## ESPERANZA MATEMATICA o VALOR ESPERADO



### INTRODUCCION

El concepto de esperanza aparece inicialmente, en Cálculo de Probabilidades, ligado a los juegos de azar y a las "estadísticas" de población. En el primer caso hacía referencia a la cantidad de dinero que un jugador esperaba obtener de su participación en determinado juego. En el otro caso el concepto "esperanza de vida", significaba la vida media de un individuo a la vista de los datos considerados en la "estadística" en cuestión.

Vayamos al caso del jugador y supongamos que éste participa en el siguiente juego: Se lanza un dado, si la cara que presenta el dado es un número par el jugador recibe tantas pesetas como puntos indica la cara, si por el contrario es impar el número de puntos el jugador deberá pagar dicha cantidad en pesetas. ¿Qué ganancia espera obtener el jugador en cada lanzamiento?

En el supuesto de un dado correcto si el jugador lleva a cabo un elevado número de lanzamientos, $N$, las distintas caras del dado deben aparecer, muy aproximadamente, el mismo número de veces, a saber, $N/6$. Así pues, en todas estas jugadas el jugador habrá ganado (o perdido) lo siguiente:

$$(-1)\cdot\frac{N}{6} + 2\cdot\frac{N}{6} + (-3)\cdot\frac{N}{6} + 4\cdot\frac{N}{6} + (-5)\cdot\frac{N}{6} + 6\cdot\frac{N}{6}$$

por tanto en cada jugada esperará obtener una cantidad $E$, que se obtendrá de dividir por $N$ lo anterior, así pues

$$E = (-1)\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + (-3)\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + (-5)\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{1}{6}\left[2+4+6-(1+3+5)\right] =$$

$$= \frac{1}{6}(12-9) = \frac{3}{6} = 0.5$$

es decir, espera ganar 50 céntimos en cada jugada (de haber obtenido una cantidad negativa habría que interpretarlo como una pérdida).

Replanteemos ahora el problema de esta otra forma. Definamos una variable aleatoria $X$ cuyos valores van a ser los pagos o las ganancias del jugador en cada jugada, de acuerdo con esto $X$ tomará los siguientes valores (positivos ganancias, negativos pagos)

$$X = \{-1,2,-3,4,-5,6\}$$

además, teniendo en cuenta la naturaleza del experimento probabilístico asociado al juego, tendremos:

$$P(X=x_i) = \frac{1}{6} \quad , \quad x_i = -1,2,-3,4,-5,6$$

de acuerdo con esto, $E$ podría también obtenerse mediante la siguiente expresión



$$E = \sum_{i=1}^{6} x_i\cdot P(X=x_i) = \frac{1}{6}\left[\sum_{i=1}^{6} x_i\right] = 0.5$$

a la expresión $\sum_{i=1}^{6} x_i\cdot P(X=x_i)$ se la designa mediante $E(X)$ y se la denomina "esperanza de la variable aleatoria $X$". Todavía cabe un nuevo replanteamiento de la cuestión. $X$, como variable aleatoria que es, supone una aplicación del espacio de probabilidad $(\Omega,\mathcal{A},P)$ asociado al experimento de lanzar un dado y el espacio medible $(R,\beta)$, es decir, la recta real con su tribu de Borel, como tal aplicación, $X$ puede ser descrita de la siguiente manera:

$$X = \sum_{i=1}^{6} x_i\cdot \varphi_{A_i} \qquad A_i = \{w\in\Omega \,;\, X(w) = x_i\} \qquad x_i = -1,2,-3,4,-5,6$$

es decir, se trata de una función simple. Recordemos que la integral de una función simple, siempre que fuera finita (o no negativa la función), se escribía

$$\int_\Omega s\, d\mu = \sum_{i=1}^{n} a_i\cdot\mu(A_i) \quad , \text{ si } \quad s = \sum_{i=1}^{n} a_i\varphi_{A_i}$$

aplicado a $X$, tendremos

$$\int_\Omega X\, dP = \sum_{i=1}^{6} x_i\cdot P(A_i)$$

pero cada $A_i = \{w_i\}$, siendo $w_i$ el suceso "han salido $i$ puntos en la cara del dado" y por tanto $P(A_i) = \frac{1}{6}$, $i=1,\dots,6$, y en definitiva, teniendo en cuenta la definición de esperanza de la variable $X$ que hemos dado, tendremos

$$E(X) = \int_\Omega X\, dP.$$

Esta última igualdad es la que se utiliza como definición del concepto de esperanza matemática de una variable aleatoria del tipo que sea. Definición que hemos visto su significado en el caso discreto y que se adopta como tal para el caso continuo, pero haciendo algunas consideraciones de las que nos vamos a ocupar a continuación.

**DEFINICION.-** Sea $X$ una variable aleatoria definida del espacio de probabilidad $(\Omega,\mathcal{A},P)$ en $(R,\beta,P')$ ($P'$ probabilidad inducida) se define la esperanza matemática de $X$, o el valor esperado de $X$, $E(X)$, mediante

$$E(X) = \int_\Omega X\, dP \quad , \text{ siempre que sea finita.}$$

**Observación**.- A la vista de la definición de esperanza que hemos dado tiene sentido el hablar de la existencia o no de ésta para una variable aleatoria, puesto que la definición viene condicionada por la finitud de la integral que en ella aparece.

Recordemos que si $X$ es no negativa la integral

$$\int_\Omega X\,dP$$

está siempre definida, aunque puede no ser finita. Para el caso de $X$ cualquiera se dice que $X$ es integrable, si es finita, cosa que ocurre cuando $X^+$ y $X^-$ tienen integral finita, en cuyo caso

$$\int_\Omega X\,dP = \int_\Omega X^+\,dP - \int_\Omega X^-\,dP.$$

Si por otra parte recordamos la relación que existe entre las integrales, sobre un mismo conjunto, de $X$ y $|X|$, podremos afirmar que:

"La $E(X)$ existe si y sólo si $E(|X|)$, que siempre está definida, es finita, es decir, existe".

**La esperanza para v.a. discretas**. Una v.a. discreta, $X$, se caracteriza por la existencia de un conjunto $A\subset R$, tal que $P(X\in A)=1$, siendo a lo sumo, $A$, numerable. Ello nos permite escribir $X$ de la siguiente forma

$$X=\sum_{a_i\in A} a_i \chi_{A_i} \quad . \; A_i=\{w\in\Omega;\ X(w)=a_i\} \quad . \; \cup A_i=\Omega \quad , \; A_i\cap A_j=\phi,\ i\neq j$$

así pues, si existe

$$E(X)=\int_\Omega X\,dP = \int_\Omega\Big(\sum_{a_i\in A} a_i \chi_{A_i}\Big)dP = \int_{\cup A_i}\Big(\sum_{a_i\in A} a_i\chi_{A_i}\Big)dP = \sum_{a_i\in A}\int_{A_i} a_i\,dP =$$

$$=\sum_{a_i\in A} a_i\int_{A_i}dP = \sum_{a_i\in A} a_i . P(A_i) = \sum_{a_i\in A} a_i . P(X=a_i).$$

**La esperanza para v.a. continuas**. En primer lugar enunciaremos, sin demostrar, un importante teorema que se aplica a los cambios de variable en una integral. Supongamos un espacio de medida $(\Omega,\mathcal{A},\mu)$ y un espacio medible $(\Omega',\mathcal{A}')$ y entre ellos una aplicación $T$. Sea además $f$ una función medible de $(\Omega',\mathcal{A}')$ en $(R,\beta)$. Sobre $(\Omega',\mathcal{A}')$ inducimos una medida, $\mu'$, mediante la aplicación medible $T$, como sigue

$$\mu'(A')=\mu(T^{-1}(A')).$$

¿Cómo podemos trasladar la integral de la función medible $f\circ T$ sobre $\Omega$, a una integral sobre $\Omega'$? El teorema que enunciamos nos da la solución



