Distribuciones Bidimensionales. Regresión y Correlación

01. Introducción. Variable aleatoria bidimensional

Consideremos el espacio probabilístico (U,Φ,p) donde es U el espacio muestral de sucesos, Φ es la sigma álgebra asociada y p es la medida de probabilidad.

Una variable aleatoria bidimensional sobre el espacio probabilizable (U,Φ) es una aplicación

$$(X,Y):U\rightarrow R^2 \text{ tal que } [X\leq x,Y\leq y]\in\Phi$$

o sea, siendo $[X\leq x,Y\leq y]=\{u\in U / X(u)\leq x,Y(u)\leq y\}\in\Phi$

Teorema 1.
La condición necesaria y suficiente para que el par (X,Y) sea una variable aleatoria bidimensional sobre el espacio probabilizable (U,Φ) es que ambos elementos del par, por separado, sean variables aleatorias sobre el mismo espacio probabilizable.
Demostración:

- Veamos que si (X,Y) es variable aleatoria, entonces X es variable aleatoria e Y es también variable aleatoria:

Consideremos la sucesión $\{A_n\}_{n\geq 1}$, donde son $A_n=[X\leq x,Y\leq n], n=1,2,...,\forall x\in R$. Como es Φ una sigma-álgebra, se tiene que la suma infinita pertenece a la misma, por lo que $\bigcup_{n=1}^{\infty}A_n=\bigcup_{n=1}^{\infty}[X\leq x,Y\leq n]=[X\leq x]\in\Phi$, lo que implica que X es variable aleatoria.
Consideremos la sucesión $\{B_n\}_{n\geq 1}$, donde son $B_n=[X\leq n,Y\leq y], n=1,2,...,\forall y\in R$. Como es Φ una sigma-álgebra, se tiene que la suma infinita pertenece a la misma, por lo que $\bigcup_{n=1}^{\infty}B_n=\bigcup_{n=1}^{\infty}[X\leq n,Y\leq y]=[Y\leq y]\in\Phi$, lo que implica que Y es variable aleatoria.

- Veamos ahora que si cada elemento, por separado es variable aleatoria, entonces el par (X,Y) es variable aleatoria:

Trivialmente, pues $[X\leq x]\in\Phi\wedge[Y\leq y]\in\Phi\rightarrow[X\leq x]\bigcap[Y\leq y]=[X\leq x,Y\leq y]\in\Phi$

02. Función de distribución bidimensional

Llamaremos distribución bidimensional al par constituido por un espacio probabilístico (U,Φ,p) y una variable aleatoria bidimensional (X,Y). Es decir, al par $\{(U,\Phi,p),(X,Y)\}$.

Puesto que, por el teorema 1, las variables X e Y son variables aleatorias, se tiene que son distribuciones monodimensionales los pares

$$\{(U,\Phi,p),X\} \text{ y } \{(U,\Phi,p),Y\}$$

que se denominan *distribuciones marginales* de la distribución bidimensional.

Definición 1:

Se llama Función de Distribución de la distribución bidimensional $\{(U,\Phi,p),(X,Y)\}$ a la función $F : R^2 \to R$ definida por:

$$\forall (x,y)\in R^2,\ F(x,y) = p[X \le x, Y \le y]$$

A fin de probar a continuación algunas de las propiedades de la Función de Distribución es conveniente analizar un lema elemental sobre variables aleatorias.

Lema 1:

Sean las variables aleatorias $M = [X \le x], \forall x \in R$ y $N = [Y \le y], \forall y \in R$ sobre el mismo espacio probabilistico (U,Φ,p). Se verifican las relaciones:

1) $M = \left(\left(\bigcup_{k=0}^{\infty}\left(M \cap [-(k+1) < Y \le -k]\right)\right)\right) \cup \left(M \cap [Y > 0]\right)$

$$M = \left(M \cap [Y \le 0]\right) \cup \left(\bigcup_{k=0}^{\infty}\left(M \cap [k < Y \le k+1]\right)\right)$$

2) $N = \left(\left(\bigcup_{k=0}^{\infty}\left(N \cap [-(k+1) < X \le -k]\right)\right)\right) \cup \left(N \cap [X > 0]\right)$

$$N = \left(N \cap [X \le 0]\right) \cup \left(\bigcup_{k=0}^{\infty}\left(N \cap [k < X \le k+1]\right)\right)$$

Demostración:

1) Puesto que $M = M \cap [-\infty < Y \le +\infty]$, se tiene:

$$M = M \cap \left([Y \le 0] \cup [y > 0]\right) = \left(M \cap [Y \le 0]\right) \cup \left(M \cap [Y > 0]\right) =$$

$$= \left(M \cap \left(\bigcup_{k=0}^{\infty}[-(k+1) < Y \le -k]\right)\right) \cup \left(M \cap [Y > 0]\right) =$$

$$= \left(\left(\bigcup_{k=0}^{\infty}\left(M \cap [-(k+1) < Y \le -k]\right)\right)\right) \cup \left(M \cap [Y > 0]\right)$$

Puesto que $M = M \cap [-\infty < Y \le +\infty]$, se tiene:

$$M = M \cap ([Y \le 0] \cup [y > 0]) = (M \cap [Y \le 0]) \cup (M \cap [Y > 0]) =$$
$$= (M \cap [Y \le 0]) \cup \left(M \cap \left(\bigcup_{k=0}^{\infty} [k < Y \le k+1] \right) \right) =$$
$$= (M \cap [Y \le 0]) \cup \left(\bigcup_{k=0}^{\infty} (M \cap [k < Y \le k+1]) \right)$$

2) Análoga al caso anterior.

Consecuencia del lema 1:
Si se toman probabilidades en las expresiones anteriores, y teniendo en cuenta que se trata de sucesos disjuntos:

1) $p(M) = p(M \cap [Y > 0]) + \sum_{k=0}^{\infty} p(M \cap [-(k+1) < Y \le -k])$

$$p(M) = p(M \cap [Y \le 0]) + \sum_{k=0}^{\infty} p(M \cap [k < Y \le k+1])$$

2) $p(N) = p(N \cap [X > 0]) + \sum_{k=0}^{\infty} p(N \cap [-(k+1) < X \le -k])$

$$p(N) = p(N \cap [X \le 0]) + \sum_{k=0}^{\infty} p(N \cap [k < X \le k+1])$$

Como es obvio, las series que intervienen son convergentes.

Teorema 2:
Propiedades de la Función de Distribución:

1) $F(x,y)$ está definida, $\forall (x,y) \in R^2$

Demostración:
Trivialmente, pues se trata de una probabilidad, que está obviamente definida: $F(x,y) = p[X \le x, Y \le y]$

2) $0 \le F(x,y) \le 1, \ \forall (x,y) \in R^2$

Demostración:
Puesto que el rango de valores de la probabilidad es el intervalo $[0,1]$ y siendo *F(x,y)* una probabilidad, será $0 \le F(x,y) \le 1, \ \forall (x,y) \in R^2$

3) Si $x_1 \le x_2 \ y \ y_1 \le y_2 \to F(x_1,y_1) \le F(x_2,y_2)$, es decir, *F(x,y)* es monótona creciente.

Demostración:
$$x_1 \le x_2 \ y \ y_1 \le y_2 \to [X \le x_1, Y \le y_1] \subseteq [X \le x_2, Y \le y_2] \to p[X \le x_1, Y \le y_1] \le$$
$$\le p[X \le x_2, Y \le y_2] \to F(x_1,y_1) \le F(x_2,y_2)$$

4) Se verifica que

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

Demostración:

1) Puesto que la serie $\sum_{k=0}^{\infty} p(M \cap [-(k+1) < Y \le -k])$ es convergente, se tiene:

$$\forall \varepsilon > 0, \exists n_0 \in N / \forall n > n_0, \sum_{k=n}^{\infty} p(M \cap [-(k+1) < Y \le -k]) < \varepsilon \quad (*)$$

es decir $F(x,-n) < \varepsilon$, de lo cual $F(x,-n) \to 0, si \ n \to \infty$, por tanto:

$$\lim_{n \to \infty} F(x,-n) = \lim_{n \to -\infty} F(x,n) \equiv \lim_{y \to -\infty} F(x,y) = 0$$

2) De forma análoga $\sum_{k=0}^{\infty} p(N \cap [-(k+1) < X \le -k])$ es convergente, se tiene:

$\forall \varepsilon > 0, \exists n_0 \in N / \forall n > n_0, \sum_{k=n}^{\infty} p(N \cap [-(k+1) < X \le -k]) < \varepsilon$ (*)

es decir $F(-n,y) < \varepsilon$, de lo cual $f(-n,y) \to 0, si\ n \to \infty$, por tanto:

$$\lim_{n \to \infty} F(-n,y) = \lim_{n \to -\infty} F(n,y) \equiv \lim_{x \to -\infty} F(x,y) = 0$$

(*) la justificación puede verse en el apéndice A, al final del texto.

