterior, $\forall t \in \mathcal{R}$,

$$F_X(t) = P(X \leq t) = P(F^{-1}(U) \leq t) = P(U \leq F(t)) = F(t). \qquad \spadesuit$$

Este resultado es la base de muchos procedimientos de simulación aleatoria porque permite obtener valores de cualquier variable aleatoria a partir de valores de una Uniforme, los valores de la Uniforme son a su vez generados con facilidad por los ordenadores. A fuer de ser rigurosos, debemos precisar que los ordenadores no generan exactamente valores de una Uniforme, lo que generan son valores pseudoaleatorios que gozan de propiedades semejantes a los de una Uniforme.

Proposición 1.4 (Obtención de una $\mathbf{U(0,1)}$) *Si F_X es continua, $U = F_X(X) \sim U(0,1)$.*

Demostración.- Hagamos $F = F_X$. Para $x \in [0,1]$, por la proposición 1.2 a), $P(U \geq x) = P(F(X) \geq x) = P(X \geq F^{-1}(x))$. La continuidad de F y la proposición 1.2 c) hacen el resto,

$$P(U \geq x) = P(X \geq F^{-1}(x)) = 1 - F(F^{-1}(x)) = 1 - x. \qquad \spadesuit$$

1.6.2. Caso multivariante

Para $X = (X_1, \ldots, X_k)$, vector aleatorio k-dimensional, abordaremos el problema solamente para el caso continuo. La obtención de la densidad de la nueva variable o vector resultante en función de $f_X(x_1, \ldots, x_k)$ plantea dificultades en el caso más general, pero bajo ciertas condiciones, equivalentes a las impuestas para el caso univariante, es posible disponer de una expresión relativamente sencilla.

Teorema 1.3 *Sea $X = (X_1, \ldots, X_k)$ es un vector aleatorio continuo con soporte D_X y sea $g = (g_1, \ldots, g_k) : \mathcal{R}^k \to \mathcal{R}^k$ una función vectorial que verifica:*

1. *g es uno a uno sobre D_X,*
2. *el Jacobiano de g, $J = \frac{\partial(g_1,\ldots,g_k)}{\partial(x_1,\ldots,x_k)}$, es distinto de cero $\forall x \in D_X$, y*
3. *existe $h = (h_1, \ldots, h_k)$ inversa de g.*

Entonces, $Y = g(X)$ es un vector aleatorio continuo cuya densidad conjunta, para $y = (y_1, \ldots, y_k) \in g(D_X)$, viene dada por

$$f_Y(y_1, \ldots, y_k) = f_X\left(h_1(y_1, \ldots, y_k), \ldots, h_k(y_1, \ldots, y_k)\right) |J^{-1}|, \tag{1.36}$$

donde $J^{-1} = \frac{\partial(h_1,\ldots,h_k)}{\partial(y_1,\ldots,y_k)}$ es el Jacobiano de h.

Este teorema no es más que el teorema del cambio de variable en una integral múltiple y su demostración rigurosa, de gran dificultad técnica, puede encontrarse en cualquier libro de Análisis Matemático. Un argumento heurístico que justifique (1.36) puede ser el siguiente. Para cada y,

$$\begin{aligned} f_Y(y_1, \ldots, y_k)dy_1 \cdots dy_k &\approx P\left\{Y \in \prod_{i=1}^{k}(y_i, y_i + dy_i)\right\} \\ &= P\left\{X \in h\left(\prod_{i=1}^{k}(y_i, y_i + dy_i)\right)\right\} \\ &= f_X(h(y)) \times vol\left[h\left(\prod_{i=1}^{k}(y_i, y_i + dy_i)\right)\right], \end{aligned}$$

Pero $vol\left[h\left(\prod_{i=1}^{k}(y_i, y_i + dy_i)\right)\right]$ es precisamente $|J^{-1}|dy_1, \ldots, dy_k$.

Veamos el interés del resultado a través de los siguientes ejemplos.

Ejemplo 1.17 (Continuación del ejemplo 1.11) *En la sección 1.3.6 estudiábamos el vector aleatorio determinado por las coordenadas de un punto elegido al azar en el círculo unidad. La densidad conjunta venía dada por*

$$f_{XY}(x, y) = \begin{cases} \dfrac{1}{\pi}, & si\ (x, y) \in C_1 \\ \\ 0, & en\ el\ resto. \end{cases}$$

Consideremos ahora las coordenadas polares del punto, $R = \sqrt{X^2 + Y^2}$ y $\Theta = \arctan Y/X$. Para obtener su densidad conjunta, necesitamos las transformaciones inversas, $X = R\cos\Theta$ e $Y = R\sin\Theta$. El correspondiente jacobiano vale $J_1 = R$ y la densidad conjunta,

$$f_{R\Theta}(r, \theta) = \begin{cases} \dfrac{r}{\pi}, & si\ (r, \theta) \in [0, 1] \times [0, 2\pi] \\ \\ 0, & en\ el\ resto. \end{cases}$$

Con facilidad se obtienen las marginales correspondientes, que resultan ser

$$f_R(r) = \begin{cases} 2r, & si\ r \in [0, 1] \\ \\ 0, & en\ el\ resto, \end{cases}$$

y para Θ*,*

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi}, & si\ \theta \in [0, 2\pi] \\ \\ 0, & en\ el\ resto. \end{cases}$$

Como $f_{R\Theta}(r,\theta) = f_R(r) f_\Theta(\theta),\ \forall (r,\theta)$*,* R *y* Θ son independientes.

Ejemplo 1.18 (Suma y producto de dos variables aleatorias) *Sea* $X = (X_1, X_2)$ *un vector aleatorio bidimensional con densidad conjunta* $f_X(x_1, x_2)$*. Definimos* $U = X_1 + X_2$ *y queremos obtener su densidad. Para poder utilizar el resultado anterior la transformación debe de ser también bidimensional, cosa que conseguimos si definimos una nueva variable* $V = X_1$*. Con* $Y = (U, V)$ *podemos aplicar el teorema, siendo la inversa* $X_1 = V$ *y* $X_2 = U - V$*, cuyo Jacobiano es* $J^{-1} = -1$*. Tendremos pues,*

$$f_Y(u, v) = f_X(v, u - v),$$

y para obtener la densidad marginal de la suma, U*,*

$$f_U(u) = \int_{-\infty}^{+\infty} f_X(v, u - v)\, dv. \tag{1.37}$$

Para obtener la densidad de $W = X_1 X_2$*, definimos* $T = X_1$ *y actuamos como antes. Con* $Y = (T, W)$ *y transformaciones inversas* $X_1 = T$ *y* $X_2 = W/T$*, el Jacobiano es* $J^{-1} = 1/T$ *y la densidad conjunta de* Y*,*

$$f_Y(t, w) = \frac{1}{|t|} f_X\left(t, \frac{w}{t}\right).$$

La marginal del producto se obtiene integrando respecto de la otra componente,

$$f_W(w) = \int_{-\infty}^{+\infty} \frac{1}{|t|} f_X\left(t, \frac{w}{t}\right) dt. \tag{1.38}$$

Hubieramos podido también proceder utilizando la transformación bidimensional $Y = (Y_1, Y_2)$*, con* $Y_1 = X_1 + X_2$ *e* $Y_2 = X_1 X_2$*, lo que en teoría nos hubiera hecho ganar tiempo; pero sólo en teoría, porque en la práctica las inversas hubieran sido más complicadas de manejar que las anteriores.*

Capítulo 2

Esperanza. Desigualdades. Función característica

2.1. Esperanza de una variable aleatoria

En el capítulo precedente hemos visto que la descripción completa de una variable o de un vector aleatorio nos la proporciona cualquiera de las funciones allí estudiadas. Es cierto que unas son de manejo más sencillo que otras, pero todas son equivalentes para el cometido citado.

En ocasiones no necesitamos un conocimiento tan exhaustivo y nos basta con una idea general. Ciertas características numéricas ligadas a las variables o los vectores aleatorios pueden satisfacernos. Estas cantidades son muy importantes en Teoría de la Probabilidad y sus aplicaciones, y su obtención se lleva a cabo a partir de las correspondientes distribuciones de probabilidad.

Entre estas constantes, sin duda las que denominaremos *esperanza matemática* y *varianza* son las de uso más difundido. La primera juega el papel de centro de gravedad de la distribución y nos indica alrededor de qué valor se situa nuestra variable o vector. La segunda completa la información indicándonos cuan dispersos o agrupados se presentan los valores alrededor de aquella. Existen también otras constantes que proporcionan información acerca de la distribución de probabilidad, son los llamados *momentos*, de los cuales esperanza y varianza son casos particulares. Los momentos pueden llegar a aportarnos un conocimiento exhaustivo de la variable aleatoria.

Definición 2.1 (Esperanza de una variable aleatoria discreta) *Sea X aleatoria discreta, f_X su función de cuantía y D_X su soporte. Si g es una función medible definida de $(\mathcal{R}, \beta)$ en $(\mathcal{R}, \beta)$, tal que $\sum_{x_i \in D_X} |g(x_i) f_X(x_i)| < +\infty$, decimos que existe la esperanza de $g(X)$, $E[g(X)]$, cuyo valor es*

$$E[g(X)] = \sum_{x_i \in D_X} g(x_i) f_X(x_i). \tag{2.1}$$

En particular, si $g(X) = X$,

$$E(X) = \sum_{x_i \in D_X} x_i f_X(x_i).$$

Definición 2.2 (Esperanza de una variable aleatoria continua) *Sea X aleatoria discreta, f_X su función de densidad. Si g es una función medible definida de $(\mathcal{R}, \beta)$ en $(\mathcal{R}, \beta)$, tal*

que $\int_{\mathcal{R}} |g(x)f_X(x)dx| < +\infty$*, decimos que existe la esperanza de* $g(X)$*,* $E[g(X)]$*, cuyo valor es,*

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx. \tag{2.2}$$

En particular, si $g(X) = X$*,*

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx.$$

2.1.1. Momentos de una variable aleatoria

Formas particulares de $g(X)$ dan lugar lo que denominamos *momentos de* X. En la tabla resumimos los distintos tipos de momentos y la correspondiente función que los origina, siempre que ésta sea integrable pues de lo contrario la esperanza no existe.

	de orden k	**absoluto de orden k**
Respecto del origen	X^k	$\|X\|^k$
Respecto de a	$(X-a)^k$	$\|X-a\|^k$
Factoriales	$X(X-1)\dots(X-k+1)$	$\|X(X-1)\dots(X-k+1)\|$

Tabla 1.- Forma de $g(X)$ para los distintos momentos de X

Respecto de la existencia de los momentos se verifica el siguiente resultado.

Proposición 2.1 *Si* $E(X^k)$ *existe, existen todos los momentos de orden inferior.*

La comprobación es inmediata a partir de la desigualdad $|X|^j \leq 1 + |X|^k,\ j \leq k$.

Ya hemos dicho en la introducción que el interés de los momentos de una variable aleatoria estriba en que son características numéricas que resumen su comportamiento probabilístico. Bajo ciertas condiciones el conocimiento de todos los momentos permite conocer completamente la distribución de probabilidad de la variable.

Especialmente relevante es el caso $k = 1$, cuyo correspondiente momento coincide con $E(X)$ y recibe también el nombre de *media*. Suele designarse mediante la letra griega μ (μ_X, si existe riesgo de confusión). Puesto que μ es una constante, en la tabla anterior podemos hacer $a = \mu$, obteniendo así una familia de momentos respecto de μ que tienen nombre propio: los *momentos centrales de orden* k, $E[(X-\mu)^k]$. De entre todos ellos cabe destacar la *varianza*,

$$Var(X) = \sigma_X^2 = E[(X-\mu)^2].$$

Propiedades de $E(X)$ **y** $V(X)$

Un primer grupo de propiedades no merecen demostración dada su sencillez. Conviene señalar que todas ellas derivan de las propiedades de la integral.

1. Propiedades de $E(X)$.

   PE1) La esperanza es un operador lineal,

   $$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)].$$

   En particular,

   $$E(aX + b) = aE(X) + b.$$

   PE2) $P(a \leq X \leq b) = 1 \Longrightarrow a \leq E(X) \leq b.$

PE3) $P(g(X) \leq h(X)) = 1 \Longrightarrow E[g(X)] \leq E[h(X)]$.

PE4) $|E[g(X)]| \leq E[|g(X)|]$.

2. Propiedades de $V(X)$.

PV1) $V(X) \geq 0$.

PV2) $V(aX + b) = a^2 V(X)$.

PV3) $V(X) = E(X^2) - [E(X)]^2$.

PV4) $V(X)$ hace mínima $E\left[(X-a)^2\right]$.

En efecto,

$$\begin{aligned} E\left[(X-a)^2\right] &= E\left[(X - E(X) + E(X) - a)^2\right] \\ &= E\left[(X-E(X))^2\right] + E\left[(E(X)-a)^2\right] + 2E\left[(X - E(X)(E(X)-a)\right] \\ &= V(X) + (E(X)-a)^2. \end{aligned}$$

El siguiente resultado nos ofrece una forma alternativa de obtener la $E(X)$ cuando X es no negativa.

Proposición 2.2 *Si para $X \geq 0$, existe $E(X)$, entonces*

$$E(X) = \int_0^{+\infty} P(X \geq x)\, dx = \int_0^{+\infty} P(X > x)\, dx = \int_0^{+\infty} (1 - F_X(x))\, dx. \tag{2.3}$$

2.1.2. Momentos de algunas variables aleatorias conocidas

Binomial

Si $X \sim B(n,p)$,

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \\ &= \sum_{x=0}^{n} x \frac{n(n-1)\ldots(n-x+1)}{x!} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^{n} \frac{(n-1)\ldots(n-x+1)}{(x-1)!} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-y-1} = np \end{aligned}$$

Para obtener $V(X)$, observemos que $E[(X(X-1)] = E(X^2) - E(X)$, y de aquí $V(X) = E[(X(X-1)] + E(X) - [E(X)]^2$. Aplicando un desarrollo análogo al anterior se obtiene $E[X(X-1)] = n(n-1)p^2$ y finalmente

$$V(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

Poisson

Si $X \sim P(\lambda)$,

$$E(X) = \sum_{x\geq 0} xe^{-\lambda}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}\sum_{x-1\geq 0}\frac{\lambda^{x-1}}{(x-1)!} = \lambda.$$

Por otra parte,

$$E[X(X-1)] = \sum_{x\geq 0} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x-2\geq 0}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2.$$

De aquí,

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Uniforme

Si $X \sim U(0,1)$,

$$E(X) = \int_{-\infty}^{+\infty} x f_X\, dx = \int_0^1 x\, dx = \frac{1}{2}$$

Para obtener $V(X)$ utilizaremos la expresión alternativa, $V(X) = E(X^2) - [E(X)]^2$,

$$E(X^2) = \int_0^1 x^2\, dx = \frac{1}{3},$$

y de aquí,

$$V(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Normal tipificada

Si $X \sim N(0,1)$, como su función de densidad es simétrica respecto del origen,

$$E(X^k) = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \begin{cases} 0, & si\ k = 2n+1 \\ m_{2n}, & si\ k = 2n. \end{cases}$$

Ello supone que $E(X) = 0$ y $V(X) = E(X^2)$. Para obtener los momentos de orden par,

$$m_{2n} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\, dx = \frac{2}{\sqrt{2\pi}}\int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\, dx.$$

Integrando por partes,

$$\int_0^{+\infty} x^{2n} e^{-\frac{x^2}{2}}\, dx =$$

$$-x^{2n-1}e^{-\frac{x^2}{2}}\Big|_0^{+\infty} + (2n-1)\int_0^{+\infty} x^{2n-2}e^{-\frac{x^2}{2}}\, dx = (2n-1)\int_0^{+\infty} x^{2n-2}e^{-\frac{x^2}{2}}\, dx,$$

lo que conduce a la fórmula de recurrencia $m_{2n} = (2n-1)m_{2n-2}$ y recurriendo sobre n,

$$m_{2n} = (2n-1)(2n-3)\cdots 1 = \\ = \frac{2n(2n-1)(2n-2)\cdots 2\cdot 1}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n n!}.$$

La varianza valdrá por tanto,

$$V(X) = E(X^2) = \frac{2!}{2\cdot 1!} = 1.$$

Normal con parámetros μ y σ^2

Si $Z \sim N(0,1)$ es fácil comprobar que la variable definida mediante la expresión $X = \sigma Z + \mu$ es $N(\mu, \sigma^2)$. Teniendo en cuenta las propiedades de la esperanza y de la varianza,

$$E(X) = \sigma E(X) + \mu = \mu, \quad var(X) = \sigma^2 var(Z) = \sigma^2,$$

que son precisamente los parámetros de la distribución.

2.2. Esperanza de un vector aleatorio

Sea $X = (X_1, \ldots, X_k)$ un vector aletorio y sea g una función medible de $\mathcal{R}^k$ en $\mathcal{R}$, la expresión de su esperanza depende, como en el caso unidimensional, del carácter del vector.

Vector aleatorio discreto.- Si D_X es el soporte del vector y f_X su función de cuantía conjunta, la esperanza se obtiene a partir de

$$E(g(X)) = \sum_{(x_1,\ldots,x_k)\in D_X} g(x_1, \ldots, x_k) f_X(x_1, \ldots, x_k).$$

Vector aleatorio continuo.- Si f_X es la función de densidad conjunta,

$$E(g(X)) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_k) f_X(x_1, \ldots, x_k)\, dx_1 \ldots dx_k.$$

En ambos casos la existencia de la esperanza esta supeditada a que $|g(x)f(x)|$ sea absolutamente sumable o integrable, respectivamente.

2.2.1. Momentos de un vector aleatorio

Como ya hemos visto en el caso de una variable aleatoria, determinadas formas de la función g dan lugar a los llamados *momentos* que se definen de forma análoga a como lo hicimos entonces. Las situaciones de mayor interés son ahora:

Momento conjunto.- El *momento conjunto de orden* $(n_1, \ldots, n_k)$ se obtiene, siempre que la esperanza exista, para

$$g(X_1, \ldots, X_k) = X_1^{n_1} \ldots X_k^{n_k},\ n_i \geq 0, \tag{2.4}$$

lo que da lugar a $E(X_1^{n_1} \ldots X_k^{n_k})$. Obsérvese que los momentos de orden k respecto del origen para cada componente pueden obtenerse como casos particulares de (2.4) haciendo $n_i = k$ y $n_j = 0,\ j \neq i$, pues entonces $E(X_1^{n_1} \ldots X_k^{n_k}) = E(X_i^k)$.

Momento conjunto central.- El *momento conjunto central de orden* $(n_1, \ldots, n_k)$ se obtienen, siempre que la esperanza exista, para

$$g(X_1, \ldots, X_k) = (X_1 - E(X_1))^{n_1} \ldots (X_k - E(X_k))^{n_k},\ n_i \geq 0,$$

2.2.2. Covarianza. Aplicaciones

De especial interés es el momento conjunto central obtenido para $n_i = 1$, $n_j = 1$ y $n_l = 0,\ l \neq (i,j)$. Recibe el nombre de *covarianza de X_i y X_j* y su expresión es,

$$cov(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))] = E(X_i X_j) - E(X_i)E(X_j).$$

La covarianza nos informa acerca del grado y tipo de dependencia existente entre ambas variables mediante su magnitud y signo, porque a diferencia de lo que ocurría con la varianza, la covarianza puede tener signo negativo.

Covarianza en la Normal bivariante

Si (X,Y) tiene una distribución Normal bivariante de parámetros μ_X, μ_y, σ_x, σ_y y ρ, ya vimos en la página 33 que $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$. Para simplificar la obtención de $cov(X,Y)$ podemos hacer uso de la invarianza por traslación de la covarianza (se trata de una propiedad de fácil demostración que ya comprobamos para la varianza). De acuerdo con ella, $cov(X,Y) = cov(X-\mu_x, Y-\mu_y)$ y podemos suponer que X e Y tienen media 0, lo que permite expresar la covarianza como $cov(X,Y) = E(XY)$. Procedamos a su cálculo.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_{XY}(x,y)dxdy \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_x}\right)^2-2\rho\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right)+\left(\frac{y}{\sigma_y}\right)^2\right\}}dxdy \\ &= \int_{-\infty}^{+\infty}\frac{y}{\sigma_y\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2}\left[\int_{-\infty}^{+\infty}\frac{x}{\sigma_x\sqrt{2\pi(1-\rho^2)}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x}{\sigma_x}-\rho\frac{y}{\sigma_y}\right)^2}dx\right]dy. \end{aligned} \tag{2.5}$$

La integral interior en (2.5) es la esperanza de una $N(\rho\sigma_x y/\sigma_y, \sigma_x^2(1-\rho^2))$ y su valor será por tanto $\rho\sigma_x y/\sigma_y$. Sustituyendo en (2.5)

$$E(XY) = \frac{\rho\sigma_x}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(\frac{y}{\sigma_y}\right)^2 e^{-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2}dy = \frac{\rho\sigma_x\sigma_y}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} z^2 e^{-\frac{1}{2}z^2}dy = \rho\sigma_x\sigma_y.$$

El coeficiente de correlación entre ambas variables es

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_x\sigma_y} = \rho.$$

Todos los parámetros de la Normal bivariante adquieren ahora significado.

La covarianza nos permite también obtener la varianza asociada a la suma de variables aleatorias, como vemos a continuación.

Esperanza y varianza de una suma

La linealidad de la esperanza aplicada cuando $g(X) = X_1 + \cdots + X_n$, permite escribir

$$E(X_1 + \cdots + X_k) = \sum_{i=1}^{k} E(X_i).$$

Si las varianzas de las variables $X_1, \ldots, X_k$ existen, la varianza de $S = \sum_{i=1}^{k} a_i X_i$, donde los a_i son reales cualesquiera, existe y viene dada por

$$\begin{aligned} V(S) &= E[(S-E(S))^2] = \\ &= E\left[\left(\sum_{i=1}^{k} a_i(X_i - E(X_i))\right)^2\right] \\ &= \sum_{i=1}^{k} a_i^2 V(X_i) + 2a_i a_j \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} cov(X_i, X_j). \end{aligned} \tag{2.6}$$

Independencia y momentos de un vector aleatorio

Un resultado interesante acerca de los momentos conjuntos se recoge en la proposición que sigue.

Proposición 2.3 *Si las variables aleatorias $X_1, \ldots, X_k$ son independientes, entonces*

$$E(X_1^{n_1} \ldots X_k^{n_k}) = \prod_{i=1}^{k} E(X_i^{n_i}).$$

Demostración.- Si suponemos continuo el vector, la densidad conjunta puede factorizarse como producto de las marginales y

$$\begin{aligned} &E(X_1^{n_1} \ldots X_k^{n_k}) = \\ &\quad = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_1^{n_1} \ldots x_k^{n_k} f_X(x_1, \ldots, x_k)\, dx_1 \ldots dx_k = \\ &\quad = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \left[\prod_{i=1}^{k} x_i^{n_i} f_i(x_i)\right] dx_1 \ldots dx_k = \\ &\quad = \prod_{i=1}^{k} \left[\int_{-\infty}^{+\infty} x_i^{n_i} f_i(x_i)\, dx_i\right] = \prod_{i=1}^{k} E(X_i^{n_i}). \end{aligned}$$

El caso discreto se demuestra análogamente. ♠

Observación 2.1 *El anterior resultado admite una formulación más general. Si las funciones $g_i,\ i = 1, \ldots, k$ son medibles, las $g_i(X_i)$ también son variables independientes y podemos escribir*

$$E\left[\prod_{i=1}^{k} g_i(X_i)\right] = \prod_{i=1}^{k} E[g_i(X_i)]. \tag{2.7}$$

Corolario 2.1 *Si las variables $X_1, \ldots, X_k$ son independientes, entonces $cov(X_i, X_j) = 0,\ \forall i, j$.*

Corolario 2.2 *Si las variables $X_1, \ldots, X_k$ son independientes y sus varianzas existen, la varianza de $S = \sum_{i=1}^{k} a_i X_i$, donde los a_i son reales cualesquiera, existe y viene dada por $V(S) = \sum_{i=1}^{k} a_i^2 V(X_i)$.*

Una aplicación de los anteriors resultados permite la obtención de la esperanza y la varianza de algunas conocidas variables de manera mucho más sencilla a como lo hicimos anteriormente. Veamos algunos ejemplos.

Ejemplo 2.1 (La Binomial y la Hipergeométrica como suma de Bernoullis) *Si recordamos las características de una Binomial fácilmente se comprueba que si $X \sim B(n, p)$ entonces $X = \sum_{i=1}^{n} X_i$ con $X_i \sim B(1, p)$ e independientes. Como $\forall i$,*

$$E(X_i) = p, \quad y \quad var(X_i) = p(1-p),$$

tendremos que

$$E(X) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = np,$$

y

$$var(X) = \sum_{i=1}^{n} var(X_i) = np(1-p). \tag{2.8}$$

Si $X \sim H(n, N, r)$ *también podemos escribir* $X = \sum_{i=1}^{n} X_i$ *con* $X_i \sim B(1, r/N)$ *pero a diferencia de la Binomial las variables* X_i *son ahora dependientes. Tendremos pues*

$$E(X) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = n\frac{r}{N},$$

y aplicando (2.6)

$$var(X) = \sum_{i=1}^{n} var(X_i) + 2\sum_{i=1}^{k}\sum_{j>i}^{k} cov(X_i, X_j) = n\ \frac{r}{N}\ \frac{N-r}{N} + n(n-1)cov(X_1, X_2), \tag{2.9}$$

puesto que todas las covarianzas son iguales.

Para obtener $cov(X_1, X_2)$,

$$\begin{aligned} cov(X_1, X_2) &= E(X_1X_2) - E(X_1)E(X_2) \\ &= P(X_1 = 1, X_2 = 1) - \left(\frac{r}{N}\right)^2 \\ &= P(X_2 = 1|X_1 = 1)P(X_1 = 1) - \left(\frac{r}{N}\right)^2 \\ &= \frac{r-1}{N-1}\ \frac{r}{N} - \left(\frac{r}{N}\right)^2 \\ &= -\frac{r(N-r)}{N^2(N-1)}. \end{aligned}$$

Sustituyendo en (2.9)

$$var(X) = n\ \frac{r}{N}\ \frac{N-r}{N}\left(1 - \frac{n-1}{N-1}\right). \tag{2.10}$$

Es interesante comparar (2.10) con (2.8), para $p = r/N$. *Vemos que difieren en el último factor, factor que será muy próximo a la unidad si* $n \ll N$. *Es decir, si la* fracción de muestreo *(así se la denomina en Teoría de Muestras) es muy pequeña. Conviene recordar aquí lo que dijimos en la página 16 al deducir la relación entre ambas distribuciones.*

Ejemplo 2.2 (La Binomial Negativa como suma de Geométricas) *Si* $X \sim BN(r, p)$ *y recordamos su definición, podemos expresarla como* $X = \sum_{i=1}^{r} X_i$, *donde cada* $X_i \sim BN(1, p)$ *e independiente de las demás y representa las pruebas Bernoulli necesarias después del* $(i-1)$*-ésimo éxito para alcanzar el i-ésimo.*

Obtendremos primero la esperanza y la varianza de una variable Geométrica de parámetro p.

$$E(X_i) = \sum_{n\geq 0} np(1-p)^n = p\sum_{n\geq 1} n(1-p)^n = \frac{1-p}{p},\ \forall i.$$

Para calcular la varianza necesitamos conocer $E(X_i^2)$,

$$E(X_i^2) = \sum_{n\geq 0} n^2p(1-p)^n = p\sum_{n\geq 1} n^2(1-p)^n = \frac{1-p}{p^2}(2-p),\ \forall i.$$

y de aquí,

$$var(X) = \frac{1-p}{p^2}(2-p) - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}.$$

La esperanza y la varianza de la Binomial Negativa de parámetros r y p, valdrán

$$E(X) = \sum_{i=1}^{r} E(X_i) = \frac{r(1-p)}{p}, \qquad var(X) = \sum_{i=1}^{r} var(X_i) = \frac{r(1-p)}{p^2}.$$

Covarianzas en una Multinomial

Si $X \sim M(n; p_1, p_2, \ldots, p_k)$, sabemos que cada componente $X_i \sim B(n, p_i)$, $i = 1, \ldots, k$, y puede por tanto expresarse como suma de n Bernoullis de parámetro p_i. La covarianza entre dos cualesquiera puede expresarse como,

$$cov(X_i, X_j) = cov\left(\sum_{k=1}^{n} X_{ik}, \sum_{l=1}^{n} X_{jl}\right).$$

Se demuestra fácilmente que

$$cov\left(\sum_{k=1}^{n} X_{ik}, \sum_{l=1}^{n} X_{jl}\right) = \sum_{k=1}^{n}\sum_{l=1}^{n} cov(X_{ik}, X_{jl}). \tag{2.11}$$

Para calcular $cov(X_{ik}, X_{jl})$ recordemos que

$$X_{ik} = \begin{cases} 1, & \text{si en la prueba } k \text{ ocurre } A_i; \\ 0, & \text{en cualquier otro caso,} \end{cases}$$

y

$$X_{jl} = \begin{cases} 1, & \text{si en la prueba } l \text{ ocurre } A_j; \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

En consecuencia, $cov(X_{ik}, X_{jl}) = 0$ si $k \neq l$ porque las pruebas de Bernoulli son independientes y

$$cov(X_{ik}, X_{jk}) = E(X_{ik}X_{jk}) - E(X_{ik})E(X_{jk}) = 0 - p_i p_j,$$

donde $E(X_{ik}X_{jk}) = 0$ porque en una misma prueba, la k-ésima, no puede ocurrir simultáneamente A_i y A_j. En definitiva,

$$cov(X_i, X_j) = \sum_{k=1}^{n} cov(X_{ik}, X_{jk}) = -np_i p_j.$$

El coeficiente de correlación entre ambas variables vale,

$$\rho_{ij} = -\frac{np_i p_j}{\sqrt{np_i(1-p_i)}\sqrt{np_j(1-p_j)}} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}.$$

El valor negativo de la covarianza y del coeficiente de correlación se explica por el hecho de que siendo el número total de pruebas fijo, n, a mayor número de ocurrencias de A_i, menor número de ocurrencias de A_j.

2.3. Esperanza condicionada

Sea (X, Y) un vector aleatorio definido sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y denotemos por $P_{X|Y=y}$ la distribución de probabilidad de X condicionada a $Y = y$. Si g es una función medible definida de $(\mathcal{R}, \beta)$ en $(\mathcal{R}, \beta)$, tal que $E(g(X))$ existe, la *esperanza condicionada* de $g(X)$ dado Y, $E[g(X)|Y]$, es una variable aleatoria que para $Y = y$ toma el valor

$$E[g(X)|y] = \int_{\Omega} g(X)\, dP_{X|Y=y}. \tag{2.12}$$

La forma de (2.12) depende de las características de la distribución del vector (X, Y).

(X, Y) **es discreto.-** Ello supone que el soporte de la distribución, D, es numerable y

$$E[g(X)|y] = \sum_{x \in D_y} g(x) P(X = x|Y = y) = \sum_{x \in D_y} g(x) f_{X|Y}(x|y),$$

donde $D_y = \{x; (x, y) \in D\}$ es la sección de D mediante y.

(X, Y) **es continuo.-** Entonces

$$E[g(X)|y] = \int_{\mathcal{R}} g(x) f_{X|Y}(x|y) dx.$$

Una definición similar puede darse para $E[h(Y)|X]$ siempre que $E[h(Y)]$ exista.