**TEOREMA I**.- Con los elementos anteriores y siempre que $f$ sea no negativa o bien integrable, tendremos

$$\int_\Omega f\circ T\,d\mu = \int_{\Omega'} f\,d\mu'.$$

Este teorema tiene aplicación inmediata a nuestra situación. En efecto, hagamos $T\equiv X$, variable aleatoria del tipo continuo con la que vamos a trabajar, en esta situación $(\Omega',\mathcal{A}')\equiv(R,\beta)$, $\mu\equiv P$ y $\mu'\equiv P'$, probabilidad inducida. Finalmente sea $f$ la identidad. Tendremos pues que $f\circ T\equiv X$ y podremos escribir, si $E(X)$ existe

$$E(X)=\int_\Omega X\,dP = \int_R x\,dP'$$

Por otra parte recordemos que la v.a. continua se caracteriza por la existencia de una función, la f.d.p. que permite escribir

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

siendo $F$ la función de distribución, y también

$$P'(A)=\int_A f\,d\lambda$$

siendo $A\in\beta$ y $\lambda$ la medida de Lebesgue. A las medidas que gozan de esta propiedad (poseer una densidad respecto de otra medida; en nuestro caso $f$ es la densidad de $P'$ respecto a la medida de Lebesgue) es de aplicación el siguiente teorema, que también enunciamos y no demostramos

**TEOREMA II**.- Si la medida $\nu$ tiene una densidad $g$ respecto de la medida $\mu$, entonces para toda función $h$, medible, no negativa o integrable, tenemos

$$\int_\Omega h\,d\nu = \int_\Omega h.g.d\mu.$$

Este teorema aplicado a nuestra situación concreta con $h\equiv x$, medible e integrable, $\nu\equiv P'$ y $\mu\equiv\lambda$, medida de Lebesgue y la densidad $g\equiv f$, da lugar a:

$$E(X) = \int_\Omega X \, dP \underset{\text{por Teorema I}}{=} \int_R x \, dP' \underset{\text{por Teorema II}}{=} \int_R x\, f(x)\, d\lambda$$

Finalmente si la función $x.f(x)$ es integrable Riemann, como es normalmente el caso, el valor de su integral coincidirá con la integral de Lebesgue escrita y obtendremos finalmente para la $E(X)$, siempre que exista

$$\boxed{E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx}$$

Esperanza de una función de una variable aleatoria. Si transformamos una variable aleatoria $X$ mediante una aplicación $g$ medible, la nueva variable aleatoria $g(X)$ tiene una esperanza que se define como sigue:

Definición.- Sea $g$ una aplicación medible y $X$ una variable aleatoria, si $E[|g(X)|] < +\infty$ definimos la esperanza de $g(X)$ mediante

$$E(g(X)) = \int_\Omega g(X)\, dP$$

siendo $(\Omega, \mathcal{A}, P)$ el espacio de probabilidad sobre el que está definida $X$.

Un razonamiento análogo al utilizado en los párrafos anteriores nos lleva a las siguientes expresiones de la $E(g(X))$, según que $X$ sea discreta o continua.

$X$ discreta:

$$E(g(X)) = \sum_{a_i \in A} g(a_i).P(X = a_i)$$

$X$ continua:

$$E(g(X)) = \int_R g(x).f(x)\, d\lambda$$

y si $g(x).f(x)$ es tal que resulta ser integrable Riemann, entonces

$$E(g(X)) = \int_R g(x)\, f(x)\, dx.$$



Momentos de una variable aleatoria

a) Momentos absolutos respecto del origen.

Si hacemos $g(X) = |X|^k$, entonces $E(g(X)) = E(|X|^k)$ que se denomina momento absoluto de orden $k$ respecto del origen. Obsérvese que si el momento absoluto de orden $k$ respecto del origen existe para una v.a. $X$, entonces para $j \le k$, tendremos

$|x|^j \le 1 + |x|^k$ y de aquí

$$\int_R |X|^j dP' \le \int_R (1+|x|^k)\, dP' = 1 + E(|X|^k) < +\infty$$

y por tanto $E(|X|^j)$ también existe.

b) Momentos respecto del origen.

Si hacemos $g(X) = (X)^k$, entonces $E(g(X)) = E(X^k)$ que se denomina momento de orden $k$ respecto del origen.

Obsérvese que para $k=1$ se obtiene la esperanza de la variable aleatoria como momento de orden uno respecto del ~~~~ origen.

c) Momentos respecto de una constante $c$.

Haciendo $g(X) = |X-c|^k$ o bien $g(X) = (X-c)^k$ se obtienen los momentos de orden $k$, absolutos o no, respecto de una constante $c$. Hay un caso especial de este tipo de momentos que es aquel para el cual $c = E(X)$, entonces a los momentos de esta manera obtenidos se les denomina momentos centrales de orden $k$ (absolutos o no). De entre los momentos centrales tiene especial interés el ~~momento~~ central de segundo orden, a saber

$$E(X - E(X))^2$$

al que se le denomina varianza de la variable aleatoria.

A continuación nos vamos a ocupar de estudiar con detalle la esperanza y la varianza de una variable aleatoria a través de sus propiedades más relevantes.

Propiedades de la Esperanza de una variable aleatoria

Hay una serie de propiedades que se derivan de inmediato de la definición de esperanza, son:

1) $E(c) = c$, donde $c$ es una constante
2) $E[cg(X)] = c\,E[g(X)]$, en particular $E[cX] = c\,E[X]$.
3) $E[g(X)+d] = E[g(X)]+d$, $d$ constante, en particular $E[X+d] = E(X)+d$

de la combinación de las dos últimas, llegamos a

4) $E[ag(X)+b] = a\,E[g(X)]+b$ , con $a$ y $b$ constantes.

5) $E\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i . E[g_i(X)]$ , $c_i$ constantes $i=1,\dots,n$.

6) Si $X \geq 0$ , $E(X) \geq 0$.

7) De 6) se deriva $E(X) \geq E(Y)$ si $X \geq Y$, siendo ambas variables aleatorias.

Otras propiedades requieren una demostración explícita. Por ejemplo

Propiedad 8.- Sea $X$ una v.a. no negativa con función de distribución $F$. Entonces

$$E(X) = \int_0^\infty (1-F(x))\,dx ,$$

en el sentido de que si alguno de los miembros de la igualdad existe también lo hace el otro y sus valores coinciden.

Demostración.- a) $X$ es una v.a. continua.

Sabemos que la $E(X)$ cuando existe viene dada por

$$E(X) = \int_0^\infty x f(x)\,dx = \lim_{n\to\infty} \int_0^n x f(x)\,dx$$

Integrando por partes obtendremos

$$\int_0^n x f(x)\,dx = [x F(x)]_0^n - \int_0^n F(x)\,dx = n.F(n) - \int_0^n F(x)\,dx$$

sumando y restando $n$, tendremos

$$\int_0^n x f(x)\,dx = -n + nF(n) + n - \int_0^n F(x)\,dx = -n[1-F(n)] + \int_0^n [1-F(x)]\,dx .$$

Pero

$$n[1-F(n)] = n\int_n^\infty f(x)\,dx < \int_n^\infty x f(x)\,dx$$

y puesto que $E(|X|) < \infty$ , entonces

$$\lim_{n\to\infty} \int_n^\infty x f(x)\,dx = 0 \longrightarrow \lim_{n\to\infty} n[1-F(n)] = 0 ,$$

tenemos pues finalmente

$$E(X) = \lim_{n\to\infty} \int_0^n x f(x)\,dx = \lim_{n\to\infty} \int_0^n [1-F(x)]\,dx = \int_0^\infty [1-F(x)]\,dx .$$

Por otra parte si partimos de $\int_0^\infty [1-F(x)]\,dx < +\infty$ , entonces

$$\int_0^n x f(x)\,dx \leq \int_0^n [1-F(x)]\,dx \leq \int_0^\infty [1-F(x)]\,dx , \ \forall n$$ y de aquí $E(|X|) < +\infty$.