5) Se verifica que los límites

$$\lim_{y \to +\infty} F(x,y) = F_1(x), \quad \lim_{x \to +\infty} F(x,y) = F_2(y)$$

son las Funciones de Distribución de las distribuciones marginales $\{(U,\Phi,p),X\}$ y $\{(U,\Phi,p),Y\}$

Demostración:

1) Puesto que la serie $\sum_{k=0}^{\infty} p(M \cap [k < Y \le k+1])$ es convergente, se tiene:

$$\forall \varepsilon > 0, \exists n_0 \in N / \forall n > n_0, \sum_{k=n}^{\infty} p(M \cap [k < Y \le k+1]) < \varepsilon$$

por tanto:

$$p(M) = p(M \cap [Y \le 0]) + \sum_{k=0}^{\infty} p(M \cap [k < Y \le k+1]) = p(M \cap [Y \le 0]) +$$

$$+ \sum_{k=0}^{n} p(M \cap [k < Y \le k+1]) + \sum_{k=n}^{\infty} p(M \cap [k < Y \le k+1])$$

es decir: $P(M) = F(x,n) + \sum_{k=n}^{\infty} p(M \cap [k < Y \le k+1])$

y en el límite:

$$P(M) = \lim_{n \to \infty} F(x,n) + 0 = \lim_{n \to \infty} F(x,n) \to p[X \le x] = F_1(x) = \lim_{n \to \infty} F(x,n)$$

en definitiva

$$\lim_{y \to +\infty} F(x,y) = F_1(x)$$

2) La prueba del otro límite es del todo análoga.

6) Se verifica que

$$\lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x,y) = 1$$

Demostración:

$$\lim_{x \to \infty} F_1(x) = 1$$

por lo que $\forall \varepsilon > 0, \exists x_0 \in R / \forall x > x_0, 1 - F_1(x) < \varepsilon/2$

Asimismo, de 5) se tiene:

$$\lim_{y \to \infty} F(x,y) = F_1(x)$$

con lo que $\forall \varepsilon > 0, \exists x_0 \in R / \forall x > x_0, F_1(x) - F(x,y) < \varepsilon/2$

Y de ambas desigualdades: $1 - F_1(x) + F_1(x) - F(x,y) < \varepsilon/2 + \varepsilon/2 \to 1 - F(x,y) < \varepsilon$

$$\lim_{x,y \to \infty} F(x,y) = 1$$

7) *F(x,y)* es continua por la derecha de cada variable.
Demostración:
Bastará probar que para $n \to \infty$ es $\lim F(x+2^{-n}, y+2^{-n}) = F(x,y)$, lo cual comprobaremos por separado para cada variable.

Consideremos la sucesión $\{A_n\}_{n\geq 1}$ con $A_n = [X \leq x + 2^{-n}, Y \leq y]$. Se tiene:

$$\lim_{n \to \infty} F(x+2^{-n}, y) = \lim_{n \to \infty} p(A_n) = p\left(\lim_{n \to \infty} A_n\right) = p\left(\bigcap_{n=1}^{\infty} A_n\right) = p[X \leq x, Y \leq y] = F(x,y)$$

Análogamente para la otra variable:

Consideremos la sucesión $\{B_n\}_{n\geq 1}$ con $B_n = [X \leq x, Y \leq y + 2^{-n}]$. Se tiene:

$$\lim_{n \to \infty} F(x, y+2^{-n}) = \lim_{n \to \infty} p(B_n) = p\left(\lim_{n \to \infty} B_n\right) = p\left(\bigcap_{n=1}^{\infty} B_n\right) = p[X \leq x, Y \leq y] = F(x,y)$$

por consiguiente

$$\lim_{n \to \infty} F(x+2^{-n}, y+2^{-n}) = F(x,y)$$

Podemos comprobar que sin embargo, no es continua a la izquierda de cada variable, es decir, que para $n \to \infty$ es $\lim F(x-2^{-n}, y-2^{-n}) \neq F(x,y)$ en general.
Consideremos la sucesión $\{C_n\}_{n\geq 1}$ con $C_n = [X \leq x - 2^{-n}, Y \leq y - 2^{-n}]$. Se tiene:

$$\lim_{n \to \infty} F(x-2^{-n}, y-2^{-n}) = \lim_{n \to \infty} p(C_n) = p\left(\lim_{n \to \infty} C_n\right) = p\left(\bigcup_{n}^{\infty} C_n\right) = p[X < x, Y < y] =$$

$$= p[X \leq x, Y \leq y] - p[X = x, Y = y] = F(x,y) - p[X = x, Y = y]$$

Obviamente, sería continua a la izquierda de cada variable si $p[X = x, Y = y] = 0$