Ejemplo 2.3 *Sean X e Y variables aleatorias independientes ambas con distribución $B(n, p)$. La distribución de $X + Y$ se obtiene fácilmente a partir de*

$$\begin{aligned}
f_{X+Y}(m) &= P\left(\bigcup_{k=0}^{m} \{X = k, Y = m - k\}\right) \\
&= \sum_{k=0}^{m} P(X = k, Y = m - k) \\
&= \sum_{k=0}^{m} P(X = k) P(Y = m - k) \\
&= \sum_{k=0}^{m} \binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)} \\
&= p^m (1-p)^{2n-m} \sum_{k=0}^{m} \binom{n}{k} \binom{n}{m-k} \\
&= \binom{2n}{m} p^m (1-p)^{2n-m},
\end{aligned}$$

de donde $X + Y \sim B(2n, p)$.

La distribución condicionada de $Y|X+Y=m$ *es*

$$\begin{aligned} P(Y=k|X+Y=m) &= \frac{P(Y=k, X+Y=m)}{P(X+Y=m)} \\ &= \frac{P(Y=k, X=m-k)}{P(X+Y=m)} \\ &= \frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-(m-k)}}{\binom{2n}{m}p^m(1-p)^{2n-m}} \\ &= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}, \end{aligned}$$

es decir, $Y|X+Y=m \sim H(m,2n,n)$*. La* $E(Y|X+Y=m)$ *valdrá*

$$E(Y|X+Y=m) = \frac{nm}{2n} = \frac{m}{2}.$$

La definición la esperanza condicionada goza de todas las propiedades inherentes al concepto de esperanza anteriormente estudiadas. A título de ejemplo podemos recordar,

PEC1) La esperanza condicionada es un operador lineal,

$$E[(ag_1(X)+bg_2(X))|Y] = aE[g_1(X)|Y] + bE[g_2(X)|Y].$$

En particular,

$$E[(aX+b)|Y] = aE(X|Y) + b.$$

PEC2) $P(a \leq X \leq b) = 1 \Longrightarrow a \leq E(X|Y) \leq b.$

PEC3) $P(g_1(X) \leq g_2(X)) = 1 \Longrightarrow E[g_1(X)|Y] \leq E[g_2(X)|Y].$

PEC4) $E(c|Y) = c$, para c constante.

Momentos de todo tipo de la distribución condicionada se definen de forma análoga a como hicimos en el caso de la distribución absoluta y gozan de las mismas propiedades. Existen, no obstante, nuevas propiedades derivadas de las peculiares características de este tipo de distribuciones y de que $E[g(X)|Y)]$, en tanto que función de Y, es una variable aleatoria. Veámoslas.

Proposición 2.4 *Si* $E(g(X))$ *existe, entonces*

$$E\left(E[g(X)|Y)]\right) = E(g(X)). \tag{2.13}$$

Demostración.- Supongamos el caso continuo.

$$\begin{aligned} E\left(E[g(X)|Y)]\right) &= \int_{\mathcal{R}} E[g(X)|y)]f_y(y)dy \\ &= \int_{\mathcal{R}} \left[\int_{\mathcal{R}} g(x)f_{X|Y}(x|y)dx\right] f_y(y)dy \\ &= \int_{\mathcal{R}} \left[\int_{\mathcal{R}} g(x)\frac{f_{XY}(x,y)}{f_y(y)}dx\right] f_y(y)dy \\ &= \int_{\mathcal{R}} g(x)\left[\int_{\mathcal{R}} f_{XY}(x,y)dy\right] dx \\ &= \int_{\mathcal{R}} g(x)f_X(x)dx = E(g(X)). \end{aligned}$$

La demostración se haría de forma análoga para el caso discreto ♠

Ejemplo 2.4 *Consideremos el vector* (X,Y) *con densidad conjunta*

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{x}, & 0 < y \leq x \leq 1 \\ \\ 0, & \textit{en el resto.} \end{cases}$$

Fácilmente obtendremos que $X \sim U(0,1)$ *y que* $Y|X = x \sim U(0,x)$*. Aplicando el resultado anterior podemos calcular* $E(Y)$*,*

$$E(Y) = \int_0^1 E(Y|x)f_X(x)dx = \int_0^1 \frac{1}{2}xdx = \frac{1}{4}.$$

Ejemplo 2.5 *Un trabajador está encargado del correcto funcionamiento de n máquinas situadas en linea recta y distantes una de otra l metros. El trabajador debe repararlas cuando se averían, cosa que sucede con igual probabilidad para todas ellas e independientemente de una a otra. El operario puede seguir dos estrategias:*

1. *acudir a reparar la máquina estropeada y permanecer en ella hasta que otra máquina se avería, desplazándose entonces hacia ella, o*

2. *situarse en el punto medio de la linea de máquinas y desde allí acudir a la averiada, regresando nuevamente a dicho punto cuando la avería está resuelta.*

Si X *es la distancia que recorre el trabajador entre dos averías consecutivas, ¿cuál de ambas estrategias le conviene más para andar menos?*

Se trata, en cualquiera de las dos estrategias, de obtener la $E(X)$ *y elegir aquella que la proporciona menor. Designaremos por* $E_i(X)$ *la esperanza obtenida bajo la estrategia* $i = 1, 2$*.*

> ***Estrategia 1.-*** *Sea* A_k *el suceso,* el operario se encuentra en la máquina k*. Para obtener* $E_1(X)$ *recurriremos a la propiedad anterior, pero utilizando como distribución condicionada la que se deriva de condicionar respecto del suceso* A_k*. Tendremos* $E_1(X) = E(E(X|A_k))$*.*

Para obtener $E(X|A_k)$ *tengamos en cuenta que si* i *es la próxima máquina averiada,* $P(A_i) = 1/n,\ \forall i$ *y el camino a recorrer será*

$$X|A_k = \begin{cases} (k-i)l, & si\ i \leq k, \\ \\ (i-k)l, & si\ i > k. \end{cases}$$

Así pues,

$$E(X|A_k) = \frac{1}{n}\left(\sum_{i=1}^{k}(k-i)l + \sum_{i=k+1}^{n}(i-k)l\right) = \frac{l}{2n}[2k^2 - 2(n+1)k + n(n+1)].$$

Utilizando

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6},$$

obtenemos para $E_1(X)$

$$E_1(X) = E(E(X|A_k)) = \frac{1}{n}\sum_{k} E(X|A_k) = \frac{l(n^2-1)}{3n}$$

Estrategia 2.- *Para facilitar los cálculos supongamos que* n *es impar, lo que supone que hay una máquina situada en el punto medio de la linea, la* $\frac{n+1}{2}$*-ésima. Si la próxima máquina averiada es la* i *la distancia a recorrer será,*

$$X = \begin{cases} 2(\frac{n+1}{2} - i)l, & si\ i \leq \frac{n+1}{2}, \\ \\ 2(i - \frac{n+1}{2})l, & si\ i > \frac{n+1}{2}, \end{cases}$$

donde el 2 *se justifica porque el operario en esta segunda estrategia regresa siempre al punto medio de la linea de máquinas. La esperanza viene dada por,*

$$E_2(X) = \frac{2}{n}\sum_{i=1}^{n}\left|\frac{n+1}{2} - i\right| l = \frac{l(n-1)}{2}.$$

Como $E_1(X) \leq E_2(X) \longrightarrow (n+1)/3n \leq 1/2 \longrightarrow n \geq 2$*, podemos deducir que la primera estrategia es mejor, salvo que hubiera una sola máquina.*

También es posible relacionar la varianza absoluta con la varianza condicionada, aunque la expresión no es tan directa como la obtenida para la esperanza.

Proposición 2.5 *Si* $E(X^2)$ *existe, entonces*

$$var(X) = E(var[X|Y]) + var(E[X|Y]). \tag{2.14}$$

Demostración.- Haciendo del anterior resultado,

$$\begin{aligned} var(X) &= E\left[(X-E(X))^2\right] = E\left\{E\left[(X-E(X))^2|Y\right]\right\} \\ &= E\left\{E\left[\left(X^2+(E(X))^2-2XE(X)\right)|Y\right]\right\} \\ &= E\left\{E[X^2|Y]+E(X)^2-2E(X)E[X|Y]\right\} \\ &= E\left\{E[X^2|Y]-(E[X|Y])^2+(E[X|Y])^2+E(X)^2-2E(X)E[X|Y]\right\} \\ &= E\left\{var[X|Y]+(E[X|Y]-E(X))^2\right\} \\ &= E\left\{var[X|Y]\right\}+E\left\{(E[X|Y]-E(X))^2\right\} \\ &= E(var[X|Y])+var(E[X|Y]). \end{aligned}$$ ♠

Corolario 2.3 *Si $E(X^2)$ existe, por la propiedad anterior*

$$var(X) \geq var(E[X|Y]).$$

Si se verifica la igualdad, de (2.14) deducimos que $E(var[X|Y]) = 0$, y como $var[X|Y] \geq 0$, tendremos que $P(var[X|Y] = 0) = 1$, es decir

$$P\left(E\left\{(X-E[X|Y])^2|Y\right\}=0\right)=1,$$

y como $(X-E[X|Y])^2 \geq 0$, aplicando de nuevo el anterior razonamiento concluiremos que

$$P\left(X=E[X|Y]\right)=1,$$

lo que supone que, con probabilidad 1, X es una función de Y puesto que $E[X|Y]$ lo es.

2.4. Desigualdades

Si para $X \geq 0$ existe su esperanza, sea $\varepsilon > 0$ y escribamos (2.3) de la forma,

$$E(X)=\int_0^{+\infty} P(X\geq x)\,dx=\int_0^{\varepsilon} P(X\geq x)\,dx+\int_{\varepsilon}^{+\infty} P(X\geq x)\,dx.$$

Como la segunda integral es no negativa y la función $P(X \geq x)$ es decreciente,

$$E(X)\geq\int_0^{\varepsilon} P(X\geq x)\,dx\geq\int_0^{\varepsilon} P(X\geq \varepsilon)\,dx=\varepsilon P(X\geq\varepsilon),$$

y de aquí,

$$P(X\geq\varepsilon)\leq\frac{E(X)}{\varepsilon}. \tag{2.15}$$

Este resultado da lugar a dos conocidas desigualdades generales que proporcionan cotas superiores para la probabilidad de ciertos conjuntos. Estas desigualdades son válidas independientemente de cuál sea la distribución de probabilidad de la variable involucrada.

Desigualdad de Markov.- La primera de ellas se obtiene al sustituir en (2.15) X por $|X|^k$ y ε por ε^k,

$$P\left(|X|\geq\varepsilon\right)=P\left(|X|^k\geq\varepsilon^k\right)\leq\frac{1}{\varepsilon^k}E\left(|X|^k\right), \tag{2.16}$$

y es conocida como la *desigualdad de Markov.*

Desigualdad de Chebyshev.- Un caso especial de (2.16) se conoce como la *desigualdad de Chebyshev* y se obtiene para $k = 2$ y $X = X - E(X)$,

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} Var(X). \tag{2.17}$$

Un interesante resultado se deriva de esta última desigualdad.

Proposición 2.6 *Si $V(X) = 0$, entonces X es constante con probabilidad 1.*

Demostración.- Supongamos $E(X) = \mu$ y consideremos los conjuntos $A_n = \{|X - \mu| \geq 1/n\}$, aplicando (2.17)

$$P(A_n) = P\left(|X - \mu| \geq \frac{1}{n}\right) = 0, \ \forall n$$

y de aquí $P\left(\cup_n A_n\right) = 0$ y $P\left(\cap_n A_n^c\right) = 1$. Pero

$$\bigcap_{n\geq 1} A_n^c = \bigcap_{n\geq 1} \left\{|X - \mu| < \frac{1}{n}\right\} = \{X = \mu\},$$

luego $P(X = \mu) = 1$. ♠

Desigualdad de Jensen.- Si $g(X)$ es convexa sabemos que $\forall a, \exists \lambda_a$ tal que $g(x) \geq g(a) + \lambda_a(x - a), \forall x$. Si hacemos ahora $a = E(X)$,

$$g(X) \geq g\left(E(X)\right) + \lambda_a(X - E(X)),$$

y tomando esperanzas obtenemos la que se conoce como desigualdad de Jensen,

$$E(g(X)) \geq g(E(X)).$$

Teorema 2.1 (Desigualdad de Cauchy-Schwarz) *Sean X e Y variables aleatorias con varianzas finitas. Entonces $cov(X, Y)$ existe y se verifica*

$$[E(XY)]^2 \leq E(X^2)E(Y^2),$$

verificándose la igualdad si y solo si *existe un real α tal que $P(\alpha X + Y = 0) = 1$.*

Demostración.- Para cualesquiera números reales a y b se verifica

$$|ab| \leq \frac{a^2 + b^2}{2},$$

lo que significa que $E(XY) < \infty$ si $E(X^2) < \infty$ y $E(Y^2) < \infty$. Por otra parte, para cualquier real α, se tiene

$$E[(\alpha X + Y)^2] = \alpha^2 E(X^2) + 2\alpha E(XY) + E(Y^2) \geq 0,$$

lo que supone que la ecuación de segundo grado tiene a lo sumo una raíz y su discriminante será no positivo. Es decir,

$$[E(XY)]^2 \leq E(X^2)E(Y^2).$$

Si se diera la igualdad, la ecuación tendría una raíz doble, α_0, y $E[(\alpha_0 X + Y)^2] = 0$. Tratándose de una función no negativa, esto implica que $P(\alpha_0 X + Y = 0) = 1$. ♠

El coeficiente de correlación

El coeficiente de correlación entre dos componentes cualesquiera de un vector aleatorio, X y Y, se define como la covarianza de dichas variables tipificadas[1].

$$\rho_{XY} = cov(X_t, Y_t) = E\left[\left(\frac{X - E(X)}{\sigma_X}\right)\left(\frac{Y - E(Y)}{\sigma_Y}\right)\right] = \frac{cov(X,Y)}{\sigma_X \sigma_Y}.$$

De la desigualdad de Cauchy-Schwarz se desprende que $\rho_{XY}^2 \leq 1$ y en particular que $-1 \leq \rho_{XY} \leq 1$. Recordemos que cuando en Cauchy-Schwarz se daba la igualdad X e Y estaban relacionadas linealmente con probabilidad 1, $P(\alpha_0 X + Y = 0) = 1$, pero por otra parte dicha igualdad implica $\rho_{XY}^2 = 1$. El valor 0 es otro valor de interés para ρ_{XY}. Si $\rho_{XY} = 0$ decimos que las variables están *incorreladas* y además $cov(X, Y) = 0$. Hay entonces una ausencia total de relación lineal entre ambas variables. Podemos decir que el valor absoluto del coeficiente de correlación es una medida del grado de linealidad entre las variables medido, de menor a mayor, en una escala de 0 a 1.

2.4.1. La distribución Normal multivariante

La expresión de la densidad de la Normal bivariante que dábamos en la página 33 admite una forma alternativa más compacta a partir de lo que se conoce como la matriz de covarianzas de la distribución, $\mathbf{\Sigma}$, una matriz 2×2 cuyos elementos son las varianzas y las covarianzas del vector (X_1, X_2),

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

La nueva expresión de la densidad es

$$f(x_1, x_2) = \frac{|\mathbf{\Sigma}|^{-\frac{1}{2}}}{2\pi} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (x_1, x_2) \in \mathcal{R}^2,$$

donde $|\mathbf{\Sigma}|$ es el determinante de $\mathbf{\Sigma}$, cuyo valor es

$$|\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2(1 - \rho_{12}^2),$$

$\mathbf{x}' = (x_1\ x_2)$ es el vector de variables y $\boldsymbol{\mu}' = (\mu_1\ \mu_2)$ es el vector de medias.

Si el vector tiene n componentes, $X_1, X_2, \ldots, X_n$, la extensión a n dimensiones de la expresión de la densidad es ahora inmediata con esta notación. La matriz de covarianzas es una matriz $n \times n$ con componentes

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1(n-1)} & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2(n-1)} & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{1(n-1)} & \sigma_{2(n-1)} & \cdots & \sigma_{n-1}^2 & \sigma_{(n-1)n} \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{(n-1)n} & \sigma_n^2 \end{pmatrix},$$

el vector de medias es $\boldsymbol{\mu}' = (\mu_1\ \mu_2\ \ldots\ \mu_n)$ y la densidad tiene por expresión

$$f(x_1, x_2, \ldots, x_n) = \frac{|\mathbf{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (x_1, x_2, \ldots, x_n) \in \mathcal{R}^n. \tag{2.18}$$

[1] Una variable tipificada es la que resulta de transformar la original restándole la media y dividiendo por la desviación típica, $X_t = (X - \mu_X)/\sigma_X$. Como consecuencia de esta transformación $E(X_t) = 0$ y $var(X_t) = 1$.

Cuando las componentes del vector son independientes, las covarianzas son todas nulas y $\mathbf{\Sigma}$ es una matriz diagonal cuyos elementos son las varianzas de cada componente, por tanto

$$|\mathbf{\Sigma}|^{-1} = \frac{1}{\sigma_1^2\sigma_2^2\cdots\sigma_n^2}.$$

Además, la forma cuadrática que aparece en el exponente de (2.18) se simplifica y la densidad adquiere la forma

$$f(x_1,x_2,\ldots,x_n) = \frac{1}{\prod_{i=1}^n(2\pi\sigma_i^2)^{\frac{1}{2}}}e^{-\frac{1}{2}\sum_{i=1}^n\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2} = \prod_{i=1}^n\left[\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}\right],$$

que no es más que el producto de las densidades de cada una de las componentes.

Añadamos por último que la matriz $\mathbf{\Sigma}$ es siempre definida positiva y lo es también la forma cuadrática que aparece en el exponente de (2.18).

Transformación lineal de una Normal multivariante

A partir del vector $\boldsymbol{X}$ definimos un nuevo vector $\mathbf{Y}$ mediante una transformación lineal cuya matriz $\mathbf{A}$ es invertible. Tendremos

$$\mathbf{Y} = \mathbf{AX}, \qquad \text{y} \qquad \mathbf{X} = \mathbf{A^{-1}Y},$$

y dada la linealidad de la esperanza, la misma relación se verifica para los vectores de medias,

$$\boldsymbol{\mu_Y} = \mathbf{A}\boldsymbol{\mu_X}, \qquad \text{y} \qquad \boldsymbol{\mu_X} = \mathbf{A^{-1}}\boldsymbol{\mu_Y},$$

Para conocer la distribución del vector $\mathbf{Y}$ recurriremos a la fórmula del cambio de variable (1.36). El jacobiano de la transformación inversa es precisamente $|J^{-1}| = |\mathbf{A^{-1}}|$, con lo que la densidad conjunta de $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ valdrá,

$$\begin{aligned}
f_{\mathbf{Y}}(y_1,y_2,\ldots,y_n) &= \frac{|\mathbf{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\mathbf{A^{-1}y}-\mathbf{A^{-1}}\boldsymbol{\mu_Y})'\mathbf{\Sigma}^{-1}(\mathbf{A^{-1}y}-\mathbf{A^{-1}}\boldsymbol{\mu_Y})}|\mathbf{A^{-1}}| \\
&= \frac{|\mathbf{A^{-1}}||\mathbf{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}[\mathbf{A^{-1}}(\mathbf{y}-\boldsymbol{\mu_Y})]'\mathbf{\Sigma}^{-1}[\mathbf{A^{-1}}(\mathbf{y}-\boldsymbol{\mu_Y})]} \\
&= \frac{|\mathbf{A^{-1}}||\mathbf{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu_Y})[\mathbf{A^{-1}}]'\mathbf{\Sigma}^{-1}\mathbf{A^{-1}}(\mathbf{y}-\boldsymbol{\mu_Y})} \\
&= \frac{|\mathbf{A\Sigma A'}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu_Y})'[\mathbf{A\Sigma A'}]^{-1}(\mathbf{y}-\boldsymbol{\mu_Y})},
\end{aligned}$$

que es la densidad de una Normal multivariante con vector de medias $\boldsymbol{\mu_Y} = \boldsymbol{A\mu_X}$ y matriz de covarianzas $\mathbf{\Sigma_Y} = \mathbf{A\Sigma A'}$

2.5. Función característica

La *función característica* es una herramienta de gran utilidad en Teoría de la Probabilidad, una de sus mayores virtudes reside en facilitar la obtención de la distribución de probabilidad de la suma de variables aleatorias y la del límite de sucesiones de variables aleatorias, situaciones ambas que aparecen con frecuencia en Inferencia Estadística.

El concepto de *función característica*, introducido por Lyapunov en una de la primeras versiones del Teorema Central del Límite, procede del Análisis Matemático donde se le conoce con el nombre de *transformada de Fourier*.

Definición 2.3 *Sea X una variable aleatoria y sea $t \in R$. La función característica de X, $\phi_X(t)$, se define como $E(e^{itX}) = E(\cos tX) + iE(\sin tX)$.*

Como $|e^{itX}| \leq 1$, $\forall t$, $\phi_X(t)$ existe siempre y está definida $\forall t \in R$. Para su obtención recordemos que,

Caso discreto.- Si X es una v. a. discreta con soporte D_X y función de cuantía $f_X(x)$,

$$\phi_X(t) = \sum_{x \in D_X} e^{itx} f_X(x). \tag{2.19}$$

Caso continuo.- Si X es una v. a. continua con función de densidad de probabilidad $f_X(x)$,

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx. \tag{2.20}$$

De la definición se derivan, entre otras, las siguientes propiedades:

Pϕ1) $\phi_X(0) = 1$

Pϕ2) $|\phi_X(t)| \leq 1$

Pϕ3) $\phi_X(t)$ es uniformemente continua

En efecto,

$$\phi_X(t+h) - \phi_X(t) = \int_\Omega e^{itX}(e^{ihX} - 1)\, dP.$$

Al tomar módulos,

$$|\phi_X(t+h) - \phi_X(t)| \leq \int_\Omega |e^{ihX} - 1|\, dP, \tag{2.21}$$

pero $|e^{ihX} - 1| \leq 2$ y (2.21) será finito, lo que permite intercambiar integración y paso al límite, obteniendo

$$\lim_{h \to 0} |\phi_X(t+h) - \phi_X(t)| \leq \int_\Omega \lim_{h \to 0} |e^{ihX} - 1|\, dP = 0.$$

Pϕ4) Si definimos $Y = aX + b$,

$$\phi_Y(t) = E(e^{itY}) = E\left(e^{it(aX+b)}\right) = e^{itb}\phi_X(at)$$

Pϕ5) Si $E(X^n)$ existe, la función característica es n veces diferenciable y $\forall k \leq n$ se verifica $\phi_X^{(k)}(0) = i^k E(X^k)$

La propiedad 5 establece un interesante relación entre las derivadas de $\phi_X(t)$ y los momentos de X cuando estos existen, relación que permite desarrollar $\phi_X(t)$ en serie de potencias. En efecto, si $E(X^n)$ existe $\forall n$, entonces,

$$\phi_X(t) = \sum_{k \geq 0} \frac{i^k E(X^k)}{k!} t^k. \tag{2.22}$$

2.5.1. Función característica e independencia

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y definimos $Y = X_1+X_2+\cdots+X_n$, por la observación (2.7) de la página 51 tendremos,

$$\phi_Y(t) = E\left(e^{it(X_1+X_2+\cdots+X_n)}\right) = E\left(\prod_{k=1}^{n} e^{itX_k}\right) = \prod_{k=1}^{n} E\left(e^{itX_k}\right) = \prod_{k=1}^{n} \phi_{X_k}(t), \qquad (2.23)$$

expresión que permite obtener con facilidad la función característica de la suma de variables independientes y cuya utilidad pondremos de manifiesto de inmediato.

2.5.2. Funciones características de algunas distribuciones conocidas

Bernoulli.- Si $X \sim B(1,p)$

$$\phi_X(t) = e^0 q + e^{it}p = q + pe^{it}.$$

Binomial.- Si $X \sim B(n,p)$

$$\phi_X(t) = \prod_{k=1}^{n}(q + pe^{it}) = (q + pe^{it})^n.$$

Poisson.- Si $X \sim P(\lambda)$

$$\phi_X(t) = \sum_{x\geq 0} e^{itx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x\geq 0}\frac{(\lambda e^{it})^x}{x!} = e^{\lambda(e^{it}-1)}.$$

Normal tipificada.- Si $Z \sim N(0,1)$, sabemos que existen los momentos de cualquier orden y en particular, $E(Z^{2n+1}) = 0,\ \forall n$ y $E(Z^{2n}) = \frac{(2n)!}{2^n n!},\ \forall n$. Aplicando (2.22),

$$\phi_Z(t) = \sum_{n\geq 0}\frac{i^{2n}(2n)!}{2^n(2n)!n!}t^{2n} = \sum_{n\geq 0}\frac{\left(\frac{(it)^2}{2}\right)^n}{n!} = \sum_{n\geq 0}\frac{\left(-\frac{t^2}{2}\right)^n}{n!} = e^{-\frac{t^2}{2}}.$$

Para obtener $\phi_X(t)$ si $X \sim N(\mu,\sigma^2)$, podemos utilizar el resultado anterior y P4. En efecto, recordemos que X puede expresarse en función de Z mediante $X = \mu + \sigma Z$ y aplicando P4,

$$\phi_X(t) = e^{i\mu t}\phi_Z(\sigma t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

Observación 2.2 *Obsérvese que $Im(\phi_Z(t)) = 0$. El lector puede comprobar que no se trata de un resultado exclusivo de la Normal tipificada, si no de una propiedad que poseen todas las v. a. con distribución de probabilidad simétrica, es decir, aquellas que verifican (P(X ≥ x)=P(X ≤ −x)).*

Gamma.- Si $X \sim G(\alpha,\lambda)$, su función de densidad de probabilidad viene dada por

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & \text{si } x > 0, \\ \\ 0 & \text{si } x \leq 0, \end{cases}$$

por lo que aplicando (2.20),

$$\phi_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-\lambda x} dx,$$

que con el cambio $y = x(1 - it/\lambda)$ conduce a

$$\phi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}.$$

Hay dos casos particulares que merecen ser mencionados:

Exponencial.- La distribución exponencial puede ser considerada un caso particular de $G(\alpha, \lambda)$ cuando $\alpha = 1$. A partir de aquí,

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

Chi-cuadrado.- Cuando $\alpha = n/2$ y $\lambda = 1/2$, decimos que X tiene una distribución χ^2 con n grados de libertad, $X \sim \chi^2_n$. Su función característica será

$$\phi_X(t) = (1 - 2it)^{-\frac{n}{2}}.$$

2.5.3. Teorema de inversión. Unicidad

Hemos obtenido la función característica de una v. a. X a partir de su distribución de probabilidad, pero es posible proceder de manera inversa por cuanto el conocimiento de $\phi_X(t)$ permite obtener $F_X(x)$.

Teorema 2.2 (Fórmula de inversión de Lévy) *Sean $\phi(t)$ y $F(x)$ las funciones característica y de distribución de la v. a. X y sean $a \leq b$ sendos puntos de continuidad de F, entonces*

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt.$$

Que puede también expresarse

$$F(x_0 + h) - F(x_0 - h) = \lim_{T\to\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin ht}{t} e^{-itx_0} \phi(t) dt,$$

donde $h > 0$ y $x_0 + h$ y $x_0 - h$ son puntos de continuidad de F.

Este resultado permite obtener $F(x)$ en cualquier x que sea punto de continuidad de F. Basta para ello que en $F(x) - F(y)$ hagamos que $y \to -\infty$ a través de puntos de continuidad. Como $F_X(x)$ es continua por la derecha, la tendremos también determinada en los puntos de discontinuidad sin más que descender hacia ellos a través de puntos de continuidad.

Si la variable es continua, un corolario del anterior teorema permite obtener la función de densidad directamente a partir de la función característica.

Corolario 2.4 *Si $\phi(t)$ es absolutamente integrable en R, entonces la función de distribución es absolutamente continua, su derivada es uniformemente continua y*

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{2\pi} \int_R e^{-itx} \phi(t) dt.$$

Este teorema tiene una trascendencia mayor por cuanto implica la *unicidad* de la función característica, que no por casualidad recibe este nombre, porque *caracteriza*, al igual que lo hacen otras funciones asociadas a X (la de distribución, la de probabilidad o densidad de probabilidad, ...), su distribución de probabilidad. Podemos afirmar que si dos variables X e Y comparten la misma función característica tienen idéntica distribución de probabilidad. La combinación de este resultado con las propiedades antes enunciadas da lugar a una potente herramienta que facilita el estudio y obtención de las distribuciones de probabilidad asociadas a la suma de variables independientes. Veámoslo con algunos ejemplos.

1) Suma de Binomiales independientes.- Si las variables $X_k \sim B(n_k, p),\ k = 1, 2, \ldots, m$ son independientes, al definir $X = \sum_{k=1}^{m} X_k$, sabemos por (2.23) que

$$\phi_X(t) = \prod_{k=1}^{m} \phi_{X_k}(t) = \prod_{k=1}^{m} (q + pe^{it})^{n_k} = (q + pe^{it})^n,$$

con $n = n_1 + n_2 + \cdots + n_m$. Pero siendo esta la función característica de una variable $B(n, p)$, podemos afirmar que $X \sim B(n, p)$.

2) Suma de Poisson independientes.- Si nuestra suma es ahora de variables Poisson independientes, $X_k \sim P(\lambda_k)$, entonces

$$\phi_X(t) = \prod_{k=1}^{m} \phi_{X_k}(t) = \prod_{k=1}^{m} e^{\lambda_k(e^{it}-1)} = e^{\lambda(e^{it}-1)},$$

con $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_m$. Así pues, $X \sim P(\lambda)$.

3) Combinación lineal de de Normales independientes.- Si $X = \sum_{k=1}^{n} c_k X_k$ con $X_k \sim N(\mu_k, \sigma_k^2)$ e independientes,

$$\begin{aligned} \phi_X(t) &= \phi_{\sum c_k X_k}(t) = \prod_{k=1}^{n} \phi_{X_k}(c_k t) \\ &= \prod_{k=1}^{n} e^{ic_k t\mu_k - \frac{\sigma_k^2 c_k^2 t^2}{2}} \\ &= e^{i\mu t - \frac{\sigma^2 t^2}{2}}, \end{aligned} \tag{2.24}$$

Se deduce de (2.24) que $X \sim N(\mu, \sigma^2)$ con

$$\mu = \sum_{k=1}^{n} c_k \mu_k \qquad \text{y} \qquad \sigma^2 = \sum_{k=1}^{n} c_k^2 \sigma_k^2.$$

4) Suma de Exponenciales independientes.- En el caso de que la suma esté formada por n variables independientes todas ellas $Exp(\lambda)$,

$$\phi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^n = \left(1 - \frac{it}{\lambda}\right)^{-n},$$

y su distribución será la de una $G(n, 1/\lambda)$.

5) Cuadrado de una $N(0,1)$**.-** Sea ahora $Y = X^2$ con $X \sim N(0,1)$, su función característica viene dada por

$$\phi_Y(t) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{x^2}{2}(1-2it)} dx = \frac{1}{(1-2it)^{\frac{1}{2}}},$$

lo que nos asegura que $Y \sim \chi^2_1$.

Algunos de estos resultados fueron obtenidos anteriormente, pero su obtención fué entonces mucho más laboriosa de lo que ahora ha resultado.