b) $X$ es una v.a. discreta

Sea $\{x_1, \dots x_n \dots\}$ el conjunto $A$ tal que $P(X \in A) = 1$ y hagamos $P(X = x_j) = p_j$ , entonces

$$E(X) = \sum_{j\geq 1} x_j . p_j .$$

Sea $I = \int_0^\infty [1-F(x)]\,dx$ . Entonces

$$I = \sum_{k\geq 1} \int_{(k-1)/n}^{k/n} P\{X > x\}\,dx$$

y puesto que $P(X>x)$ es una función no creciente de $x$, tendremos

$$P(X > \tfrac{k}{n}) \leq P(X>x) \leq P(X > \tfrac{k-1}{n}) , \quad \tfrac{k}{n} \leq x \leq \tfrac{k-1}{n} , \ \forall n$$

y de aquí, integrando y sumando sobre $k$,

$$\frac{1}{n}\sum_{k\geq 1} P(X > \tfrac{k}{n}) \leq I \leq \frac{1}{n}\sum_{k\geq 1} P(X > \tfrac{k-1}{n}) \quad , \forall n .$$

Hagamos

$$U_n = \frac{1}{n}\sum_{k\geq 1} P(X > \tfrac{k-1}{n}) \quad y \quad L_n = \frac{1}{n}\sum_{k\geq 1} P(X > \tfrac{k}{n}) .$$

Tenemos

$$L_n = \frac{1}{n}\sum_{k\geq 1}\sum_{j\geq k} P\left\{\frac{j}{n} < X \leq \frac{j+1}{n}\right\} = \frac{1}{n}\sum_{k\geq 2} (k-1) P\left\{\frac{k-1}{n} < X \leq \frac{k}{n}\right\}$$

reordenando la serie. Así

$$L_n = \sum_{k\geq 2} \frac{k}{n} P\left\{\frac{k-1}{n} < X \leq \frac{k}{n}\right\} - \sum_{k\geq 2} \frac{1}{n} P\left\{\frac{k-1}{n} < X \leq \frac{k}{n}\right\} =$$

$$= \sum_{k\geq 2} \frac{k}{n}\left[\sum_{(k-1)/n < x_j \leq \frac{k}{n}} p_j\right] - \frac{1}{n} P\left\{X > \frac{1}{n}\right\} \geq \sum_{k\geq 1}\ \sum_{(k-1)/n < x_j \leq k/n} x_j p_j - \frac{1}{n} =$$

$$= E(X) - \frac{1}{n} \quad , \forall n$$

Análogamente podemos demostrar que para cada $n$, tenemos

$$U_n \leq E(X) + \frac{1}{n} .$$

Tenemos pues

$$E(X) - \frac{1}{n} \leq L_n \leq I \leq U_n \leq E(X) + \frac{1}{n} , \ \forall n$$ y es cierto el teorema.

Una consecuencia de la propiedad 8 es esta nueva propiedad.

**Propiedad 9**.- Para una v.a. cualquiera $X$, $E(|X|)<+\infty$ si y sólo si las integrales

$$\int_{-\infty}^{0} P\{X\le x\}\,dx \quad y \quad \int_{0}^{\infty} P\{X>x\}\,dx$$ son finitas y además en este caso

$$E(X)=\int_0^\infty P\{X>x\}\,dx-\int_{-\infty}^0 P\{X\le x\}\,dx .$$

Demostración.- Si $E(|X|)<+\infty$, existe la esperanza de $X$ y sabemos además que

$$E(X)=E(X^+)-E(X^-)$$

siendo

$$X^+=\begin{cases}0, & X<0\\ X, & X\ge 0\end{cases} \quad y \quad X^-=\begin{cases}0, & X>0\\ -X, & X\le 0\end{cases}$$

Se trata en ambos casos de variables aleatorias no negativas y podemos aplicarles la propiedad anterior. Tendremos pues

$$E(X^+)=\int_0^\infty (1-G(y))\,dy=\int_0^\infty P\{X^+>y\}\,dy \text{ , donde } G \text{ es la f.d. de } X^+$$

pero teniendo en cuenta la definición de $X^+$, $P(X^+>y)=P(X>y)$, $y\ge 0$ y por tanto

$$E(X^+)=\int_0^\infty P(X>x)\,dx .$$

Por otra parte

$$E(X^-)=\int_0^\infty (1-H(y))\,dy=\int_0^\infty P\{X^->y\}\,dy \text{ , donde } H \text{ es la f.d. de } X^-$$

y teniendo en cuenta la definición de $X^-$ tenemos, $P(X^->y)=P(X<-y)=$
$=P(X\le -y)-P(X=-y)$ y por tanto

$$E(X^-)=\int_0^\infty P(X\le -x)\,dx-\int_0^\infty P(X=-x)\,dx=\int_{-\infty}^0 P(X\le x)\,dx-\int_{-\infty}^0 P(X=x)\,dx$$

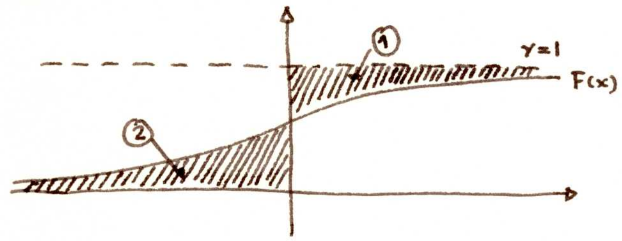
pero sea cual sea el tipo de v.a. de la que nos ocupemos la función $P(X=x)$ es distinta de cero a lo sumo en un conjunto numerable y por tanto su integral de Lebesgue es nula. En definitiva

$$E(X)=\int_0^\infty P(X>x)\,dx-\int_{-\infty}^0 P(X\le x)\,dx$$



La implicación contraria se deriva del hecho de ser $|X|=X^++X^-$ y por tanto, si $E(X^+)<+\infty$ y $E(X^-)<+\infty$, entonces $E(|X|)=E(X^+)+E(X^-)$ y además es finita.

Comentario.- Obsérvese que la anterior expresión para la $E(X)$ tiene una significación geométrica en función de las áreas que cubren las funciones $F(x)$ y $1-F(x)$ en los semiejes negativo y positivo respectivamente. Concretamente la diferencia entre las áreas ① y ② de la figura.




**Propiedades de la varianza**

1) $\sigma^2(c)=0$, si $c$ es una constante.

2) $\sigma^2(c\,g(X))=c^2\cdot\sigma^2[g(X)]$. En particular $\sigma^2(cX)=c^2\cdot\sigma^2(X)$

3) $\sigma^2(g(X)+d)=E\{[g(X)+d]-E[g(X)+d]\}^2=E\{g(X)-E[g(X)]+d-E(d)\}^2=$
$=E\{[g(X)-E(g(X))]+d-d\}^2=E\{g(X)-E(g(X))\}^2=\sigma^2(g(X))$.

Combinando 2 y 3 tenemos

4) $\sigma^2(c\,g(X)+d)=c^2\cdot\sigma^2(g(X))$

5) $\sigma^2[g(X)]=E[g(X)-E(g(X))]^2=E\{[g(X)]^2-2g(X)\cdot E(g(X))+[E(g(X))]^2\}=$
$=E[g(X)]^2-2[E(g(X))]^2+[E(g(X))]^2=E[g(X)]^2-[E(g(X))]^2$

lo que para una variable aleatoria se escribe

$$\sigma^2(X)=E[X^2]-[E(X)]^2.$$

TEOREMA.- Para una v.a. $X$, con $E|X|^2<+\infty$, la $E[X-c]^2$ se minimiza para $c=E(X)$, es decir, para la varianza.

Demostración.-

$$\sigma^2(X)=E[X-E(X)]^2=E\{X-c-[E(X)-c]\}^2=E\{[X-c]^2-2(X-c)(E(X)-c)+$$
$$+[E(X)-c]^2\}=E[X-c]^2-2(E(X)-c)^2+[E(X)-c]^2=$$
$$=E(X-c)^2-[E(X)-c]^2$$

y siendo $[E(X)-c]^2\ge 0$, entonces $\sigma^2(X)\le E[X-c]^2$, alcanzándose la igualdad para $c=E(X)$.