8) Si $x_1 < x_2$, $y_1 < y_2$ entonces:

$p[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F(x_2,y_2) + F(x_1,y_1) - F(x_1,y_2) - F(x_2,y_1)$

Demostración:
Si $x_1 < x_2$, $y_1 < y_2$, se tiene:

$$[x_1 < X \leq x_2, y_1 < Y \leq y_2] = [X \leq x_2, Y \leq y_2] - [X \leq x_2, Y \leq y_1] \cup [X \leq x_1, Y \leq y_2]$$

si tomamos probabilidades:

$$\text{p}[x_1 < X \leq x_2, y_1 < Y \leq y_2] = p[X \leq x_2, Y \leq y_2] - p([X \leq x_2, Y \leq y_1] \cup [X \leq x_1, Y \leq y_2]) =$$
$$= p[X \leq x_2, Y \leq y_2] - p[X \leq x_2, Y \leq y_1] - p[X \leq x_1, Y \leq y_2] + p[X \leq x_1, Y \leq y_1]) =$$
$$= F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$$

Definición 2:
Dos variables aleatorias, X e Y, sobre un mismo espacio probabilistico (U,Φ,p) se dice que son *estocásticamente independientes* sii

$$\forall (x,y)\in R^2,\ p[X\le x, Y\le y]=p[X\le x].p[Y\le y]$$

Teorema 3:
La condición necesaria y suficiente para que dos variables aleatorias X,Y sobre un mismo espacio probabilistico (U,Φ,p) sean estocásticamente independientes es que

$$\forall (x,y)\in R^2,\ F(x,y)=F_1(x).F_2(y)$$

Demostración:
Trivialmente, pues si ambas variables, X,Y, son estocasticamente independientes,
$\forall (x,y)\in R^2,\ F(x,y)=p[X\le x, Y\le y]=p[X\le x].p[Y\le y]=F_1(x).F_2(y)$
Recíprocamente, si se verifica la relación entre las Funciones de Distribución, será:
$\forall (x,y)\in R^2,\ p[X\le x, Y\le y]=F(x,y)=F_1(x).F_2(y)=p[X\le x].p[Y\le y]$

03. Distribuciones bidimensionales discretas

Definición 3:
Una distribución bidimensional es de tipo discreto si ambas distribuciones marginales unidimensionales son de tipo discreto.

Teorema 4:
Consideremos una distribución bidimensional discreta cuyas distribuciones marginales se encuentran en las sucesiones de puntos $\{x_1,x_2,...\}$ y $\{y_1,y_2,...\}$. Si llamamos $p[X=x_i]=p_{1i}$ y $p[Y=y_k]=p_{2k}$, respectivamente, se verifica:

1) La masa de la distribución está en los puntos $\{(x_1,y_1),(x_2,y_2),...\}$.
2) Si llamamos $p[X=x_i, Y=y_k]=p_{ik}$, se tiene que $p_{1i}=\sum_{k=1}^{\infty} p_{ik}$, $p_{2k}=\sum_{i=1}^{\infty} p_{ik}$.
3) $F(x,y)=\sum_{\substack{x_i\le x\\ y_k\le y}} p_{ik}$
4) Las variables X,Y son estocásticamente independientes sii $p_{ik}=p_{1i}.p_{2k}$.

Demostración:

1) Es evidente, pues (X,Y) solo toma los valores $\{(x_i,y_k)\}, i,k=1,2,...$
2) Llamemos $M=[X=x_i]$. Se tiene:

$$M=\bigcup_{k=1}^{\infty}(M\cap[Y=y_k])\rightarrow p_{1i}=p(M)=\sum_{k=1}^{\infty}p(M\cap[Y=y_k])=\sum_{k=1}^{\infty}p([X=x_i]\cap[Y=y_k])=$$

$$=\sum_{k=1}^{\infty}p[X=x_i, Y=y_k]=\sum_{k=1}^{\infty}p_{ik}$$

Análogamente para la otra variable: llamemos $N = [Y = y_k]$. Se tiene:

$$N = \bigcup_{i=1}^{\infty}\left(N \cap [X = x_i]\right) \to p_{2k} = p(N) = \sum_{i=1}^{\infty} p\left(N \cap [X = x_i]\right) = \sum_{i=1}^{\infty} p\left([X = x_i] \cap [Y = y_k]\right) =$$

$$= \sum_{i=1}^{\infty} p[X = x_i, Y = y_k] = \sum_{i=1}^{\infty} p_{ik}$$

3) $F(x,y) = p[X \le x, Y \le y] = p\left(\bigcup_{\substack{x_i \le x \\ y_k \le y}}^{\infty} [X = x_i, Y = y_k]\right) = \sum_{\substack{x_i \le x \\ y_k \le y}} p[X = x_i, Y = y_k] = \sum_{\substack{x_i \le x \\ y_k \le y}} p_{ik}$