2.5.4. Teorema de continuidad de Lévy

Se trata del último de los resultados que presentaremos y permite conocer la convergencia de una sucesión de v. a. a través de la convergencia puntual de la sucesión de sus funciones características.

Teorema 2.3 (Directo) *Sea* $\{X_n\}_{n\geq 1}$ *una sucesión de v. a. y* $\{F_n(x)\}_{n\geq 1}$ *y* $\{\phi_n(t)\}_{n\geq 1}$ *las respectivas sucesiones de sus funciones de distribución y características. Sea* X *una v. a. y* $F_X(x)$ *y* $\phi(t)$ *sus funciones de distribución y característica, respectivamente. Si* $F_n \xrightarrow{w} F$ *(es decir,* $X_n \xrightarrow{L} X$*), entonces*

$$\phi_n(t) \longrightarrow \phi(t), \quad \forall t \in R.$$

Resultado que se completa con el teorema inverso.

Teorema 2.4 (Inverso) *Sea* $\{\phi_n(t)\}_{n\geq 1}$ *una sucesión de funciones características y* $\{F_n(x)\}_{n\geq 1}$ *la sucesión de funciones de distribución asociadas. Sea* $\phi(t)$ *una función continua, si* $\forall t \in R$*,* $\phi_n(t) \to \phi(t)$*, entonces*

$$F_n \xrightarrow{w} F,$$

donde $F(x)$ *en una función de distribución cuya función característica es* $\phi(t)$*.*

Este resultado permite, como decíamos antes, estudiar el comportamiento límite de sucesiones de v. a. a través del de sus funciones características, generalmente de mayor sencillez. Sin duda una de sus aplicaciones más relevantes ha sido el conjunto de resultados que se conocen como *Teorema Central del Límite* (TCL), bautizados con este nombre por Lyapunov que pretendió así destacar el papel *central* de estos resultados en la Teoría de la Probabilidad.

Capítulo 3

Sucesiones de variables aleatorias. Teoremas de convergencia

3.1. Introducción

Los capítulos anteriores nos han permitido familiarizarnos con el concepto de variable y vector aleatorio, dotándonos de las herramientas que nos permiten conocer su comportamiento probabilístico. En el caso de un vector aleatorio somos capaces de estudiar el comportamiento conjunto de un número finito de variables aleatorias. Pero imaginemos por un momento los modelos probabilísticos asociados a los siguientes fenómenos:

1. lanzamientos sucesivos de una moneda,
2. tiempos transcurridos entre dos llamadas consecutivas a una misma centralita,
3. sucesión de estimadores de un parámetro cuando se incrementa el tamaño de la muestra...

En todos los casos las variables aleatorias involucradas lo son en cantidad numerable y habremos de ser capaces de estudiar su comportamiento conjunto y, tal y como siempre sucede en ocasiones similares, de conocer cuanto haga referencia al límite de la sucesión. Del comportamiento conjunto se ocupa una parte de la Teoría de la Probabilidad que dado su interés ha tomado entidad propia: la Teoría de los Procesos Estocásticos. En este capítulo nos ocuparemos de estudiar cuanto está relacionado con el límite de las sucesiones de variables aleatorias. Este estudio requiere en *primer lugar*, introducir los tipos de convergencia apropiados a la naturaleza de las sucesiones que nos ocupan, para en *segundo lugar* obtener las condiciones bajo las que tienen lugar las dos convergencias que nos interesan: la convergencia de la sucesión de variables a una constante (Leyes de los Grandes Números) y la convergencia a otra variable (Teorema Central del Límite). El estudio de esta segunda situación se ve facilitado con el uso de una herramienta conocida como *función característica* de la cual nos habremos ocupado previamente.

Dos sencillos ejemplos relacionados con la distribución Binomial no servirán de introducción y nos ayudarán a situarnos en el problema.

Ejemplo 3.1 (Un resultado de J. Bernoulli) *Si repetimos n veces un experimento cuyo resultado es la ocurrencia o no del suceso A, tal que $P(A) = p$, y si la repeticiones son independientes unas de otras, la variable X_n =*número de ocurrencias de A*, tiene una distribución $B(n, p)$. La variable X_n/n representa la frecuencia relativa de A y sabemos que*

$$E\left(\frac{X_n}{n}\right) = \frac{1}{n}E(X_n) = \frac{np}{n} = p,$$

y

$$var\left(\frac{X_n}{n}\right) = \frac{1}{n^2} var(X_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Si aplicamos la desigualdad de Chebyshev,

$$P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{var(X_n/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \xrightarrow{n} 0.$$

Deducimos que la frecuencia relativa de ocurrencias de A converge*, en algún sentido, a P(A).*

Ejemplo 3.2 (Binomial vs Poisson) *El segundo ejemplo ya fue expuesto en la página 15 y no lo repetiremos aquí. Hacía referencia a la aproximación de la distribución Binomial mediante la distribución de Poisson. Vimos que cuando tenemos un gran número de pruebas Bernoulli con una probabilidad de éxito muy pequeña de manera que* $\lim_n np_n = \lambda$, $0 < \lambda < +\infty$, *la sucesión de funciones de cuantía de las variables aleatorias* $X_n \sim B(n, p_n)$ *converge a la función de cuantía de* $X \sim Po(\lambda)$.

Dos ejemplos con sucesiones de variables Binomiales que han conducido a *límites* muy distintos. En el primero, el valor límite es la probabilidad de un suceso, y por tanto una constante; en el segundo la función de cuantía tiende a otra función de cuantía.

3.2. Tipos de convergencia

Comenzaremos por formalizar el tipo de convergencia que aparece en el primer ejemplo. Para ello, y también para el resto de definiciones, sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ consideremos la sucesión de variables aleatorias $\{X_n\}$ y la variable aleatoria X.

Definición 3.1 (Convergencia en probabilidad) *Decimos que* $\{X_n\}$ *converge a X en probabilidad,* $X_n \xrightarrow{P} X$*, si para cada* $\delta > 0$*,*

$$\lim_n P\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\} = 0.$$

No es esta una convergencia puntual como las que estamos acostumbrados a utilizar en Análisis Matemático. La siguiente sí es de este tipo.

Definición 3.2 (Convergencia casi segura o con probabilidad 1) *Decimos que* $\{X_n\}$ *converge casi seguramente*[1] *a X (o con probabilidad 1),* $X_n \xrightarrow{a.s.} X$*, si*

$$P(\{\omega : \lim_n X_n(\omega) = X(\omega)\}) = 1.$$

El último tipo de convergencia involucra a las funciones de distribución asociadas a cada variable y requiere previamente una definición para la convergencia de aquellas.

Definición 3.3 (Convergencia débil) *Sean* F_n*,* $n \geq 1$ *y F, funciones de distribución de probabilidad. Decimos que la sucesión* F_n *converge débilmente*[2] *a F,* $F_n \xrightarrow{\omega} F$*, si* $\lim_n F_n(x) = F(x)$*,* $\forall x$ *que sea punto de continuidad de F.*

[1] Utilizaremos la abreviatura *a. s.*, que corresponde a las iniciales de **a**lmost **s**urely, por ser la notación más extendida

[2] Utilizaremos la abreviatura ω, que corresponde a la inicial de **w**eakly, por ser la notación más extendida

Definición 3.4 (Convergencia en ley) *Decimos que* $\{X_n\}$ *converge en ley a* X, $X_n \xrightarrow{L} X$, *si* $F_{X_n} \xrightarrow{\omega} F_X$. *Teniendo en cuenta la definición de* F_n *y* F, *la convergencia en ley puede expresarse también*

$$X_n \xrightarrow{L} X \Longleftrightarrow \lim_n P(X_n \leq x) = P(X \leq x).$$

Definición 3.5 (Convergencia en media cuadrática) *Decimos que* $\{X_n\}$ *converge en media cuadrática a* X, $X_n \xrightarrow{m.s} X$, *si*

$$\lim_n E[(X_n - X)^2] = 0.$$

Las relaciones entre los tres tipos de convergencia se establecen en el siguiente teorema.

Teorema 3.1 (Relaciones entre convergencias) *Sean* X_n *y* X *variables aleatorias definidas sobre un mismo espacio de probabilidad entonces:*

$$\left.\begin{array}{l} X_n \overset{a.s.}{\to} X \\ \\ X_n \overset{m.s.}{\to} X \end{array}\right\} \Rightarrow X_n \overset{P}{\to} X \Rightarrow X_n \overset{L}{\to} X$$

Las convergencias *casi segura* y *en probabilidad* tienen distinta naturaleza, mientras aquella es de tipo puntual, esta última es de tipo conjuntista. El ejemplo que sigue ilustra bien esta diferencia y pone de manifiesto que la contraria de la primera implicación no es cierta.

Ejemplo 3.3 (La convergencia en probabilidad $\not\Rightarrow$ la convergencia casi segura) *Como espacio de probabilidad consideraremos el intervalo unidad dotado de la* σ*-álgebra de Borel y de la medida de Lebesgue, es decir, un espacio de probabilidad uniforme en]0,1]. Definimos la sucesión* $X_n = \mathbf{1}_{I_n}$, $\forall n$, *con* $I_n =]\frac{p}{2^q}, \frac{p+1}{2^q}]$, *siendo* p *y* q *los únicos enteros positivos que verifican,* $p + 2^q = n$ *y* $0 \leq p < 2^q$. *Obviamente* $q = q(n)$ *y* $\lim_n q(n) = +\infty$. *Los primeros términos d ela sucesión son,*

$$\begin{array}{llll} n=1 & q=0, p=0 & X_1 = \mathbf{1}_{]0,1]} \\ n=2 & q=1, p=0 & X_2 = \mathbf{1}_{]0,\frac{1}{2}]} \\ n=3 & q=1, p=1 & X_3 = \mathbf{1}_{]\frac{1}{2},1]} \\ n=4 & q=2, p=0 & X_4 = \mathbf{1}_{]0,\frac{1}{4}]} \\ n=5 & q=2, p=1 & X_5 = \mathbf{1}_{]\frac{1}{4},\frac{1}{2}]} \\ n=6 & q=2, p=2 & X_6 = \mathbf{1}_{]\frac{1}{2},\frac{3}{4}]} \\ n=7 & q=2, p=3 & X_7 = \mathbf{1}_{]\frac{3}{4},1]} \\ \ldots\ldots & & \end{array}$$

Observemos que si $X = 0$, $\forall \delta > 0$ *se tiene* $\lambda\{\omega : |X_n(\omega)| \geq \delta\} = \lambda\{\omega : |X_n(\omega)| = 1\} = \lambda(I_n) = 2^{-q}$, $2^q \leq n < 2^{q+1}$ *y* $X_n \xrightarrow{\lambda} 0$*; pero dada la construcción de las* X_n *en ningún* $\omega \in [0,1]$ *se verifica* $\lim X_n(\omega) = 0$.

Las convergencias casi y en media cuadrática no se implican mutuamente. Veámoslo con sendos contraejemplos.

Ejemplo 3.4 (La convergencia a.s. $\not\Longleftrightarrow$ la convergencia m.s) *Con las misma sucesión del ejemplo anterior, como* $X_n = \mathbf{1}_{I_n}$, *con* $I_n =]\frac{p}{2^{q(n)}}, \frac{p+1}{2^{q(n)}}]$,

$$E(X_n^2) = \frac{1}{2^{q(n)}},$$

con $\lim_n q(n) = \infty$. *En definitiva,* $E(X_n^2) \xrightarrow{n} 0$, *lo que pone de manifiesto que la convergencia en media cuadrática* $\nrightarrow$ *la convergencia casi segura.*

Definamos ahora $X_n = \sqrt{n}\mathbf{1}_{]0,1/n]}$. *Claramente* $X_n \xrightarrow{a.s.} 0$ *pero* $E(X_n^2) = 1, \forall n$ *y* $X_n \overset{a.s.}{\nrightarrow} 0$.

La convergencia en ley (débil) no implica la convergencia en probabilidad, como pone de manifiesto el siguiente ejemplo, lo que justifica el nombre de *convergencia débil* puesto que es la última en la cadena de implicaciones.

Ejemplo 3.5 *Consideremos una variable Bernoulli con* $p = 1/2$, $X \sim B(1, 1/2)$ *y definamos una sucesión de variables aleatorias,* $X_n = X$, $\forall n$. *La variable* $Y = 1 - X$ *tiene la misma distribución que* X, *es decir,* $Y \sim B(1, 1/2)$. *Obviamente* $X_n \xrightarrow{L} Y$, *pero como* $|X_n - Y| = |2X - 1| = 1$, *no puede haber convergencia en probabilidad.*

3.3. Leyes de los Grandes Números

El nombre de *leyes de los grandes números* hace referencia al estudio de un tipo especial de límites derivados de la sucesión de variables aleatorias $\{X_n\}$. Concretamente los de la forma $\lim_n \frac{S_n - a_n}{b_n}$, con $S_n = \sum_{i=1}^n X_i$ y siendo $\{a_n\}$ y $\{b_n\}$ sucesiones de constantes tales que $\lim b_n = +\infty$. En esta sección fijaremos las condiciones para saber cuando existe convergencia a.s, y como nos ocuparemos también de la convergencia en probabilidad, las leyes se denominarán *fuerte* y *débil*, respectivamente.

Teorema 3.2 (Ley débil) *Sea* $\{X_k\}$ *una sucesión de variables aleatorias independientes tales que* $E(X_k^2) < +\infty$, $\forall k$, *y* $\lim_n \frac{1}{n^2} \sum_{k=1}^n var(X_k) = 0$, *entonces*

$$\frac{1}{n} \sum_{k=1}^{n} (X_k - E(X_k)) \xrightarrow{P} 0.$$

Demostración.- Para $S_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k))$, $E(S_n) = 0$ y $var(S_n) = \frac{1}{n^2} \sum_{k=1}^n var(X_k)$. Por la desigualdad de Chebyshev, $\forall \varepsilon > 0$

$$P(|S_n| \geq \varepsilon) \leq \frac{var(S_n)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^{n} var(X_k),$$

que al pasar al límite nos asegura la convergencia en probabilidad de S_n a 0. ♠

Corolario 3.1 *Si las* X_n *son i.i.d. con varianza finita y esperanza común* $E(X_1)$, *entonces* $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} E(X_1)$.

Demostración.- Si $var(X_k) = \sigma^2$, $\forall k$, tendremos $\frac{1}{n^2} \sum_{k=1}^n var(X_k) = \frac{\sigma^2}{n}$ que tiende a cero con n. Es por tanto de aplicación la ley débil que conduce al resultado enunciado. ♠

Este resultado fué demostrado por primera vez por J. Bernoulli para variables con distribución Binomial (véase el ejemplo 3.1), versión que se conoce como la *ley de los grandes números de Bernoulli*

El siguiente paso será fijar las condiciones para que el resultado sea válido bajo convergencia a.s.

Teorema 3.3 (Ley fuerte) *Si* $\{X_k\}$ *es una sucesión de variables aleatorias i.i.d. con media finita, entonces*

$$\sum_{k=1}^{n} \frac{X_k}{n} \xrightarrow{a.s.} E(X_1).$$

Corolario 3.2 *Si $\{X_k\}$ es una sucesión de variables aleatorias i.i.d. con $E(X_1^-) < +\infty$ y $E(X_1^+) = +\infty$, entonces $\frac{S_n}{n} \xrightarrow{a.s.} \infty$.*

La demostración de la *ley fuerte* es de una complejidad, aun en su versión más sencilla debida a Etemadi, fuera del alcance y pretensiones de este texto. La primera demostración, más compleja que la de Etemadi, se debe a Kolmogorov y es el resultado final de una cadena de propiedades previas de gran interés y utilidad en sí mismas. Aconsejamos vivamente al estudiante que encuentre ocasión de hojear ambos desarrollos en cualquiera de los textos habituales (Burrill, Billingsley,...), pero que en ningún caso olvide el desarrollo de Kolmogorov.

3.4. Teorema Central de Límite

Una aplicación inmediata es el Teorema de De Moivre-Laplace, una versión temprana del TCL, que estudia el comportamiento asintótico de una $B(n,p)$.

Teorema 3.4 (De Moivre-Laplace) *Sea $X_n \sim B(n,p)$ y definamos $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$. Entonces*

$$Z_n \xrightarrow{L} N(0,1).$$

Demostración.- Aplicando los resultados anteriores, se obtiene

$$\phi_{Z_n}(t) = \left((1-p)e^{-it\sqrt{\frac{p}{n(1-p)}}} + pe^{it\sqrt{\frac{(1-p)}{np}}} \right)^n,$$

que admite un desarrollo en serie de potencias de la forma

$$\phi_{Z_n}(t) = \left[1 - \frac{t^2}{2n}(1 + R_n) \right]^n,$$

con $R_n \to 0$, si $n \to \infty$. En consecuencia,

$$\lim_{n\to\infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}}.$$

La unicidad y el teorema de continuidad hacen el resto. ♠

Observación 3.1 *Lo que el teorema afirma es que si $X \sim B(n,p)$, para n suficientemente grande, tenemos*

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x \right) \simeq \Phi(x),$$

donde $\Phi(x)$ es la función de distribución de la $N(0,1)$.

¿De qué forma puede generalizarse este resultado? Como ya sabemos $X_n \sim B(n,p)$ es la suma de n v. a. i.i.d., todas ellas Bernoulli ($Y_k \sim B(1,p)$), cuya varianza común, $var(Y_1) = p(1-p)$, es finita. En esta dirección tiene lugar la generalización: variables independientes, con igual distribución y con varianza finita.

Teorema 3.5 (Lindeberg) *Sean $X_1, X_2, \ldots, X_n$, v.a. i.i.d. con media y varianza finitas, μ y σ^2, respectivamente. Sea $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ su media muestral, entonces*

$$Y_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{var(\overline{X}_n)}} \xrightarrow{L} N(0,1).$$

Demostración.- Teniendo en cuenta la definición de $\overline{X}_n$ podemos escribir

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k,$$

con $Z_k = (X_k - \mu)/\sigma$, variables aleatorias i.i.d. con $E(Z_1) = 0$ y $var(Z_1) = 1$. Aplicando Pϕ4 y (2.23) tendremos

$$\phi_{Y_n}(t) = \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Pero existiendo los dos primeros momentos de Z_1 y teniendo en cuenta (2.22), $\phi_{Z_1}(t)$ puede también expresarse de la forma

$$\phi_{Z_1}(t) = 1 - \frac{t^2}{2n}(1 + R_n),$$

con $R_n \to 0$, si $n \to \infty$. En consecuencia,

$$\phi_{Y_n}(t) = \left[1 - \frac{t^2}{2n}(1 + R_n)\right]^n.$$

Así pues,

$$\lim_{n\to\infty} \phi_{Y_n}(t) = e^{-\frac{t^2}{2}},$$

que es la función característica de una $N(0, 1)$. ♠

Observemos que el Teorema de De Moivre-Laplace es un caso particular del Teorema de Lindeberg, acerca de cuya importancia se invita al lector a reflexionar porque lo que en él se afirma es, ni más ni menos, que sea cual sea la distribución común a las X_i, su media muestral $\overline{X}_n$, adecuadamente tipificada, converge a una $N(0, 1)$ cuando $n \to \infty$.

El teorema de Lindeberg, que puede considerarse el teorema central del límite básico, admite una generalización en la dirección de relajar la condición de equidistribución exigida a las variables. Las llamadas condiciones de Lindeberg y Lyapunov muestran sendos resultados que permiten eliminar aquella condición.

Ejemplo 3.6 (La fórmula de Stirling para aproximar n!) *Consideremos una sucesión de variables aleatorias $X_1, X_2, \ldots,$, independientes e idénticamente distribuidas, Poisson de parámetro $\lambda = 1$. La variable $S_n = \sum_{i=1}^{n} X_i$ es también Poisson con parámetro $\lambda_n = n$. Si $Z \sim N(0, 1)$, para n suficientemente grande el TCL nos permite escribir,*

$$\begin{aligned}
P(S_n = n) &= P(n - 1 < S_n \leq n) \\
&= P\left(-\frac{1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leq 0\right) \\
&\approx P\left(-\frac{1}{\sqrt{n}} < Z \leq 0\right) \\
&= \frac{1}{2\pi} \int_{-1/\sqrt{n}}^{0} e^{-x^2/2} dx \\
&\approx \frac{1}{2\pi} \frac{1}{\sqrt{n}},
\end{aligned}$$

en donde la última expresión surge de aproximar la integral entre $[-1/\sqrt{n}, 0]$ *de* $f(x) = e^{-x^2/2}$ *mediante el área del rectángulo que tiene por base el intervalo de integración y por altura el* $f(0) = 1$.

Por otra parte,

$$P(S_n = n) = e^{-n}\frac{n^n}{n!}.$$

Igualando ambos resultado y despejando $n!$ *se obtiene la llamada fórmula de Stirling,*

$$n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}.$$

3.4.1. Una curiosa aplicación del TCL: estimación del valor de π

De Moivre y Laplace dieron en primer lugar una *versión local* del TCL al demostrar que si $X \sim B(n, p)$,

$$P(X = m)\sqrt{np(1-p)} \approx \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \tag{3.1}$$

para n suficientemente grande y $x = \frac{m-np}{\sqrt{np(1-p)}}$. Esta aproximación nos va a servir para estudiar la credibilidad de algunas aproximaciones al número π obtenidas a partir del problema de la *aguja de Buffon*.

Recordemos que en el problema planteado por Buffon se pretende calcular la probabilidad de que una aguja de longitud $2l$, lanzada al azar sobre una trama de paralelas separadas entre si una distancia $2a$, con $a > l$, corte a alguna de las paralelas. Puestos de acuerdo sobre el significado de *lanzada al azar*, la respuesta es

$$P(corte) = \frac{2l}{a\pi},$$

resultado que permite obtener una aproximación de π si, conocidos a y l, sustituimos en $\pi = \frac{2l}{aP(corte)}$ la probabilidad de corte por su estimador natural *la frecuencia relativa de corte*, p, a lo largo de n lanzamientos. Podremos escribir, si en lugar de trabajar con π lo hacemos con su inverso,

$$\frac{1}{\pi} = \frac{am}{2ln},$$

donde m es el número de cortes en los n lanzamientos.

El año 1901 Lazzarini realizó 3408 lanzamientos obteniendo para π el valor 3,1415929 con ¡¡6 cifras decimales exactas!!. La aproximación es tan buena que merece como mínimo alguna pequeña reflexión. Para empezar supongamos que el número de cortes aumenta en una unidad, las aproximaciones de los inversos de π correspondientes a los m y $m+1$ cortes diferirían en

$$\frac{a(m+1)}{2ln} - \frac{am}{2ln} = \frac{a}{2ln} \geq \frac{1}{2n},$$

que si $n \approx 5000$, da lugar a $\frac{1}{2n} \approx 10^{-4}$. Es decir, un corte más produce una diferencia mayor que la precisión de 10^{-6} alcanzada. *No queda más alternativa que reconocer que Lazzarini tuvo la suerte de obtener* exactamente *el número de cortes, m, que conducía a tan excelente aproximación*. La pregunta inmediata es, *cual es la probabilidad de que ello ocurriera?*, y para responderla podemos recurrir a (3.1) de la siguiente forma,

$$P(X = m) \approx \frac{1}{\sqrt{2\pi np(1-p)}}e^{-\frac{(m-np)^2}{2np(1-p)}} \leq \frac{1}{\sqrt{2\pi np(1-p)}},$$

que suponiendo $a = 2l$ y $p = 1/\pi$ nos da para $P(X = m)$ la siguiente cota

$$P(X = m) \leq \sqrt{\frac{\pi}{2n(\pi - 1)}}.$$

Para el caso de Lazzarini n=3408 y $P(X = m) \leq 0{,}0146,\ \forall m$. *Parece ser que Lazzarini era un hombre de suerte, quizás demasiada.*

Capítulo 4

Procesos Estocásticos

4.1. Introducción

Los temas precedentes nos han dotado de las herramientas necesarias para conocer el comportamiento de una variable aleatoria o de un conjunto finito de ellas, un vector aleatorio. En cualquier caso, se trataba de un número finito de variables aleatorias. El Capítulo 3 nos ha permitido, además, estudiar el comportamiento asintótico de una sucesión de variables aleatorias a través de las llamadas Leyes de los Grandes Números y del Teorema Central del Límite. Pero ya advertíamos en la Introducción de dicho capítulo que de su comportamiento conjunto se ocupaba una parte específica de la Teoría de la Probabilidad, los Procesos Estocásticos.

Antes de introducir las definiciones básicas, quizás convenga puntualizar que el concepto de proceso estocástico abarca situaciones más generales que las sucesiones de variables aleatorias. Se trata de estudiar el comportamiento de un conjunto de variables aleatorias que puede estar indexado por un conjunto no numerables. A los ejemplos que se citaban en la mencionada Introducción del Capítulo 3, todos ellos sucesiones aleatorias, podemos añadir otros fenómenos aleatorios que generan familias aleatorias no numerables.

Un ejemplo sería el estudio de la ocurrencia de un mismo suceso a lo largo del tiempo, si dichas ocurrencias son independientes y nos ocupamos del intervalo de tiempo que transcurre entre una y otra, nuestro interés se centra en la sucesión $\{X_n\}_{n\geq 1}$, donde $X_i = \{$*tiempo transcurrido entre las ocurrencias* i-ésima *y la* i-1-ésima$\}$. Si lo que nos interesa es el número de veces que el suceso ha ocurrido en el intervalo $[0,t]$, la familia a considerar es N_t, con $t>0$, que desde luego no es numerable.

4.2. Definiciones básicas y descripción de un proceso estocástico

Definición 4.1 *Un proceso estocástico es una colección de variables aleatorias* $\{X_t, t\in T\}$ *definidas sobre un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$.

Obsérvese que el índice y el conjunto de índices se denotan mediante las letras t y T, respectivamente. La razón para ello es que en su origen los procesos estocásticos surgen del estudio de la evolución temporal de fenómenos aleatorios. Ello no presupone nada respecto a la numerabilidad de T.

Una primera clasificación de los procesos estocásticos toma como criterios el tipo de variables involucradas y la dimensión del índice. De acuerdo con ellos se pueden establecer cuatro tipos

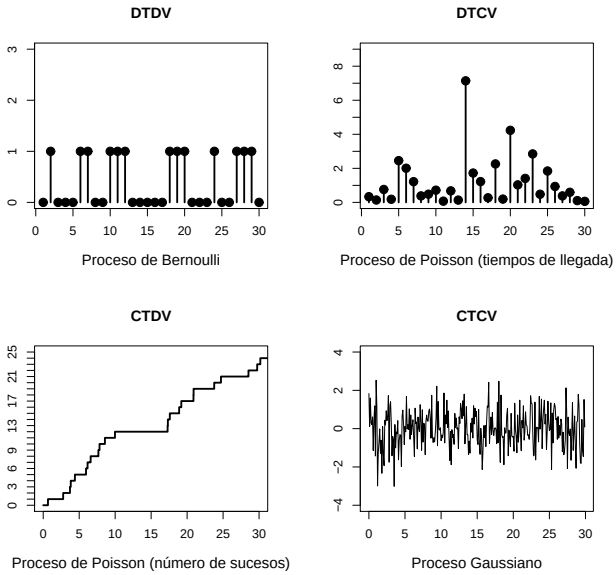


Figura 4.1: Ejemplos de los diferentes tipos de procesos estocásticos

de procesos,

- DTDV, acrónimo del inglés *Discret Time / Discret Values*, procesos con índice numerable y variables discretas.
- DTCV, acrónimo del inglés *Discret Time / Continuous Values*, procesos con índice numerable y variables continuas.
- CTDV, acrónimo del inglés *Continuous Time / Discret Values*, procesos con índice no numerable y variables discretas.
- CTCV, acrónimo del inglés *Continuous Time / Continuous Values*, procesos con índice no numerable y variables continuas.

La Figura 4.1 muestra la gráfica de un proceso de cada uno de los cuatro tipos.

4.2.1. Trayectoria de un proceso

Un proceso estocástico puede también ser visto como una función aleatoria con un doble argumento, $\{X(t,\omega), t \in T, \omega \in \Omega\}$. Si con esta notación fijamos $\omega = \omega_0$, tendremos una *realización* del proceso, $X(\cdot, \omega_0)$, cuya representación gráfica constituye lo que denominamos la *trayectoria del proceso* (*sample path*). La Figura 4.2 muestra cuatro trayectorias de un proceso de Poisson.

Por el contrario, si lo que fijamos es $t = t_0$, estamos refiriéndonos a la variable aleatoria $X_{t_0} = X(t_0, \cdot)$. La líneas verticales que aparecen en la Figura 4.2 representan a las variables aleatorias N_{20} y N_{25}, su intersección con las cuatro trayectorias del proceso son las valores que dichas variables han tomado en cada realización.

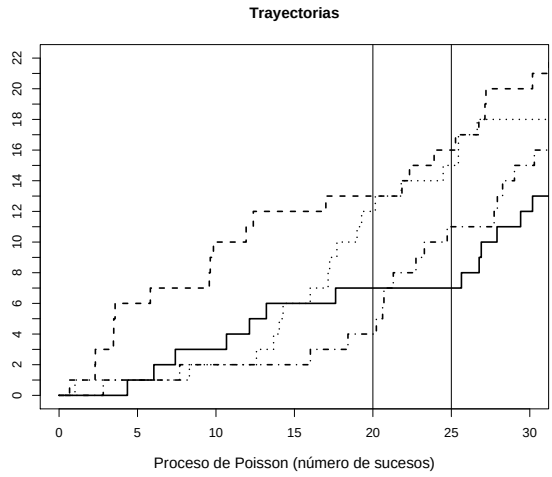


Figura 4.2: Trayectorias de un proceso de Poisson y realizaciones de las variables N_{20} y N_{25}

4.2.2. Distribuciones finito-dimensionales

Un proceso estocástico se describe a partir de las distribuciones de probabilidad que induce sobre $\mathcal{R}^n$. Si $\{t_1, t_2, \ldots, t_n\} \subset T$ es un subconjunto finito cualquiera de índices, la distribución conjunta del vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ vendrá dada por

$$F_{t_1,t_2,\ldots,t_n}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = P(X_{t_1} \leq x_{t_1}, \ldots, X_{t_n} \leq x_{t_n}), \tag{4.1}$$

que recibe el nombre de *distribución finito-dimensional* del proceso. Estas distribuciones pueden igualmente venir dadas en términos de la funciones de densidad o probabilidad conjuntas del vector.

Un proceso estocástico se describe especificando sus distribuciones finito-dimensionales, que permiten obtener la probabilidad de cualquier suceso involucrado en el proceso. Hay que advertir, no obstante, que el conjunto de distribuciones finito-dimensionales no determina por completo las características del proceso[1], por lo que en ocasiones hay que definir ciertas condiciones o propiedades adicionales. Si el proceso se puede especificar completamente a partir de dichas distribuciones, decimos que el proceso es *separable*.

Dos de estas propiedades son las que conocen como *incrementos independientes* y *de Markov*. Veamos en qué consisten.

Definición 4.2 (Incrementos independientes) *Se dice que un proceso estocásticos tiene sus incrementos independientes si para $t_1 < t_2 < \ldots < t_n$, las variables*

$$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}, \tag{4.2}$$

son independientes.

