ESPERANZA Y MOMENTOS DE UN VECTOR ALEATORIO

DEFINICIÓN.- Sea $(\Omega, \mathcal{B}, P)$ un espacio de probabilidad y $\vec{X}$ una aplicación medible de $(\Omega, \mathcal{B})$ en (R^m, β^m), es decir, un vector aleatorio. Si $\vec{X}$ es una aplicación medible y g, definida de R^m en R, también lo es, la composición $g(\vec{X})$ es una variable aleatoria, cuya esperanza se define de la forma consabida. A saber, si $g(\vec{X})$ es integrable, entonces

$$E(g(\vec{X})) = \int_\Omega g(\vec{X})\, dP.$$

Ya sabemos que la anterior expresión tiene un significado u otro en función del tipo de variable aleatoria que $g(\vec{X})$ es. Así, para el caso discreto, sabemos que

$$E(g(\vec{X})) = \sum_{\vec{x} \in A} g(\vec{x}) . P(\vec{X} = \vec{x}) \qquad \text{donde } A \text{ es tal que } P(\vec{X} \in A) = 1.$$

Si nos situamos en el caso continuo

$$E(g(\vec{X})) = \int_{R^m} g(\vec{x})\, f(\vec{x})\, d\vec{x} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1 \ldots x_m)\, f(x_1 \ldots x_m)\, dx_1 \ldots dx_m$$

donde $f(\vec{x})$ es la función de densidad de probabilidad del vector aleatorio $\vec{X}$.

Los conceptos de momentos centrales y respecto del origen así como los momentos absolutos y los momentos sin más tienen una definición paralela a la establecida para el caso de una variable aleatoria estudiada anteriormente con detalle. Surgen sin embargo, en esta ocasión, situaciones nuevas de interés. En efecto:

a) Si $g(\vec{X}) = X_1^{n_1} \ldots X_m^{n_m}$, con $n_i \geq 0$, enteros para $i = 1, \ldots, m$ a la correspondiente esperanza

$$E(g(\vec{X})) = E(X_1^{n_1} \ldots X_m^{n_m})$$

si existe, se la conoce como momento conjunto de orden $(n_1, \ldots, n_m)$. Se suele designar mediante $m_{n_1 \ldots n_m}$. Obsérvese que para $n_i = 1$, $n_j = 0$, $j \neq i$, tenemos el momento $m_{0 \ldots 1 \ldots 0}$ que coincide con la esperanza de la variable aleatoria X_i (no olvidemos que las componentes del vector aleatorio son todas variables aleatorias).

b) Análogamente, si $g(\vec{X}) = (X_1 - E(X_1))^{n_1} \ldots (X_m - E(X_m))^{n_m}$, el momento resultante al obtener la esperanza es conocido como el momento central conjunto de orden $(n_1 \ldots n_m)$.

Los correspondientes momentos absolutos surgirían de manera similar. Por ser especialmente interesante vamos a estudiar el caso bivariante.

VECTOR ALEATORIO BIDIMENSIONAL.- Para un vector aleatorio bidimensional (X, Y), el momento central de orden $(1,1)$ es conocido con el nombre de covarianza y su definición, cuando existe la correspondiente esperanza es

$$cov(X, Y) = E((X - E(X))(Y - E(Y))) = E((X - m_{10})(Y - m_{01})) = E(XY) - E(X).E(Y) =$$
$$= m_{11} - m_{10} . m_{01}$$

Veamos a continuación una interesante propiedad

DESIGUALDAD DE CAUCHY-SCHWARZ.- Sean X e Y dos variables aleatorias con varianzas finitas. Entonces la $cov(X, Y)$ existe. Además

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

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

Demostración.- Para cualquier par de números reales a y b tenemos

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

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

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

tratándose de una función no negativa, $E((\alpha X + Y)^2) \geq 0$, $\forall \alpha$, lo que supone que la ecuación de segundo grado de la derecha de la igualdad tiene a lo sumo una raíz y por tanto su discriminante será no positivo, es decir

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

Obsérvese que si $[E(XY)]^2 = E(X^2).E(Y^2)$ la ecuación tiene raíz doble, a saber $\alpha_0 = -\frac{E(XY)}{E(X^2)}$, lo que sustituido en la ecuación supone que $E[(\alpha_0 X + Y)^2] = 0$ y dadas las características de la función $(\alpha_0 X + Y)^2$ y las propiedades de la integral de una función no negativa tendremos $P\{(\alpha_0 X + Y) = 0\} = 1$.

Introducido el concepto de covarianza podemos ahora conocer el valor de la varianza de una suma finita de variables aleatorias.

VARIANZA DE UNA SUMA.- Sean $X_1, \ldots, X_n$ variables aleatorias cuyas varianzas existen. Sean $a_1, \ldots, a_n$, números reales. Definimos

$$S = \sum_{i=1}^{n} a_i X_i.$$

Entonces la varianza de S existe y viene dada por

$$var(S) = \sum_{i=1}^{n} a_i^2 . var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i a_j\, cov(X_i X_j).$$

En particular si las variables aleatorias son tales que todas las posibles covarianzas son nulas, entonces

$$var(S) = \sum_{i=1}^{n} a_i^2 . var(X_i).$$

Demostración.- El resultado se obtiene de desarrollar el cuadrado

$$\left[\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i)\right]^2$$

y después tomar esperanzas.