## Tipificación de una v.a.

Una v.a. $X$ se dice que está tipificada cuando $E(X)=0$ y $\sigma^2(X)=1$. Las propiedades anteriormente expuestas de la esperanza y de la varianza permiten obtener a partir de cualquier v.a. con $E(X)=\mu$ y $\sigma^2(X)=\sigma^2$, una nueva variable $Y$ que está tipificada. En efecto, definamos

$$Y = \frac{X - E(X)}{\sqrt{\sigma^2(X)}} = \frac{X-\mu}{\sigma}$$

entonces

$$E(Y) = \frac{1}{\sigma}\left(E(X)-\mu\right) = 0$$

$$\sigma^2(Y) = \frac{1}{\sigma^2}\sigma^2(X-\mu) = \frac{1}{\sigma^2}\sigma^2(X) = 1.$$

## Relación entre los momentos centrales y los momentos respecto del origen.

Los momentos respecto del origen se definen mediante

$$m_k = E(X^k)$$

mientras que los momentos centrales, son

$$\mu_k = E(X-\mu)^k \quad , \text{donde } \mu = E(X).$$

Recordando el desarrollo del binomio de Newton ($k$ entero positivo), tendremos

$$\mu_k = E\left[(X-\mu)^k\right] = E\left[X^k - \binom{k}{1}X^{k-1}\cdot\mu + \binom{k}{2}X^{k-2}\cdot\mu^2 + \cdots + (-1)^k\mu^k\right] =$$

$$= m_k - \binom{k}{1}\mu . m_{k-1} + \binom{k}{2}\mu^2 . m_{k-2} + \cdots + (-1)^k . \mu^k$$

Por otra parte

$$m_k = E(X^k) = E\left[\left((X-\mu)+\mu\right)^k\right] = E\left[(X-\mu)^k + \binom{k}{1}\mu .(X-\mu)^{k-1} + \cdots + \mu^k\right] =$$

$$= \mu_k + \binom{k}{1}\mu . \mu_{k-1} + \binom{k}{2}\mu^2 . \mu_{k-2} + \cdots + \mu^k.$$

Estas relaciones nos permiten obtener unos otros momentos conociendo tan solo unos de ellos.



## ALGUNAS DESIGUALDADES DE INTERES

Se trata en este apartado de obtener algunas desigualdades para los momentos de una v.a. El principal resultado es el teorema que demostraremos a continuación, que proporciona una cota superior para la cota de probabilidad en términos de algún momento de la v.a.

**TEOREMA.-** Sea $g(X)$ una función medible, no negativa de una v.a. $X$. Si existe $E(g(X))$ entonces, para todo $\varepsilon>0$, tenemos

$$P\{g(X) \geq \varepsilon\} \leq \frac{E[g(X)]}{\varepsilon}.$$

**Demostración.-** Recordemos que

$$E[g(X)] = \int_\Omega g(X)\,dP$$

recordemos también que toda integral de una función medible no negativa es una medida sobre la familia $\mathcal{A}$ de los conjuntos medibles. Sea

$$A = \{\omega\in\Omega;\ g(X(\omega)) < \varepsilon\}$$

y

$$A^c = \{\omega\in\Omega;\ g(X(\omega)) \geq \varepsilon\}.$$

Tendremos, $\Omega = A\cup A^c$, y de aquí

$$E[g(X)] = \int_\Omega g(X)\,dP = \int_{A\cup A^c} g(X)\,dP = \int_A g(X)\,dP + \int_{A^c} g(X)\,dP$$

pero siendo $g(X)\geq 0$, también $\int_A g(X)\,dP \geq 0$ y por tanto

$$E[g(X)] \geq \int_{A^c} g(X)\,dP \geq \varepsilon\int_{A^c} dP = \varepsilon . P(A^c)$$

luego

$$P(A^c) = P\{g(X)\geq\varepsilon\} \leq \frac{E[g(X)]}{\varepsilon}.$$

**COROLARIO I (Desigualdad de MARKOV).-** Hagamos $g(X)=|X|^r$ y $\varepsilon = k^r$, con $r>0, k>0$. Entonces

$$P\{|X|^r \geq k^r\} = P\{|X|\geq k\} \leq \frac{E[|X|^r]}{k^r}.$$

**COROLARIO 2** (Desigualdad de Tchebycher).- Apoyandonos en la anterior desigualdad para $g(X)=|X-E(X)|^2=|X-\mu|^2$, con $\mu=E(X)$ y para $\varepsilon=k^2\sigma^2(X)=k^2\sigma^2$, con $\sigma^2=\sigma^2(X)$, llegamos a la conocida desigualdad de Tchebycher

$$P(|X-\mu|\geq k\sigma)\leq \frac{1}{k^2}$$

que en ocasiones, si tomamos $\varepsilon=k^2$, también se nos presenta de la forma

$$P(|X-\mu|\geq k)\leq \frac{\sigma^2}{k^2}.$$

Observación.- Si $X$ es una v.a. con $E(X)=\mu$ y $\sigma^2(X)=0$, la desigualdad de Tchebycher nos conduce a

$$P(|X-\mu|\geq k)=0, \quad \forall k>0.$$

Sea $A_n=\{\omega: |X(\omega)-\mu|\geq \frac{1}{n}\}$, tendremos $P(A_n)=0$. Además, por construcción $A_n\subset A_{n+1}$ por lo que $A_n^c\supset A_{n+1}^c$, siendo $P(A_n^c)=1$. Si pasamos al límite tendremos

$$P(\cap A_n^c)=\lim_n P(A_n^c)=1$$

pero

$$\cap A_n^c=\{\omega: |X(\omega)-\mu|<\tfrac{1}{n}, \forall n\}=\{\omega: |X(\omega)-\mu|=0\}=\{\omega: X(\omega)=\mu\}$$

es decir

$$P(X=\mu)=1.$$

La v.a. es pues constante con probabilidad 1.

---

En otro orden de cosas es posible mejorar, en ocasiones, la desigualdad de Tchebycher si suponemos la existencia de momentos de orden superior. Vamos a comprobarlo, pero previamente necesitaremos el siguiente lema

LEMA.- Sea $X$ una v.a. con $E(X)=0$ y $\sigma^2(X)=\sigma^2$. Entonces

$$P(X\geq x)\leq \frac{\sigma^2}{\sigma^2+x^2}, \quad \text{si } x>0$$

y

$$P(X>x)\geq \frac{x^2}{\sigma^2+x^2}, \quad \text{si } x<0.$$

Demostración.- Sea $h(t)=(t+c)^2$, $c>0$. Se trata de una función medible no negativa. Entonces, para $0<x<t$, tenemos $h(t)\geq (x+c)^2$. De aquí se sigue

$$P(X\geq x)\leq P\{h(X)\geq (x+c)^2\}\leq \frac{E(X+c)^2}{(x+c)^2}, \quad c>0,\ x>0.$$



Ahora bien, como $E(X)=0$, entonces $E(X^2)=\sigma^2$ y por tanto

$$P(X\geq x)\leq \frac{\sigma^2+c^2}{(x+c)^2}, \quad c>0,\ x>0$$

pero $f(c)=\frac{\sigma^2+c^2}{(x+c)^2}$ tiene un mínimo para $c=\sigma^2/x$ y por tanto

$$P(X\geq x)\leq \frac{\sigma^2+\sigma^4/x^2}{\left(x+\frac{\sigma^2}{x}\right)^2}=\frac{\sigma^2(x^2+\sigma^2)}{(x^2+\sigma^2)^2}=\frac{\sigma^2}{\sigma^2+x^2}, \quad x>0.$$

La otra desigualdad se demuestra de forma similar. Además, se puede demostrar que estas dos desigualdades no pueden mejorarse.

Con la ayuda del lema podemos ahora obtener una nueva desigualdad que mejora, en ocasiones, la desigualdad de Tchebycher.