Definición 4.3 (Markov) *Se dice que un proceso estocásticos es un proceso de Markov si la evolución del proceso depende sólo del pasado inmediato, es decir, dados $t_1 < t_2 < \ldots < t_n$*

$$P(X_{t_n} \in B | X_{t_1}, X_{t_2}, \ldots, X_{t_{n-1}}) = P(X_{t_n} \in B | X_{t_{n-1}}), \tag{4.3}$$

[1] En la página 319 del libro de Billingsley, *Probability and Measure, 2nd Ed.*, hay un ejemplo relativo al proceso de Poisson

donde B es cualquier conjunto de Borel (suceso) en $\mathcal{R}$.

La condición (4.3) puede igualmente expresarse en términos de las funciones de densidad o de probabilidad, según sea la naturaleza de las variables del proceso,

$$f_{t_n|t_1,\ldots,t_{n-1}}(x_{t_n}|x_{t_1},\ldots,x_{t_{n-1}}) = f_{t_n|t_{n-1}}(x_{t_n}|x_{t_{n-1}}).$$

En el caso discreto las probabilidades $P(X_{t_n} = x_n|X_{t_{n-1}} = x_{n-1})$ se denominan probabilidades de transición, cuyo conocimiento caracteriza completamente el proceso de Markov.

Ambas propiedades están relacionadas. En efecto, los incrementos independientes implican la propiedad de Markov, pero el recíproco no es cierto. En el caso discreto la implicación demuestra fácilmente,

$$\begin{aligned} P(X_{t_n} = x_n|X_{t_1} = x_1,\ldots,X_{t_{n-1}} = x_{n-1}) & \\ = \; P(X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1}|X_{t_1} = x_1,\ldots,X_{t_{n-1}} = x_{n-1}) & \quad (4.4)\\ = \; P(X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1}|X_{t_{n-1}} = x_{n-1}) & \quad (4.5)\\ = \; P(X_{t_n} = x_n|X_{t_{n-1}} = x_{n-1}), & \quad (4.6) \end{aligned}$$

donde el paso de (4.4) a (4.5) es consecuencia de la independencia de los incrementos.

4.2.3. Funciones de momento

Las llamadas funciones de momento, obtenidas a partir de los momentos de las variables involucradas en un proceso estocástico, juegan un papel muy importante a la hora de conocer su comportamiento y en las aplicaciones del mismo. Las más relevantes son las siguientes:

Función media.- Se define como

$$\mu_X(t) = E[X_t], \quad t \in T. \tag{4.7}$$

Para su obtención tendremos en cuenta el tipo de variables que conforman el proceso. En el caso discreto,

$$\mu_X(t) = \sum_{x \in D_{X_t}} xP(X_t = x),$$

donde D_{X_t} es el soporte de X_t.

En el caso continuo,

$$\mu(t) = \int_{-\infty}^{\infty} x f_t(x)dx.$$

Función de autocorrelación.- Se define a partir del momento conjunto de dos variables asociadas a dos tiempos cualesquiera, t_1 y t_2,

$$R(t_1, t_2) = E[X_{t_1}X_{t_2}]. \tag{4.8}$$

Para el caso discreto (4.8) se obtiene mediante

$$R(t_1, t_2) = \sum_{x_1 \in D_{X_{t_1}}, x_2 \in D_{X_{t_2}}} x_1 x_2 P(X_{t_1} = x_1, X_{t_2} = x_2).$$

En el caso continuo

$$R(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{t_1 t_2}(x_1, x_2)dx_1 dx_2.$$

Función de autocovarianza.- Se define a partir del momento central conjunto de dos variables asociadas a dos tiempos cualesquiera, t_1 y t_2,

$$C(t_1, t_2) = E[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))], \tag{4.9}$$

con sus correspondientes versiones discreta y continua. Se deduce fácilmente la siguiente relación entre $R(t_1, t_2)$ y $R(t_1, t_2)$,

$$C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2).$$

El caso particular $t_1 = t_2 = t$ recibe el nombre de *función varianza*, $\sigma^2(t) = C(t, t)$. Por último, la *función de correlación* se obtiene mediante

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{\sigma^2(t_1)\sigma^2(t_2)}},$$

que como es bien sabido verifica $|\rho(t_1, t_2)| \leq 1$. Algunos autores, particularmente en el campo de las Series Temporales, reservan el nombre de *función de autocorrelación* para esta última función.

Hay que advertir que todas las definiciones anteriores tiene sentido si las correspondientes integrales o series son finitas.

Ejemplo 4.1 *Consideremos el siguiente proceso estocástico del tipo* CTCV*,*

$$X_t = A \sin(\omega_0 t + \Theta), t \in \mathcal{R},$$

donde A y Θ son variables aleatorias independientes, con $\Theta \sim U(-\pi, \pi)$. Obtengamos sus momentos.

$$\begin{aligned} \mu(t) &= E[A \sin(\omega_0 t + \Theta)] \\ &= E(A)E[\sin(\omega_0 t + \Theta)] \\ &= \mu_A \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(\omega_0 t + \theta) d\theta \\ &= \mu_A \cdot 0 = 0. \end{aligned}$$

La autocorrelación vale

$$\begin{aligned} R(t_1, t_2) &= E[X_{t_1} X_{t_2}] \\ &= E[A^2 \sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)] \\ &= E(A^2)E[\sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)] \qquad (4.10) \\ &= E(A^2) \cdot \frac{1}{2} \{E[\cos \omega_0(t_1 - t_2)] - E[\cos(\omega_0(t_1 + t_2) + 2\Theta)]\} \qquad (4.11) \\ &= \frac{1}{2} E(A^2) \cos \omega_0(t_1 - t_2). \end{aligned}$$

Para pasar de (4.10) a (4.11) hemos aplicado

$$\sin \alpha \sin \beta = \frac{\cos(\alpha - \beta) - \cos(\alpha + \beta)}{2}.$$

Obsérvese que $\mu(t) = 0$ es constante y que $R(t_1, t_2)$ depende de $t_1 - t_2$, en realidad de $|t_1 - t_2|$ porque $\cos(\alpha) = \cos(-\alpha)$. Como más tarde veremos, un proceso de estas características se dice que es estacionario en sentido amplio (wide-sense stationary, WSS*).*

4.3. Algunos procesos estocásticos de interés

4.3.1. Procesos IID

Se trata de procesos constituidos por variables aleatorias independientes e idénticamente distribuidas (IID). Como consecuencia de ello, las distribuciones finito-dimensionales se obtienen fácilmente a partir de la distribución común. Si $F(x)$ denota la función de distribución común a todas las X_t del proceso,

$$F_{t_1,t_2,\ldots,t_n}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = P(X_{t_1} \leq x_{t_1}, \ldots, X_{t_n} \leq x_{t_n}) = F(x_{t_1})F(x_{t_2}) \ldots F(x_{t_n}).$$

Las funciones de momento tiene también expresiones sencillas. Así, la función media vale

$$\mu(t) = E(X_t) = \mu,$$

donde μ es la esperanza común a todas las X_t.

La función de autocovarianza vale 0 porque,

$$C(t_1, t_2) = E[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))] = E[(X_{t_1} - \mu(t_1))]E[(X_{t_2} - \mu(t_2))] = 0,$$

y para $t_1 = t_2 = t$ da lugar a la varianza común, $\sigma^2(t) = \sigma^2$. Es decir,

$$C(t_1, t_2) = \begin{cases} 0, & \text{si } t_1 \neq t_2; \\ \sigma^2, & \text{si } t_1 = t_2. \end{cases}$$

De la misma forma, la función de autocorrelación es

$$R(t_1, t_2) = \begin{cases} \mu^2, & \text{si } t_1 \neq t_2; \\ \sigma^2 + \mu^2, & \text{si } t_1 = t_2. \end{cases}$$

De entre los procesos estocásticos IID existen algunos de especial interés.

Proceso de Bernoulli.- Se trata de una sucesión ($T = N$) de variables Bernoulli independientes, $X_n \sim B(1, p)$. La sucesión de lanzamientos de una moneda (*cara=1, cruz=0*) da lugar a un proceso de Bernoulli. La media y la varianza del proceso valen

$$\mu(n) = p, \qquad \sigma^2(n) = p(1-p).$$

La obtención de cualquier probabilidad ligada al proceso es sencilla. Por ejemplo, la probabilidad que los cuatro primeros lanzamientos sean {C++C}={1001} es

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) &= P(X_1 = 0)P(X_2 = 1)P(X_3 = 1)P(X_4 = 0) \\ &= p^2(1-p)^2. \end{aligned}$$

Procesos suma IID

La suma de procesos estocásticos es una forma de obtener nuevos procesos de interés. A partir de una sucesión ($T = N$) de variables IID definimos el *proceso suma* de la forma

$$S_n = X_1 + X_2 + \ldots + X_n, \quad n \geq 1,$$

o también

$$S_n = S_{n-1} + X_n,$$

con $S_0 = 0$.

Si el proceso original es un proceso IID, el proceso suma tiene *incrementos independientes* para intervalos de tiempo no solapados. En efecto, para el caso en que las X_k son discretas, si $n_0 < n \leq n_1$ y $n_2 < n \leq n_3$, con $n_1 \leq n_2$,

$$\begin{aligned} S_{n_1} - S_{n_0} &= X_{n_0+1} + \cdots + X_{n_1} \\ S_{n_3} - S_{n_2} &= X_{n_2+1} + \cdots + X_{n_3}. \end{aligned}$$

Cada sumando está constituido por variables que son independientes de las del otro, lo que supone la independencia de los incrementos. Observemos además que para $n_2 > n_1$,

$$S_{n_2} - S_{n_1} = X_{n_1+1} + \cdots + X_{n_2} = X_1 + X_2 + \cdots X_{n_2-n_1} = S_{n_2-n_1}.$$

Por ejemplo, si las variables son discretas lo será la suma por la expresión anterior tendremos,

$$P(S_{n_2} - S_{n_1} = y) = P(S_{n_2-n_1} = y).$$

La probabilidad depende tan sólo de la longitud del intervalo y no de su posición, por lo que decimos que el proceso suma tiene los *incrementos estacionarios*. Más adelante nos ocuparemos con detalle del concepto de estacionariedad y las propiedades que de él se derivan.

Las distribuciones finito-dimensionales del proceso suma se obtienen haciendo uso de las propiedad de incrementos estacionarios independientes. Si $n_1 < n_2 < \cdots < n_k$,

$$\begin{aligned} &P(S_{n_1} = x_1, S_{n_2} = x_2, \ldots, S_{n_k} = x_k) = \\ &= P(S_{n_1} = x_1, S_{n_2} - S_{n_1} = x_2 - x_1, \ldots, S_{n_k} - S_{n_{k-1}} = x_k - x_{k-1}) \\ &= P(S_{n_1} = x_1)P(S_{n_2} - S_{n_1} = x_2 - x_1) \ldots, P(S_{n_k} - S_{n_{k-1}} = x_k - x_{k-1}) \\ &= P(S_{n_1} = x_1)P(S_{n_2-n_1} = x_2 - x_1) \ldots, P(S_{n_k-n_{k-1}} = x_k - x_{k-1}). \end{aligned} \tag{4.12}$$

Si las variable del proceso original son continuas, es fácil obtener una expresión equivalente a (4.12) para la densidad conjunta,

$$f_{S_{n_1},S_{n_2},\ldots,S_{n_k}}(x_1, x_2, \ldots, x_k) = f_{S_{n_1}}(x_1) f_{S_{n_2-n_1}}(x_2 - x_1) \cdots f_{S_{n_k-n_{k-1}}}(x_k - x_{k-1}). \tag{4.13}$$

Las funciones de momento del proceso suma se derivan fácilmente de los momentos comunes a las variables del proceso original. La media y la varianza valen,

$$\mu(n) = E(S_n) = n\mu, \qquad \sigma^2(n) = var(S_n) = n\sigma^2,$$

donde μ y σ^2 son la media y la varianza comunes a las X_k. La función de autocovarianza es

$$\begin{aligned} C(n_1, n_2) &= E[(S_{n_1} - n_1\mu)(S_{n_2} - n_2\mu) \\ &= E\left[\left\{\sum_{i=1}^{n_1}(X_i - \mu)\right\}\left\{\sum_{j=1}^{n_2}(X_j - \mu)\right\}\right] \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} E[(X_i - \mu)(X_j - \mu)] \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} cov(X_i, X_j). \end{aligned}$$

Como X_i y X_j son i.i.d.,

$$cov(X_i, X_j) = \begin{cases} \sigma^2, & \text{si } i = j; \\ 0, & \text{si } i \neq j. \end{cases}$$

En definitiva,

$$C(n_1, n_2) = \sum_{i=1}^{\text{mín}(n_1,n_2)} \sigma^2 = \text{mín}(n_1, n_2)\sigma^2. \tag{4.14}$$

Veamos tres ejemplos interesantes de procesos suma.

Camino aleatorio.- Se trata de proceso estocástico que describe el desplazamiento aleatorio de un móvil en $\mathcal{R}^n$. En el caso unidimensional, del que nos ocuparemos por más sencillo, el móvil se desplaza a saltos unitarios e independientes por la recta real, haciéndolo hacia la derecha con probabilidad p y hacia la izquierda con probabilidad $1-p$. Al cabo de n desplazamientos, la posición del móvil viene dada por $S_n = D_1 + D_2 + \cdots + D_n$, donde D_k tiene por función de probabilidad,

$$f_k(x) = \begin{cases} 1-p, & \text{si } x = -1; \\ p, & \text{si } x = 1; \\ 0, & \text{en el resto.} \end{cases}$$

Si en los n desplazamientos k de ellos lo han sido hacia la derecha y los restantes $n-k$ en sentido contrario, la posición final será $S_n = k - (n-k) = 2k - n$ por lo que su función de probabilidad valdrá,

$$f_{S_n}(2k-n) = P(S_n = 2k-n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{4.15}$$

La media y la varianza del camino aleatorio unidimensional valen,

$$\mu(n) = E(S_n) = \sum_{k=1}^{n} E(D_k) = n(2p-1),$$

y

$$\sigma_n^2 = var(S_n) = \sum_{k=1}^{n} var(D_k) = 4np(1-p).$$

la función de autocovarianza vale, aplicando (4.14),

$$C(n_1, n_2) = \text{mín}(n_1, n_2) 4p(1-p),$$

y la de autocorrelación,

$$R(n_1, n_2) = C(n_1, n_2) + \mu(n_1)\mu(n_2) = \text{mín}(n_1, n_2) 4p(1-p) + n_1 n_2 (2p-1)^2.$$

Proceso Binomial.- Si el proceso original es un proceso Bernoulli, el proceso suma resultante es el *proceso de recuento* o *Binomial*, que proporciona el número de éxitos entre las n primeras pruebas de Bernoulli. Se llama así porque, como bien sabemos, las variables del nuevo proceso son $S_n \sim B(n, p)$. Para obtener las distribuciones finito-dimensionales podemos utilizar (4.12),

$$\begin{aligned} &P(S_{n_1} = m_1, S_{n_2} = m_2, \ldots, S_{n_k} = k_k) = \\ &\quad = P(S_{n_1} = m_1) P(S_{n_2 - n_1} = m_2 - m_1) \ldots, P(S_{n_k - n_{k-1}} = m_k - m_{k-1}) \\ &\quad = \binom{n_1}{m_1} \binom{n_2 - n_1}{m_2 - m_1} \cdots \binom{n_k - n_{k-1}}{m_k - m_{k-1}} p^{m_k} (1-p)^{n_k - m_k}. \end{aligned}$$

La media del proceso Binomial vale

$$\mu(n) = E(S_n) = np,$$

que no es constante y crece linealmente con n. Su varianza es

$$\sigma^2(n) = var(S_n) = np(1-p),$$

y las funciones de autocovarianza y autocorrelación,

$$C(n_1, n_2) = \text{mín}(n_1, n_2)p(1-p), \qquad R(n_1, n_2) = \text{mín}(n_1, n_2)p(1-p) + n_1 n_2 p^2.$$

Proceso suma Gaussiano.- Se obtiene a partir de la suma de variables $X_k \sim N(0, \sigma^2)$, lo que implica que $S_n \sim N(0, n\sigma^2)$. Utilizando (4.13) obtenemos la densidad de las distribuciones finito-dimensionales

$$\begin{aligned} & f_{S_{n_1}, S_{n_2}, \ldots, S_{n_k}}(x_1, x_2, \ldots, x_k) = \\ & = \frac{1}{\sigma\sqrt{2\pi n_1}} e^{-\frac{x_1^2}{2n_1\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi(n_k - n_{k-1})}} e^{-\frac{(x_k - x_{k-1})^2}{2(n_k - n_{k-1})\sigma^2}}. \end{aligned}$$

Las funciones de momento valen,

$$\mu(n) = 0, \quad \sigma^2(n) = n\sigma^2, \quad C(n_1, n_2) = R(n_1, n_2) = \text{mín}(n_1, n_2)\sigma^2.$$

4.3.2. Ruido blanco

Un *ruido blanco* es un proceso estocástico con media cero, $\mu(t) = 0$, varianza constante, $\sigma^2(t) = \sigma^2$ y con componentes incorreladas. Como consecuencia de ello, la función de autocovarianza y autocorrelación coinciden y valen,

$$R(t_1, t_2) = C(t_1, t_2) = \begin{cases} \sigma^2, & t_1 = t_2; \\ 0, & t_1 \neq t_2. \end{cases}$$

La definición es válida tanto si se trata de una sucesión, t discreto, como si t es continuo.

Un proceso IID es, según está definición, un ruido blanco, porque la independencia de las variables implica la incorrelación. En ocasiones se da una definición más restrictiva de este tipo de procesos, al exigir que las variables sean independientes. Como veremos a continuación, sólo en el caso de que las variables sean normales o gaussianas ambas definiciones son equivalentes.

4.3.3. Proceso Gaussiano

Se trata de un proceso en el que sus variables $X_t \sim N(\mu(t), \sigma^2(t))$ y sus distribuciones finito-dimensionales son normales multivariantes,

$$f(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{4.16}$$

donde

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1(n-1)} & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2(n-1)} & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{1(n-1)} & \sigma_{2(n-1)} & \cdots & \sigma_{n-1}^2 & \sigma_{(n-1)n} \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{(n-1)n} & \sigma_n^2 \end{pmatrix},$$

son el vector de medias y la matriz de covarianzas del vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$. Esta definición es válida tanto para tiempos discretos como continuos.

Si el proceso Gaussiano es tal que $\mu_i = 0, \forall i$, $\sigma_{ij} = 0, i \neq j$ y $\sigma_i^2 = \sigma^2, \forall i$, estamos en presencia de un caso particular de ruido blanco, el *ruido blanco Gaussiano*, que dadas las particulares propiedades de la Normal multivariante, en la que incorrelación equivale a independencia, está constituido por variables independientes.

La importancia del proceso Gaussiano reside de una parte en sus propiedades, heredadas de las propiedades de las Normal. Dos de ellas convienen ser destacadas,

1. las distribuciones finito-dimensionales del proceso están completamente especificadas con los momentos de primer y segundo orden, es decir, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$,

2. la transformación lineal del proceso da lugar a un nuevo proceso Gaussiano (ver página 61).

Por otra parte, son muchos los fenómenos relacionados con señal y ruido que pueden ser modelizados con éxito utilizando un proceso Gaussiano.

4.3.4. Proceso de Poisson

Consideremos una sucesión de ocurrencias de un determinado suceso a lo largo del tiempo. Sea X_1 el tiempo de espera necesario para que el suceso ocurra por primera vez, X_2 el tiempo transcurrido entre la primera y la segunda ocurrencia, y en general, X_i el tiempo entre las ocurrencias consecutivas $i-1$ e i. El modelo formal que describe este fenómeno es una sucesión de variables aleatorias definidas sobre un determinado espacio de probabilidad.

Otra característica de interés ligada al fenómeno es el número de sucesos que han ocurrido en un determinado intervalo de tiempo $]t_1, t_2]$. Por ejemplo, para $t_1 = 0$ y $t_2 = t$, definimos la variable $N_t = \{$*número de sucesos ocurridos hasta el tiempo* t$\}$. Es del proceso estocástico que estas variables definen del que nos vamos a ocupar. Si la sucesión de tiempos de espera verifica las siguientes condiciones iniciales,

C1) no pueden ocurrir dos sucesos simultáneamente,

C2) en cada intervalo finito de tiempo ocurren a lo sumo un número finito de suceso, y

C3) los tiempos de espera son independientes e idénticamente distribuidos según una $Exp(\lambda)$,

el proceso recibe el nombre de *proceso de Poisson* y es un proceso del tipo *CTDV*. En efecto, las variables N_t son variables discretas que toman valores en $\{0, 1, 2, \ldots\}$.

Para estudiar el comportamiento probabilístico de las variables N_t es conveniente recurrir a una nueva sucesión de variables, S_n, definidas a partir de los tiempos de espera entre sucesos,

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \geq 0,$$

con $S_0 = 0$. La variable S_n representa el tiempo transcurrido hasta la llegada del *n-ésimo* suceso. Como consecuencia de **C1** y **C2**, la sucesión es estrictamente creciente,

$$0 = S_0 < S_1 < S_2 < \cdots, \qquad \sup_n S_n = \infty.$$

La Figura 4.3 muestra gráficamente la relación entre las S_n y las X_n a través de una realización de un proceso de Poisson.

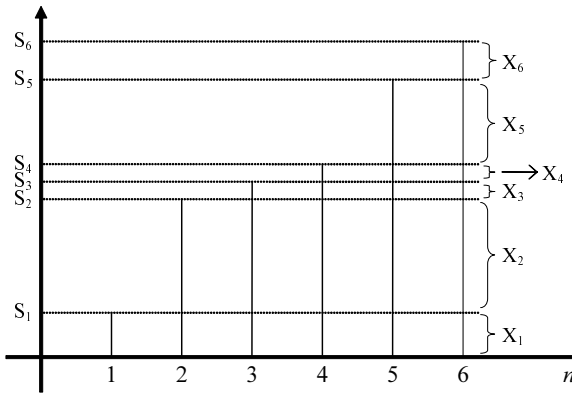


Figura 4.3: Tiempos entre sucesos, X_i, y tiempos acumulados, S_k, en un proceso de Poisson

La distribución de S_n, al ser suma de exponenciales IID con parámetro común λ, viene dada por (2.5.3) y es una $G(n, 1/\lambda)$, conocida como distribución de Erlang cuya densidad es,

$$g_n(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0,$$

y cuya función de distribución es,

$$G_n(t) = \sum_{k \geq n} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = 1 - e^{-\lambda t} \sum_{k \geq 0}^{n-1} \frac{(\lambda t)^k}{k!}, \tag{4.17}$$

como fácilmente se comprueba derivando.

El número N_t de sucesos que han ocurrido en el intervalo $]0, t]$ es el mayor n tal que $S_n \leq t$,

$$N_t = \text{máx}\{n; S_n \leq t\}.$$

Teniendo en cuenta esta relación es fácil deducir que

$$\{N_t \geq n\} = \{S_n \leq t\}, \tag{4.18}$$

y combinando (4.18) y (4.17),

$$P(N_t \geq n) = G_n(t) = \sum_{k \geq n} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Es ahora fácil obtener

$$P(N_t = n) = e^{\lambda t} \frac{(\lambda t)^n}{n!}. \tag{4.19}$$

Es decir, $N_t \sim Po(\lambda t)$, lo que justifica el nombre dado al proceso. Este resultado nos permite además darle significado al parámetro λ porque, si recordamos que $E(N_t) = \lambda t$, se deduce de aquí que λ es *el número de suceso que ocurren por unidad de tiempo*, una característica específica del fenómeno aleatorio que da lugar al proceso. Estudiemos algunas propiedades del proceso de Poisson.

Incrementos independientes y estacionarios

La sucesión de tiempos acumulados es un proceso suma obtenido a partir de variables exponenciales IID y como tal, tiene incrementos independientes y estacionarios. Dada la relación (4.18) que liga las N_t y las S_n, es lógico pensar que también el proceso de Poisson goce de la misma propiedad. La implicación no es inmediata si tenemos en cuenta que al considerar los tiempos de espera sobre un intervalo cualquiera $]t, t+s]$, el primero de ellos puede estar contenido sólo parcialmente en él, tal como se muestra en la Figura 4.4. En ella observamos que $N_{t+s} - N_t = m$ y que parte del tiempo transcurrido entre el suceso N_t y el N_{t+1}, X_{N_t+1}, está en $(t, t+s]$, en concreto la que hemos denotado mediante Y_{N_t+1}. La comprobación de la propiedad no es sencilla, razón por la cual no la presentamos en estas notas. El lector interesado puede consultarla en las páginas 310-311 del texto de Billingsley (1995) donde se demuestra también que $Y_{N_t+1} \sim Exp(\lambda)$

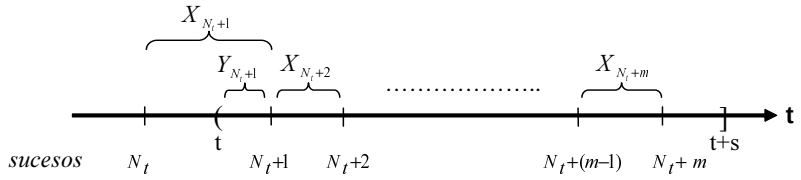


Figura 4.4: Relación entre los tiempos de espera y un intervalo arbitario $(t, t+s]$ en un proceso de Poisson

Más sencillo resulta comprobar la estacionariedad de los incrementos. En efecto, dada la independencia entre las X_i, como Y_{N_t+1} depende exclusivamente de X_{N_t+1} y de S_{N_t},

$$X_{N_t+1} - Y_{N_t+1} = t - S_{N_t},$$

Y_{N_t+1} será también independiente de todos los tiempos de espera posteriores, $X_{N_t+k}, k \geq 2$. La consecuencia inmediata es que la variable $N_{t+s} - N_t$ está relacionada con la suma de variables exponenciales IID,

$$Y_{N_t+1} + \sum_{k=2}^{m} X_{N_t+k},$$

de forma semejante a como N_t lo está con las X_i. A efectos prácticos y gracias a la propiedad de falta de memoria es como si desplazáramos el origen de tiempos a t. Así pues

$$P(N_{t+s} - N_t = m) = P(N_s = m) = e^{-\lambda s} \frac{(\lambda s)^m}{m!}, \tag{4.20}$$

y podemos escribir, en general, que si $t_2 > t_1$, $N_{t_2} - N_{t_1} \sim Po(\lambda(t_2 - t_1))$, que depende sólo del incremento de tiempos y no de su posición. Son pues estacionarios.

Distribuciones finito-dimensionales y funciones de momento

La independencia y estacionariedad de los incrementos permite obtener fácilmente las distribuciones finito-dimensionales. Para el caso de dos tiempos cualesquiera, $t_2 > t_1$,

$$\begin{aligned} P(N_{t_1} = n_1, N_{t_2} = n_2) &= P(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2 - n_1) \\ &= P(N_{t_1} = n_1) P(N_{t_2} - N_{t_1} = n_2 - n_1) \\ &= e^{-\lambda t_1} \frac{(\lambda t_1)^{n_1}}{n_1!} e^{-\lambda(t_2 - t_1)} \frac{(\lambda(t_2 - t_1))^{n_2 - n_1}}{(n_2 - n_1)!} \\ &= e^{-\lambda t_2} \frac{\lambda^{n_2} t_1^{n_1} (t_2 - t_1)^{n_2 - n_1}}{n_1!(n_2 - n_1)!}. \end{aligned}$$

Por inducción obtendríamos la expresión para cualquier número finito de tiempos.

Por lo que respecta a las funciones de momento,

Media del proceso.- Como ya vimos anteriormente

$$\mu(t) = \lambda t. \tag{4.21}$$

Función de autocorrelación.- Para dos tiempos cualesquiera $t_2 > t_1$, haciendo uso de la estacionareidad e independencia de los incrementos,

$$\begin{aligned} E[N_{t_1} N_{t_2}] &= E[(N_{t_1} + (N_{t_2} - N_{t_1}))N_{t_1}] \\ &= E[N_{t_1}^2] + E[N_{t_2} - N_{t_1}]E[N_{t_1}] \\ &= \lambda t_1 + \lambda^2 t_1^2 + \lambda(t_2 - t_1)\lambda t_1 \\ &= \lambda t_1 + \lambda^2 t_1 t_2. \end{aligned}$$

Si $t_1 > t_2$ intercambiaríamos t_1 y t_2 en la expresión anterior. En consecuencia, la función de autocorrelación vale

$$R(t_1, t_2) = E[N_{t_1} N_{t_2}] = \lambda \min(t_1, t_2) + \lambda^2 t_1 t_2. \tag{4.22}$$

Función de autocovarianza.- De la relación entre $R(t_1, t_2)$ y $C(t_1, t_2)$ se deduce

$$C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2) = \lambda \min(t_1, t_2). \tag{4.23}$$

Para $t_1 = t_2 = t$ obtendremos la **varianza** del proceso que, como en el caso de la variable Poisson, coincide con la media

$$\sigma^2(t) = \lambda t. \tag{4.24}$$

Proceso de Poisson y llegadas al azar

El proceso de Poisson surge de la necesidad de modelizar fenómenos aleatorios consistentes en las ocurrencias sucesivas a lo largo del tiempo de un determinado suceso. Las partículas radiactivas que llegan a un contador Geiger son un buen ejemplo. En este tipo de fenómenos es costumbre hablar de *"llegadas al azar"* del suceso. Para entender el significado de la expresión y al mismo tiempo justificarla, supongamos que observamos el proceso en un intervalo $]0, t]$ a través de su variable asociada, N_t, y que el número de llegadas ha sido n, $N_t = n$. Consideremos un subintervalo cualquiera de $]0, t]$, sin pérdida de generalidad podemos hacer coincidir los

orígenes de ambos intervalos, $]0,t_1] \subset [0,t]$, y vamos obtener la probabilidad de que k de los n sucesos hayan ocurrido en $]0,t_1]$. Haciendo uso de las propiedades del proceso,

$$\begin{aligned} P(N_{t_1}=k|N_t=n) &= \frac{P(N_{t_1}=k, N_t=n)}{P(N_t=n)} \\ &= \frac{P(N_{t_1}=k, N_{t-t_1}=n-k)}{P(N_t=n)} \\ &= \frac{[e^{-\lambda t_1}(\lambda t_1)^k/k!][e^{-\lambda(t-t_1)}(\lambda(t-t_1))^{n-k}/(n-k)!]}{e^{-\lambda t}(\lambda t)^n/n!} \\ &= \frac{n!}{k!(n-k)!} \times \frac{e^{-\lambda t}\lambda^n t_1^k (t-t_1)^{n-k}}{e^{-\lambda t}\lambda^n t^n} \\ &= \binom{n}{k}\left(\frac{t_1}{t}\right)^k \left(\frac{t-t_1}{t}\right)^{n-k}. \end{aligned}$$

Es decir, $N_{t_1}|N_t = n \sim B(n,p)$ con $p = t_1/t$, lo que significa que la probabilidad de que cualquiera de los n sucesos ocurra en $]0,t_1]$ es proporcional a la longitud del subintervalo o, equivalentemente, los n sucesos se distribuyen uniformemente, al azar, en el intervalo $]0,t]$.

Este resultado tiene una aplicación inmediata en la simulación de un proceso de Poisson de parámetro λ. En efecto, un sencillo algoritmo para simular las llegadas en un intervalo $]0,t]$ es

1. Generar un valor de una variable $Po(\lambda t)$, lo que nos dará el número de llegadas.
2. Si el valor anteriormente simulado es n_0, generar n_0 valores de una $U(0,t)$, que determinarán los tiempos de llegada de los n_0 sucesos.