ALGUNAS DESIGUALDADES DE INTERES

Lema.- Para a y b reales cualesquiera, tenemos

$$|a+b|^r \leq C_r\left(|a|^r + |b|^r\right), \text{ donde } C_r = 1, \text{ si } 0 \leq r \leq 1 \text{ y } C_r = 2^{r-1} \text{ si } r > 1.$$

Demostración.- Para $r=0$ y $r=1$ la desigualdad es trivialmente cierta. Observemos por otra parte que bastará comprobar la desigualdad para $0 < a \leq b$. Así pues, $0 < a \leq b$ y hagamos $x = a/b$. Entonces

$$\frac{(a+b)^r}{a^r + b^r} = \frac{(1+x)^r}{1+x^r}$$

Sea $f(x) = (1+x)^r/(1+x^r)$, su derivada vale

$$f'(x) = \frac{r(1+x)^{r-1}}{(1+x^r)^2}\left(1 - x^{r-1}\right), \quad 0 < x \leq 1, \ r > 0.$$

Se sigue que $f'(x) > 0$, si $r > 1$, para $r = 1$, entonces $f'(x) = 0$ y para $r < 1$, entonces $f'(x) \leq 0$. En consecuencia

$$\max_{0 \leq x \leq 1} f(x) = f(0) = 1, \quad \text{si } r \leq 1$$

$$\max_{0 \leq x \leq 1} f(x) = f(1) = 2^{r-1}, \quad \text{si } r \geq 1.$$

Con lo que queda demostrado el lema.

Obsérvese que hubiéramos podido obtener, de inmediato, una cota del tipo

$$|a+b|^r \leq 2^r\left(|a|^r + |b|^r\right)$$

puesto que

$$|a+b| \leq \max\left(2|a|, 2|b|\right)$$

y de aquí al elevar a la r-ésima potencia obtenemos la desigualdad anterior.

Como una aplicación de esta desigualdad obtenida en el lema tenemos la siguiente propiedad.

TEOREMA.- Sean X e Y variables aleatorias y $r > 0$ un número fijo. Si $E(|X|^r)$ y $E(|Y|^r)$ son ambas finitas, entonces también lo es $E(|X+Y|^r)$.

Demostración.- Basta tomar $a = X$, $b = Y$ y aplicar la desigualdad anterior, luego tomamos esperanzas y llegamos al resultado enunciado en el teorema.

DESIGUALDAD DE HÖLDER.- Para x e y reales cualesquiera y para p y q, dos números reales positivos tales que $p > 1$ y $1/p + 1/q = 1$, tenemos

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$$

Demostración.- Observemos que para $x > 0$, la función $w = \log x$ es cóncava, por tanto para $0 < t < 1$ y para $x_1 > 0$, $x_2 > 0$, tenemos

$$\log\left[t x_1 + (1-t) x_2\right] \geq t \log x_1 + (1-t) \log x_2.$$

tomando antilogaritmos, tenemos

$$x_1^t \cdot x_2^{(1-t)} \leq t x_1 + (1-t) x_2$$

Elegimos ahora $x_1 = |x|^p$, $x_2 = |y|^q$, $t = 1/p$, $1-t = 1/q$, con $p > 1$ y $\frac{1}{p} + \frac{1}{q} = 1$, llegando entonces a

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}$$

desigualdad conocida como la desigualdad de Hölder

Aplicando la desigualdad de Hölder obtenemos otra conocida desigualdad en términos de

~~DESIGUALDAD DE MINKOWSKI.- Para $p \geq 1$,~~

$$\cancel{\left\{E\left[|X+Y|^p\right]\right\}^{1/p} \leq \left\{E(|X|^p)\right\}^{1/p} + \left\{E(|Y|^p)\right\}^{1/p}}$$

~~Demostración.- El caso $p=1$ es consecuencia inmediata del hecho $|X+Y| \leq |X| + |Y|$. Sea $p > 1$, tenemos~~

$$\cancel{|X+Y|^p = |X+Y| \cdot |X+Y|^{p-1} \leq |X| \cdot |X+Y|^{p-1} + |Y| \cdot |X+Y|^{p-1}}$$

~~tomando esperanzas y utilizando la desigualdad de~~

esperanzas que es, a menudo, conocida también como la desigualdad de Hölder.

DESIGUALDAD DE HÖLDER PARA ESPERANZAS.- Sean $p > 1$, $q > 1$ tales que $\frac{1}{p} + \frac{1}{q} = 1$. Entonces

$$E(|XY|) \leq \left\{E(|X|^p)\right\}^{1/p} \left\{E(|Y|^q)\right\}^{1/q}.$$

Demostración.- Utilizando la desigualdad de Hölder, con $x = X\left\{E(|X|^p)\right\}^{-1/p}$ e $y = Y\left\{E(|Y|^q)\right\}^{-1/q}$ llegamos a

$$|XY| \leq p^{-1}|X|^p \left\{E(|X|^p)\right\}^{1/p - 1}\left\{E(|Y|^q)\right\}^{1/q} +$$

$$+ q^{-1}|Y|^q \left\{E(|Y|^q)\right\}^{1/q - 1}\left\{E(|X|^p)\right\}^{1/p}$$

tomando esperanzas llegamos a la desigualdad buscada.

Un corolario importante de esta última desigualdad nos conduce a la ya conocida desigualdad de Cauchy-Schwarz. En efecto, para $p=q=2$, tenemos

$$E(|XY|) \le \{E(X^2)\}^{1/2}\cdot\{E(Y^2)\}^{1/2},$$

que puede también escribirse bajo la forma más conocida

$$\{E(|XY|)\}^2 \le E(X^2)\cdot E(Y^2).$$

Finalizamos con una desigualdad debida a Minkowski

DESIGUALDAD DE MINKOWSKI.- Para $p\ge 1$,

$$\{E(|X+Y|^p)\}^{1/p} \le \{E(|X|^p)\}^{1/p} + \{E(|Y|^p)\}^{1/p}.$$

Demostración.- Para $p=1$, la desigualdad es consecuencia inmediata de $|X+Y| \le |X|+|Y|$.