TEOREMA.- Sea $X$ una v.a. tal que $E(|X|^4)<+\infty$, $E(X)=0$, $E(X^2)=\sigma^2$. Entonces

$$P\{|X|\geq k\sigma\}\geq \frac{\mu_4-\sigma^4}{\mu_4+\sigma^4k^4-2k^2\sigma^4}, \quad \text{para } k>1, \text{ donde } \mu_4=E(X^4).$$

Demostración.- En la primera de las desigualdades del lema anterior hagamos $X=(X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)$ y $x=1$. Entonces

$$P\{X^2-\sigma^2\geq k^2\sigma^2-\sigma^2\}\geq \frac{\sigma^2\{(X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)\}}{1+\sigma^2\{(X^2-\sigma^2)/(k^2\sigma^2-\sigma^2)\}}=$$

$$=\frac{E\{[(X^2-\sigma^2)-E(X^2-\sigma^2)]^2\}}{\sigma^4(k^2-1)^2+E\{[(X^2-\sigma^2)-E(X^2-\sigma^2)]^2\}}=\frac{E\{(X^2-\sigma^2)^2\}}{\sigma^4(k^2-1)^2+E\{(X^2-\sigma^2)^2\}}=$$

$$=\frac{\mu_4-\sigma^4}{\sigma^4(k^2-1)^2+\mu_4-\sigma^4}=\frac{\mu_4-\sigma^4}{\mu_4+\sigma^4k^4-2k^2\sigma^4}, \quad k>1.$$

Puede demostrarse que esta última cota mejora la de Tchebycher si $k^2\geq \mu_4/\sigma^4$ mientras que si $1\leq k^2<\mu_4/\sigma^4$ es peor que aquella.

Una propiedad que asegura la existencia de los momentos

Ya hemos visto que la existencia de la esperanza de una determinada función aleatoria $g(X)$ está supeditada a la existencia de la correspondiente integral, es decir, a la integrabilidad de la función. El teorema que demostramos a continuación nos asegura la existencia de todos los momentos de una variable si ésta verifica ciertas condiciones:

**TEOREMA**.- Sea una v.a. $X$ satisfaciendo

$$\frac{P\{|X|>\alpha k\}}{P\{|X|>k\}} \longrightarrow 0, \text{ cuando } k\to\infty, \quad \forall \alpha>1.$$

Entonces $X$ posee todos los momentos. (Obsérvese que si $\alpha=1$, el límite es 1, mientras que si $\alpha<1$, $P\{|X|>\alpha k\} \geq P\{|X|>k\}$ y el límite no puede ser 0).

Demostración.- Sea $\varepsilon>0$ (cuyo valor fijaremos más tarde), elegimos $k_0$ de manera que

$$\frac{P\{|X|>\alpha k\}}{P\{|X|>k\}} < \varepsilon. \quad \forall k\geq k_0.$$

Asimismo elegimos $k_1$ de manera que $P\{|X|>k\}<\varepsilon$, $\forall k\geq k_1$.
Sea $N=\max(k_1,k_0)$. Para cada entero $r$, fijo, tenemos

$$\frac{P\{|X|>\alpha^r k\}}{P\{|X|>k\}} = \prod_{p=1}^{r} \frac{P\{|X|>\alpha^p k\}}{P\{|X|>\alpha^{p-1}k\}}, \quad k\geq N$$

como ~~[illegible]~~ $k'=\alpha^{p-1}k > k \geq k_0$, tendremos que

$$\frac{P\{|X|>\alpha^p k\}}{P\{|X|>\alpha^{p-1}k\}} = \frac{P\{|X|>\alpha k'\}}{P\{|X|>k'\}} < \varepsilon \quad , p=1,\dots,r$$

y por tanto

$$\frac{P\{|X|>\alpha^r k\}}{P\{|X|>k\}} < \varepsilon^r \longrightarrow P\{|X|>\alpha^r k\} < \varepsilon^{r+1} \quad . k\geq N.$$

Por otra parte, como la variable aleatoria $|X|^n$ es no negativa, utilizando una propiedad antes demostrada tenemos

$$E\left[|X|^n\right] = \int_0^\infty \left[1-F_{|X|^n}(y)\right] dy$$

que en términos de la f.d. de la v.a. $|X|$ y tras hacer el pertinente cambio de variable, adopta la forma:

$$E\left[|X|^n\right] = n\int_0^\infty x^{n-1}.P\{|X|>x\}\,dx =$$



$$= n\int_0^N x^{n-1}.P\{|X|>x\}\,dx + n\int_N^\infty x^{n-1}.P\{|X|>x\}\,dx$$

puesto que la primera integral es finita, nos bastará con demostrar la finitud de la segunda para comprobar la existencia de la $E(X^n)$, entonces

$$\int_N^\infty x^{n-1}.P\{|X|>x\}\,dx = \sum_{r=1}^\infty \int_{\alpha^{r-1}N}^{\alpha^r N} x^{n-1}.P\{|X|>x\}\,dx \leq$$

$$\leq \sum_{r=1}^\infty (\alpha^r N)^{n-1}.P\{|X|>\alpha^{r-1}N\}\left[\alpha^r N-\alpha^{r-1}N\right] \leq$$

$$\leq \sum_{r=1}^\infty (\alpha^r N)^{n-1}.\varepsilon^r.2\alpha^r N = 2N^n \sum_{r=1}^\infty (\alpha^n\varepsilon)^r =$$

$$= 2N^n.\frac{\alpha^n\varepsilon}{1-\alpha^n\varepsilon} < \infty$$

Siempre que $\varepsilon$ haya sido adecuadamente elegido $(\varepsilon\alpha^n<1)$. Tenemos pues que $E(|X|^n)<\infty$, $\forall n$, y la función posee todos sus momentos.

## MOMENTOS DE ALGUNAS VARIABLES ALEATORIAS

### Caso discreto

1.- Bernouilli. Recordemos que se trata de una v.a. discreta que toma los valores 0, 1 con probabilidad $p$ y $q$ respectivamente, de manera que $p+q=1$.

$$E(X) = 0.p+1.q = q$$

$$E(X^2) = 0^2.p+1^2.q = q \qquad var(X) = E(X^2)-\left[E(X)\right]^2 = q-q^2 = q(1-q) = q.p.$$

2.- Binomial. Sea $X$ una v.a. Binomial con parámetros $n,p$. Veamos cuál es su esperanza.

$$E(X) = \sum_{x=0}^n x.P(X=x) = \sum_{x=0}^n x.\binom{n}{x}p^x.q^{n-x} = \sum_{x=0}^n x\,\frac{n(n-1)\dots(n-x+1)}{x!}\,p^x.q^{n-x} =$$

$$= np\sum_{x=1}^n \frac{(n-1)\dots(n-x+1)}{(x-1)!}\,p^{x-1}.q^{n-x} = n.p\sum_{x-1=0}^{n-1}\binom{n-1}{x-1}p^{x-1}.q^{n-x} = np$$

Para obtener la $var(X)$, observemos que

$$E(X(X-1)) = E(X^2-X) = E(X^2)-E(X)$$

y de aquí

$$\operatorname{var}(X) = E\big(X(X-1)\big) + E(X) - [E(X)]^2.$$

ahora bien

$$E\big(X(X-1)\big) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} x(x-1)\,\frac{n(n-1)\dots(n-x+1)}{x!}\,p^x q^{n-x} =$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)\dots(n-x+1)}{(x-2)!}\,p^{x-2} q^{n-x} = n(n-1)p^2 \sum_{x-2=0}^{n-2}\binom{n-2}{x-2} p^{x-2} q^{n-x} =$$

$$= n(n-1)p^2.$$

Por tanto

$$\operatorname{var}(X) = n(n-1)p^2 + np - n^2p^2 = np\,[(n-1)p + 1 - np] = np(1-p) = n.p.q$$

Nota. Los momentos del tipo $E[X(X-1)\dots(X-n+1)]$ reciben el nombre de momentos factoriales de $n$-simo orden. Su utilidad ha quedado de manifiesto en la obtención de la varianza anterior.