Las funciones para genera valores de variables Poisson y Uniformes están disponibles en cualquier ordenador.

Derivación alternativa del proceso de Poisson

Existe una forma alternativa de obtener el proceso de Poisson basada en resultados elementales de Teoría de la Probabilidad y estableciendo condiciones iniciales para el fenómeno. Con la notación habitual, estas condiciones son:

CA1) si $t_1 < t_2 < t_3$, los sucesos $\{N_{t_2-t_1}=n\}$ y $\{N_{t_3-t_2}=m\}$ son independientes, para cualesquiera valores no negativos de n y m,

CA2) los sucesos $\{N_{t_2-t_1}=n\}$, $n=0,1,\ldots$, constituyen una partición del espacio muestral y $P(N_{t_2-t_1}=n)$ depende sólo de la diferencia t_2-t_1,

CA3) si t es suficientemente pequeño, entonces $P(N_t \geq 2)$ es despreciablemente pequeña comparada con $P(N_t=1)$, es decir

$$\lim_{t\downarrow 0} \frac{P(N_t \geq 2)}{P(N_t=1)} = \lim_{t\downarrow 0} \frac{1-P(N_t=0)-P(N_t=1)}{P(N_t=1)} = 0,$$

lo que equivale a

$$\lim_{t\downarrow 0} \frac{1-P(N_t=0)}{P(N_t=1)} = 1.$$

El desarrollo de esta alternativa, del que no nos ocuparemos en estas notas, es intuitivo, interesante y sencillo. Aconsejamos al lector de estas notas consultarlo en Stark y Woods (2002), Gnedenko (1979) o en el material complementario Montes (2007).

Generalización del proceso de Poisson

El proceso de Poisson admite generalizaciones en varios sentidos. La más inmediata es aquella que suprime la estacionariedad admitiendo que su intensidad λ, número de ocurrencias por unidad de tiempo, dependa del tiempo, $\lambda(t) > 0$, $\forall t \geq 0$. Tenemos entonces un proceso de Poisson *no uniforme* o *no homogéneo*, en el que su media vale

$$\mu(t) = \int_0^t \lambda(x)dx, \quad \forall t \geq 0.$$

4.3.5. Señal telegráfica aleatoria (RTS)

El proceso conocido como RTS, Random Telegraph Signal en inglés, es un proceso relacionado con el proceso de Poisson. Se trata de un proceso $CTDV$ en el que la variables X_t toman sólo dos valores simétricos, $\{-a, a\}$, de acuerdo con el siguiente esquema.

1. $X_0 = \pm a$ con igual probabilidad $p = 1/2$,

2. X_t cambia de signo con cada ocurrencia de un suceso en un proceso de Poisson de parámetro λ.

Un proceso de estas características surge cuando toda la información de interés en una señal aleatoria está contenida en los puntos donde se produce un cambio de signo, los puntos donde se cruza el eje de las $X's$. Como las variaciones de amplitud carecen e interés, el esquema anterior proporciona un modelo sencillo para el estudio de este tipo de fenómenos.

La función de probabilidad de X_t depende del valor N_t, más concretamente de su paridad. En efecto,

$$P(X_t = \pm a) = P(X_t = \pm a|X_0 = a)P(X_0 = a) + P(X_t = \pm a|X_0 = -a)P(X_0 = -a). \quad (4.25)$$

De la definición del proceso se deduce que X_t y X_0 tomarán el mismo valor si $N_t =$ par, en caso contrario tomarán valores opuestos. Así,

$$\begin{aligned} P(X_t = \pm a|X_0 = \pm a) &= P(N_t = \text{par}) \\ &= \sum_{n\geq 0} e^{-\lambda t}\frac{(\lambda t)^{2n}}{(2n)!}. \end{aligned} \quad (4.26)$$

Para sumar la serie observemos que

$$e^{\lambda t} + e^{-\lambda t} = \sum_{k\geq 0} \frac{(\lambda t)^k + (-\lambda t)^k}{k!} = 2\sum_{n\geq 0} \frac{(\lambda t)^{2n}}{(2n)!}.$$

Sustituyendo en (4.26)

$$P(X_t = \pm a|X_0 = \pm a) = e^{-\lambda t}\frac{e^{\lambda t} + e^{-\lambda t}}{2} = \frac{1}{2}(1 + e^{-2\lambda t}). \quad (4.27)$$

Análogamente,

$$P(X_t = \pm a|X_0 = \mp a) = e^{-\lambda t}\frac{e^{\lambda t} - e^{-\lambda t}}{2} = \frac{1}{2}(1 - e^{-2\lambda t}). \quad (4.28)$$

Sustituyendo en (4.25)

$$\begin{aligned} P(X_t = +a) &= \frac{1}{2}\left[\frac{1}{2}(1+e^{-2\lambda t}) + \frac{1}{2}(1-e^{-2\lambda t})\right] = \frac{1}{2}, \\ P(X_t = -a) &= \frac{1}{2}\left[\frac{1}{2}(1-e^{-2\lambda t}) + \frac{1}{2}(1+e^{-2\lambda t})\right] = \frac{1}{2}, \end{aligned}$$

lo que no dice que el proceso RTS toma cualquiera de los dos valores con igual probabilidad en cualquier instante de tiempo.

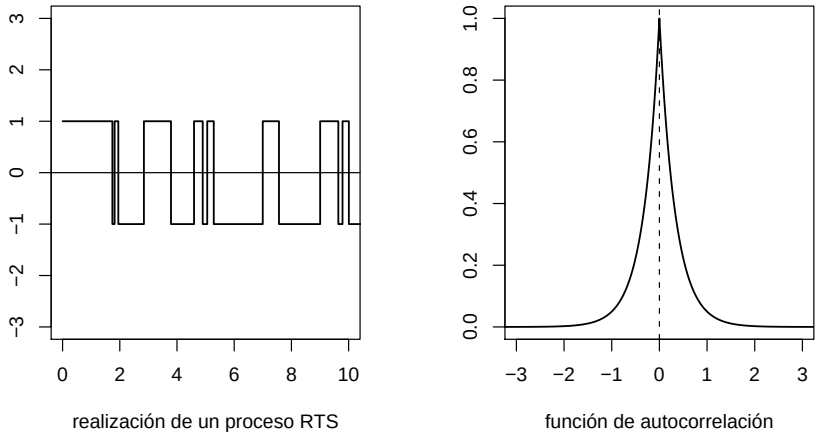


Figura 4.5: Realización de un proceso RTS con $a = 1$ y $\lambda = 1{,}5$ y su correspondiente función de autocorrelación

Distribuciones finito-dimensionales y funciones de momento

Para dos tiempos cualesquiera t y s, si $t > s$,

$$f_{st}(x,y) = P(X_s = x, X_t = y) = P(X_s = x | X_t = y) P(X_t = x),$$

si $s > t$

$$f_{st}(x,y) = P(X_s = x, X_t = y) = P(X_t = y | X_s = x) P(X_s = y).$$

En cualquiera de los dos casos lo que determina el valor es la paridad de $N_{|t-s|}$,

$$f_{st}(x,y) = \begin{cases} \frac{1}{2}P(N_{|t-s|} = \text{par}) = \frac{1}{2} \times \frac{1}{2}\left(1 + e^{-2\lambda|t-s|}\right), & \text{si } x = y; \\ \frac{1}{2}P(N_{|t-s|} = \text{impar}) = \frac{1}{2} \times \frac{1}{2}\left(1 - e^{-2\lambda|t-s|}\right), & \text{si } x \neq y. \end{cases} \tag{4.29}$$

La media del proceso RTS es nula dada la simetría de las variables X_t, $\mu(t) = 0$. La varianza vale

$$\sigma^2(t) = E(X_t^2) = a^2 \times \frac{1}{2} + (-a)^2 \times \frac{1}{2} = a^2.$$

Las funciones de autocorrelación y autocovarianza coinciden por ser la media nula,

$$\begin{aligned} C(t_1,t_2) &= E(X_{t_1}X_{t_2}) \\ &= a^2P(X_{t_1}=X_{t_2}) - a^2P(X_{t_1}\neq X_{t_2}), \end{aligned}$$

probabilidades que se obtienen a partir de (4.29),

$$\begin{aligned} C(t_1,t_2) &= a^2P(X_{t_1}=X_{t_2}) - a^2P(X_{t_1}\neq X_{t_2}) \\ &= \frac{a^2}{2}\big(1+e^{-2\lambda|t_1-t_2|}\big) - \frac{a^2}{2}\big(1-e^{-2\lambda|t_1-t_2|}\big) \\ &= a^2e^{-2\lambda|t_1-t_2|}. \end{aligned} \tag{4.30}$$

La función se amortigua a medida que aumenta la diferencia entre ambos tiempos, disminuyendo así la correlación entre ambas variables. En la Figura 4.5 se muestran las gráficas de una realización de un proceso RTS y de su función de autocorrelación, en la que se aprecia el efecto de amortiguación mencionado.

4.3.6. Modulación por desplazamiento de fase (PSK)

La transmisión de datos binarios a través de ciertos medios, la linea telefónica sería un ejemplo, exige una modulación de dichos datos. La modulación por desplazamiento de fase (PSK del inglés *Phase-Shift Keying*) es un método básico que consiste transformar los datos, una sucesión de variables aleatorias independientes todas ellas Bernoulli uniformes, $B_n \sim B(1,1/2)$, en una sucesión de ángulo-fase Θ_n que se utiliza para modular el $\cos(2\pi f_0 t)$ de la portadora de la señal. Para ello definimos,

$$\Theta_n = \begin{cases} +\pi/2, & \text{si } B_n = 1; \\ -\pi/2, & \text{si } B_n = 0, \end{cases} \tag{4.31}$$

y el ángulo del proceso mediante

$$\Theta_t = \Theta_k, \quad \text{para } kT \leq t < (k+1)T, \tag{4.32}$$

donde T es una constante que denota el tiempo de transmisión de un bit, que suele elegirse como múltiplo de $1/f_0$ para tener así un número entero de ciclos en el tiempo de transmisión del bit. El inverso de T se denomina la *tasa de baudios*.

El proceso de la señal modulada es el proceso PSK, cuya expresión es

$$X_t = \cos(2\pi f_0 t + \Theta_t). \tag{4.33}$$

Para la obtención de las funciones de momento del proceso es conveniente hacer uso de las funciones,

$$h^{(c)}(t) = \begin{cases} \cos(2\pi f_0 t), & 0 \leq t < T; \\ 0, & \text{en el resto}, \end{cases} \tag{4.34}$$

y

$$h^{(s)}(t) = \begin{cases} \sin(2\pi f_0 t), & 0 \leq t < T; \\ 0, & \text{en el resto}. \end{cases} \tag{4.35}$$

Aplicando la fórmula del coseno del ángulo suma y teniendo en cuenta la relación (4.32), podemos escribir

$$\begin{aligned}\cos(2\pi f_0 t + \Theta_t) &= \cos(\Theta_t)\cos(2\pi f_0 t) - \sin(\Theta_t)\sin(2\pi f_0 t) \\ &= \sum_{k=-\infty}^{k=+\infty} \cos(\Theta_k) h^{(c)}(t - kT) - \sum_{k=-\infty}^{k=+\infty} \sin(\Theta_k) h^{(s)}(t - kT) \\ &= -\sum_{k=-\infty}^{k=+\infty} \sin(\Theta_k) h^{(s)}(t - kT), \end{aligned} \tag{4.36}$$

donde el primer sumatorio se anula porque todos sus términos son 0, puesto que si recordamos (4.31), $\Theta_k = \pm\pi/2, \forall k$ y en consecuencia $\cos(\Theta_k) = 0, \forall k$. La expresión (4.36) permite obtener fácilmente $\mu(t) = 0$ dado que $\sin(\Theta_k) = \pm 1$ con igual probabilidad.

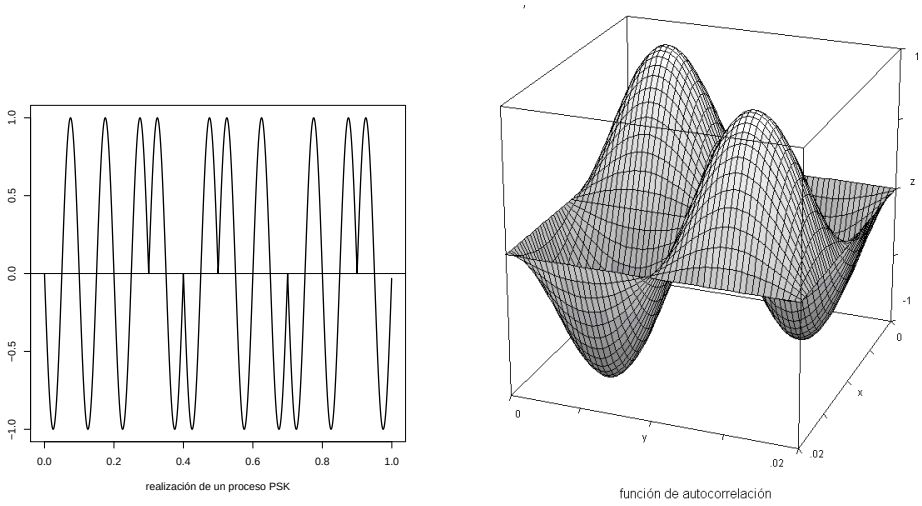


Figura 4.6: Realización de un proceso PSK y su correspondiente función de autocorrelación

Las funciones autocovarianza y autocorrelación coinciden, para calcularlas recurriremos nuevamente a (4.36) y a la versión numerable de la covarianza de una suma (expresión 2.11 en la página 53).

$$R(t_1, t_2) = E(X_{t_1} X_{t_2}) = \sum_{k,l} E[\sin(\Theta_k)\sin(\Theta_l)] h^{(s)}(t_1 - kT) h^{(s)}(t_2 - lT),$$

pero teniendo en cuenta la independencia de la sucesión original y que $\Theta_k = \pm\pi/2, \forall k$,

$$E[\sin(\Theta_k)\sin(\Theta_l)] = \begin{cases} 0, & \text{si } k \neq l; \\ 1, & \text{si } k = l, \end{cases}$$

con lo que

$$R(t_1, t_2) = \sum_{k=-\infty}^{k=+\infty} h^{(s)}(t_1 - kT) h^{(s)}(t_2 - kT).$$

Por definición, el soporte de $h^{(s)}(t)$ es el intervalo $[0,T]$, por lo que el anterior sumatorio será siempre nulo a menos que $\exists k_0$ tal que t_1 y t_2 estén ambos en $[(k_0-1)T, k_0T[$. Si denotamos por $t^* = t \bmod T$, la expresión final de $R(t_1,t_2)$ será

$$R(t_1,t_2) = \begin{cases} h^{(s)}(t_1^*)h^{(s)}(t_2^*), & \text{si } \lfloor t_1/T \rfloor = \lfloor t_2/T \rfloor \ ; \\ 0, & \text{en el resto,} \end{cases} \tag{4.37}$$

donde $\lfloor \cdot \rfloor$ representa la parte entera por defecto del argumento. La Figura 4.6 muestra una realización de un proceso PSK y su función de autocorrelación.

4.3.7. Proceso de Wiener. Movimiento Browniano

Límite de un camino aleatorio

El camino aleatorio que describimos en la página 82 era la suma de desplazamientos independientes a derecha e izquierda, de forma que sólo saltos unitarios estaban permitidos en uno u otro sentido en cada intervalo unitario de tiempo. Imaginemos ahora desplazamientos pequeños de longitud Δ que se producen en intervalos de tiempo pequeños de longitud τ, cuando desplazamiento y tiempo tiendan a cero obtendremos un proceso cuyas realizaciones serán funciones continuas en el tiempo. Debemos precisar en que condiciones tienden a cero ambas cantidades.

Consideremos una partícula situada en el origen. En cada intervalo de tiempo τ se desplaza una cantidad aleatoria Z de manera que

$$P(Z = +\Delta) = p, \qquad P(Z = -\Delta) = 1 - p = q,$$

siendo los distintos desplazamientos independientes. Se trata de una variable dicotómica con media y varianza

$$\mu_Z = (p-q)\Delta, \qquad \sigma_Z^2 = 4pq\Delta^2,$$

y cuya función característica vale

$$\varphi_Z(u;\Delta) = E(e^{iuZ}) = pe^{iu\Delta} + qe^{-iu\Delta}.$$

En un tiempo t se habrán producido $n = \lfloor t/\tau \rfloor$ desplazamientos, siendo X_t la posición final de la partícula. Dicha posición es por tanto la suma $X_t = \sum_{i=1}^{n} Z_i$, con las Z_i i.i.d como la Z anterior. Así, la función característica de X_t valdrá

$$\varphi_{X_t}(u;\tau,\Delta) = E(e^{iuX_t}) = (pe^{iu\Delta} + qe^{-iu\Delta})^n = (pe^{iu\Delta} + qe^{-iu\Delta})^{\lfloor t/\tau \rfloor}. \tag{4.38}$$

La media y la varianza de X_t se obtienen fácilmente a partir de μ_Z y σ_Z^2,

$$\mu_{X_t} = \lfloor t/\tau \rfloor (p-q)\Delta, \qquad \sigma_Z^2 = \lfloor t/\tau \rfloor 4pq\Delta^2.$$

Nuestro objetivo es $\Delta \to 0$ y $\tau \to 0$ de manera tal que obtengamos un resultado *razonable*. Por ejemplo, por razonable podríamos entender que media y varianza de X_t, para $t = 1$, fueran finitas e iguales a μ y σ^2, respectivamente. Ello supondría que Δ y τ deben tender a cero de forma tal que

$$\frac{(p-q)\Delta}{\tau} \to \mu \quad \text{y} \quad \frac{4pq\Delta^2}{\tau} \to \sigma^2.$$

Para ello debe cumplirse,

$$\Delta = \sigma\sqrt{\tau}, \quad p = \frac{1}{2}\left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right), \quad q = \frac{1}{2}\left(1 - \frac{\mu\sqrt{\tau}}{\sigma}\right). \tag{4.39}$$

Estas relaciones suponen que tanto p como q deben ser valores muy próximos a $1/2$ si queremos evitar degeneraciones en el proceso límite y que Δ es de un orden de magnitud mucho mayor que τ puesto que como infinitésimo $\Delta = O(\tau^{1/2})$.

La distribución de probabilidad límite de las X_t podemos conocerla a través del comportamiento límite de su función característica, para ello sustituiremos (4.39) en (4.38)

$$\varphi_{X_t}(u;\tau) = \left[\frac{1}{2}\left(1+\frac{\mu\sqrt{\tau}}{\sigma}\right)e^{iu\sigma\sqrt{\tau}} + \frac{1}{2}\left(1-\frac{\mu\sqrt{\tau}}{\sigma}\right)e^{-iu\sigma\sqrt{\tau}}\right]^{t/\tau}, \tag{4.40}$$

y desarrollaremos en potencias de τ la expresión entre corchetes. Hecho esto y haciendo $\tau \to 0$, tendremos

$$\varphi_{X_t}(u) = \exp(\mu t u - \frac{1}{2}\sigma^2 t u),$$

que es la función característica de una $N(\mu t, \sigma^2 t)$.

Observemos además que el proceso formado por la variables X_t hereda las propiedades del camino aleatorio que lo genera, al fin y al cabo lo único que hemos hecho es aplicar el teorema Central de Límite a la suma de variable dicotómicas que definen X_t. De entre ellas conviene destacar la propiedad de *incrementos independientes* y *estacionarios*. Así, si $t_1 < t_2$, $X_{t_2} - X_{t_1} \sim N(\mu(t_2 - t_1), \sigma^2(t_2 - t_1))$.

Para obtener ahora las distribuciones finito-dimensional del proceso límite podemos recurrir a la expresión (4.13), que proporciona la densidad conjunta a partir de las densidades de los incrementos. Si $t_1 < t_2 < \ldots < t_n$,

$$\begin{aligned}
f_{t_1,t_2,\ldots,t_n}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) &= f_{t_1}(x_{t_1}) f_{t_2-t_1}(x_{t_2} - x_{t_1}) \cdots f_{t_n - t_{n-1}}(x_{t_n} - x_{t_{n-1}}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2 t_1}} e^{-\frac{1}{2}\left(\frac{(x_{t_1}-\mu t_1)^2}{\sigma^2 t_1}\right)} \cdots \\
&\quad \cdots \frac{1}{\sqrt{2\pi\sigma^2(t_n - t_{n-1})}} e^{-\frac{1}{2}\left(\frac{[(x_{t_n}-x_{t_{n-1}})-(\mu t_n - \mu t_{n-1})]^2}{\sigma^2(t_2 - t_{n-1})}\right)} \\
&= \frac{e^{-\frac{1}{2}\left\{\frac{(x_{t_1}-\mu t_1)^2}{\sigma^2 t_1} + \cdots + \frac{[(x_{t_n}-x_{t_{n-1}})-(\mu t_n - \mu t_{n-1})]^2}{\sigma^2(t_2 - t_{n-1})}\right\}}}{\sqrt{(2\pi\sigma^2)^n t_1 \cdots (t_n - t_{n-1})}},
\end{aligned}$$

que es la densidad conjunta de una Normal multivariate. El proceso límite es un proceso Gaussiano con incrementos independientes.

Proceso de Wiener

El proceso de Wiener es un proceso límite como el descrito en el párrafo anterior con $p = q = 1/2$.

Definición 4.4 (Proceso de Wiener) *Un proceso de Wiener es un proceso estocástico W_t, que verifica,*

1. *Su valor inicial es 0, $P(W_0 = 0) = 1$.*
2. *Sus incrementos son independientes.*
3. *Para $0 \le t_1 < t_2$, el incremento $W_{t_2} - W_{t_1} \sim N(0, \sigma^2(t_2 - t_1))$.*

Los fundamentos matemáticos del proceso fueron establecidos por Norbert Wiener, matemático americano especialista en inferencia estadística y teoría de la comunicación. El proceso es conocido también como movimiento browniano, expresión utilizada para describir el movimiento aleatorio de las moléculas de un gas al entrechocar entre sí, que recibe este nombre en honor de Robert Brown, un botánico del siglo diecinueve que lo describió por primera vez.

La media y la varianza del proceso de Wiener son ya conocidas,

$$\mu(t) = 0, \qquad \sigma^2(t) = \sigma^2 t.$$

Autocovarianza y autocorrelación coinciden. Para su obtención, como los incrementos son independientes procederemos como hicimos en (4.22) y (4.23), obteniendo,

$$C(t_1, t_2) = R(t_1, t_2) = \sigma^2 \min(t_1, t_2). \tag{4.41}$$

Observemos que las funciones de autocovarianza del proceso de Wiener, (4.41), y el proceso de Poisson, (4.23), son iguales a pesar de tratarse de dos procesos de naturaleza muy distinta. Las trayectorias del segundo son funciones escalonadas mientras que las del primero son continuas. Esta igualdad pone en evidencia que las funciones de momento son sólo descripciones parciales del proceso.

4.3.8. Cadenas de Markov

Un proceso estocástico X_t es un *proceso de Markov* si verifica la propiedad de Markov, (4.3), enunciada en la página 77. Recordemos que dicha propiedad supone que la evolución del proceso depende tan sólo de su pasado inmediato. En términos de probabilidad, si las variables del proceso son discretas, la propiedad se expresa mediante,

$$P(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}, \ldots, X_{t_1} = x_{t_1}) = P(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}). \tag{4.42}$$

Si las variables del proceso son continuas su expresión es

$$P(X_{t_{n+1}} \leq x_{t_{n+1}} | X_{t_n} \leq x_{t_n}, \ldots, X_{t_1} \leq x_{t_1}) = P(X_{t_{n+1}} \leq x_{t_{n+1}} | X_{t_n} \leq x_{t_n}), \tag{4.43}$$

o sus equivalentes en términos de las funciones de probabilidad o de distribución, respectivamente.

Algunos de los procesos estudiados hasta ahora son procesos de Markov. En efecto, los procesos suma IID y el proceso de Poisson tenían incrementos independientes y como vimos en (4.6), ello implica que poseen la propiedad de Markov.

Para el proceso RTS, recordemos que el valor de $X_{t_{n+1}} = \pm a$ y que su valor depende exclusivamente del signo que tenga X_{t_n} y de la paridad de $N_{t_{n+1}-t_n}$, número de cambios (sucesos) del proceso de Poisson subyacente, y es por tanto un proceso de Markov.

Hay, obviamente, procesos que no cumplen la condición. Veamos un ejemplo.

Ejemplo 4.2 (Un proceso que no es de Markov) *Consideremos el siguiente proceso,*

$$Y_n = \frac{1}{2}(X_n + X_{n-1}),$$

donde las X_k son independientes, ambas Bernoulli con $p = 1/2$. Un proceso de estas características recibe el nombre de proceso de medias móviles de orden 2, puesto que se define a partir de la media aritmética de las dos últimas realización de otro proceso.

La función de probabilidad de Y_n se calcula fácilmente. Por ejemplo,

$$P(Y_n = 0) = P(X_n = 0, X_{n-1} = 0) = \frac{1}{4}.$$

Para el resto,

$$P(Y_n = 1/2) = \frac{1}{2}, \qquad P(Y_n = 1) = \frac{1}{4}.$$

Obtengamos ahora la probabilidad condicionada para dos valores consecutivos de Y_n*,*

$$\begin{aligned} P(Y_n = 1|Y_{n-1} = 1/2) &= \frac{P(Y_n = 1, Y_{n-1} = 1/2)}{P(Y_{n-1} = 1/2)} \\ &= \frac{P(X_n = 1, X_{n-1} = 1, X_{n-2} = 0)}{P(Y_{n-1} = 1/2)} \\ &= \frac{1/8}{1/2} = \frac{1}{4}. \end{aligned}$$

Si tenemos información de un instante anterior, por ejemplo, $Y_{n-2} = 1$*, podemos calcular,*

$$P(Y_n = 1|Y_{n-1} = 1/2, Y_{n-2} = 1) = \frac{P(Y_n = 1, Y_{n-1} = 1/2, Y_{n-2} = 1)}{P(Y_{n-1} = 1/2, Y_{n-2} = 1)},$$

pero el suceso $\{Y_n = 1, Y_{n-1} = 1/2\} = \{X_n = 1, X_{n-1} = 1, X_{n-2} = 0\}$ *y el suceso* $\{Y_{n-2} = 1\} = \{X_{n-2} = 1, X_{n-2} = 1\}$*, con lo que ambos sucesos son incompatibles, por tanto,*

$$P(Y_n = 1|Y_{n-1} = 1/2, Y_{n-2} = 1) = 0 \neq P(Y_n = 1|Y_{n-1} = 1/2)$$

y el proceso no puede ser Markov.

Cadenas de Markov discretas

Si el proceso de markov es DTDV, recibe el nombre de *cadena de Markov discreta*. Este tipo de procesos son de un gran interés práctico y vamos a ocuparnos de ellos con cierto detalle.

Es costumbre designar mediante X_0 el inicio de la cadena. El soporte de las X_n, $S = \{s_0, s_1, s_2, \ldots\}$, se denomina *espacio de estados* de la cadena, que puede eventualmente ser finito. Por razones de simplicidad nos referiremos a los valores de s_i solamente mediante su subíndice i.

Probabilidades iniciales y de transición.- Las *probabilidades iniciales* son

$$\pi_i = P(X_0 = i), \quad \pi_i \geq 0, \forall i, \quad \sum_i \pi_i = 1.$$

Las *probabilidades de transición de un paso*, p_{ij}, designan

$$p_{ij} = P(X_{n+1} = j|X_n = i)$$

y supondremos que no varían con n. Diremos que son probabilidades de transición *homogéneas*.

Las distribuciones finito-dimensionales se obtienen fácilmente a partir de las p_{ij} si hacemos uso de la propiedad de Markov y del teorema de factorización (1.2). En efecto,

$$\begin{aligned} &P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \\ &\quad = P(X_n = i_n|X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \cdots P(X_1 = i_1|X_0 = i_0)P(X_0 = i_0) \\ &\quad = P(X_n = i_n|X_{n-1} = i_{n-1}) \cdots P(X_1 = i_1|X_0 = i_0)P(X_0 = i_0) \\ &\quad = p_{i_{n-1},i_n} p_{i_{n-2},i_{n-1}} \cdots p_{i_0,i_1} \pi_{i_0}. \end{aligned}$$

La cadena de Markov discreta, X_n, queda completamente especificada con las probabilidades iniciales y la llamada *matriz de transición de un paso*, que recoge las probabilidades del mismo nombre,

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p_{i0} & p_{i1} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{4.44}$$

De acuerdo con el significado de p_{ij} las filas de la matriz suman la unidad,

$$\sum_j p_{ij} = 1, \quad \forall i.$$

Probabilidades de transición de n pasos.- Las *probabilidades de transición de* n *pasos*,

$$p_{ij}(n) = P(X_{n+k} = j | X_k = i), \quad \forall n \geq 0$$

proporcionan las probabilidades de pasar del estado i al estado j en n pasos o transiciones. Como en la anterior expresión las probabilidades de transición no dependen de k,

$$p_{ij}(n) = P(X_{n+k} = j | X_k = i) = P(X_n = j | X_0 = i), \quad \forall n \geq 0, \quad \forall k \geq 0.$$

Vamos a obtenerlas recursivamente. Para obtener $p_{ij}(2)$, consideremos la partición del espacio muestral constituida por los sucesos $\{X_1 = k\}, k \in S$, tendremos que $\Omega = \cup_k \{X_1 = k\}$. Podemos escribir,

$$\begin{aligned} P(X_2 = j | X_0 = i) &= P\left(\{X_2 = j\} \bigcap [\cup_k \{X_1 = k\}] \, | X_0 = i\right) \\ &= \sum_k P(X_2 = j, X_1 = k | X_0 = i) \\ &= \sum_k \frac{P(X_2 = j, X_1 = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_k \frac{P(X_2 = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) P(X_0 = i)}{P(X_0 = i)} \\ &= \sum_k P(X_2 = j | X_1 = k) P(X_1 = k | X_0 = i) \\ &= \sum_k p_{ik} p_{kj}, \end{aligned} \tag{4.45}$$

pero la expresión(4.45) no es más que el producto de la fila i por la columna j de la matriz $\mathbf{P}$, que siendo válida para $\forall i, j$, permite escribir

$$\mathbf{P}(2) = \mathbf{P}\mathbf{P} = \mathbf{P}^2.$$

Aplicando recursivamente el razonamiento anterior llegaremos a

$$\mathbf{P}(n) = \mathbf{P}(n-1)\mathbf{P} = \mathbf{P}^n. \tag{4.46}$$

Esta ecuación puede también escribirse

$$\mathbf{P}(n+m) = \mathbf{P}^n\mathbf{P}^m,$$

que término a término se expresa de la forma

$$p_{ij}(n+m) = \sum_k p_{ik}(n)p_{kj}(m), \quad n,m > 0, \forall (i,j) \in S. \tag{4.47}$$

Expresiones conocidas como las *ecuaciones de Chapman-Kolmogorov*.