4) Si son estocásticamente independientes, se tendrá:

$$p[X \le x, Y \le y] = p[X \le x].p[Y \le y] \to \sum_{\substack{x_i \le x \\ y_k \le y}} p_{ik} = \sum_{x_i \le x} p_{1i} . \sum_{y_k \le y} p_{2k} = \sum_{\substack{x_i \le x \\ y_k \le y}} p_{1i} . p_{2k} \to$$

$$\to p_{ik} = p_{1i} . p_{2k}$$

Recíprocamente:

$$p_{ik} = p_{1i} . p_{2k} \to p[X = x_i, Y = y_k] = p[X = x_i].p[Y = y_k] \to \sum_{\substack{x_i \le x \\ y_k \le y}} p[X = x_i, Y = y_k] =$$

$$= \sum_{\substack{x_i \le x \\ y_k \le y}} p[X = x_i].p[Y = y_k] = \sum_{\substack{x_i \le x \\ y_k \le y}} p[X = x_i]. \sum_{\substack{x_i \le x \\ y_k \le y}} p[Y = y_k] = p[X \le x].p[Y \le y]$$

04. Distribuciones bidimensionales absolutamente continuas

Definición 4:
Una distribución bidimensional no discreta se dice que es *absolutamente continua* si existe una función $f(x,y)$ no negativa, $f(x,y) \ge 0$, medible Lebesgue, llamada *función de densidad de la distribución,* que verifica:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v).du.dv$$

es decir,

$$\frac{\partial^2 F(x,y)}{\partial x . \partial y} = f(x,y)$$

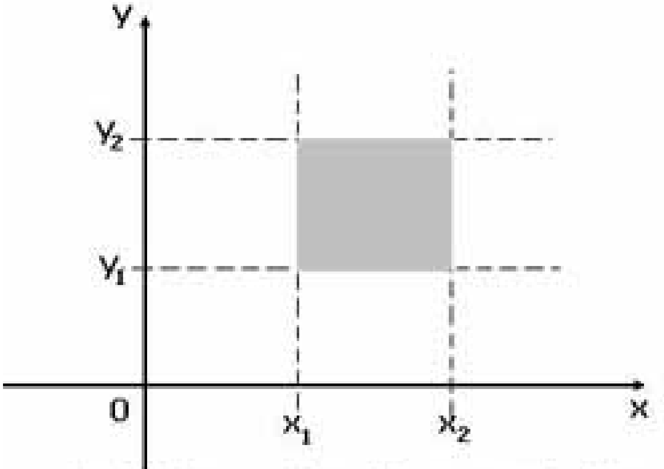
Teorema 5:
Se verifica que

$$p[x_1 < X \le x_2, y_1 < Y \le y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u,v).du.dv$$

Demostración:
Del teorema 2, 8), se tiene:

$$p[x_1 < X \le x_2, y_1 < Y \le y_2] = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) =$$
$$= p[X \le x_2, Y \le y_2] + p[X \le x_1, Y \le y_1] - (p[X \le x_1, Y \le y_1] + p[X \le x_1, y_1 < Y \le y_2]) -$$
$$-(p[X \le x_1, Y \le y_1] + p[x_1 < X \le x_2, Y \le y_1]) = p[X \le x_2, Y \le y_2] - p[X \le x_1, y_1 < Y \le y_2] -$$
$$- p[x_1 < X \le x_2, Y \le y_1] - p[X \le x_1, Y \le y_1]$$

Resulta, gráficamente, la zona sombreada en la figura



A p[x1 <X<=x2 ,y1 <Y<=y2] le correspondería la zona sombreada

Es decir, $p[x_1 < X \le x_2, y_1 < Y \le y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u,v).du.dv$

En general, para todo conjunto B medible Borel en R^2 se verifica que

$$p[(X,Y) \in B] = \iint_B f(u,v).du.dv$$

Teorema 6:

Si es *(X,Y)* una variable aleatoria bidimensional absolutamente continua y es $f(x,y)$ su función de densidad, se verifica:

1) Las variables aleatorias X e Y son absolutamente continuas y sus respectivas funciones de densidad son:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,v).dv, \quad f_2(y) = \int_{-\infty}^{+\infty} f(u,y).du$$

2) X e Y estocásticamente independientes $\leftrightarrow f_1(x).f_2(y) = f(x,y)$

Demostración:

1) Por el teorema 2, 5):

$$F_1(x) = \lim_{y \to +\infty} F(x,y) = \lim_{y \to +\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v).dudv = \int_{-\infty}^{x} \left[\lim_{y \to +\infty} \int_{-\infty}^{y} f(u,v).dv \right] du = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(u,v).dv \right] du =$$

$= \int_{-\infty}^{x} f_1(u).du$, por tanto $F_1(x) = \int_{-\infty}^{x} f_1(u).du$ y $f_1(x) = \int_{-\infty}^{+\infty} f(x,v).dv$

Análogamente para la variable y:

$$F_2(y)=\lim_{x\to+\infty}F(x,y)=\lim_{x\to+\infty}\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v).dudv=\int_{-\infty}^{y}\left[\lim_{x\to+\infty}\int_{-\infty}^{x}f(u,v).du\right]dv=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty}f(u,v).du\right]dv=$$

$=\int_{-\infty}^{y}f_2(v).dv$, por tanto $F_2(y)=\int_{-\infty}^{y}f_2(v).dv$ y $f_2(x)=\int_{-\infty}^{+\infty}f(u,y).du$

2) a) Veamos que si son independientes entonces $f(x,y)=f_1(x).f_2(y)$:

X,Y independientes $\Rightarrow F(x,y)=F_1(x).F_2(y)\Rightarrow f(x,y)=\dfrac{\partial^2F(x,y)}{\partial x\partial y}=$

$$=\frac{\partial^2(F_1(x).F_2(y))}{\partial x\partial y}=\frac{\partial F_1(x)}{\partial x}\frac{\partial F_2(y)}{\partial y}=f_1(x).f_2(y)\Rightarrow f(x,y)=f_1(x).f_2(y)$$

b) Veamos ahora que si $f_1(x).f_2(y)=f(x,y)$ entonces son independientes:

$$f(x,y)=f_1(x).f_2(y)\to F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v).dudv=\int_{-\infty}^{x}\int_{-\infty}^{y}f_1(u).f_2(v)dudv=$$

$$=\int_{-\infty}^{x}f_1(u)du\int_{-\infty}^{y}f_2(v)dv=F_1(x).F_2(y)$$

05. Momentos en las distribuciones bidimensionales

Definición 5:
Si son $M_1=E[X],M_2=E[Y]$, las respectivas medias (esperanzas matemáticas) de las variables unidimensionales X e Y, se dice que el par (M_1,M_2) es el *centro de gravedad* de la distribución bidimensional correspondiente a la variable aleatoria (X,Y).

Definición 6:
Sea una función $g(x,y)$ integrable Stieltjes-Lebesgue en R^2. Se define su valor medio o esperanza matemática $E[g(X,Y)]$ respecto a la probabilidad bidimensional $p(B)$ como la integral

$$E[g(X,Y)]=\int_{R^2}g(x,y).dp(B)$$

Si la distribución bidimensional es discreta:

$$E[g(X,Y)]=\sum_{i,k}g(x_i,y_k).p_{ik}$$

Si la distribución bidimensional es absolutamente continua:

$$E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y).f(x,y).dxdy$$

Teorema 7:

Si existen las $E[g_k(X,Y)], k=1,2,...,h$, se verifica que

$$\forall c_k \in R, k=1,2,...,h,\ E\left[\sum_{k=1}^{h} c_k g_k(X,Y)\right] = \sum_{k=1}^{h} c_k E[g_k(X,Y)]$$

Demostración:
Trivial.

Definición 7:
Dados los enteros positivos, $i,k \in Z^+$, y una distribución bidimensional, (X,Y), se define:
a) M*omento de órden (i,k),* M_{ik}, como la esperanza matemática de $X^i.Y^k$:

$$M_{ik} = E\left[X^i.Y^k\right]$$

b) *Momento central de orden (i,k),* m_{ik}, como la esperanza matemática del producto de las desviaciones $(X-M_1)^i.(Y-M_2)^k$:

$$m_{ik} = E\left[(X-M_1)^i.(Y-M_2)^k\right]$$

Teorema 8:
Se verifican las relaciones siguientes:
1) $M_{i0} = E\left[X^i\right],\ M_{0k} = E\left[Y^k\right]$
2) $m_{10} = m_{01} = 0,\ m_{20} = D_1^2 = E\left[(X-M_1)^2\right],\ m_{02} = D_2^2 = E\left[(Y-M_2)^2\right]$
3) $m_{20} = M_{20} - M_1^2,\quad m_{11} = M_{11} - M_1 M_2,\quad m_{02} = M_{02} - M_2^2$