Sea $p>1$, entonces

$$|X+Y|^p = |X+Y|\cdot|X+Y|^{p-1} \le |X|\cdot|X+Y|^{p-1} + |Y|\cdot|X+Y|^{p-1}$$

tomando esperanzas y utilizando la correspondiente desigualdad de Hölder, en la que hacemos $Y = |X+Y|^{p-1}$ $(p>1)$, tenemos

$$E(|X+Y|^p) \le \{E(|X|^p)\}^{1/p}\cdot\{E(|X+Y|^{(p-1)q})\}^{1/q} + \{E(|Y|^p)\}^{1/p}\cdot\{E(|X+Y|^{(p-1)q})\}^{1/q} =$$

$$= \left[\{E(|X|^p)\}^{1/p} + \{E(|Y|^p)\}^{1/p}\right]\cdot\{E(|X+Y|^{(p-1)q})\}^{1/q}.$$

Excluyendo el caso trivial en que $E(|X+Y|^p)=0$, y teniendo en cuenta que $(p-1)q=p$, reagrupando adecuadamente en la anterior desigualdad llegamos al resultado buscado.

ESPERANZA CONDICIONADA

Recordemos que si $\vec{X}$ es un vector aleatorio al que fraccionamos en dos subvectores $\vec{X}_1, \vec{X}_2$, para $\vec{X}_1=\vec{x}_1$ tal que $f_{\vec{X}_1}(\vec{x}_1)>0$, la función

$$f_{\vec{X}_2/\vec{X}_1}(\vec{x}_2) = \frac{f(\vec{x})}{f_{\vec{X}_1}(\vec{x}_1)}, \quad \text{con } \vec{x}=(\vec{x}_1,\vec{x}_2) \text{ y } f \text{ la f.d.p. conjunta}$$

define una f.d.p.; más concretamente define una distribución de probabilidad para el subvector $\vec{X}_2$ condicionado a que el otro subvector $\vec{X}_1$ tome un valor fijo $\vec{x}_1$.

Tendrá pues sentido estudiar la esperanza de este subvector condicionado, utilizando para ello la correspondiente f.d.p. y la definición usual ya establecida para el concepto de esperanza. Por razones de simplicidad vamos a centrarnos en el caso de un vector bidimensional en el que los subvectores serán cada una de las variables que lo componen; la generalización a dimensiones mayores es inmediata.

DEFINICIÓN.- Sea $\vec{X}=(X,Y)$ un vector aleatorio definido sobre el espacio de probabilidad $(\Omega,\mathcal{A},P)$ y sea h una función medible definida de R en R. Supongamos que existe la $E(h(X))$. Entonces la esperanza condicionada de $h(X)$, dado Y, que se escribe $E(h(X)/Y)$, es una variable aleatoria que toma el valor $E(h(X)/y)$, definido mediante

$$E(h(X)/y) = \int_\Omega h(X)\,dP_{Y=y}$$

donde $P_{Y=y}$ quiere significar la distribución de probabilidad condicionada al hecho de que Y toma el valor y. Según el tipo de vector aleatorio que sea $\vec{X}=(X,Y)$ aquella expresión adopta la forma:

1) $$E(h(X)/y) = \sum_{x\in A} h(x)\cdot P(X=x/Y=y) = \sum_{x\in A} h(x)\cdot\frac{f(x,y)}{f_2(y)}$$

donde $A=B_y$, la sección de B mediante y, siendo B tal que $P(\vec{X}\in B)=1$. Esto para el caso en que $\vec{X}$ sea un vector aleatorio del tipo discreto

2) En el caso continuo

$$E(h(X)/y) = \int_R h(x)\cdot f_{X/Y=y}(x)\,dx = \int_R h(x)\,\frac{f(x,y)}{f_2(y)}\,dx.$$

Obsérvese que en ambos casos $E(h(X)/y)$ es, efectivamente, una función del valor y.

Obviamente, una definición similar puede darse para $E(h(Y)/X)$, siempre que exista $E(h(Y))$.

Dada la definición establecida es evidente que la esperanza condicionada goza de todas las propiedades inherentes al concepto de esperanza y que ya hemos estudiado. Recordemos a título de ejemplo:

PROPIEDADES DE LA ESPERANZA CONDICIONADA

a) $E(c/Y) = c$, donde c es constante.

b) $E((aX+b)/Y) = a\,E(X/Y) + b$, a y b constantes.

c) Si h_1 y h_2 son funciones medibles, de manera que sus esperanzas, $E(h_i(X))$ existen, tenemos

$$E\{[a_1 h_1(X) + a_2 h_2(X)]/Y\} = a_1 E(h_1(X)/Y) + a_2 E(h_2(X)/Y).$$

d) Si $X \geq 0$, entonces $E(X/Y) \geq 0$.

e) Si $X_1 \geq X_2$, entonces $E(X_1/Y) \geq E(X_2/Y)$.

Momentos de todo tipo de una distribución de probabilidad condicional se definen de forma análoga a como se establecieron en un principio y gozan también, claro está, de las mismas propiedades. Sin embargo, las características especiales de este tipo de distribuciones dan lugar a algunas propiedades que les son características. Veámoslas.