3.- Poisson. Si $X$ es una v.a. Poisson con parámetro $\lambda$

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

$$E\big(X(X-1)\big) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x-2=0}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2$$

De aquí

$$\operatorname{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Obsérvese que para una v.a. $P(\lambda)$, los momentos factoriales de orden $n$, valen $\lambda^n$. En efecto

$$E\big(X(X-1)\dots(X-n+1)\big) = \sum_{x=0}^{\infty} x(x-1)\dots(x-n+1)\,e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^n e^{-\lambda}\sum_{x-n=0}^{\infty}\frac{\lambda^{x-n}}{(x-n)!} = \lambda^n.$$

Caso Continuo

1.- Uniforme. Sea $X$ una v.a. uniforme en el intervalo $[a,b]$, entonces

$$E(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \int_a^b x.\frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{2}(a+b)$$



$$E(X^2) = \int_a^b x^2\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}\,\frac{b^3-a^3}{b-a}$$

$$\operatorname{var}(X) = \frac{b^3-a^3}{3(b-a)} - \left\{\frac{1}{2}(b+a)\right\}^2 = \frac{1}{3}\cdot\frac{b^3-a^3}{b-a} - \frac{1}{4}\cdot(b+a)^2 = \frac{4(b^3-a^3) - 3(b^2+a^2+2ab)(b-a)}{12(b-a)} =$$

$$= \frac{b^3 - 3b^2a + 3ba^2 - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

2.- Normal tipificada. Si $X$ es una v.a. $N(0,1)$, entonces $E(X^{2n+1}) = 0$ y $E(X^{2n}) = \frac{(2n)!}{2^n(n!)}$, $n \geq 0$.

En efecto:

$$E(X^{2n+1}) = \int_{-\infty}^{+\infty} x^{2n+1} f(x)\,dx = \int_{-\infty}^{+\infty} x^{2n+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} x^{2n+1}e^{-x^2/2}\,dx +$$

$$+ \int_0^{+\infty} x^{2n+1}e^{-x^2/2}\,dx = -\frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x^{2n+1}e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x^{2n+1}e^{-x^2/2}\,dx = 0$$

Por otra parte

$$E(X^{2n}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2n}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}.2\int_0^{\infty} x^{2n}e^{-x^2/2}\,dx$$

ahora bien

$$\int_0^{\infty} x^{2n}.e^{-x^2/2}\,dx = \int_0^{\infty} -x^{2n-1}(-x)\,e^{-x^2/2}\,dx = \int_0^{\infty} -x^{2n-1}\,d\,e^{-x^2/2} = \left[-x^{2n-1}.e^{-x^2/2}\right]_0^{\infty} +$$

$$+ (2n-1)\int_0^{\infty} x^{2n-2}.e^{-x^2/2}\,dx = (2n-1)\int_0^{\infty} x^{2n-2}.e^{-x^2/2}\,dx$$

Si designamos $E(X^{2n})$ por $m_{2n}$, tendremos

$$m_{2n} = (2n-1)\,m_{2n-2}$$
$$m_{2n-2} = (2n-3)\,m_{2n-4}$$
$$\text{- - - - - - - -}$$
$$m_2 = 1.\,m_0$$
$$m_0 = 1 \quad \text{, pues} \quad m_0 = E(X^0) = E(1) = 1$$

y de aquí

$$\operatorname{var}(X) = E\big(X(X-1)\big) + E(X) - [E(X)]^2.$$

ahora bien

$$E\big(X(X-1)\big) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x \cdot q^{n-x} = \sum_{x=0}^{n} x(x-1)\,\frac{n(n-1)\dots(n-x+1)}{x!}\, p^x \cdot q^{n-x} =$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)\dots(n-x+1)}{(x-2)!}\, p^{x-2}\cdot q^{n-x} = n(n-1)p^2 \sum_{x-2=0}^{n-2} \binom{n-2}{x-2} p^{x-2} q^{n-x} =$$

$$= n(n-1)p^2.$$

Por tanto

$$\operatorname{var}(X) = n(n-1)p^2 + np - n^2p^2 = np[(n-1)p + 1 - np] = np(1-p) = n.p.q$$

Nota. Los momentos del tipo $E[X(X-1)\dots(X-n+1)]$ reciben el nombre de momentos factoriales de n-simo orden. Su utilidad ha quedado de manifiesto en la obtención de la varianza anterior.

3.- Poisson. Si $X$ es una v.a. Poisson con parámetro $\lambda$

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

$$E\big(X(X-1)\big) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x-2=0}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2$$

De aquí

$$\operatorname{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

obsérvese que para una v.a. $P(\lambda)$, los momentos factoriales de orden $n$, valen $\lambda^n$. En efecto

$$E\big(X(X-1)\dots(X-n+1)\big) = \sum_{x=0}^{\infty} x(x-1)\dots(x-n+1)\,e^{-\lambda}\frac{\lambda^x}{x!} = \lambda^n e^{-\lambda}\sum_{x-n=0}^{\infty}\frac{\lambda^{x-n}}{(x-n)!} = \lambda^n.$$

Caso Continuo

1.- Uniforme. Sea $X$ una v.a. uniforme en el intervalo $[a,b]$, entonces

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\,dx = \int_a^b x.\frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{2}(a+b)$$



$$E(X^2) = \int_a^b x^2\,\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}\,\frac{b^3-a^3}{b-a}$$

$$\operatorname{var}(X) = \frac{b^3-a^3}{3(b-a)} - \left\{\frac{1}{2}(b+a)\right\}^2 = \frac{1}{3}\cdot\frac{b^3-a^3}{b-a} - \frac{1}{4}\cdot(b+a)^2 = \frac{4(b^3-a^3) - 3(b^2+a^2+2ab)(b-a)}{12(b-a)} =$$

$$= \frac{b^3 - 3b^2a + 3ba^2 - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

2.- Normal tipificada. Si $X$ es una v.a. $N(0,1)$, entonces $E(X^{2n+1}) = 0$ y $E(X^{2n}) = \frac{(2n)!}{2^n (n!)}$, $n \geq 0$.

En efecto:

$$E(X^{2n+1}) = \int_{-\infty}^{+\infty} x^{2n+1} f(x)\,dx = \int_{-\infty}^{+\infty} x^{2n+1}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} x^{2n+1} e^{-x^2/2}\,dx +$$

$$+ \int_0^{+\infty} x^{2n+1} e^{-x^2/2}\,dx = -\frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x^{2n+1} e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x^{2n+1} e^{-x^2/2}\,dx = 0$$

Por otra parte

$$E(X^{2n}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2n} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}.2\int_0^{\infty} x^{2n} e^{-x^2/2}\,dx$$

ahora bien

$$\int_0^{\infty} x^{2n}.e^{-x^2/2}\,dx = \int_0^{\infty} -x^{2n-1}(-x)\,e^{-x^2/2}\,dx = \int_0^{\infty} -x^{2n-1}\,de^{-x^2/2} = \left[-x^{2n-1}.e^{-x^2/2}\right]_0^{\infty} +$$

$$+ (2n-1)\int_0^{\infty} x^{2n-2}.e^{-x^2/2}\,dx = (2n-1)\int_0^{\infty} x^{2n-2}.e^{-x^2/2}\,dx$$

Si designamos $E(X^{2n})$ por $m_{2n}$, tendremos

$m_{2n} = (2n-1)\,m_{2n-2}$

$m_{2n-2} = (2n-3)\,m_{2n-4}$

- - - - - - - - -

$m_2 = 1.\,m_0$

$m_0 = 1$, pues $m_0 = E(X^0) = E(1) = 1$

multiplicando ambos lados de las igualdades tenemos:

$$m_{2n} = (2n-1)(2n-3) \cdots 1 = \frac{2n(2n-1)(2n-2)(2n-3)\cdots 2.1}{2n(2n-2)\cdots 2} =$$

$$= \frac{(2n)!}{2(n).2(n-1)\cdots 2.1} = \frac{(2n)!}{2^n.(n!)}.$$

Entonces para la $E(X)$ y para la $var(X)$, tendremos

$$E(X) = 0$$

$$var(X) = E(X^2) = \frac{2!}{2^1\, 1!} = \frac{2}{2} = 1.$$

<u>Observación</u>.- Si queremos obtener la Esperanza y la Varianza de una v.a. $N(\mu, \sigma^2)$, tendremos que al definir una nueva variable, a partir de ésta, mediante

$$Z = \frac{X-\mu}{\sigma}$$

la variable $Z$ resultaba ser $N(0,1)$, por tanto

$$E(Z) = 0$$

pero

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\left[E(X)-\mu\right] = 0 \rightarrow E(X) = \mu$$

además

$$var(Z) = 1 = var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\, var(X) \rightarrow var(X) = \sigma^2$$

Así pues, la esperanza y la varianza de una $N(\mu, \sigma^2)$ coinciden con los parámetros de la distribución.