Demostración:
1) $M_{i0} = \int_{R^2} x^i y^0 dp(B) = \int_{R^2} x^i dp(B) = E\left[X^i\right],\ M_{0k} = \int_{R^2} x^0 y^k dp(B) = \int_{R^2} y^k dp(B) = E\left[Y^k\right]$

2) $m_{10} = \int_{R^2} (x-M_1)^1.(y-M_2)^0 dp(B) = \int_{R^2} (x-M_1) dp(B) = \int_{R^2} x dp(B) - M_1 \int_{R^2} dp(B) =$
$= M_1 - M_1 = 0$

$$m_{01} = \int_{R^2} (x-M_1)^0.(y-M_2)^1 dp(B) = \int_{R^2} (y-M_2) dp(B) = \int_{R^2} y dp(B) - M_2 \int_{R^2} dp(B) =$$
$$= M_2 - M_2 = 0$$

$$m_{20} = \int_{R^2} (x-M_1)^2.(y-M_2)^0 dp(B) = \int_{R^2} (x-M_1)^2 dp(B) = D_1^2 = E\left[(X-M_1)^2\right]$$

$$m_{02} = \int_{R^2} (x-M_1)^0.(y-M_2)^2 dp(B) = \int_{R^2} (y-M_2)^2 dp(B) = D_2^2 = E\left[(Y-M_2)^2\right]$$

3) $m_{20} = \int_{R^2} (x-M_1)^2 dp(B) = \int_{R^2} (x^2 - 2xM_1 + M_1^2) dp(B) = \int_{R^2} x^2 dp(B) - 2M_1 \int_{R^2} x dp(B) +$
$+ M_1^2 \int_{R^2} dp(B) = M_{20} - 2M_1 M_1 + M_1^2 = M_{20} - M_1^2$

$$m_{02} = \int_{R^2} (y-M_2)^2 dp(B) = \int_{R^2} (y^2 - 2yM_2 + M_2^2) dp(B) = \int_{R^2} y^2 dp(B) - 2M_2 \int_{R^2} y dp(B) +$$
$$+ M_2^2 \int_{R^2} dp(B) = M_{02} - 2M_2 M_2 + M_2^2 = M_{02} - M_2^2$$

$$m_{11} = \int_{R^2} (x-M_1)^1.(y-M_2)^1 dp(B) = \int_{R^2} (xy - M_1 y - M_2 x + M_1 M_2) dp(B) = \int_{R^2} xy dp(B) -$$

$$-M_1\int_{R^2} ydp(B) - M_2\int_{R^2} xdp(B) + M_1M_2\int_{R^2} dp(B) = M_{11} - M_1M_2 - M_2M_1 + M_1M_2 =$$
$$= M_{11} - M_1M_2$$

Definición 8:

a) Al momento central m_{11} se le llama *momento mixto o covarianza* de la distribución.

b) Se denomina *matriz de momentos centrales* de la distribución a la matriz:

$$m = \begin{pmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{pmatrix}$$

Teorema 9:

1) $\det(m) = m_{20}.m_{02} - m_{11}^2 \geq 0$

2) Si es $r = rang(m)$, se tiene:

$r = 0 \leftrightarrow$ La masa está concentrada en un único punto.

$r = 1 \leftrightarrow$ La masa está concentrada en una recta.

$r = 2 \leftrightarrow$ La masa esta dispersa.

Demostración:

1) Considerando que la esperanza matemática de una expresión al cuadrado es positiva, se tiene que $E\left[(l(X-M_1)+h(Y-M_2))^2\right] \geq 0,\ \forall l,h \in R$, (**) por lo cual

$$E\left[(l(X-M_1)+h(Y-M_2))^2\right] = E\left[l^2(X-M_1)^2 + 2lh(X-M_1)(Y-M_2) + h^2(Y-M_2)^2\right] =$$
$$= l^2E\left[(X-M_1)^2\right] + 2lhE\left[(X-M_1)(Y-M_2)\right] + h^2E\left[(Y-M_2)^2\right] = l^2m_{20} + 2lhm_{11} + h^2m_{02} \geq 0$$