PROPIEDAD 1.- Si $E(h(X))$ existe, entonces

$$E[E(h(X)/Y)] = E(h(X)).$$

Demostración.- Si $\vec{X}$ es un vector aleatorio continuo, entonces la v.a. $E(h(X)/Y)$ viene definida mediante

$$g(y) = E(h(X)/y) = \int_R h(x)\,\frac{f(x,y)}{f_2(y)}\,dx$$

como función medible de la variable aleatoria Y que es, su esperanza vendrá dada por

$$E[g(Y)] = \int_R g(y)\cdot f_2(y)\,dy = \int_R\left[\int_R h(x)\cdot\frac{f(x,y)}{f_2(y)}\,dx\right] f_2(y)\,dy =$$

$$= \int_R\left[\int_R h(x)\cdot f(x,y)\,dx\right] dy = \int_R h(x)\left[\int_R f(x,y)\,dy\right] dx =$$

$$= \int_R h(x)\, f_1(x)\,dx = E(h(X)).$$

Para el caso discreto la demostración es análoga.

PROPIEDAD 2.- Si $E(X^2) < +\infty$, entonces

$$\operatorname{var}(X) = \operatorname{var}(E(X/Y)) + E(\operatorname{var}(X/Y)).$$

Demostración.- Haciendo uso de la propiedad anterior, tenemos que

$$\operatorname{var}(X) = E((X-E(X))^2) = E\left[E((X-E(X))^2/Y)\right] = E\left[E\{(X^2+(E(X))^2-2XE(X))/Y\}\right] =$$

$$= E\left[E(X^2/Y) + (E(X))^2 - 2E(X)\cdot E(X/Y)\right] = E\left[E(X^2/Y) - (E(X/Y))^2 + (E(X/Y))^2 +\right.$$

$$\left. + (E(X))^2 - 2E(X)\cdot E(X/Y)\right] = E\left[\operatorname{var}(X/Y) + (E(X/Y) - E(X))^2\right] =$$

$$= E(\operatorname{var}(X/Y)) + E\left[(E(X/Y) - E(X))^2\right] = E(\operatorname{var}(X/Y)) + \operatorname{var}(E(X/Y)).$$

COROLARIO.- Si $E(X^2) < +\infty$, entonces de la propiedad 2 se sigue de inmediato que

$$\operatorname{var}(X) \geq \operatorname{var}(E(X/Y)).$$

Obsérvese por otra parte que cuando se verifica la igualdad $\operatorname{var}(X) = \operatorname{var}(E(X/Y))$, ello supone que

$$E(\operatorname{var}(X/Y)) = 0$$

es decir

$$E((X - E(X/Y))^2/Y) = 0$$

y de acuerdo con las propiedades de las integrales de funciones medibles no negativas, ello supone que la función

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

es nula casi por todas partes, o sea

$$P(X = E(X/Y)) = 1$$

lo que supone que X es, casi por todas partes, una función de Y, por cuanto $E(X/Y)$ lo es.

EL PRINCIPIO DE LOS MINIMOS CUADRADOS

Sean X e Y dos variables aleatorias de manera que existe entre ellas una relación de dependencia que deseamos determinar. Supongamos ~~que la relación funcional entre ambas es y = h(x) y además~~ que $E(Y^2)$ y $E[(h(X))^2]$ son finitas. Vamos a tratar de encontrar una función $h(x)$ que aproxime el valor de la relación funcional entre X e Y.

El principio de los mínimos cuadrados consiste en la elección de $h(x)$ de tal suerte que la cantidad

$$E\{(Y-h(X))^2\}$$

sea mínima. ¿Cuál ha de ser entonces $h(X)$?

Observemos que, si (X,Y) es un vector aleatorio continuo, tenemos

$$E\{(Y-h(X))^2\} = \int_{R^2} [y-h(x)]^2 f(x,y)\, dx\, dy = \int_{R^2} [y-h(x)]^2 f_{Y/X}(y) \cdot f_X(x)\, dy\, dx =$$

$$= \int_R f_X(x) \left\{ \int_R [y-h(x)]^2 f_{Y/X}(y)\, dy \right\} dx,$$

que sea mínima esta expresión supone que se minimice lo que figura entre llaves en el integrando y de acuerdo con las propiedades estudiadas para la varianza, dicha cantidad se minimizará cuando $h(x) = E(Y/x)$.

Un resultado análogo se verifica para el caso en que (X,Y) es discreto.

DEFINICION. – La relación $y = E(Y/x)$ se conoce como la regresión de Y sobre X, análogamente $x = E(X/y)$ es la regresión de X sobre Y.

LINEA DE REGRESION. – En ocasiones, por motivos de sencillez fundamentalmente, se está interesado en aproximar la relación entre X e Y mediante una línea recta, $y = ax+b$. En esta situación y haciendo uso del principio de los mínimos cuadrados antes citado, a y b deberán ser elegidos de manera que

$$L = E\{(Y-aX-b)^2\}$$

sea mínimo. Siempre bajo el supuesto de la existencia de $E(Y^2)$ y $E(X^2)$. La obtención de a y b se reduce a un problema de máximos y mínimos, bastando con derivar parcialmente L respecto a ambos parámetros e igualar, posteriormente, a cero dichas derivadas. Se obtienen así para a y b, las expresiones

$$a = \frac{cov(X,Y)}{var(X)}, \quad b = E(Y) - E(X) \cdot \frac{cov(X,Y)}{var(X)}$$

siendo la ecuación de la recta, conocida como línea de regresión de Y sobre X,

$$y - E(Y) = \frac{cov(X,Y)}{var(X)} (x - E(X)). \quad (1)$$

Una expresión análoga se obtiene para la línea de regresión de X sobre Y, a saber

$$x - E(X) = \frac{cov(X,Y)}{var(Y)} (y - E(Y)). \quad (2)$$

Obsérvese que la relación (2) no se obtiene a partir de la (1) despejando en aquella $x - E(X)$, puesto que en ambas relaciones los papeles de x e y aparecen cambiados. La relación (1) se obtiene considerando a x como variable casual, mientras que en la (2) este papel viene desempeñado por y.