Probabilidades de los estados después de n pasos.- Las probabilidades π_i constituyen la distribución inicial sobre el espacio de estados, S. La distribución después de n pasos será

$$\begin{aligned}
\pi_i^{(n)} &= P(X_n = i) \\
&= P\left(\{X_n = i\} \bigcap [\cup_j \{X_{n-1} = j\}]\right) \\
&= \sum_j P\left(\{X_n = i\} \cap \{X_{n-1} = j\}\right) \\
&= \sum_j P(X_n = i | X_{n-1} = j) P(X_{n-1} = j) \\
&= \sum_j \pi_j^{(n-1)} p_{ji}
\end{aligned}$$

resultado que podemos expresar en forma matricial,

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-1)}\mathbf{P}. \tag{4.48}$$

Aplicando recursivamente (4.48) podemos relacionar $\boldsymbol{\pi}^{(n)}$ con $\boldsymbol{\pi} = \boldsymbol{\pi}^{(0)}$,

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}\mathbf{P}^n. \tag{4.49}$$

Ejemplo 4.3 (Paquetes de voz) *Un modelo de Markov para la transmisión de paquetes de voz supone que si el n-ésimo paquete contiene silencio, la probabilidad de que siga un silencio es $1-\alpha$ y α si lo que sigue es voz. En el caso contrario dichas probabilidades son $1-\beta$, para el paso de voz a voz y β par el paso de voz a silencio. Si representamos los dos estados con voz $= 1$ y silencio $= 0$ el esquema de las transiciones posibles sería el que muestra la Figura 4.7.*

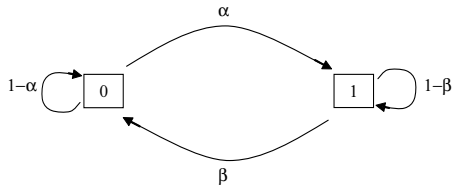


Figura 4.7: Probabilidades de transición en el modelo de transmisión de paquetes de voz

Si X_n es la variable que indica la actividad en un paquete de voz en el tiempo n, se trata de una cadena de Markov discreta con espacio de estados $S = \{0, 1\}$ y matriz de transición

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

la matriz de transición de orden n *vale*

$$\mathbf{P}^n = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix},$$

cuyo límite es

$$\lim_{n\to\infty} \mathbf{P}^n = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}.$$

Por ejemplo, si $\alpha = 1/10$ y $\beta = 2/10$, en el límite

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}.$$

El proceso goza además de una interesante propiedad. Si la distribución inicial es

$$\pi_0 = p_0, \qquad \pi_1 = 1 - p_0,$$

aplicando (4.49) y pasando al límite podemos obtener la distribución límite del espacio de estados,

$$\lim_n \boldsymbol{\pi}^{(\boldsymbol{n})} = \boldsymbol{\pi} \lim \mathbf{P}^n = \begin{bmatrix} p_0 & 1-p_0 \end{bmatrix}\begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}.$$

que no depende de la distribución inicial.

Equilibrio de una cadena de Markov.- La propiedad que acabamos de ver en el ejemplo significa que la cadena de Markov tiende a una situación estacionaria a medida que n aumenta.

Observemos que si, en general,

$$\lim_n p_{ij}(n) = \pi_j^{(e)}, \quad \forall i,$$

la consecuencia es que

$$\lim_n \boldsymbol{\pi}_{\boldsymbol{j}}^{(\boldsymbol{n})} = \sum_i \pi_j^{(e)}\pi_i = \pi_j^{(e)}, \quad \forall j.$$

La distribución sobre el espacio de estados tiende a una distribución fija independiente de la distribución inicial $\boldsymbol{\pi}$. Se dice que la cadena ha alcanzando el *equilibrio* o una *situación estable*. La distribución alcanzada, $\boldsymbol{\pi}^{(\boldsymbol{e})}$ recibe el nombre de *distribución de equilibrio* o *estacionaria*.

La distribución de equilibrio puede obtenerse, si existe, sin necesidad de pasar al límite. Para ello basta recurrir a (4.48) y tener en cuenta que $\boldsymbol{\pi}^{(\boldsymbol{n})}$ y $\boldsymbol{\pi}^{(\boldsymbol{n-1})}$ tienen por límite común la distribución estacionaria. Bastará con resolver el sistema de ecuaciones,

$$\boldsymbol{\pi}^{(\boldsymbol{e})} = \boldsymbol{\pi}^{(\boldsymbol{e})}\mathbf{P}, \tag{4.50}$$

con la condición adicional $\sum_i \pi_i^{(e)} = 1$.

Si nuestro sistema se inicia con la distribución de equilibrio, por (4.49) y (4.50)

$$\boldsymbol{\pi}^{(\boldsymbol{n})} = \boldsymbol{\pi}^{(\boldsymbol{e})}\mathbf{P}^n = \boldsymbol{\pi}^{(\boldsymbol{e})}.$$

Ejemplo 4.4 (Continuación del ejemplo 4.3) *La distribución de equilibrio podríamos haberla obtenido utilizando (4.50). Si* $\boldsymbol{\pi^{(e)}} = [\pi_0^{(e)}\ \pi_1^{(e)}]$*, el sistema a resolver es*

$$\begin{aligned}\pi_0^{(e)} &= (1-\alpha)\pi_0^{(e)} + \beta\pi_1^{(e)} \\ \pi_1^{(e)} &= \alpha\pi_0^{(e)} + (1-\beta)\pi_1^{(e)},\end{aligned}$$

que da lugar a una única ecuación, $\alpha\pi_0^{(e)} = \beta\pi_1^{(e)}$ *que junto con* $\pi_0^{(e)} + \pi_1^{(e)} = 1$ *conduce a*

$$\pi_0^{(e)} = \frac{\beta}{\alpha+\beta} \qquad \pi_1^{(e)} = \frac{\alpha}{\alpha+\beta}.$$

No todas las cadenas de Markov gozan de esta propiedad de equilibrio. El ejemplo siguiente lo demuestra.

Ejemplo 4.5 (Proceso Binomial) *El proceso Binomial que describíamos en la página 82 es una cadena de Markov puesto que se trata de un proceso suma de Bernoullis independientes,* $B \sim (1, p)$*. Dada su definición, en una transición el proceso permanece en el mismo estado o lo incrementa en una unidad, según que hayamos obtenido un fracaso o un éxito en la prueba de Bernoulli correspondiente. La matriz de transición será,*

$$\mathbf{P} = \begin{bmatrix} 1-p & p & 0 & 0 & \cdots \\ 0 & 1-p & p & 0 & \cdots \\ 0 & 0 & 1-p & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Se comprueba que $\forall j,\ \pi_j(n) \to 0$ *cuando* $n \to \infty$*.*

Cadenas de Markov continuas en el tiempo

Si el proceso de markov es CTDV, recibe el nombre de *cadena de Markov continua en el tiempo*. Como en el caso de las cadenas discretas, también ahora la matriz de transición especifica completamente la cadena.

La distribución finito-dimensional para $n+1$ tiempos arbitrarios $t_1 < t_2 < \ldots < t_n < t_{n+1}$, viene dada por

$$\begin{aligned}&P(X_{t_{n+1}} = x_{t_{n+1}}, X_{t_n} = x_{t_n}, \ldots, X_{t_1} = x_{t_1}) = \\ &\quad = P(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}) \cdots P(X_{t_2} = x_{t_2} | X_{t_1} = x_{t_1}) P(X_{t_1} = x_{t_1}),\end{aligned} \tag{4.51}$$

que exige conocer las probabilidades de transición entre dos tiempos cualesquiera t y $t+s$,

$$P(X_{t+s} = j | X_t = i), \quad \forall s \geq 0.$$

Supondremos que dichas probabilidades son *homogéneas* y dependen tan sólo de la diferencia de tiempos,

$$P(X_{t+s} = j | X_t = i) = P(X_s = j | X_0 = i) = p_{ij}(s), \quad \forall s > 0, \forall t.$$

La matriz de transición viene ahora referida a un intervalo de tiempo t, de forma que $\mathbf{P}(t) = [p_{ij}(t)]$ denota la *matriz de probabilidades de transición en un intervalo de tiempo* t.

Ejemplo 4.6 (Probabilidades de transición para el proceso de Poisson) *Para el proceso de Poisson de parámetro λ,*

$$p_{ij}(t) = P(N_t = j - i) = e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}, \quad j \geq i.$$

La matriz de transición para un intervalo de duración t será

$$\mathbf{P}(t) = \begin{bmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2 & \cdots & \cdots \\ 0 & e^{-\lambda t} & \lambda t e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2 & \cdots \\ 0 & 0 & e^{-\lambda t} & \lambda t e^{-\lambda t} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Cuando $t \to 0$, podemos sustituir $e^{-\lambda t}$ por los primeros términos de su desarrollo en serie de potencias, despreciando las potencias mayores o igual a 2, con lo que la matriz de transición adopta la forma,

$$\mathbf{P}(\Delta t) = \begin{bmatrix} 1-\lambda\Delta t & \lambda\Delta t & 0 & 0 & \cdots \\ 0 & 1-\lambda\Delta t & \lambda\Delta t & 0 & \cdots \\ 0 & 0 & 1-\lambda\Delta t & \lambda\Delta t & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Tiempo de permanencia en un estado.- El proceso RTS modelizaba el cambio de signo de una señal aleatoria y es una cadena de Markov continua en el tiempo que sólo presenta dos estado, $S = \{-a, +a\}$. El cambio se signo estado se produce con cada llegada de un suceso en el proceso de Poisson subyacente. Como los tiempos de llegada son exponenciales, se deduce que el tiempo T_i que la cadena permanece en le estado i es $Exp(\lambda)$.

Esta propiedad, inmediata, en el caso del proceso RTS, es una propiedad general de este tipo de cadenas de Markov. En efecto, si desde que el proceso alcanzó el estado i ha transcurrido un tiempo s, la probabilidad de que su estancia en él se prolongue por un tiempo r vale

$$P(T_i > r + s | T_i > s) = P(T_i > r + s | X_t = i), \quad 0 \leq t \leq s,$$

pero por la propiedad de Markov, lo único relevante es donde estaba la cadena en el tiempo s, lo que equivale a reestablecer el origen de tiempos en s y por tanto,

$$P(T_i > r + s | T_i > s) = P(T_i > r).$$

El tiempo de ocupación carece pues de memoria, pero esta es una propiedad exclusiva de la distribución exponencial y la conclusión es que $T_i \sim Exp(\lambda_i)$ y

$$P(T_i > r) = e^{-\lambda_i r}.$$

El tiempo medio de permanencia del proceso en el estado i será $E(T_i) = 1/\lambda_i$.

Obsérvese que el parámetro depende del estado i, pudiendo ocurrir, como sucede en el proceso RTS, que $\lambda_i = \lambda, \forall i$. Una vez el proceso ha entrado en i se incia el tiempo T_i de su permanencia en él, pero independientemente de lo que esta dure, la transición a un nuevo estado j se lleva con probabilidad $\tilde{p}_{ij}$, que podemos asociar a una cadena de Markov discreta que decimos está *empotrada* en la original.

4.4. Procesos estacionarios

En la página 81 introducíamos el concepto de estacionariedad al comprobar que las distribución ligada a los incrementos dependía tan sólo de la diferencia de tiempos y no de su posición.

El concepto de estacionariedad recoge una propiedad común a muchos procesos consistente en que su comportamiento probabilístico no cambia con el tiempo. La definición formal es la siguiente.

Definición 4.5 (Proceso estacionario) *Un proceso estocástico X_t se dice que es estacionario si sus distribuciones finito-dimensionales son invariantes por traslación. Es decir,*

$$F_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n) = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(x_1,x_2,\ldots,x_n), \quad \forall(t_1,t_2,\ldots,t_n),\ \forall\tau. \tag{4.52}$$

la primera consecuencia de esta definición es que las distribuciones individuales de las variables que componen el proceso no dependen de t, puesto que de (4.52) se deduce

$$F_t(x) = F_{t+\tau}(x) = F(x), \quad \forall t,\tau,$$

y como consecuencia

$$\mu(t) = E(X_t) = \mu, \quad \forall t$$

y

$$\sigma^2(t) = E[(X_t-\mu)^2] = \sigma^2.$$

Las distribución conjunta de (X_{t_1}, X_{t_2}) dependerá tan sólo de la diferencia de tiempos, $t_2 - t_1$. Basta para ello hacer $\tau = t_1$ en (4.52),

$$F_{t_1,t_2}(x_1,x_2) = F_{0,t_2-t_1}(x_1,x_2), \quad \forall t_1,t_2.$$

La consecuencia es que los momentos de segundo orden y las funciones correspondientes dependen también, solamente, de dicha diferencia.

$$R(t_1,t_2) = R(t_2-t_1), \qquad C(t_1,t_2) = C(t_2-t_1), \quad \forall t_1,t_2.$$

Algunos de los procesos antes definidos gozan de esta propiedad. Así, el proceso IID es estacionario porque

$$\begin{aligned} F_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n) &= F(x_1)\cdots F(x_n) \\ &= = F_{t_1+\tau,t_2+\tau,\ldots,t_n+\tau}(x_1,x_2,\ldots,x_n), \end{aligned}$$

$\forall\tau,t_1,t_2,\ldots,t_n$. Sin embargo, el proceso suma IID no es estacionario porque recordemos que su media y varianza valían $\mu(n) = n\mu$ y $\sigma^2(n) = n\sigma^2$, que crecen con n y no son constantes.

Ejemplo 4.7 (Estacionariedad del proceso RTS) *El proceso RTS se estudió en la Sección 4.3.5. Recordemos que el proceso toma sólo dos valores simétricos, $\{-a,a\}$, de acuerdo con el siguiente esquema.*

1. *$X_0 = \pm a$ con igual probabilidad $p = 1/2$,*
2. *X_t cambia de signo con cada ocurrencia de un suceso en un proceso de Poisson de parámetro λ.*

Obtengamos la distribución finito-dimensional, para lo que haremos uso de la propiedad de incrementos independientes que el proceso RTS hereda del proceso de Poisson subyacente.

$$\begin{aligned} &P(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) = \\ &\quad = \quad P(X_{t_1} = x_1)P(X_{t_2} = x_2 | X_{t_1} = x_1) \cdots P(X_{t_n} = x_{t_n} | X_{t_{n-1}} = x_{t_{n-1}}). \end{aligned} \tag{4.53}$$

Las probabilidades condicionadas que aparecen en la expresión dependen del número de cambios que se han producido entre los dos tiempos implicados, más concretamente de la paridad de $N_{|t_i - t_j|}$, numero de sucesos ocurridos en el intervalo de tiempo. Aplicando (4.29)

$$P(X_{t_j} = x_j | X_{t_i} = x_i) = \begin{cases} \frac{1}{2}P(N_{|t_i-t_j|} = par) = \frac{1}{2}\left(1 + e^{-2\lambda|t_i - t_j|}\right), & si\ x_i = x_j; \\ \\ \frac{1}{2}P(N_{|t_i-t_j|} = impar) = \frac{1}{2}\left(1 - e^{-2\lambda|t_i - t_j|}\right), & si\ x_i \neq x_j. \end{cases}$$

Si efectuamos ahora una traslación τ de todos los tiempos, $|t_i - t_j| = |(t_i - \tau) - (t_j - \tau)|$, y

$$P(X_{t_j} = x_j | X_{t_i} = x_i) = P(X_{t_j+\tau} = x_j | X_{t_i+\tau} = x_i),$$

y como $P(X_{t_1} = x_1) = 1/2, \forall t_1, \forall x_1$, se deduce la estacionariedad del proceso RTS.

Si generalizamos el proceso RTS de manera que $P(X_0 = +a) = p$ y $P(X_0 = -a) = 1 - p$ la estacionariedad se pierde. En efecto, en (4.53) las probabilidades condicionadas no cambian al efectuar la traslación τ, pero sí pueden hacerlo $P(X_{t_1} = x_1)$. Por ejemplo, si $x_1 = a$,

$$\begin{aligned} P(X_{t_1} = a) &= P(X_{t_1} = a | X_0 = a)P(X_0 = a) + P(X_{t_1} = a | X_0 = -a)P(X_0 = -a) \\ &= pP(N_{t_1} = par) + (1 - p)P(N_{t_1} = impar) \\ &= \frac{1}{2}(1 - e^{-2\lambda t_1} + 2pe^{-2\lambda t_1}). \end{aligned}$$

Al efectuar la traslación,

$$\begin{aligned} P(X_{t_1+\tau} = a) &= P(X_{t_1+\tau} = a | X_0 = a)P(X_0 = a) + P(X_{t_1+\tau} = a | X_0 = -a)P(X_0 = -a) \\ &= pP(N_{t_1+\tau} = par) + (1 - p)P(N_{t_1+\tau} = impar) \\ &= \frac{1}{2}(1 - e^{-2\lambda(t_1+\tau)} + 2pe^{-2\lambda(t_1+\tau)}). \end{aligned}$$

Sólo cuando desde el inicio hay equiprobabilidad, $p = 1/2$, $P(X_t = \pm a) = 1/2, \forall t$ (véase la demostración en la Sección 4.3.5) ambas probabilidades coinciden.

4.4.1. Estacionariedad en sentido amplio (WSS)

Definición 4.6 *Decimos que un procesos estocástico es* estacionario en sentido amplio *(WSS, Wide-Sense Stationary) si su media es constante y su función de autocorrelación (autocovarianza) es invariante por traslación. Es decir,*

$$\mu(t) = \mu, \quad \forall t, \quad \text{y} \quad R(t_1, t_2) = R(t_1 + \tau, t_2 + \tau) = R(\tau), \quad \forall t_1, t_2.$$

La WWS es una propiedad más débil que la estacionariedad. Esta última implica a aquella pero no al revés, como pone de manifiesto el siguiente contraejemplo.

Ejemplo 4.8 (WSS $\nrightarrow$ la estacionariedad) *El proceso X_n esta formado por dos procesos intercalados de variables independientes, si $n = 2k$, $X_n = \pm 1$ con probabilidad 1/2; si $n = 2k+1$,*

X_n toma los valores 1/3 y -3 con probabilidad 9/10 y 1/10, respectivamente. El proceso no puede ser estacionario porque su función de probabilidad varía con n.

Por otra parte, se comprueba fácilmente que

$$\mu(n) = 0, \quad \forall n,$$

y

$$C(n_1, n_2) = \begin{cases} E(X_{n_1})E(X_{n_2}) = 0, & si\ n_1 \neq n_2; \\ E(X_{n_1}^2), & si\ n_1 = n_2. \end{cases}$$

El proceso es WSS.

Propiedades de la función de autocorrelación de un proceso WSS

En un proceso WSS la función de autocorrelación tiene una serie de propiedades que por su interés posterior vamos a deducir.

PA1) Para $\tau = 0$

$$R(0) = R(t,t) = E(X_t^2), \quad \forall t,$$

que es la *potencia media* del proceso.

PA2) La función de autocorrelación es una función par. En efecto,

$$R(\tau) = E(X_{t+\tau}X_t),$$

y haciendo $s = t + \tau$,

$$R(\tau) = E(X_{t+\tau}X_t) = E(X_s X_{s-\tau}) = R(-\tau).$$

PA3) La función de autocorrelación mide la tasa de cambio del proceso en términos de probabilidad. Si consideramos el cambio que ha sufrido el proceso entre t y $t + \tau$,

$$\begin{aligned} P(|X_{t+\tau} - X_t| > \varepsilon) &= P(|X_{t+\tau} - X_t|^2 > \varepsilon^2) && (4.54) \\ &\leq \frac{E[(X_{t+\tau} - X_t)^2]}{\varepsilon^2} && (4.55) \\ &= \frac{E[X_{t+\tau}^2 + X_t^2 - 2X_{t+\tau}X_t]}{\varepsilon^2} \\ &= \frac{2[R(0) - R(\tau)]}{\varepsilon^2}. && (4.56) \end{aligned}$$

El paso de (4.54) a (4.55) se hace en virtud de la desigualdad de Markov.

La expresión (4.56) indica que si $R(\tau)$ crece despacio, la probabilidad de cambio para el proceso es pequeña.

PA4) La función de autocorrelación alcanza su máximo en 0. Aplicando la desigualdad de Schwartz, $[E(XY)]^2 \leq E(X^2)E(Y)^2)$, a $X_{t+\tau}$ y X_t,

$$R(\tau)^2 = [E(X_{t+\tau}X_t)]^2 \leq E(X_{t+\tau}^2)E(X_t^2) = R(0)^2.$$

y por tanto $|R(\tau)| \leq R(0)$.

PA5) Si $\exists d; R(d) = R(0)$, la función de autocorrelación es periódica con periodo d. Aplicando de nuevo la desigualdad de Schwartz a $X_{t+\tau+d} - X_{t+\tau}$ y X_t,

$$\{E[(X_{t+\tau+d} - X_{t+\tau})X_t]\}^2 \leq E[(X_{t+\tau+d} - X_{t+\tau})^2]E[X_t^2],$$

que en términos de $R(\tau)$ podemos escribir

$$[R(\tau + d) - R(\tau)]^2 \leq 2[R(0) - R(d)]R(0).$$

Como $R(d) = R(0)$ entonces $R(\tau + d) = R(\tau), \forall\tau$ con lo que la función es periódica de periodo d.

Por otra parte de

$$E[(X_{t+d} - X_t)^2] = 2[R(0) - R(d)] = 0,$$

se deduce que el proceso es periódico en media cuadrática.

PA6) Si $X_t = m + N_t$ donde N_t es un proceso de media cero con $R_N(\tau) \stackrel{\tau\to\infty}{\longrightarrow} 0$, entonces

$$\begin{aligned} R_X(\tau) = E[(m + N_{t+\tau})(m + N_t)] &= m^2 + 2mE(N_t) + R_N(\tau) \\ &= m^2 + R_N(\tau) \stackrel{\tau\to\infty}{\longrightarrow} m^2. \end{aligned}$$

Deducimos de estas propiedades que la función de autocorrelación puede tener tres tipos de componentes,

1. una componente que se aproxima a 0 cuando $\tau \to \infty$,
2. una componente periódica, y
3. una componente con media no nula.

Processos Gaussianos y estacionariedad

Hemos visto que la WSS $\nrightarrow$ estacionariedad, pero este resultado tiene un excepción en el caso del proceso Gaussiano. Como vimos en la Sección 4.3.3, las distribuciones finito-dimensionales del proceso están completamente especificadas mediante su vector de medias y la matriz de covarianzas.

Si el proceso Gaussiano es WSS, sabemos que $\mu(t) = \mu, \forall t$ y $C(t_1, t_2) = g(|t_1 - t_2|)$. La consecuencia inmediata es que las distribución conjunta de $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ será invariante por traslación y el proceso será estacionario.

Ejemplo 4.9 (Un proceso Gaussiano de medias móviles) *Definamos*

$$Y_n = \frac{X_n + X_{n-1}}{2},$$

donde X_n es un proceso Gaussiano IID con media 0 y varianza σ^2.

La media de Y_n es también 0 y su covarianza

$$\begin{aligned} C_Y(n_1, n_2) &= \frac{1}{4}E[(X_{n_1} + X_{n_1-1})(X_{n_2} + X_{n_2-1}) \\ &= \frac{1}{4}E[X_{n_1}X_{n_2} + X_{n_1}X_{n_2-1} + X_{n_1-1}X_{n_2} + X_{n_1-1}X_{n_2-1}) \\ &= \begin{cases} \sigma^2/2, & \textit{si } n_1 - n_2 = 0; \\ \sigma^2/4, & |n_1 - n_2| = 1; \\ 0, & \textit{en el resto.} \end{cases} \end{aligned}$$

El proceso Y_n es por tanto WSS y además es Gaussiano por ser combinación lineal de variables Gaussianas. Las distribuciones finito dimensionales del proceso están especificadas con el vector de medias nulo y la matriz de covarianzas que define $C_Y(n_1, n_2)$.

4.4.2. Procesos cicloestacionarios

Son muchos los procesos que surgen de la repetición periódica de un experimento o fenómeno aleatorio. Es lógico pensar que la periodicidad del proceso influya en su comportamiento probabilístico. Surge así la noción de *cicloestacionariedad*.

Definición 4.7 (Proceso cicloestacionario (CE)) *Decimos que el proceso X_t es cicloestacionario, CE si sus distribuciones finito-dimensionales son invariantes por traslación mediante múltiplos enteros de un cierto período T. Es decir, $\forall(t_1, t_2, \ldots, t_n)$ y $\forall k \in \mathbb{Z}$,*

$$F_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n) = F_{t_1+kT,t_2+kT,\ldots,t_n+kT}(x_1, x_2, \ldots, x_n). \tag{4.57}$$

El concepto de WSS también tiene ahora su equivalente.

Definición 4.8 (Proceso cicloestacionario en sentido amplio (WSC)) *Decimos que el proceso X_t es cicloestacionario en sentido amplio, WSC si su media y su función de autocovarianza son invariantes por traslación mediante múltiplos enteros de un cierto período T,*

$$\mu(t + kT) = \mu(t), \qquad C(t_1 + kT, t_2 + kT) = C(t_1, t_2). \tag{4.58}$$

Es fácil comprobar que CE $\rightarrow$ WSC, pero no al contrario.

Ejemplo 4.10 *Consideremos el proceso*

$$X_t = A\cos\left(\frac{2\pi t}{T}\right).$$

La distribución conjunta de $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ es

$$\begin{aligned} F_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n) &= P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n) \\ &= P[A\cos(2\pi t_1/T) \le x_1, \ldots, A\cos(2\pi t_n/T) \le x_n] \\ &= P[A\cos(2\pi(t_1 + kT)/T) \le x_1, \ldots, A\cos(2\pi(t_n + kT)/T) \le x_n] \\ &= F_{t_1+kT,t_2+kT,\ldots,t_n+kT}(x_1, x_2, \ldots, x_n), \end{aligned}$$

y el proceso es cicloestacionario.

El proceso del ejemplo anterior es periódico en el sentido que todas sus trayectorias lo son. No debe por ello sorprendernos que el proceso sea cicloestacionario. Hay procesos con un comportamiento cíclico que no tienen ese tipo de trayectorias y que no pudiendo por ello ser cicloestacionarios, son WSC. Veamos un ejemplo.

Ejemplo 4.11 *Un modem transmite señales binarias 0 y 1 IID de la siguiente forma,*

- *para transmitir un 1, emite una señal rectangular de amplitud 1 y duración T,*
- *para transmitir un 0, emite una señal rectangular de amplitud -1 y duración T.*

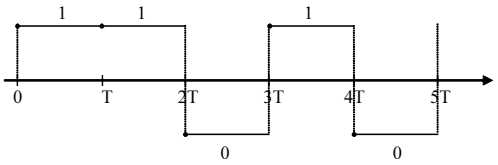


Figura 4.8: Una trayectoria del proceso de pulsos modulados

La Figura 4.8 muestra la trayectoria correspondiente a la secuencia de datos 11010 ....

El proceso puede ser descrito mediante

$$X_n = \sum_{n=-\infty}^{+\infty} A_n \mathbf{1}_{[0,T[}(t - nT),$$

donde $A_n = \pm 1$ según el valor a transmitir y $\mathbf{1}_{[0,T[}(\cdot)$ es la función indicatriz del intervalo $[0, T]$.

La media del proceso es 0 porque $E(A_n) = 0, \forall n$. La función de autocovarianza vale

$$\begin{aligned} C(t_1, t_2) &= E\left[\left(\sum_{n=-\infty}^{+\infty} A_n \mathbf{1}_{[0,T[}(t_1 - nT)\right)\left(\sum_{m=-\infty}^{+\infty} A_m \mathbf{1}_{[0,T[}(t_2 - mT)\right)\right] && (4.59) \\ &= E\left[\sum_{n=-\infty}^{+\infty} A_n^2 \mathbf{1}_{[0,T[}(t_1 - nT)\mathbf{1}_{[0,T[}(t_2 - nT)\right] && (4.60) \\ &= \begin{cases} 1, & \text{si } (t_1, t_2) \in [nT, (n-1)T[\times[nT, (n-1)T[; \\ 0, & \text{en el resto.} \end{cases} \end{aligned}$$

El paso de (4.59) a (4.60) se debe a la independencia de las A_n. La Figura 4.9 muestra la región del plano (celdas en gris) en la que $C(t_1, t_2)$ vale la unidad, de donde se deduce que $C(t_1 + kT, t_2 + kT) = C(t_1, t_2)$ y el proceso es WSC.

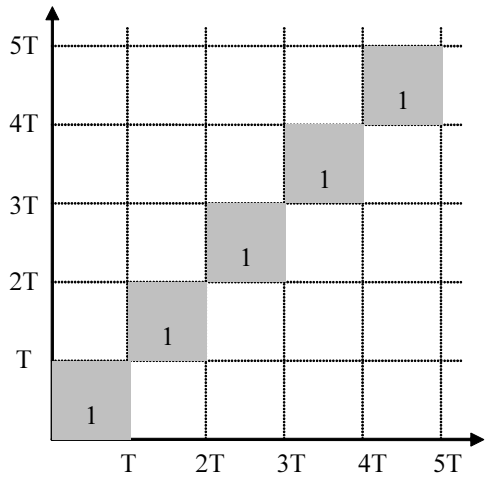


Figura 4.9: La función de autocovarianza del proceso de pulsos modulados

Capítulo 5

Transformación lineal de un proceso estacionario

5.1. Densidad espectral de potencia (PSD) de un proceso WSS

El desarrollo en serie de de Fourier de una función determinista permite expresarla como una suma ponderada de funciones periódicas. Si la función no presenta cambios bruscos y varía suavemente, los coeficientes (pesos) de la serie correspondientes a las sinusoides de baja frecuencia presentan valores elevados frente a los de las altas frecuencias. Funciones muy variables muestras una distribución de coeficientes opuesta. Ello supone que la tasa de variación de la función está relacionada con el valor de los distintos coeficientes, que considerados ellos mismos como una función recibe el nombre de *espectro* de la función original.

La idea que subyace en este tipo de descomposición es representar la función en el dominio de las frecuencias y es trasladable a los procesos estocásticos, convirtiéndose en una herramienta de gran utilidad como a continuación veremos. Pero la traslación no puede hacerse de forma directa debido a la diferencia existente entre los fenómenos deterministas y aleatorios. Como ya sabemos, las distintas realizaciones de un proceso dan lugar a diferentes funciones (trayectorias) y por esta razón el *espectro* de un proceso estocástico expresa la variación media de todas aquellas trayectorias a lo largo del tiempo.

Por otra parte, conviene recordar ahora las propiedades que la función de autocorrelación de un proceso WSS poseía y que están enumeradas y demostradas en la página 104. En concreto la PA3), en virtud de la expresión (4.56), mostraba cómo la función de autocorrelación mide la tasa de cambio del proceso en términos de probabilidad.

$$P(|X_{t+\tau} - X_t| > \varepsilon) \leq \frac{2[R(0) - R(\tau)]}{\varepsilon^2}.$$

De la síntesis entre esta propiedad y la descomposición en serie de Fourier surge el concepto de *densidad espectral de potencia* (PSD de sus siglas en inglés) que introducimos y estudiamos a continuación, distinguiendo entre procesos discretos y continuos en el tiempo.