3.- <u>Gamma</u>. Sea $X$ una v.a. Gamma con parámetros $\alpha$ y $\beta$.

$$E(X) = \int_0^\infty x\, f(x)\, dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x.x^{\alpha-1} e^{-x/\beta} dx = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha . \frac{e^{-x/\beta}}{-\beta} dx =$$

$$= \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha\, d e^{-x/\beta} = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\left\{\left[x^\alpha . e^{-x/\beta}\right]_0^\infty - \alpha\int_0^\infty x^{\alpha-1}.e^{-x/\beta} dx\right\} =$$

$$= \frac{\alpha\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^{\alpha-1}.e^{-x/\beta} dx = \alpha\beta \quad , \text{ pues } \int_0^\infty x^{\alpha-1}.e^{-x/\beta} dx = \Gamma(\alpha).\beta^\alpha$$



Análogamente

$$E(X^2) = \frac{1}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^{\alpha+1}.e^{-x/\beta} dx = \beta^2\alpha(\alpha+1).$$

Para la varianza

$$var(X) = E(X^2) - [E(X)]^2 = \beta^2\alpha(\alpha+1) - \beta^2\alpha^2 = \beta^2\alpha.$$

<u>Casos particulares</u>:

a) Si $X$ es una $\chi^2_r$, tendremos que entonces $\alpha = r/2$, $\beta = 2$, con $r$ entero, por tanto

$$E(X) = \frac{r}{2}.2 = r \qquad , \; var(X) = \frac{r}{2}.4 = 2r$$

b) Si $X$ es una <u>exponencial negativa</u> con parámetro $\lambda$, entonces $\alpha = 1$, $\beta = 1/\lambda$ y por tanto

$$E(X) = 1.1/\lambda = \frac{1}{\lambda} \qquad , \; var(X) = 1.\frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

4.- <u>Cauchy</u>. Esta situación es interesante por cuanto se trata de una variable que carece de momentos. En efecto

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \frac{\sigma}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{\sigma^2 + (x-\mu)^2}$$

podemos suponer $\mu = 0$. $\sigma = 1$, entonces

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{1+x^2} = \frac{1}{2\pi}\left[\log(1+x^2)\right]_{-\infty}^{+\infty} \text{ que es una indeterminación.}$$

De cualquier forma obsérvese que

$$E(|X|) = \frac{1}{\pi}\int_{-\infty}^{+\infty}|x|.\frac{dx}{1+x^2} = \frac{2}{\pi}\int_0^\infty\frac{x\,dx}{1+x^2} = \frac{1}{\pi}\left[\log(1+x^2)\right]_0^{+\infty} = +\infty$$

y por tanto no existe la esperanza.

<u>Observación</u>.- El <u>valor principal de Cauchy</u> para una integral impropia se define mediante

$$\lim_{b\to\infty}\int_{-b}^{b} g(x)\, dx$$

obsérvese que en nuestro caso, el V.P.C de la integral anterior es finito, concretamente 0, aunque la integral no es convergente (desde luego en el caso de convergencia ambos valores coinciden).

multiplicando ambos lados de las igualdades tenemos:

$$m_{2n} = (2n-1)(2n-3)\cdots 1 = \frac{2n(2n-1)(2n-2)(2n-3)\cdots 2.1}{2n(2n-2)\cdots 2} =$$

$$= \frac{(2n)!}{2(n).2(n-1)\cdots 2.1} = \frac{(2n)!}{2^n.(n!)}.$$

Entonces para la $E(X)$ y para la $var(X)$, tendremos

$$E(X)=0$$

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

Observación.- Si queremos obtener la Esperanza y la Varianza de una v.a. $N(\mu,\sigma^2)$, tendremos que al definir una nueva variable, a partir de ésta, mediante

$$Z = \frac{X-\mu}{\sigma}$$

la variable $Z$ resultaba ser $N(0,1)$, por tanto

$$E(Z)=0$$

pero

$$E(Z) = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\left[E(X)-\mu\right] = 0 \rightarrow E(X)=\mu$$

además

$$var(Z) = 1 = var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} var(X) \rightarrow var(X) = \sigma^2$$

Así pues, la esperanza y la varianza de una $N(\mu,\sigma^2)$ coinciden con los parámetros de la distribución.

3.- Gamma. Sea $X$ una v.a. Gamma con parámetros $\alpha$ y $\beta$.

$$E(X) = \int_0^\infty x f(x)\,dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x.x^{\alpha-1}e^{-x/\beta}\,dx = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha.\frac{e^{-x/\beta}}{-\beta}\,dx =$$

$$= \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^\alpha\,de^{-x/\beta} = \frac{-\beta}{\Gamma(\alpha).\beta^\alpha}\left\{\left[x^\alpha.e^{-x/\beta}\right]_0^\infty - \alpha\int_0^\infty x^{\alpha-1}.e^{-x/\beta}\,dx\right\} =$$

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



Análogamente

$$E(X^2) = \frac{1}{\Gamma(\alpha).\beta^\alpha}\int_0^\infty x^{\alpha+1}.e^{-x/\beta}\,dx = \beta^2\alpha(\alpha+1).$$

Para la varianza

$$var(X) = E(X^2) - [E(X)]^2 = \beta^2\alpha(\alpha+1) - \beta^2\alpha^2 = \beta^2\alpha.$$

Casos particulares:

a) Si $X$ es una $\chi^2_r$, tendremos que entonces $\alpha = r/2$, $\beta=2$, con $r$ entero, por tanto

$$E(X) = \frac{r}{2}.2 = r \quad , \quad var(X) = \frac{r}{2}.4 = 2r$$

b) Si $X$ es una exponencial negativa con parámetro $\lambda$, entonces $\alpha=1$, $\beta=1/\lambda$ y por tanto

$$E(X) = 1.1/\lambda = \frac{1}{\lambda} \quad , \quad var(X) = 1.\frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

4.- Cauchy. Esta situación es interesante por cuanto se trata de una variable que carece de momentos. En efecto

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \frac{\sigma}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{\sigma^2+(x-\mu)^2}$$

podemos suponer $\mu=0$, $\sigma=1$, entonces

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x\,dx}{1+x^2} = \frac{1}{2\pi}\left[\log(1+x^2)\right]_{-\infty}^{+\infty}$$ que es una indeterminación.

De cualquier forma obsérvese que

$$E(|X|) = \frac{1}{\pi}\int_{-\infty}^{+\infty}|x|.\frac{dx}{1+x^2} = \frac{2}{\pi}\int_0^\infty\frac{x\,dx}{1+x^2} = \frac{1}{\pi}\left[\log(1+x^2)\right]_0^{+\infty} = +\infty$$

y por tanto no existe la esperanza.

Observación.- El valor principal de Cauchy para una integral impropia se define mediante

$$\lim_{b\to\infty}\int_{-b}^{b} g(x)\,dx$$

Obsérvese que en nuestro caso, el V.P.C de la integral anterior es finito, concretamente 0 aunque la integral no es convergente (desde luego en el caso de convergencia ambos valores coinciden).

## OTRAS CARACTERISTICAS NUMERICAS DE UNA DISTRIBUCION DE PROBABILIDAD



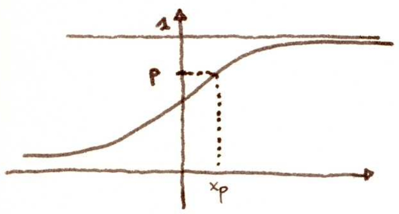
El concepto de esperanza nos ha permitido obtener los momentos de diversos tipos de una distribución de probabilidad (concretamente de la v.a. correspondiente), que, en definitiva, no son más que características numéricas que nos permiten un conocimiento, sino exhaustivo, bastante aproximado y, en cualquier caso, de forma rápida de la distribución de probabilidad asociada a la v.a. Pero no son estas las únicas características numéricas de interés, a continuación introducimos otras, conocidas genéricamente como <u>cuantiles</u>.