Un procedimiento análogo permitiría aproximar la relación entre X e Y mediante un polinomio de grado n, por ejemplo, $y = a + bx + cx^2 + dx^3$, para $n=3$.

DEFINICION. – Si $E(X^2)$ y $E(Y^2)$ existen, a las cantidades

$$\gamma_{X/Y} = \frac{cov(X,Y)}{var(Y)} \quad y \quad \gamma_{Y/X} = \frac{cov(X,Y)}{var(X)}$$

se les designa coeficientes de regresión lineal de Y sobre X y X sobre Y respectivamente, pero al revés.

DEFINICION. – Si $E(X^2)$ y $E(Y^2)$ existen, se define el coeficiente de correlación lineal entre X e Y, mediante

$$\rho = \frac{cov(X,Y)}{(var(X) \cdot var(Y))^{1/2}}$$

a

$$\rho^2 = \gamma_{X/Y} \cdot \gamma_{Y/X}$$

se le denomina en ocasiones coeficiente de determinación. Obsérvese que el signo de ρ es el mismo que el de la $cov(X,Y)$.

DEFINICION. – Decimos que dos variables X e Y son incorreladas si y sólo si $\rho = 0$.

Obviamente si $cov(X,Y) = 0$, entonces $\rho = 0$ y las variables son incorreladas. Posteriormente, cuando introduzcamos el concepto de independencia entre variables, volveremos sobre este concepto.

A continuación estudiaremos algunas propiedades básicas del coeficiente de correlación.

PROPIEDAD 1. – a) El coeficiente de correlación lineal entre X e Y satisface la relación

$$|\rho| \leq 1.$$

b) La igualdad $\rho = \pm 1$ se verifica si existen constantes a y b tales que

$$P\{Y = aX + b\} = 1.$$

Demostración. – a) Recordemos que de acuerdo con la desigualdad de Cauchy-Schwarz

$$[cov(X,Y)]^2 \leq var(X) \cdot var(Y)$$

y de aquí la desigualdad del enunciado.

b) Además, recordemos que si $\rho = \pm 1$, entonces $[cov(X,Y)]^2 = var(X) \cdot var(Y)$

EL PRINCIPIO DE LOS MINIMOS CUADRADOS

Sean X e Y dos variables aleatorias de manera que existe entre ellas una relación de dependencia que deseamos determinar. Supongamos ~~que la relación funcional entre ambas es $y = h(x)$ y además~~ que $E(Y^2)$ y $E[(h(X))^2]$ son finitas. Vamos a tratar de encontrar una función $h(x)$ **que aproxime el valor de la relación funcional entre X e Y.**

El principio de los mínimos cuadrados consiste en la elección de $h(x)$ de tal suerte que la cantidad

$$E\{(Y-h(X))^2\}$$

sea mínima. ¿Cuál ha de ser entonces $h(X)$?

Observemos que, si (X,Y) es un vector aleatorio continuo, tenemos

$$E\{(Y-h(X))^2\} = \int_{R^2} [y-h(x)]^2 f(x,y)\,dx\,dy = \int_{R^2} [y-h(x)]^2 f_{Y/X}(y)\cdot f_X(x)\,dy\,dx =$$

$$= \int_R f_X(x)\left\{\int_R [y-h(x)]^2 f_{Y/X}(y)\,dy\right\} dx,$$

que sea mínima esta expresión supone que se minimice lo que figura entre llaves en el integrando y de acuerdo con las propiedades estudiadas para la varianza, dicha cantidad se minimizará cuando $h(x) = E(Y/x)$.

Un resultado análogo se verifica para el caso en que (X,Y) es discreto.

DEFINICION.– La relación $y = E(Y/x)$ se conoce como la regresión de Y sobre X, análogamente $x = E(X/y)$ es la regresión de X sobre Y.

LINEA DE REGRESION.– En ocasiones, por motivos de sencillez fundamentalmente, se está interesado en aproximar la relación entre X e Y mediante una línea recta, $y = ax+b$. En esta situación y haciendo uso del principio de los mínimos cuadrados antes citado, a y b deberán ser elegidos de manera que

$$L = E\{(Y-aX-b)^2\}$$

sea mínimo. Siempre bajo el supuesto de la existencia de $E(Y^2)$ y $E(X^2)$, la obtención de a y b se reduce a un problema de máximos y mínimos, bastando con derivar parcialmente L respecto a ambos parámetros e igualar, posteriormente, a cero dichas derivadas. Se obtienen así para a y b, las expresiones

$$a = \frac{cov(X,Y)}{var(X)}, \qquad b = E(Y) - E(X)\cdot\frac{cov(X,Y)}{var(X)}$$

siendo la ecuación de la recta, conocida como línea de regresión de Y sobre X,

$$y - E(Y) = \frac{cov(X,Y)}{var(X)}\,(x - E(X)). \qquad (1)$$

Una expresión análoga se obtiene para la línea de regresión de X sobre Y, a saber

$$x - E(X) = \frac{cov(X,Y)}{var(Y)}\,(y - E(Y)). \qquad (2)$$

Obsérvese que la relación (2) no se obtiene a partir de la (1) despejando en aquella $x - E(X)$, puesto que en ambas relaciones los papeles de x e y aparecen cambiados. La relación (1) se obtiene considerando a x como variable causal, mientras que en la (2) este papel viene desempeñado por y.