5.1.1. PSD para procesos estocásticos WSS discretos en el tiempo

Definición 5.1 *Sea X_t un proceso estocástico WSS discreto en el tiempo, su espectro o densidad espectral de potencia, $P(\omega)$, es la transformada de Fourier de su función de autocorrelación*

$R(k)$,

$$P(\omega) = \mathcal{F}[R(k)] = \sum_{k=-\infty}^{k=+\infty} R(k) \exp(-i2\pi\omega k), \quad -\infty < \omega < \infty. \tag{5.1}$$

Si tenemos en cuenta que $R(k) = R(-k)$ y $e^{-ix} + e^{ix} = 2\cos x$, la expresión (5.1) adopta la forma

$$P(\omega) = R(0) + 2\sum_{k\geq 1} R(k) \cos 2\pi\omega k, \tag{5.2}$$

lo que implica que $P(\omega)$ es una función real par, $P(\omega) = P(-\omega)$. De esta última expresión para $P(\omega)$ se deduce que es periódica con periodo 1, por lo que sólo consideraremos el rango de frecuencias $-1/2 < \omega \leq 1/2$; pero dada su paridad bastará con definir $P(\omega)$ para $\omega \in [0, 1/2]$.

Podemos definir el *espectro normalizado* dividendo (5.2) por $R(0)$, obtendremos

$$P^*(\omega) = 1 + 2\sum_{k\geq 1} \frac{R(k)}{R(0)} \cos 2\pi\omega k. \tag{5.3}$$

En el caso de tratarse de un proceso centrado con $\mu(n) = 0, \forall n$, el cociente

$$\frac{R(k)}{R(0)} = \frac{C(k)}{C(0)} = \rho(k),$$

es la función de correlación y (5.3) se escribe,

$$P^*(\omega) = 1 + 2\sum_{k\geq 1} \rho(k) \cos 2\pi\omega k. \tag{5.4}$$

Ejemplo 5.1 (Espectro de un ruido blanco) *Recordemos que si el proceso $\{X_t\}$ es un ruido blanco, $\mu(t) = 0$, $\sigma^2(t) = \sigma^2$, $\forall t$ y las variables son incorreladas (algunos autores exigen independencia). Se trata, evidentemente, de un proceso WSS en el que*

$$C(k) = R(k) = \begin{cases} \sigma^2, & si\ k = 0; \\ 0, & si\ k \neq 0. \end{cases}$$

Sustituyendo en (5.2) tendremos

$$P(\omega) = \sigma^2,$$

o bien su versión normalizada,

$$P^*(\omega) = 1.$$

En cualquiera de sus formas, el espectro de un ruido blanco es constante para cualquier frecuencia ω, a semejanza de lo que ocurre con el espectro óptico plano de la luz blanca, lo que justifica el nombre que reciben este tipo de procesos.

PSD versus R(k)

La definición de $P(\omega)$ como $\mathcal{F}[R(k)]$ permite recuperar la función de autocorrelación si $P(\omega)$ es conocida. Basta para ello recurrir a la transformada inversa de Fourier,

$$R(k) = \mathcal{F}^{-1}[P(\omega)] = \int_{-1/2}^{1/2} P(\omega) \exp(i2\pi k\omega) d\omega = 2\int_0^{1/2} P(\omega) \cos(2\pi k\omega) d\omega. \tag{5.5}$$

Un resultado conocido como el Teorema de Wold (pueden consultarse detalles en el texto de Priestley (1981)), afirma que cualquier función $P(\omega)$ no negativa definida sobre $[0, 1/2]$ define un espectro y, en consecuencia, $R(k)$ es una función de autocorrelación si y solo si puede expresarse de la forma (5.5) a partir de una de estas $P(\omega)$. Se deriva de ello que las condiciones a cumplir para ser función autocorrelación son mucho más restrictivas que la exigidas para el espectro.

Como herramientas matemáticas $P(\omega)$ y $R(k)$ son equivalentes por lo que respecta a la información que contienen del proceso subyacente. No obstante, como ya señalábamos al comienzo del Capítulo, el espectro permite conocer e interpretar las componentes cíclicas del proceso. Veamos un ejemplo que evidencia este aspecto.

Ejemplo 5.2 (Espectro de un proceso autoregresivo de orden uno, AR(1)) *Un proceso estocástico AR(1) se define mediante,*

$$X_t = \alpha X_{t-1} + Z_t, \tag{5.6}$$

donde $\{Z_t\}$ es una sucesión de ruido blanco y $|\alpha| < 1$. La restricción para el valor de α es necesaria para que el proceso sea WSS (ver la Sección 5.1 de Montes (2007)).

Para calcular $R(k)$ multiplicamos ambas partes de (5.6) por X_{t-k} y al tomar esperanzas se obtiene,

$$R(k) = \alpha R(k-1),$$

y recursivamente,

$$R(k) = \alpha^k R(0) = \alpha^k \gamma^2,$$

con $R(k) = R(-k), k = 0, 1, 2, \ldots$, siendo γ^2 la varianza del proceso. Su espectro vale

$$\begin{aligned} P(\omega) &= \gamma^2 \sum_{-\infty}^{+\infty} \alpha^k \exp(-i2\pi\omega k) \\ &= \gamma^2 + \gamma^2 \sum_{k\geq 1} \alpha^k \exp(-i2\pi\omega k) + \gamma^2 \sum_{k\geq 1} \alpha^k \exp(i2\pi\omega k) \\ &= \gamma^2 \left(1 + \frac{\alpha \exp(-i2\pi\omega)}{1 - \alpha \exp(-i2\pi\omega)} + \frac{\alpha \exp(i2\pi\omega)}{1 - \alpha \exp(i2\pi\omega)}\right). \end{aligned}$$

Operando y teniendo en cuenta que $e^{-ix}e^{ix} = 1$ y que $e^{-ix} + e^{ix} = 2\cos x$ se obtiene finalmente,

$$P(\omega) = \frac{\gamma^2(1-\alpha^2)}{1 - 2\alpha\cos 2\pi\omega + \alpha^2}. \tag{5.7}$$

A partir de (5.6) podemos obtener γ^2 en función de $\sigma^2 = var(Z_t)$,

$$\gamma^2 = \alpha^2\gamma^2 + \sigma^2 \quad \Longrightarrow \quad \gamma^2 = \frac{\sigma^2}{1-\alpha^2}.$$

Sustituyendo en (5.7) se obtiene como expresión alternativa para $P(\omega)$,

$$P(\omega) = \frac{\sigma^2}{1 - 2\alpha\cos 2\pi\omega + \alpha^2}. \tag{5.8}$$

En la Figura 5.1 hemos representado (5.8) para $\sigma^2 = 1$ y $\alpha = -0{,}5; 0{,}5; 0{,}9$ y una trayectoria de cada uno de estos procesos. El espectro para $\alpha = -0{,}5$ (superior) es creciente lo que significa un mayor peso de las altas frecuencias, la potencia se concentra en ellas, y ello se corresponde con una trayectoria que alterna valores positivos y negativos con rápidas oscilaciones entre unos y otros. Para valores positivos de α los espectros son decrecientes, indicando predominio de bajas frecuencias, más acusado para $\alpha = 0{,}9$ (inferior). Las trayectorias correspondientes muestran oscilaciones menos frecuentes con largas permanencias de valores de un mismo signo.

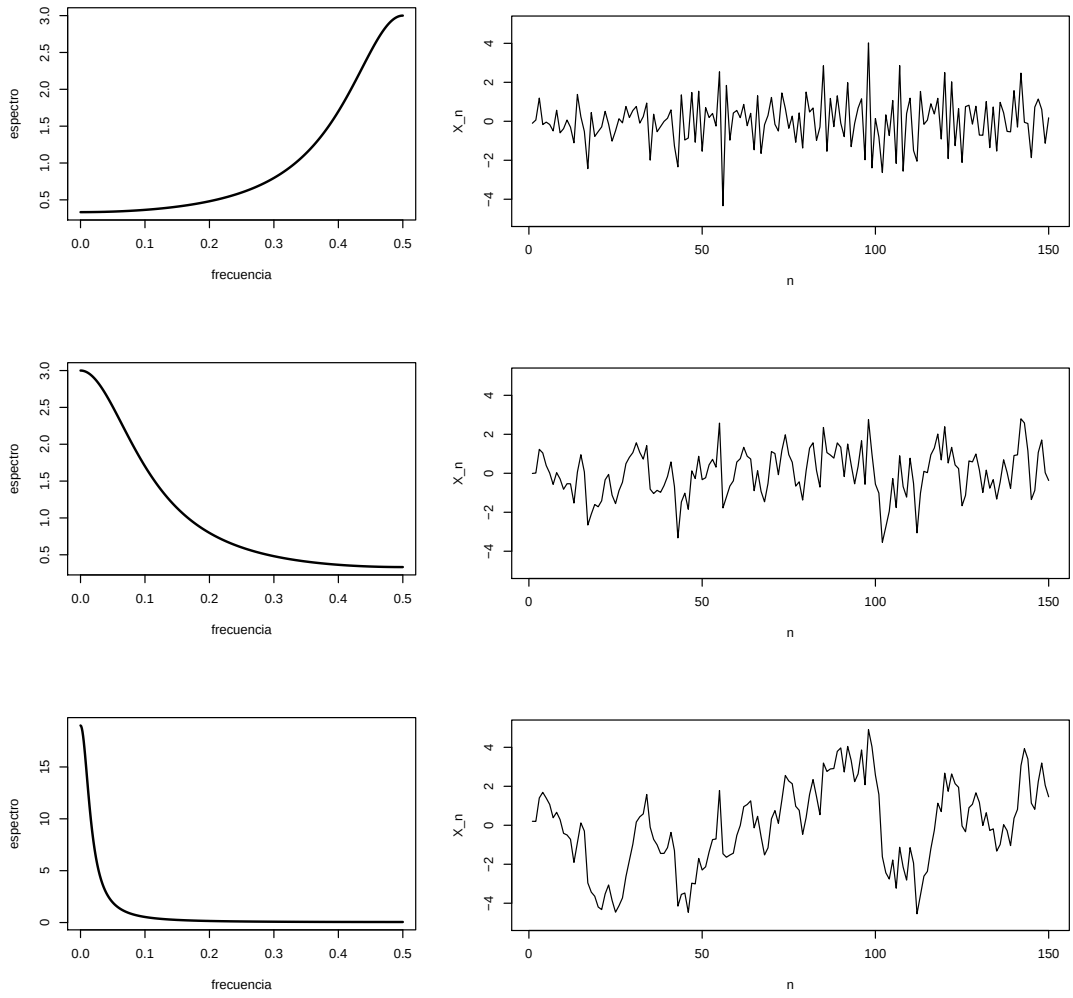


Figura 5.1: Espectro y trayectoria del proceso AR(1) para $\alpha = -0{,}5$ (superior), $\alpha = 0{,}5$ (central) y $\alpha = 0{,}9$ (inferior)

La existencia del espectro está supeditada a la convergencia de la serie en (5.1). Hay procesos en lo que ello no ocurre. Veamos un ejemplo y cómo resolver el problema.

Ejemplo 5.3 (Un espectro divergente) *Consideremos el proceso*

$$X_t = A\cos\theta t + B\sin\theta t + Z_t, \tag{5.9}$$

con $\{Z_t\}$ *una sucesión de ruido blanco con varianza* σ^2*, y* A *y* B *sendas variables aleatorias IID com media 0 y varianza* γ^2 *e independientes de* Z_t*. La media del proceso será por tanto* $E(X_t) = 0$ *y su varianza,*

$$var(X_t) = E(A^2)\cos^2\theta t + E(B^2)\sin^2\theta t + E(Z_t^2) = \gamma^2 + \sigma^2.$$

La función de autocorrelación, para $t \neq s$,

$$\begin{aligned} R(t,s) &= E\{[A\cos\theta t + B\sin\theta t + Z_t][A\cos\theta s + B\sin\theta s + Z_s]\} \\ &= E(A^2)\cos\theta t\cos\theta s + E(B^2)\sin\theta t\sin\theta s \\ &= \gamma^2\cos\theta(t-s). \end{aligned}$$

En definitiva,

$$R(k) = \begin{cases} \gamma^2 + \sigma^2, & si\ k = 0; \\ \gamma^2\cos k\theta, & k \neq 0, \end{cases}$$

y el proceso es WSS. Su espectro vale,

$$\begin{aligned} P(\omega) &= \gamma^2 + \sigma^2 + 2\gamma^2\sum_{k\geq 1}\cos(k\theta)\cos(2\pi\omega k) \\ &= \gamma^2 + \sigma^2 + 2\gamma^2\sum_{k\geq 1}[\cos k(\theta + 2\pi\omega) + \cos k(\theta - 2\pi\omega)], \end{aligned}$$

que para $\omega = \theta/2\pi$ diverge y no existe.

Si se observa la expresión (5.1) constataremos que la definición de $\mathcal{F}[R(k)]$ sólo es válida si la serie es convergente. Si la serie es divergente, como es ahora el caso, el problema se resuelve recurriendo a una definición más general de $\mathcal{F}[R(k)]$ que involucra el concepto de distribución, y una desarrollo matemático fuera del alcance y las pretensiones de estas notas. Un proceso discreto en el tiempo puede expresarse mediante una suma infinita de δ's de Dirac y podemos aplicarle la definición general de transformada de Fourier para distribuciones, puesto que la delta de Dirac es una distribución. Ambas definiciones coinciden si la serie es convergente, como era de esperar para una definición que queremos coherente. A efectos prácticos, en una situación como la que nos encontramos admitiremos que $P(\omega)$ puede ser infinito en valores aislados que concentran la potencia del proceso. Lo expresamos recurriendo a la función δ,

$$P(\omega) = \gamma^2 + \sigma^2 + 2\gamma^2\delta(\omega - \theta/2\pi).$$

Filtros lineales

Un *filtro* es simplemente una transformación de un proceso X_t para dar lugar a otro proceso Y_t. Si la transformación es de la forma,

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}, \tag{5.10}$$

hablamos de un *filtro lineal*. Cada elemento del proceso resultante es una combinación lineal, eventualmente infinita, de elementos del proceso original. Si hay un número finito de a_j distintos de cero y X_t es WSS, también lo será Y_t, pero en el caso de que la combinación lineal sea estrictamente infinita, la WSS de Y_t depende de los a_j.

Si ambos procesos son WSS, es interesante establecer la relación entre $R_X(k)$ y $R_Y(k)$ y $P_X(\omega)$ y $P_Y(\omega)$. Comencemos por la función de autocorrelación,

$$\begin{aligned} R_Y(k) &= E(Y_t Y_{t-k}) \\ &= E\left[\left(\sum_{l=-\infty}^{\infty} a_l X_{t-l}\right)\left(\sum_{j=-\infty}^{\infty} a_j X_{t-k-j}\right)\right] \\ &= \sum_{l=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-l). \end{aligned} \tag{5.11}$$

Para obtener la relación entre las funciones de densidad espectral bastará calcular la transformada de Fourier de ambas partes de (5.11),

$$\begin{aligned} P_Y(\omega) &= \sum_{k=-\infty}^{k=+\infty} R_Y(k)\exp(-i2\pi\omega k) \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-l)e^{-i2\pi\omega k} \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_l a_j R_X(k+j-i)e^{-i2\pi\omega l}e^{i2\pi\omega j}e^{-i2\pi\omega(k+j-l)} \\ &= \sum_{l=-\infty}^{\infty} a_l e^{-i2\pi\omega l}\sum_{j=-\infty}^{\infty} a_j e^{i2\pi\omega j}\sum_{k=-\infty}^{\infty} R_X(k+j-i)e^{-i2\pi\omega(k+j-l)}. \end{aligned} \quad (5.12)$$

Si hacemos $h(\omega)=\sum_{l=-\infty}^{\infty} a_l e^{-i2\pi\omega l}$ y un adecuado cambio de índices en el último sumatorio, (5.12) puede escribirse,

$$P_Y(\omega) = |h(\omega)|^2 P_X(\omega). \quad (5.13)$$

La función $h(\omega)$ recibe el nombre de *función de transferencia* del filtro lineal. Obsérvese que la relación entre las funciones de densidad espectral es mucho más sencilla que la que liga las funciones de autocorrelación, (5.11).

Ejemplo 5.4 (Un proceso MA(3) de un proceso AR(1)) *A partir del proceso AR(1),*

$$X_t = \alpha X_{t-1} + Z_t,$$

con Z_t un ruido blanco de varianza σ^2, definimos un nuevo proceso

$$Y_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1}),$$

un proceso de medias móviles de orden 3, MA(3), que no es más que un filtro lineal con $a_j = 1/3, j=-1,0,1$ y $a_j = 0$ para cualquier otro j.

Para obtener $P_Y(\omega)$, recordemos el espectro del proceso AR(1) que obtuvimos en el ejemplo 5.2,

$$P_X(\omega) = \frac{\sigma^2}{1-2\alpha\cos 2\pi\omega + \alpha^2}.$$

Por otra parte, la función de transferencia vale,

$$h(\omega) = \frac{1}{3}(e^{i2\pi\omega} + 1 + e^{-i2\pi\omega}) = \frac{1}{3}(1+2\cos 2\pi\omega),$$

y

$$|h(\omega)|^2 = \frac{1}{9}(1+4\cos 2\pi\omega + 4\cos^2 2\pi\omega) = \frac{1}{9}(3+4\cos 2\pi\omega + 2\cos 4\pi\omega).$$

Finalmente,

$$P_Y(\omega) = \frac{\sigma^2(3+4\cos 2\pi\omega + 2\cos 4\pi\omega)}{9(1-2\alpha\cos 2\pi\omega+\alpha^2)}.$$

La Figura 5.2 muestra los espectros de ambos procesos y la función de transferencia que los relaciona, para $\sigma^2 = 1$ y $\alpha = -0{,}5$. El efecto del filtrado del proceso original es disminuir la varianza, menor área bajo el espectro, como resultado del suavizado que las medias móviles implican. El suavizado se evidencia también en la mayor presencia de bajas frecuencias en el espectro de Y_t.

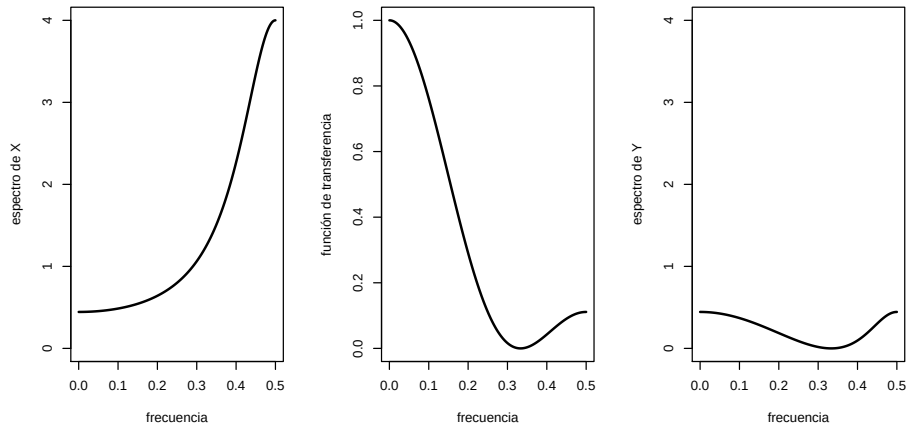


Figura 5.2: Espectros y función de transferencia de los procesos MA(3) y AR(1)

Filtrado lineal de un ruido blanco: proceso lineal general

Un caso particular de (5.10) es el que resulta de filtrar una sucesión de ruido blanco, Z_t, con varianza σ^2,

$$X_t = \sum_{j=0}^{\infty} a_j Z_{t-j}. \tag{5.14}$$

Recordemos que $P_Z(\omega) = \sigma^2$ y

$$h(\omega) = \sum_{j=0}^{\infty} a_j e^{-i2\pi\omega j}.$$

El espectro del nuevo proceso valdrá

$$\begin{aligned}
P_X(\omega) &= \sigma^2 |h(\omega)|^2 \\
&= \sigma^2 \left(\sum_{j=0}^{\infty} a_j e^{-i2\pi\omega j} \sum_{k=0}^{\infty} a_k e^{i2\pi\omega k} \right) \\
&= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + \sum \sum_{j \neq k} a_j a_k e^{-i2\pi\omega(k-j)} \right) \\
&= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + 2 \sum_{k \geq 1} \sum_{j<k} a_j a_k \cos 2\pi\omega(k-j) \right) \\
&= \sigma^2 \left(\sum_{j=0}^{\infty} a_j^2 + 2 \sum_{m \geq 1} \sum_{k>m} a_{k-m} a_k \cos 2\pi\omega m \right) \\
&= b_0 + \sum_{m \geq 1} b_m \cos 2\pi\omega m,
\end{aligned} \tag{5.15}$$

con $b_0 = \sigma^2 \sum_{\geq 0} a_j^2$ y $b_m = \sum_{k>m} a_{k-m} a_k$.

Este proceso se denomina *proceso lineal general*, y el caso de mayor interés en la práctica es el que se obtiene cuando sólo un número finito de coeficientes a_j son distintos de 0. Los procesos ARMA, autoregresivos de medias móviles, son sin duda los más interesantes de entre los que cumplen esta condición. En la Sección 5.1 de Montes (2007) nos ocupamos de ellos.

La WSS de X_t no está garantizada aún cuando Z_t lo sea. Sólo lo estaría en el caso de una combinación lineal finita. El caso más general requiere una condición sobre los a_j. Para obtenerla,

$$\begin{aligned} R(t,s) &= E(X_t X_s) \\ &= E\left[\sum_{j=0}^{\infty} a_j Z_{t-j} \sum_{l=0}^{\infty} a_l Z_{s-l}\right] \\ &= \sum_{j=0}^{\infty}\sum_{l=0}^{\infty} a_j a_l E(Z_{t-j} Z_{s-l}). \end{aligned} \tag{5.16}$$

pero siendo Z_t un ruido blanco, $E(Z_{t-j}Z_{s-l}) \neq 0$ sólo si $t - j = s - l$, por lo que (5.16) se escribirá, si $t \leq s$,

$$R(t,s) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+s-t}. \tag{5.17}$$

Por (5.14) $E(X_t) = 0, \forall t$, y la expresión (5.17) evidencia que $R(t,s) = R(s-t) = R(k)$, con lo que el proceso será WSS si la $\sum_{j=0}^{\infty} a_j a_{j+k}$ es convergente $\forall k$. Ello supone que para $k = 0$

$$\sigma_{X_t}^2 = R(t,t) = R(0) = \sum_{j=0}^{\infty} a_j^2 < \infty. \tag{5.18}$$

Por otra parte, si $R(0)$ es finito, como $|\rho_{X_t X_{t-k}}| \leq 1$, tendremos

$$|\rho_{X_t X_{t-k}}| = \left|\frac{R(k)}{\sqrt{\sigma_{X_t}^2 \sigma_{X_{t-k}}^2}}\right| = \left|\frac{R(k)}{R(0)}\right| \leq 1,$$

y despejando

$$|R(k)| \leq |R(0)| < \infty.$$

En definitiva, *la condición necesaria y suficiente para que el proceso lineal general sea WSS es que* $\sum_{j=0}^{\infty} a_j^2 < \infty$.

5.1.2. PSD para procesos estocásticos WSS continuos en el tiempo

Definición 5.2 *Sea X_t un proceso estocástico WSS continuo en el tiempo, su densidad espectral de potencia, $P(\omega)$, es la transformada de Fourier de su función de autocorrelación $R(\tau)$,*

$$P(\omega) = \mathcal{F}[R(\tau)] = \int_{-\infty}^{+\infty} R(\tau) \exp(-i2\pi\omega\tau) d\tau, \quad -\infty < \omega < \infty. \tag{5.19}$$

En realidad $P(\omega)$ se define como

$$\lim_{T\to\infty} \frac{1}{T} E\left[\left|\int_{-T/2}^{T/2} X_t \exp(-i2\pi\omega\tau) d\tau\right|^2\right], \tag{5.20}$$

pero un resultado conocido como el teorema de Einstein-Wiener-Khinchin establece la igualdad entre (5.20) y (5.19). De ambas definiciones y de las propiedades de $R(\tau)$ se derivan las siguientes propiedades para $P(\omega)$, análogas, como no podía ser de otra forma, a las que hemos visto en el caso discreto.

PSDc1) De (5.20) se deduce que

$$P(\omega) \geq 0, \quad \forall \omega.$$

PSDc2) En la propiedad **PA2)** de la página 104, vimos que la función de autocorrelación de un proceso WSS es una función par, $R(\tau) = R(-\tau)$, en consecuencia la PSD es una función real,

$$\begin{aligned} P(\omega) &= \int_{-\infty}^{+\infty} R(\tau) \exp(-i2\pi\omega\tau) d\tau \\ &= \int_{-\infty}^{+\infty} R(\tau)(\cos 2\pi\omega\tau - i \sin 2\pi\omega\tau) d\tau \qquad (5.21) \\ &= 2\int_{0}^{+\infty} R(\tau) \cos 2\pi\omega\tau d\tau, \qquad (5.22) \end{aligned}$$

donde el paso de (5.21) a (5.22) se justifica porque la función *seno* es una función impar.

PSDc3) La expresión (5.22) y la paridad de la $R(\tau)$ y de la función *coseno* implican la paridad de PSD, $P(\omega) = P(-\omega)$.

Recordemos la relación entre las funciones de autocovarianza y autocorrelación en un proceso WSS,

$$R(\tau) = C(\tau) + \mu^2.$$

Al calcular la PSD,

$$\begin{aligned} P(\omega) &= \mathcal{F}[R(\tau)] = \mathcal{F}[C(\tau) + \mu^2] \\ &= \mathcal{F}[C(\tau)] + \mathcal{F}[\mu^2] \\ &= \mathcal{F}[C(\tau)] + \mu^2\delta(\omega), \end{aligned}$$

donde $\delta(\omega)$ es la función *delta* de Dirac, que es la transformada de Fourier de una función constante e igual a 1 en todo $\mathcal{R}$. El comentario que hicimos al final del Ejemplo 5.3 es válido también ahora.

Observación 5.1 (El fenómeno de *aliasing*) *Al enumerar las propiedades de la PSD para el caso continuo hemos mencionado su analogía con las del caso discreto, pero se observará que ahora la periodicidad está ausente. En la expresión (5.19) los límites de la integral son $-\infty$ y $+\infty$. La razón para ello es que cuando observamos un proceso en tiempo continuo es posible identificar cualquier componente por elevada que sea su frecuencia, por el contrario, si la observación se lleva a cabo a intervalos de tiempo, los efectos debidos a frecuencias superiores a $1/2$ o inferiores a $-1/2$ son indistinguibles de los efectos de baja frecuencia. Este fenómeno se conoce con el nombre de "aliasing". Para ilustrarlo consideremos la función*

$$x(t) = \cos 2\pi t\omega,$$

con $\omega = \omega_0 + 1/2$ *y* $0 < \omega_0 < 1/2$. *La función es observada sólo para tiempos enteros. Si tenemos en cuenta que* $\cos x = \cos(x + 2\pi t)$ *para* t *entero, y que* $\cos x = \cos(-x)$, *las distintas observaciones de* $x(t)$ *cumplen,*

$$\begin{aligned} x(t) &= \cos 2\pi t\omega \\ &= \cos(2\pi t\omega - 2\pi t) \\ &= \cos[2\pi t(\omega_0 + 1/2) - 2\pi t] \\ &= \cos(2\pi t\omega_0 - \pi t) \\ &= \cos(\pi t - 2\pi t\omega_0) \\ &= \cos(2\pi t\omega'), \end{aligned}$$

con $\omega' = 1/2 - \omega_0$ *que verifica* $0 < \omega' < 1/2$. *En definitiva, las frecuencias* $\omega > 1/2$ *no se distinguen de las frecuencias* $\omega' < 1/2$. *La Figura 5.3 muestra el fenómeno, se observa en ella que al muestrear* s *veces por segundo, las muestras de una y otra sinuoside se confunden. En procesamiento de señales se dice que cada una de las sinusoides se convierte en un* alias *para la otra, de ahí el nombre que este fenómeno recibe.*

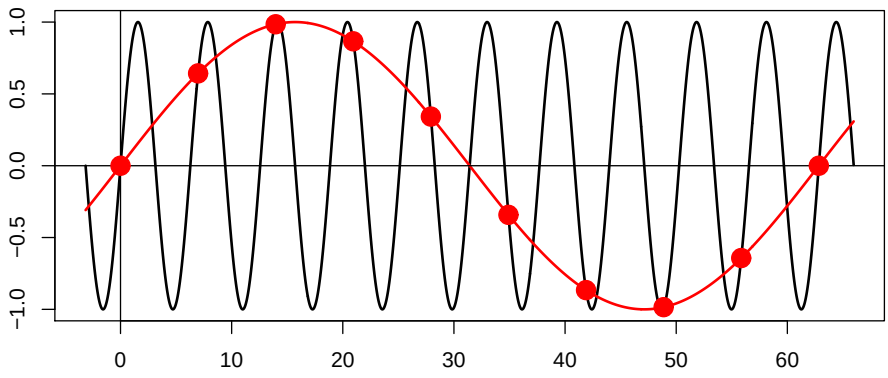


Figura 5.3: El fenómeno de *aliasing*

PSD versus R(k)

Si la PSD es conocida, la transformada inversa de Fourier permite recuperar la función de autocorrelación,

$$R(\tau) = \mathcal{F}^{-1}[P(\omega)] = \int_{-\infty}^{+\infty} P(\omega)\exp(i2\pi\omega\tau)d\omega. \tag{5.23}$$

De la definición de $R(\tau)$ se deducía que $R(0) = E(X_t^2)$. Aplicando (5.23) podemos calcular esta esperanza, conocida como *potencia media de* X_t, mediante,

$$E(X_t^2) = R(0) = \int_{-\infty}^{+\infty} P(\omega)d\omega. \tag{5.24}$$

A continuación se obtiene la PSD para algunos de los procesos estudiados en el Capítulo anterior.

Ejemplo 5.5 (PSD del proceso RTS) *Recordemos que el proceso RTS es estacionario, y por tanto WSS, y que su función de autocorrelación veien dada por (4.30),*

$$R(\tau) = a^2 e^{-2\lambda|\tau|},$$

donde λ es el parámetro del proceso de Poisson subyacente. Para obtener $P(\omega)$,

$$\begin{aligned} P(\omega) &= \int_{-\infty}^{+\infty} a^2 e^{-2\lambda|\tau|} e^{-i2\pi\omega\tau} d\tau \\ &= a^2 \int_{-\infty}^{0} e^{\tau(2\lambda - i2\pi\omega)} d\tau + a^2 \int_{0}^{+\infty} e^{-\tau(2\lambda + i2\pi\omega)} d\tau \\ &= \frac{a^2}{2\lambda - i2\pi\omega} + \frac{a^2}{2\lambda + i2\pi\omega} \\ &= \frac{\lambda a^2}{\lambda^2 + \pi^2\omega^2}. \end{aligned} \tag{5.25}$$

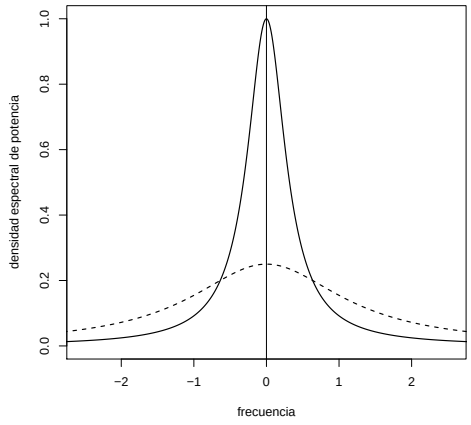


Figura 5.4: Densidad espectral de potencia del proceso RTS para $\lambda = 1$ (—) y $\lambda = 4$ (- - -)

La Figura 5.4 muestra la gráfica de la densidad espectral de potencia para sendos procesos RTS con $\lambda = 1$ y $\lambda = 4$. Como el máximo de (5.25) se alcanza en $\omega = 0$ y su valor es $P(0) = a^2/\lambda$, para valores de λ pequeños los mayores valores de la función se concentran en las bajas frecuencias, pero a medida que λ aumenta las frecuencias altas adquieren mayor peso. Todo ello se corresponde con lo que λ representa en el proceso. Valores pequeños de λ suponen una tasa baja de cambios por unidad de tiempo y por tanto una menor frecuencia, lo contrario de lo que implican elevados valores de λ.