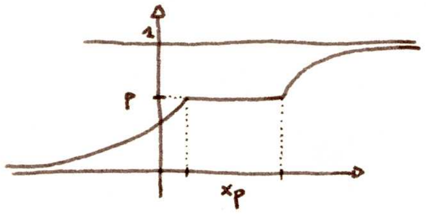
<u>Definición</u>.- Sea $X$ una v.a. con f.d. $F$ y sea $p$, tal que $0<p<1$. Toda cantidad $x_p$ que satisface simultaneamente

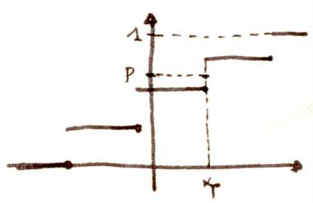
$$P(X \le x_p) = F(x_p) \geq p \quad \text{y} \quad P(X \geq x_p) \geq 1-p$$

se la denomina <u>cuantil de orden p</u> de la v.a. $X$ (o de la distribución de probabilidad correspondiente).

Dadas la definición y las propiedades de la f.d. un cuantil de orden $p$ existe siempre, pero no tiene porque ser único. En particular todas las soluciones de las ecuaciones antes planteadas son cuantiles de orden $p$, pudiendo su número ser no numerable; por ejemplo, si la recta $y=p$ y la f.d. $F(x)$ tienen un intervalo común, todo el intervalo está constituido por cuantiles del orden deseado. Por contra si la función de distribución es strictamente creciente la solución es única y único el cuantil. Vease en las figuras algunas de las situaciones que pueden surgir.




Algunos cuantiles tienen nombre propio. Si $p=0.1, 0.2, \ldots, 0.9$ a los $x_p$ correspondientes se les denomina <u>deciles</u>, si $p=0.25, 0.5, 0.75$ a los correspondientes $x_p$ se les conoce como <u>cuartiles</u> de la distribución y de estos tiene especial importancia la <u>mediana</u>, que es el segundo cuartil, es decir el correspondiente a $p=1/2$.

Como ocurre con todos los cuantiles la mediana no tiene porque ser única, pero si la distribución de probabilidad es simetrica, el centro de simetria de la misma coincide, desde luego, con la mediana. El papel de la mediana es importante, como medida de centralización, especialmente en aquellas v.a. que carecen de media, puesto que, como ya hemos advertido, nunca carecerán de mediana. Esta posee además otra importante propiedad:

<u>Propiedad de la mediana</u>.- Sea $X$ una v.a. continua. El momento absoluto $E(|X-c|)$ alcanza un mínimo si $c$ se elige igual a la mediana de la distribución.



Demostración.- Observemos en primer lugar que si $X$ es una v.a. continua aún cuando la mediana no sea única, verifica las igualdades de las condiciones que la definen. Es decir, si $m$ es una mediana, tenemos

$$F(m) = \frac{1}{2} \quad \text{y} \quad 1-F(m) = \frac{1}{2}.$$

Hecha esta observación, tenemos

$$E(|X-c|) = \int_{-\infty}^{+\infty} |x-c| f(x)dx = \int_{-\infty}^{c} (c-x) f(x)dx + \int_{c}^{+\infty} (x-c) f(x)dx =$$

$$= \int_{-\infty}^{m} (c-m+m-x) f(x)dx + \int_{m}^{c} (c-x) f(x)dx + \int_{c}^{+\infty} (x-m+m-c) f(x)dx =$$

$$= \int_{-\infty}^{m} (m-x) f(x)dx + \int_{c}^{+\infty} (x-m) f(x)dx + \int_{-\infty}^{m} (c-m) f(x)dx + \int_{m}^{c} (c-x) f(x)dx +$$

$$+ \int_{c}^{+\infty} (m-c) f(x)dx = E(|X-m|) - \int_{m}^{c} (x-m) f(x)dx + \int_{m}^{c} (c-x) f(x)dx +$$

$$+ \int_{-\infty}^{m} (c-m) f(x)dx + \int_{m}^{c} (m-c) f(x)dx - \int_{m}^{c} (m-c) f(x)dx + \int_{c}^{+\infty} (m-c) f(x)dx =$$

$$= E(|X-m|) + \int_{-\infty}^{m} (c-m) f(x)dx + \int_{m}^{+\infty} (m-c) f(x)dx + \int_{m}^{c} (c-x-x+m-m+c) f(x)dx =$$

$$= E(|X-m|) + 2\int_{m}^{c} (c-x) f(x)dx + (c-m)F(m) + (m-c)[1-F(m)] =$$

$$= E(|X-m|) + 2\int_{m}^{c} (c-x) f(x)dx \qquad c>m$$

Analogamente, para $m>c$, $E(|X-c|) = E(|X-m|) + 2\int_{c}^{m} (x-c) f(x)dx$. En ambos casos el segundo sumando del segundo miembro de la igualdad es positivo para $c \neq m$, y por tanto es cierta la propiedad enunciada.

Otra característica de interés es la MODA cuya definición damos a continuación.

**Definición**.- Sea $X$ una v.a. con función de densidad de probabilidad o de cuantía, $f$, según sea continua o discreta. Una moda de $f$, si existe, es cualquier número que maximiza $f(x)$.

De acuerdo con la definición, si $f$ es diferenciable dos veces, la moda puede obtenerse mediante diferenciación de la función, todo se reduce a un problema clásico de máximos de una función. Sin embargo esta actuación no es posible en el caso discreto. Para el caso de la Binomial y de la distribución de Poisson los dos teoremas que vienen a continuación nos dan la solución correspondiente.

**TEOREMA** (moda de la Binomial).- Sea $X$ una v.a. $B(n,p)$. Consideremos el número $(n+1)p$ y sea $[(n+1)p]=m$, la parte entera de $(n+1)p$. Entonces si $(n+1)p \neq [(n+1)p]$, $f(x)$ tiene una única moda para $x=m$. Si $(n+1)p=[(n+1)p]$, entonces las modas son dos, $x_1=m$ y $x_2=m-1$.

Demostración.- Tenemos

$$\frac{f(x)}{f(x-1)} = \frac{\binom{n}{x} p^x q^{n-x}}{\binom{n}{x-1} p^{x-1} q^{n-x+1}} = \frac{\frac{n!}{x!(n-x)!}\, p^x \cdot q^{n-x}}{\frac{n!}{(x-1)!(n-x+1)!}\, p^{x-1} \cdot q^{n-x+1}} = \frac{n-x+1}{x} \cdot \frac{p}{q}$$

por tanto $f(x) \geq f(x-1)$, creciente, si y sólo si

$(n-x+1)p \geq x(1-p)$, o bien $np-xp+p \geq x-xp$, o bien $(n+1)p \geq x$.

Así pues, si $(n+1)p$ no es entero, $f(x)$ aumenta para $x \leq [(n+1)p]$ y decrece a continuación. La moda será pues $m=[(n+1)p]$.

Si $(n+1)p$ es entero, entonces de los anteriores cálculos se deduce que para $x=(n+1)p$ $f(x)=f(x-1)$, luego en $x=(n+1)p-1$ hay también una moda.

**TEOREMA** (moda de una v.a. Poisson).- Sea $X$ una v.a. $P(\lambda)$. Entonces si $\lambda$ no es entero, $f(x)$ tiene una única moda en $x=[\lambda]$. Si $\lambda$ es entero $f(x)$ tiene dos modas en $x_1=\lambda$, $x_2=\lambda-1$

Demostración.- Tenemos

$$\frac{f(x)}{f(x-1)} = \frac{e^{-\lambda} \cdot \lambda^x / x!}{e^{-\lambda} \cdot \lambda^{x-1}/(x-1)!} = \frac{\lambda}{x}$$

por tanto $f(x) \geq f(x-1)$ si y sólo si $\lambda \geq x$. Así si $\lambda$ no es entero la función aumenta mientras $x \leq [\lambda]$ y después disminuye. El máximo se alcanza en $x=[\lambda]$. Si $\lambda$ es entero, entonces para $x=\lambda$, $f(x)=f(x-1)$, luego el máximo se alcanza en $x=\lambda$ y en $x=\lambda-1$.