Un procedimiento análogo permitiría aproximar la relación entre X e Y mediante un polinomio de grado n, por ejemplo, $y = a + bx + cx^2 + dx^3$, para $n=3$.

DEFINICION.– Si $E(X^2)$ y $E(Y^2)$ existen, a las cantidades

$$\gamma_{X/Y} = \frac{cov(X,Y)}{var(Y)} \quad y \quad \gamma_{Y/X} = \frac{cov(X,Y)}{var(X)}$$

se les designa coeficientes de regresión lineal de Y sobre X y X sobre Y respectivamente, pero al revés.

DEFINICION.– Si $E(X^2)$ y $E(Y^2)$ existen, se define el coeficiente de correlación lineal entre X e Y, mediante

$$\rho = \frac{cov(X,Y)}{(var(X)\cdot var(Y))^{1/2}}$$

a

$$\rho^2 = \gamma_{X/Y}\cdot\gamma_{Y/X}$$

se le denomina en ocasiones coeficiente de determinación. Obsérvese que el signo de ρ es el mismo que el de la $cov(X,Y)$.

DEFINICION.– Decimos que dos variables X e Y son incorreladas si y sólo si $\rho = 0$.

Obviamente si $cov(X,Y) = 0$, entonces $\rho = 0$ y las variables son incorreladas. Posteriormente, cuando introduzcamos el concepto de independencia entre variables, volveremos sobre este concepto.

A continuación estudiaremos algunas propiedades básicas del coeficiente de correlación.

PROPIEDAD 1.– a) El coeficiente de correlación lineal entre X e Y satisface la relación

$$|\rho| \le 1.$$

b) La igualdad $\rho = \pm 1$ se verifica si existen constantes a y b tales que

$$P\{Y = aX + b\} = 1.$$

Demostración.– a) Recordemos que de acuerdo con la desigualdad de Cauchy-Schwarz

$$[Cov(X,Y)]^2 \le var(X)\cdot var(Y)$$

y de aquí la desigualdad del enunciado.

b) Además, recordemos que si $\rho = \pm 1$, entonces $[Cov(X,Y)]^2 = var(X)\cdot var(Y)$

y ello ocurre si y sólo si $P\{aX+Y=0\}=1$, que es de la forma $P\{Y=aX+b\}=1$, mediante una adecuada traslación de Y.

COROLARIO.- Las dos líneas de regresión (1) y (2) coinciden si y sólo si $\rho=1$ o $\rho=-1$.

PROPIEDAD 2.- Sean $E(X^2)<+\infty$, $E(Y^2)<+\infty$ y hagamos $U=aX+b$, $V=cY+d$, entonces

$$\rho_{X,Y}=\pm\rho_{U,V};$$

donde $\rho_{X,Y}$ y ρ_{UV} son, respectivamente, los coeficientes de correlación lineal de X e Y y de U y V.

Demostración.- La propiedad se deduce de las propiedades de los momentos ya estudiados.

EJEMPLOS

(1) Sea (X,Y) un vector aleatorio con función de densidad conjunta

$$f(x,y)=\begin{cases} x+y, & 0<x<1,\ 0<y<1 \\ 0 & \text{en el resto.}\end{cases}$$

Entonces

$$E(X^l Y^m)=\int_0^1\int_0^1 x^l y^m (x+y)\,dx\,dy=\int_0^1\int_0^1 x^{l+1}y^m\,dx\,dy+\int_0^1\int_0^1 x^l y^{m+1}\,dx\,dy$$

$$=\frac{1}{(l+2)(m+1)}+\frac{1}{(l+1)(m+2)};\quad \text{con } l \text{ y } m \text{ enteros positivos.}$$

A partir de este resultado, tenemos

$l=1,\ m=0,\quad E(X)=\frac{7}{12}\quad;\quad l=0,\ m=1,\quad E(Y)=\frac{7}{12}.$

$l=2,\ m=0,\quad E(X^2)=\frac{5}{12}\quad;\quad l=0,\ m=2,\quad E(Y^2)=\frac{5}{12}$

$$\mathrm{var}(X)=\frac{5}{12}-\frac{49}{144}=\frac{11}{144}=\mathrm{var}(Y)$$

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{3}-\frac{49}{144}=\frac{48-49}{144}=-\frac{1}{144}.$$

(2) Recordemos el ejemplo 5 (pag 8. de vectores aleatorios). El vector aleatorio (X,Y), con f.d.p conjunta

$$f(x,y)=\begin{cases} 2, & 0<x<y<1 \\ 0 & \text{en resto}\end{cases}$$

tenía por densidades marginales $f_1(x)=\begin{cases}2(1-x), & 0<x<1\\ 0, & \text{en el resto}\end{cases}$

y $f_2(y)=\begin{cases}2y, & 0<y<1\\ 0, & \text{en el resto.}\end{cases}$

Asimismo las funciones de densidad condicionadas venían dadas por (véase ejemplo 7, pag. 13 del mismo capítulo)

$$f_{X/y}(x)=\frac{1}{y},\quad 0<x<y$$

$$f_{Y/x}(y)=\frac{1}{1-x},\quad x<y<1.$$

En ambos casos se trata de distribuciones de probabilidad uniformes en el correspondiente intervalo. A partir de aquí

$$E(Y/x)=\int_x^1 y f_{Y/x}(y)\,dy=\int_x^1 y\cdot\frac{1}{1-x}\,dy=\frac{1}{1-x}\cdot\frac{1-x^2}{2}=\frac{1+x}{2},\quad 0<x<1.$$

$$E(X/y)=\int_0^y x f_{X/y}(x)\,dx=\int_0^y x\cdot\frac{1}{y}\,dx=\frac{1}{y}\cdot\frac{y^2}{2}=\frac{y}{2},\quad 0<y<1.$$