Ejemplo 5.6 (PSD de un ruido blanco continuo) *La gráfica de la izquierda de la Figura 5.5 muestra la PSD de un ruido blanco recortada a un rango de frecuencias $-\omega_0 \leq \omega \leq \omega_0$. Se*

observa que $P(\omega)$ es constante en todo en intervalo. Ya comentábamos en el Ejemplo 5.1 que el ruido blanco recibe este nombre por similitud con la luz blanca, cuyo espectro es constante para todas la frecuencias.

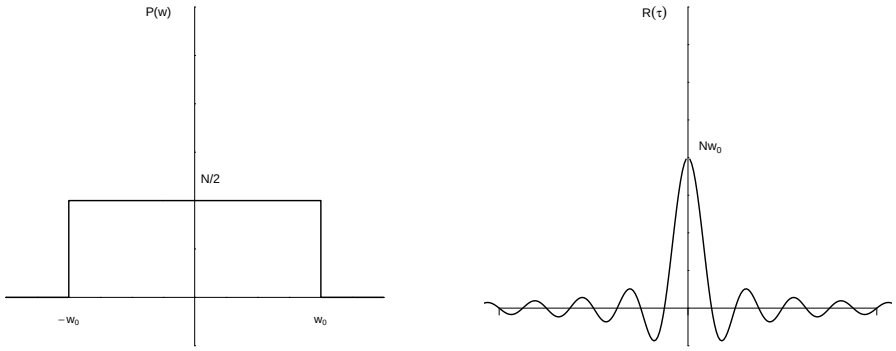


Figura 5.5: $P(\omega)$ con frecuencias acotadas para un ruido blanco (izquierda) y su correspondiente $R(\tau)$ (derecha)

Para obtener la función de autocorrelación recurrimos a (5.23),

$$\begin{aligned} R(\tau) &= \frac{N}{2}\int_{\omega_0}^{\omega_0} \exp(i2\pi\omega\tau)d\omega. \\ &= \frac{N}{2} \times \frac{e^{-i2\pi\omega_0\tau} - e^{i2\pi\omega_0\tau}}{-i2\pi\tau} \\ &= \frac{N \sin 2\pi\omega_0\tau}{2\pi\tau}. \end{aligned}$$

La gráfica de la derecha en la Figura 5.5 corresponde a $R(\tau)$, que anula en los puntos $\tau = \pm k/2\omega_0, k = 1, 2, \ldots$, lo que implica que las variables X_t y $X_{t+\tau}$ son incorreladas para dichos valores de τ.

La potencia del proceso vale,

$$E(X_t^2) = R(0) = \frac{N}{2}\int_{\omega_0}^{\omega_0} d\omega = N\omega_0,$$

que coincide con el máximo de la función $R(\tau)$, tal como se observa en la gráfica de la Figura 5.5.

A medida que $\omega_0 \to \infty$, el efecto sobre $R(\tau)$ es el que se muestra en las gráficas de la Figura 5.6. En el límite, $P(\omega)$ será constante para cualquier frecuencia, hecho que caracteriza a un ruido blanco, y su función de autocorrelación acumulará toda la potencia en 0 y valdrá por tanto infinito. Utilizando la función δ lo expresaríamos,

$$R(\tau) = \frac{N}{2}\delta(\tau).$$

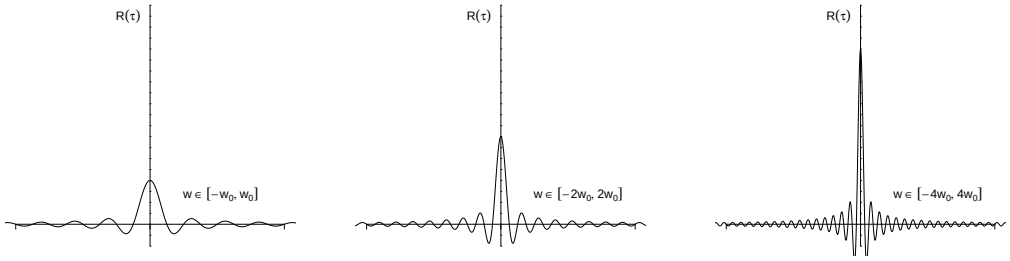


Figura 5.6: Efecto del crecimiento de ω_0 sobre $R(\tau)$

Discretización temporal de un ruido blanco continuo

El ejemplo anterior ilustra la obtención de $R(\tau)$ para un ruido blanco como límite, cuando $\omega_0 \to \infty$, de las $R_{\omega_0}(\tau)$ correspondientes a ruidos blancos filtrados del original mediante una banda de frecuencias limitadas por $|\omega| \leq \omega_0$. Pero tiene un interés añadido, se trata de un modelo utilizado habitualmente en el tratamiento de señales porque todos los sistemas físicos, con raras excepciones, presentan una banda de frecuencias limitada. Esta limitación real simplifica la descripción matemática de los procesos sin sacrificar su realismo.

Pero las simplificaciones no acaban con un filtrado paso bajo. Una segunda e inevitable simplificación es la que supone muestrear un proceso continuo para obtener un *proceso discreto* en el tiempo con el que realmente trabajaremos. Hemos pasado del proceso X_t a la sucesión X_n. Si originalmente X_t era un ruido blanco, ¿cuáles son las características de la X_n?, ¿es también una sucesión de ruido blanco?

El nuevo proceso X_n, antes del filtrado del original, puede representarse de la forma,

$$X_n = X_t|_{t=nT}, \quad -\infty < n < +\infty,$$

donde $T = f_0^{-1}$ es el periodo de muestreo y f_0 la frecuencia de muestreo. Si queremos reconstruir la señal original a partir de la muestreada, un conocido teorema exige que f_0 sea mayor o igual que $2\omega_0$, la frecuencia de Nyquist.

Es inmediato comprobar que al igual que el proceso original, $E(X_n) = 0$. Calculemos ahora $R(k)$ para comprobar si la sucesión es también WSS.

$$\begin{aligned} R_{X_n}(k) &= E(X_n X_{n+k}) \\ &= E(X_{nT} X_{(n+k)T}) \\ &= E(X_{nT} X_{nT+kT}) \\ &= R_{X_t}(kT), \end{aligned} \tag{5.26}$$

que al no depender de n supone que X_n es WSS. Además, de (5.26) se deduce que $R_{X_n}(k)$ es, a su vez, el resultado de muestrear la autocorrelación del proceso original con el mismo periodo T con el que se obtuvo X_n. En concreto, si $f_0 = 2\omega_0$ el periodo valdrá $T = (2\omega_0)^{-1}$ y

$$R_{X_n}(k) = R_{X_t}[k(2\omega_0)^{-1}],$$

pero en ejemplo anterior vimos que para un ruido blanco con densidad espectral de potencia

constante e igual a $N/2$,

$$R_{X_t}(\tau) = \frac{N\omega_0 \sin 2\pi\omega_0\tau}{2\pi\omega_0\tau},$$

que como ya dijimos se anula en $\tau = \pm k(2\omega_0)^{-1}$, $k = 1, 2, \ldots$, que son lo puntos donde está definida R_{X_n}. En definitiva,

$$R_{X_n}(k) = \begin{cases} N\omega_0, & \text{si } k = 0; \\ 0, & k \neq 0, \end{cases}$$

que es la autocorrelación de una sucesión de ruido blanco con $\sigma^2 = N\omega_0$.

Ejemplo 5.7 (Señal binaria asíncrona (ABS)) *Una señal binaria asíncrona es una sucesión constituida por sucesivos impulsos de duración T y de amplitud aleatoria X_n, una variable que toma los valores $\pm a$ con igual probabilidad. La sucesión es asíncrona porque el instante de inicio, D, es una variable $U \sim [-T/2, T/2[$. Como consecuencia de ello, dos tiempos cualesquiera, $t_1 < t_2$, tales que $t_2 - t_1 < T$, pueden no coincidir en el mismo impulso.*

A partir de una sucesión de estas características se define el proceso ABS, continuo en el tiempo, mediante la expresión

$$X_t = \sum_n X_n \mathbf{1}_{[D+nT-\frac{T}{2}, D+nT+\frac{T}{2}]}(t), \tag{5.27}$$

donde $\mathbf{1}_A(\cdot)$ es la función característica del conjunto A, de manera que

$$\mathbf{1}_A(s) = \begin{cases} 1, & si\ s \in A; \\ 0, & si\ s \in A^c. \end{cases}$$

Una trayectoria del proceso se muestra en la Figura 5.7.

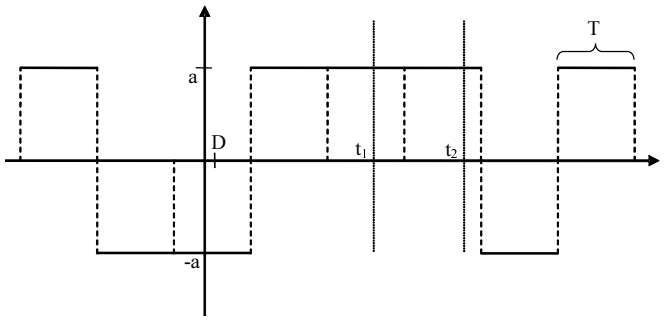


Figura 5.7: Trayectoria de un proceso ABS con $D \neq 0$

La función de autocorrelación del proceso vale,

$$\begin{aligned} R_X(t_1, t_2) &= E(X_{t_1} X_{t_2}) \\ &= E\left[\sum_n X_n \mathbf{1}_{[D+nT-\frac{T}{2}, D+nT+\frac{T}{2}]}(t_1) \sum_m X_m \mathbf{1}_{[D+mT-\frac{T}{2}, D+mT+\frac{T}{2}]}(t_2) \right]. \end{aligned}$$

Suponemos que las amplitudes de los impulsos son independientes entre sí y del desplazamiento inicial D. En consecuencia, $E(X_n X_m) = E(X_n)E(X_m) = 0$ si $n \neq m$, $E(X_n X_m) = E(X_n^2) =$

a^2 *si* $n = m$, *y*

$$\begin{aligned} R_X(t_1,t_2) &= a^2 \sum_n E\left[\mathbf{1}_{[D+nT-\frac{T}{2},D+nT+\frac{T}{2}]}(t_1)\mathbf{1}_{[D+nT-\frac{T}{2},D+nT+\frac{T}{2}]}(t_2)\right] \\ &\quad + \sum_{n\neq m}\sum_m E(X_n)E(X_m)E\left[\mathbf{1}_{[D+nT-\frac{T}{2},D+nT+\frac{T}{2}]}(t_1)\mathbf{1}_{[D+mT-\frac{T}{2},D+mT+\frac{T}{2}]}(t_2)\right] \\ &= a^2 \sum_n E\left[\mathbf{1}_{[D+nT-\frac{T}{2},D+nT+\frac{T}{2}]}(t_1)\mathbf{1}_{[D+nT-\frac{T}{2},D+nT+\frac{T}{2}]}(t_2)\right]. \end{aligned}$$

Observemos que $\mathbf{1}_A(t)$ *es una variable aleatoria discreta que toma solamente los valores 0 y 1, por tanto*

$$E\left[\mathbf{1}_A(t_1)\mathbf{1}_A(t_2)\right] = P[\mathbf{1}_A(t_1) = 1, \mathbf{1}_A(t_2) = 1],$$

lo que supone que tanto t_1 *como* t_2 *pertenecen al mismo y único intervalo* $A = [D + n_0T - \frac{T}{2}, D + n_0T + \frac{T}{2}]$. *Si* $t_1 < t_2$, *entonces* D *ha tomado un valor tal que el anterior intervalo cubre al intervalo que ambos definen,* $[t_1, t_2]$. *Si* $t_2 - t_1 = \tau$, *para que esto ocurra* $D \in [-T/2, T/2 - \tau]$. *El razonamiento es igualmente válido para el caso* $t_2 < t_1$. *En definitiva*

$$E\left[\mathbf{1}_{[D+n_0T-\frac{T}{2},D+n_0T+\frac{T}{2}]}(t_1)\mathbf{1}_{[D+n_0T-\frac{T}{2},D+n_0T+\frac{T}{2}]}(t_2)\right] = P(D \in [-T/2, T/2 - |\tau|]) = \frac{T-|\tau|}{T} = 1 - \frac{|\tau|}{T},\ |\tau| \leq T,$$

mientras que la esperanza será 0 para cualquier otro n. *La expresión final para* $R_X(\tau)$, *con* $|\tau| = t_2 - t_1$ *será*

$$R(\tau) = \begin{cases} a^2\left(1 - \frac{|\tau|}{T}\right), & |\tau| \leq T; \\ 0, & |\tau| \geq T. \end{cases}$$

Se trata de una función triangular cuya gráfica se muestra en la Figura 5.8 para $a = 1$ *y* $T = 1$.

Conocida la función de autocorrelación, podemos calcular ahora la densidad espectral de potencia del proceso ABS y, en general, de cualquier proceso que posea una $R(\tau)$ *que sea triangular (supondremos* $a = 1$*).*

$$\begin{aligned} P(\omega) &= \int_{-\infty}^{+\infty} R(\tau)\exp(-i2\pi\omega\tau)d\tau \\ &= \int_{-T}^{+T}\left(1 - \frac{|\tau|}{T}\right)\exp(-i2\pi\omega\tau)d\tau \\ &= 2\int_0^T \left(1 - \frac{|\tau|}{T}\right)\cos 2\pi\omega\tau d\tau \\ &= T\left(\frac{\sin \pi\omega T}{\pi\omega T}\right)^2. \end{aligned} \tag{5.28}$$

El resultado sugiere una forma más sencilla de obtener (5.28) si observamos que se trata del cuadrado de una función sinc. *Los pulsos rectangulares tienen por transformada de Fourier una función de estas características. Así, si la autocorrelación es un pulso rectangular unitario,*

$$R(\tau) = \begin{cases} 1, & si\ |\tau| \leq 1/2; \\ 0, & si\ |\tau| > 1/2, \end{cases}$$

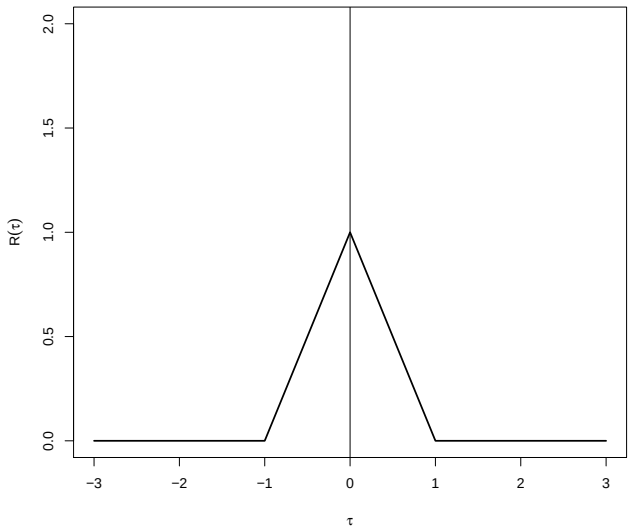


Figura 5.8: Función de autocorrelación de un proceso ABS con $a = 1$ y $T = 1$

la $P(\omega)$ asociada vale

$$P(\omega) = 2\int_0^1 \cos 2\pi\omega\tau d\tau = \frac{\sin \pi\omega}{\pi\omega}.$$

Calculemos ahora la convolución de dos pulsos pulsos rectangulares unitarios.

$$R_c(\nu) = \int_{-1/2}^{+1/2} R_1(\tau)R_2(\nu - \tau)d\tau.$$

Si observamos la Figura 5.9 vemos que para un $\nu \geq 0$, $R_2(\nu - \tau)$ ha desplazado su soporte a $[-1/2+\nu, 1/2+\nu]$ y el producto $R_1(\tau)R_2(\nu-\tau)$ es igual a la unidad en el intervalo $[-1/2+\nu, 1/2]$. El valor de $R_c(\nu)$ será el área del recinto punteado, $A = 1 - \nu$. El razonamiento es igualmente válido para $\nu < 0$ y la conclusión es que $R_c(\tau)$ es una función triangular,

$$R_c(\nu) = \begin{cases} 1 - |\nu|, & si\ |\nu| \leq 1; \\ 0, & si\ |\nu| > 1. \end{cases}$$

La conclusión es que la autocorrelación triangular es la convolución de sendas autocorrelaciones pulsos rectangulares. Pero es sabido que

$$\mathcal{F}[R_1 * R_2] = \mathcal{F}[R_1]\mathcal{F}[R_2].$$

En nuestro caso si $R_1(\tau) = R_2(\tau)$ son pulsos rectangulares de soporte $[-T/2, T/2]$ y altura $1/\sqrt{T}$, la expresión (5.28) se obtiene de inmediato.

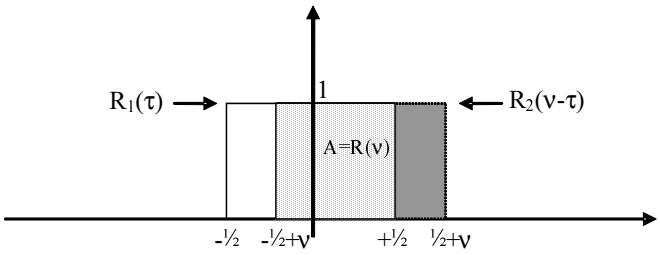


Figura 5.9: Convolución de dos pulsos rectangulares

Densidad espectral de potencia cruzada

Cuando tenemos dos procesos estocásticos, X_t e Y_t, ambos WSS, se define la función de autocorrelación cruzada mediante,

$$R_{XY}(\tau) = E(X_{t+\tau}Y_t).$$

Análogamente, podemos definir su *densidad espectral cruzada*, $P_{XY}(\omega)$, como $\mathcal{F}[R_{XY}(\tau)]$, que es en general un función compleja.

5.2. Estimación de la densidad espectral de potencia

La Inferencia Estadística nos enseña que la estimación de cualquier característica ligada a un fenómeno aleatorio exige la observación repetida del mismo, que proporciona lo que denominamos una muestra aleatoria. Así, la media poblacional, μ, de una variable aleatoria X podemos estimarla mediante la media muestral, $\bar{X}_n = \sum_{i=1}^{n} X_i$, obtenida a partir de una muestra aleatoria de tamaño n, $X_1, \ldots, X_n$. Lo esencial de este procedimiento es la repetibilidad del experimento, única forma de conseguir la n observaciones que componen la muestra.

Nuestro objetivo es ahora estimar características ligadas a un proceso estocástico, en concreto su PSD, y el problema que se nos plantea es la imposibilidad de obtener una muestra del proceso. Lo que sí está a nuestro alcance es poder efectuar una única observación del proceso a lo largo del tiempo, incluso durante un periodo largo de tiempo. ¿Cómo obtener estimaciones en tales condiciones? Necesitamos introducir el concepto de ergodicidad.

5.2.1. Ergodicidad

La ergodicidad es una propiedad por la cual las medias a lo largo del tiempo convergen a los valores esperados poblacionales. Existen diferentes tipos de ergodicidad, según la característica poblacional involucrada. Definiremos algunas de ellas, que nos han de ser útiles en cuanto sigue.

Definición 5.3 (Ergodicidad en media) *Decimos que un proceso WSS es* ergódico en media *si la media temporal converge en media cuadrática a la media del proceso, μ,*

$$\hat{X}_T = \frac{1}{2T}\int_{-T}^{+T} X_t dt \xrightarrow{m.s.} \mu, \text{ cuando } T \to \infty.$$

Aún cuando consideramos fuera del alcance de estas notas las demostraciones de las propiedades de los estimadores ergódicos, y de las relaciones entre los distintos de ergodicidad, sí las mencionaremos. El estimador $\hat{X}_T$ es un estimador insesgado de μ y su varianza depende de la función de autocovarianza.

Definición 5.4 (Ergodicidad en autocorrelación) *Decimos que un proceso WSS es* ergódico en autocorrelación *si*

$$\frac{1}{2T}\int_{-T}^{+T} X_{t+\tau}X_t dt \xrightarrow{m.s.} R(\tau), \quad \forall\tau.$$

Si definimos un nuevo proceso para cada k mediante

$$Y_\tau(t) = X_{t+\tau}X_t,$$

la ergodicidad en autocorrelación admite una caracterización alternativa, como nos muestra el siguiente teorema.

Teorema 5.1 *Un proceso estocástico WSS, X_t, es* ergódico en autocorrelación *si y solo si*

$$\frac{1}{2T}\int_{-T}^{+T}\left(1-\frac{|u|}{2T}\right)C_{Y_\tau}(u)du \xrightarrow{m.s.} 0, \quad \forall\tau.$$

5.2.2. Periodograma: definición y propiedades

La estimación de la densidad espectral de potencia puede llevarse a cabo de dos formas distintas. La primera se basa en la definición de la PSD como $\mathcal{F}[\mathcal{R}(\tau)]$, supone estimar en primer lugar la función de autocorrelación y estimar a continuación la PSD mediante la transformada de Fourier de dicha estimación. La segunda obtiene la PSD directamente de los datos a partir de lo que se conoce como el *periodograma*.

Definición 5.5 (Periodograma) *Sean $X_0, X_1, \ldots, X_{n-1}$,* n *observaciones de un proceso WSS, X_t, discreto en el tiempo. Sea $\tilde{x}_n(\omega)$ la transformada de Fourier discreta de dichas observaciones,*

$$\tilde{x}_n(\omega) = \sum_{j=0}^{n-1} X_j \exp(-i2\pi\omega j). \tag{5.29}$$

Se define el periodograma, como el cuadrado medio de dicha transformada,

$$I_n(\omega) = \frac{1}{n}|\tilde{x}_n(\omega)|^2. \tag{5.30}$$

Si tenemos en cuenta que $|\tilde{x}_n(\omega)|^2$ es una medida de la energía del proceso en la frecuencia ω, $I_n(\omega)$, la media temporal de dicha energía, es una estimación de la potencia en ω.

La justificación de porqué el periodograma es un estimador adecuado para la PSD la encontramos al calcular la esperanza de (5.30). Tengamos primero en cuenta que

$$|\tilde{x}_n(\omega)|^2 = \left[\sum_{j=0}^{n-1} X_j \exp(-i2\pi\omega j)\right]\left[\sum_{k=0}^{n-1} X_j \exp(i2\pi\omega k)\right],$$

por tratarse de una variable aleatoria compleja. Calculemos ahora la esperanza.

$$\begin{aligned} E[I_n(\omega)] &= \frac{1}{n} E\left\{ \left[\sum_{j=0}^{n-1} X_j e^{-i2\pi\omega j} \right] \left[\sum_{k=0}^{n-1} X_k e^{i2\pi\omega k} \right] \right\} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} E(X_j X_k) e^{-i2\pi\omega(j-k)} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} R_X(j-k) e^{-i2\pi\omega(j-k)}. \end{aligned} \tag{5.31}$$

La expresión (5.31) puede simplificarse si tenemos en cuenta que el argumento $j-k$ toma valores entre $-(n-1)$ y $n-1$, tantos como diagonales tiene el retículo $[0,1,\ldots,n-1]\times[0,1,\ldots,n-1]$, puesto que en cada una de esas diagonales $j-k$ toma el mismo valor. Así pues, haciendo $j-k=m$,

$$\begin{aligned} E[I_n(\omega)] &= \frac{1}{n} \sum_{m=-(n-1)}^{n-1} (n-|m|) R_X(m) e^{-i2\pi\omega m} \\ &= \sum_{m=-(n-1)}^{n-1} \left(1 - \frac{|m|}{n}\right) R_X(m) e^{-i2\pi\omega m}. \end{aligned} \tag{5.32}$$

La consecuencia inmediata de (5.32) es que $I_n(\omega)$ es un estimador asintóticamente insesgado de $P_X(\omega)$, lo que justifica su elección. Es deseable que la varianza del estimador sea pequeña, en particular que sea asintóticamente 0. No es el caso del periodograma, puede demostrarse que para un proceso WSS,

$$var[I_n(\omega)] \approx P^2(\omega). \tag{5.33}$$

Se trata de un estimador inconsistente, lo que puede conducir a estimaciones con gran variabilidad. En los textos de Diggle (1990) y Stark y Woods (2002) puede el lector encontrar un desarrollo de estos aspectos.

Si el proceso X_t es un proceso continuo en el tiempo, cuanto hemos dicho para el caso discreto puede extenderse de forma inmediata. Si observamos el proceso en el intervalo $[0,T]$, definimos el periodograma mediante

$$I_T(\omega) = \frac{1}{T} |\tilde{x}_T(\omega)|^2,$$

con

$$\tilde{x}_n(\omega) = \int_0^T X_t e^{-i2\pi\omega t} dt.$$

Un desarrollo semejante al que condujo a (5.32) conduce ahora a

$$E[I_T(\omega)] = \int_{-T}^{+T} \left(1 - \frac{\tau}{T}\right) R_X(\tau) e^{-i2\pi\omega\tau} d\tau,$$

que para $T \to \infty$ implica

$$E[I_T(\omega)] \longrightarrow P_X(\omega).$$

Periodograma medio. Aproximación de Bartlett

Una forma de disminuir la varianza del periodograma, cuando se dispone del suficiente número de observaciones, consiste en dividir éstas en subconjuntos consecutivos y estimar un periodograma en cada uno de ellos. Con todos ellos podemos calcular un periodograma medio que posee menor variabilidad.

Si m es un divisor de n, número de observaciones, y k el correspondiente cociente, definimos los k subconjuntos de la forma,

$$X^{[l]} = \{X_{j+(l-1)m}, 0 \leq j \leq m-1\}, \quad 1 \leq l \leq k.$$

El periodograma correspondiente a $X^{[l]}$ se define mediante (5.30),

$$I_m^{[l]}(\omega) = \frac{1}{m}|\tilde{x}_m^{[l]}(\omega)|^2 = \frac{1}{m}\left|\sum_{j=(l-1)m}^{lm-1} X_j \exp(-i2\pi\omega j)\right|^2.$$

Definimos ahora el periodograma medio como aquél cuyo valor en cada frecuencia, ω, es la media aritmética de los $I_m^{[l]}(\omega)$,

$$B(\omega) = \frac{1}{k}\sum_{l=1}^{k} I_m^{[l]}(\omega). \tag{5.34}$$

Este periodograma recibe el nombre de periodograma de Bartlett, que fue quién lo introdujo (Bartlett, 1950). La linealidad de la esperanza implica que este nuevo periodograma es también un estimador insesgado de $P(\omega)$. Por lo que respecta a su varianza, podemos obtenerla a partir de (5.33) teniendo en cuenta que para m lo suficientemente grande y $l_1 \neq l_2$, $I_m^{[l_1]}(\omega)$ y $I_m^{[l_2]}(\omega)$ son incorreladas,

$$var[B(\omega)] = \frac{1}{k^2}\sum_{l=1}^{k} var[I_m^{[l]}(\omega)] = \frac{1}{k}var[I_m^{[1]}(\omega)] \approx \frac{1}{k}P^2(\omega).$$

Ejemplo 5.8 *En el ejemplo 5.3 de Montes (2007) hemos obtenido la densidad espectral de potencia del proceso ARMA(2,2)*

$$(1 - 1{,}2B + 0{,}4B^2)X_t = (1 - 0{,}8B + 0{,}1B^2)Z_t.$$

Hemos generado 1024 observaciones del proceso que aparecen representadas en el gráfico de la Figura 5.10.

Para ver el efecto del periodograma medio de Bartlett hemos estimado el periodograma de las 1024 simulaciones. Hemos calculado sendos periodogramas medios para $k = 4$ y $k = 16$. Las gráficas de todos ellos, junto con la densidad espectral de potencia del modelo ARMA(2,2) se muestran conjuntamente en la Figura 5.11. Se aprecia claramente la mejora de la estimación a medida que k aumenta y también la disminución de la variabilidad, basta para ello observar las diferentes escalas para las ordenadas.

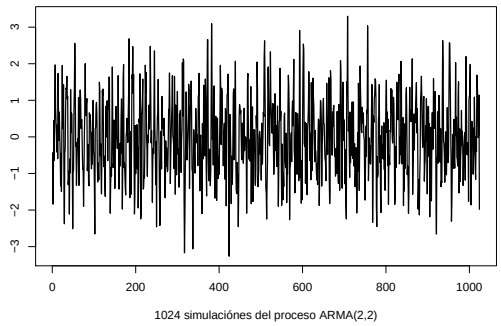


Figura 5.10: Simulaciones del proceso ARMA(2,2)

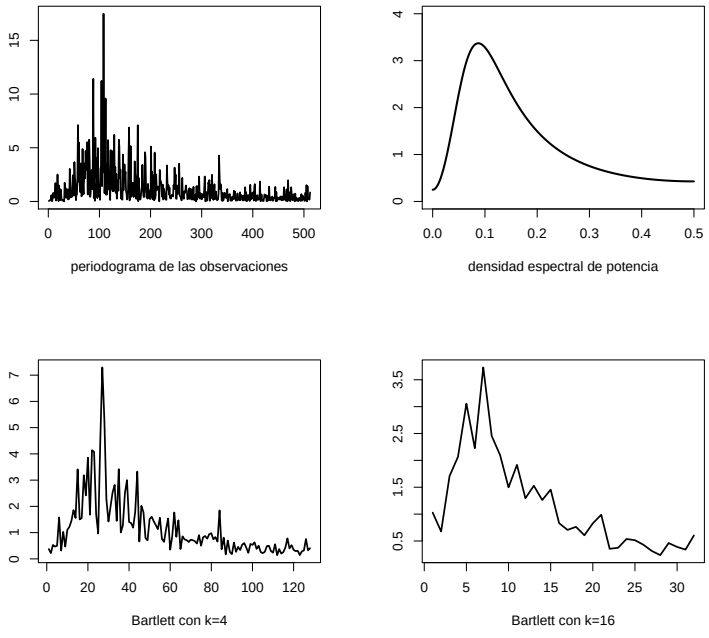


Figura 5.11: Periodogramas medios para distintos valores de k

Bibliografía

Bartlett, M. S. (1950). Periodogram analysis and continous spectra. *Biometrika*. **37**, 1–16

Billingsley, P. (1995). *Probability and Measure. 3rd Edition.* Wiley, N.Y.

Diggle, P. (1990). *Time Series. A Biostatistical Introduction.* Oxford University Press, N.Y.

Gnedenko, B. V. (1979). *Teoría de la Probabilidad*. Editorial Mir, Moscú.

Montes, F. (2007). *Procesos Estocásticos para Ingenieros: Teoría y Aplicaciones. Materiales complementarios*. Dpt. d'estadística i I. O. Universitat de València.

Priestley, M. B. (1981). *Spectral analysis and time series.* Academic Press, London.

Stark, H. y Woods, J. W. (2002). *Probability and random processes: with applications to signal processing. 3rd Edition.* Prentice Hall, N. J.