Análogamente

$$E(X^2/y)=\int_0^y x^2\cdot\frac{1}{y}\,dx=\frac{1}{y}\cdot\frac{y^3}{3}=\frac{y^2}{3},\quad 0<y<1$$

y por tanto

$$\mathrm{var}(X/y)=E(X^2/y)-[E(X/y)]^2=\frac{y^2}{3}-\frac{y^2}{4}=\frac{y^2}{12},\quad 0<y<1$$

Por otra parte

$$E(X)=\int_0^1 x f_1(x)\,dx=\int_0^1 x\,2(1-x)\,dx=\int_0^1(2x-2x^2)\,dx=\left[x^2-\frac{2x^3}{3}\right]_0^1=$$

$$=1-\frac{2}{3}=\frac{1}{3}$$

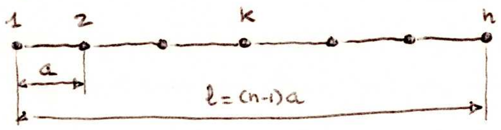
pero

$$E(E(X/Y))=\int_0^1 E(X/y)f_2(y)\,dy=\int_0^1\frac{y}{2}\cdot 2y\,dy=\int_0^1 y^2\,dy=\left[\frac{y^3}{3}\right]_0^1=\frac{1}{3},$$

que coincide, como ya sabíamos, con $E(X)$.

(3) La igualdad $E(X) = E(E(X/Y))$ es en ocasiones muy útil para obtener la primera de las esperanzas a partir de la segunda, cuando la obtención directa de aquella es complicada. Veamos el siguiente ejemplo:

"Un trabajador opera con determinado tipo de máquinas en un número n. Las máquinas están alineadas y separadas entre sí una distancia a. Supongamos que el operador se traslada de una máquina a otra según un orden de prioridad totalmente aleatorio. Encontrar la longitud media del camino que recorre el operador para atender dos máquinas consecutivamente".



Numeramos las máquinas de izquierda a derecha, de 1 a n, y designamos por B_k el suceso, "el operador está atendiendo la máquina k". Puesto que las máquinas son todas del mismo tipo, y el orden de prioridad es aleatorio, la probabilidad p_i^k, de que la próxima máquina en requerir los servicios del operador sea la i, es $1/n$ $(1 \le i \le n)$. La longitud del camino que deberá recorrer el operador para atender dicha máquina, vendrá dada por

$$\lambda_i^k = \begin{cases} (k-i)a, & k \ge i \\ (i-k)a, & k < i \end{cases}$$

designando por λ la v.a. correspondiente, tendremos

$$E(\lambda/B_k) = \sum_{i=1}^{n} \lambda_i^k \cdot p_i^k = \frac{1}{n}\left[\sum_{i=1}^{k}(k-i)a + \sum_{i=k+1}^{n}(i-k)a\right] =$$

$$= \frac{a}{n}\left[\frac{k(k-1)}{2} + \frac{(n-k)(n-k+1)}{2}\right] = \frac{a}{2n}\left[2k^2 - 2k(n+1) + n(n+1)\right]$$

Por otra parte la probabilidad de que el operador esté sirviendo a la máquina k, es la misma para todos los k, es decir $P(B_k) = \frac{1}{n}$, $1 \le k \le n$, de acuerdo con esto

$$E(\lambda) = E\left[E(\lambda/B_k)\right] = \sum_{k=1}^{n} E(\lambda/B_k) \cdot P(B_k) =$$

$$= \sum_{k=1}^{n} \frac{a}{2n^2}\left[2k^2 - 2k(n+1) + n(n+1)\right] = \frac{a}{2n^2}\left[2\sum_{k=1}^{n} k^2 - 2(n+1)\sum_{k=1}^{n} k + n^2(n+1)\right]$$

$$= \frac{a}{2n^2}\left[2 \cdot \frac{n(n+1)(2n+1)}{6} - 2(n+1) \cdot \frac{n(n+1)}{2} + n^2(n+1)\right] =$$

$$= \frac{a}{2n^2}\left[\frac{n(n+1)(2n+1)}{3} - n(n+1)\right] = \frac{a}{2n^2}\left[n(n+1)\left\{\frac{2n+1}{3} - 1\right\}\right] =$$

$$= \frac{a}{2n^2} \cdot \frac{n(n+1)}{3} \cdot (2n-2) = \frac{2an}{6n^2}(n^2-1) = \frac{a(n^2-1)}{3n} = \frac{\ell(n+1)}{3n} = \boxed{\frac{\ell}{3}\left(1 + \frac{1}{n}\right)}$$

(4) Sea (X,Y) un vector aleatorio con densidad conjunta dada por

$$f(x,y) = \begin{cases} e^{-y}, & 0 < x < y < +\infty \\ 0, & \text{en el resto}. \end{cases}$$

Las correspondientes marginales vendrán dadas por

$$f_1(x) = \int_R f(x,y)\,dy = \int_x^{+\infty} e^{-y}\,dy = \left(-e^{-y}\right)_x^{+\infty} = e^{-x}, \quad 0 < x < +\infty$$
$$= 0, \quad \text{en el resto}$$

$$f_2(y) = \int_R f(x,y)\,dx = \int_0^y e^{-y}\,dx = e^{-y} \cdot y, \quad 0 < y < +\infty$$
$$= 0, \quad \text{en el resto}.$$

A partir de aquí las densidades condicionales se obtienen fácilmente,

$$f_{X/Y}(x) = \frac{f(x,y)}{f_2(y)} = \frac{e^{-y}}{e^{-y} \cdot y} = \frac{1}{y}, \quad 0 < x < y$$
$$= 0, \quad \text{en el resto}.$$

$$f_{Y/X}(y) = \frac{f(x,y)}{f_1(x)} = \frac{e^{-y}}{e^{-x}} = e^{x-y}, \quad x < y < +\infty$$
$$= 0, \quad \text{en el resto}.$$

Entonces

$$E(X/y) = \int_R x\, f_{X/Y}(x)\,dx = \int_0^y x \cdot \frac{1}{y}\,dx = \frac{1}{y} \cdot \frac{y^2}{2} = \frac{y}{2}, \quad 0 < y < +\infty$$

$$E(Y/x) = \int_R y\, f_{Y/X}(y)\,dy = \int_x^{+\infty} y \cdot e^{x-y}\,dy = \int_0^{+\infty} (u+x)e^{-u}\,du = x + 1, \quad 0 < x < +\infty$$

En este caso la relación entre X e Y, por ser lineal en ambos casos, coincidirá con las líneas de regresión de X sobre Y e Y sobre X. De acuerdo con esto los coeficientes de regresión valdrán

$$\gamma_{X/Y} = \frac{1}{2} \quad , \quad \gamma_{Y/X} = 1 .$$

Por tanto

$$e^2 = \gamma_{X/Y} \cdot \gamma_{Y/X} = \frac{1}{2}$$

y

$$e = +\sqrt{1/2} ,$$

estando el signo determinado a partir del signo de la covarianza. La covarianza viene dada por

$$\operatorname{cov}(X,Y) = E(XY) - E(X).E(Y)$$

Se computa que

$$E(X,Y) = \int_0^{+\infty}\int_0^{y} xy\, e^{-y}\, dx\, dy = 3$$

$$E(X) = \int_0^{+\infty} x e^{-x}\, dx = 1 \quad , \quad E(Y) = \int_0^{+\infty} y^2 e^{-y}\, dy = 2 .$$

y por tanto

$$\operatorname{cov}(X,Y) = 3 - 2 = 1 .$$

MATRIZ DE COVARIANZAS DE UN VECTOR ALEATORIO

Ya hemos dicho que para un vector aleatorio, $\vec{X}$, de dimensión n, la esperanza $E\left((X_1-E(X_1))^{p_1}(X_2-E(X_2))^{p_2}\cdots\cdots(X_n-E(X_n))^{p_n}\right)$ se conoce como el momento central conjunto de orden $(p_1, p_2, \ldots, p_n)$. En ocasiones el orden del momento suele asociarse a la suma de los exponentes, en este caso diríamos que el momento es de orden $p_1+p_2+\cdots+p_n$. De acuerdo con esta nomenclatura tienen especial interés todos los <u>momentos de segundo orden</u>, centrales, desde luego, de un vector aleatorio. En efecto, los momentos de segundo orden de $\vec{X}$, supone que de los p_i, $i=1,\ldots,n$, a lo sumo 2 son distintos de cero, y todos los restantes son nulos, así pues, puede ocurrir que $p_i=1$, $p_j=1$, $p_k=0$, $k\neq i$, $k\neq j$, o bien, $p_i=2$, $p_j=0$, $j\neq i$. En la primera de estas situaciones lo que obtendremos será la covarianza entre las variables X_i y X_j, en la segunda obtendremos la varianza de i. Designemos por σ_{ij} la covarianza citada y por σ_{ii} la varianza. Es costumbre disponer estos valores en una matriz cuadrada de dimensión n, a la que se suele denotar mediante Σ, y que suele ser conocida

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \sigma_{21} & & \sigma_{2n} \\ \vdots & \sigma_{ij} & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix}$$

como la <u>matriz de covarianzas</u> del vector aleatorio $\vec{X}$. Obsérvese que los elementos de Σ son las covarianzas entre las variables X_i y X_j (fila y columna respectivamente del elemento σ_{ij}). Los elementos de la diagonal principal σ_{ii}, son las varianzas de la correspondiente variable, a saber, X_i. Además, dada la definición de covarianza, la matriz es, evidentemente, simétrica, de forma que $\sigma_{ij}=\sigma_{ji}$. La matriz de covarianzas es un parámetro muy importante en el conocimiento de las características de la distribución de probabilidad de un vector aleatorio, supone, en cierta manera, la generalización del concepto de varianza para el caso unidimensional.

Por ejemplo, en el anterior ejemplo nº 1 de la página, tenemos un vector bidimensional, cuya matriz de covarianzas, de 2x2, en este caso viene dada por

$$\Sigma = \begin{bmatrix} \frac{11}{144} & -\frac{1}{144} \\ -\frac{1}{144} & \frac{11}{144} \end{bmatrix} .$$

La matriz de covarianzas goza de otras interesantes propiedades de las que en su momento nos ocuparemos.