Es decir, $l^2m_{20} + 2lhm_{11} + h^2m_{02} = (l,h)\begin{pmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{pmatrix}\begin{pmatrix} l \\ h \end{pmatrix} \geq 0$, lo que indica que la matriz hermitiana m es semidefinida positiva.

$$\det\begin{pmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{pmatrix} = m_{20}.m_{02} - m_{11}^2 \geq 0$$

2) Si $r = 0 \rightarrow m_{20} = m_{02} = m_{11} = 0$, es decir $E\left[(X-M_1)^2\right] = 0,\ E\left[(Y-M_2)^2\right] = 0,$ $E\left[(X-M_1)(Y-M_2)\right] = 0$, lo que indica que las desviaciones con respecto a las medias (M_1,M_2) son nulas, con lo cual toda la masa de la distribución está concentrada en el único punto (M_1,M_2). Recíprocamente, si toda la masa está en el centro de gravedad (M_1,M_2) todas las desviaciones de las medias son nulas, esto es, $m_{20} = E\left[(X-M_1)^2\right] = 0,\ m_{02} = E\left[(Y-M_2)^2\right] = 0,\ m_{11} = E\left[(X-M_1)(Y-M_2)\right] = 0$, de lo cual $r=rang(m)=0$.

Si $r = 1 \rightarrow m_{20}m_{02} - m_{11}^2 = 0$, con lo que $\exists l_0,h_0 \in R/(l_0,h_0)\begin{pmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{pmatrix}\begin{pmatrix} l_0 \\ h_0 \end{pmatrix} = 0 \rightarrow$

$\rightarrow l_0^2m_{20} + 2l_0h_0m_{11} + h_0^2m_{02} = 0$, es decir $E\left[(l_0(X-M_1)+h_0(Y-M_2))^2\right] = 0 \rightarrow$

$\rightarrow l_0(X-M_1) + h_0(Y-M_2) = 0$, que es la ecuación de una línea recta que pasa por el centro de gravedad (M_1,M_2) de la distribución. Luego, toda la masa de la distribución se encuentra en una línea recta que pasa por su centro de gravedad. Recíprocamente, si la masa de la distribución se encuentra concentrada en una línea recta, ésta ha de pasar obviamente por su centro de gravedad, por lo cual se tiene que $\exists l_0,h_0 \in R/l_0(X-M_1) + h_0(Y-M_2) = 0$, con lo que será:

$$E\left[(l_0(X-M_1)+h_0(Y-M_2))^2\right]=0\rightarrow \exists l_0,h_0\in R/(l_0,h_0)\begin{pmatrix} m_{20} & m_{11}\\ m_{11} & m_{02}\end{pmatrix}\begin{pmatrix} l_0\\ h_0\end{pmatrix}=0\rightarrow$$

$$\rightarrow \det\begin{pmatrix} m_{20} & m_{11}\\ m_{11} & m_{02}\end{pmatrix}=m_{20}.m_{02}-m_{11}^2=0\rightarrow r=rang(m)=1$$

3) Si $r=2$ la masa no está concentrada en un único punto (pues entonces sería $r=0$), ni en una línea recta (pues entonces sería $r=1$). Recíprocamente, si la masa no está concentrada en un único punto, ni en una sola recta, no pueden ser $r=0$ ni $r=1$, luego necesariamente es $r=2$.

06. Distribuciones condicionales de una distribución bidimensional

06.1. Definiciones

Dada una distribución bidimensional $\{(U,\Phi,p),(X,Y)\}$, si una de las dos variables, ya sea X o Y, está sometida a algún tipo de restricción, la distribución de la otra se dice condicionada a tal restricción.

La distribución de la variable X condicionada a que $[Y\in S]$ es el par $\{(U,\Phi,p_c),X\}$, donde $p_c:\Phi\rightarrow R$ está definida por

$$\forall A\in\Phi,\ p_c(A)=\frac{p(A\cap[Y\in S])}{p[Y\in S]}$$

La distribución de la variable Y condicionada a que $[X\in S]$ es el par $\{(U,\Phi,p_c^{'}),Y\}$, donde $p'_c:\Phi\rightarrow R$ está definida por

$$\forall A\in\Phi,\ p'_c(A)=\frac{p(A\cap[X\in S])}{p[X\in S]}$$

Función de distribución condicional:

$$F_c(x/Y\in S)=\frac{p([X\leq x]\cap[Y\in S])}{p[Y\in S]}=\frac{p[X\leq x,Y\in S]}{p[Y\in S]}$$

$$F_c(y/X\in S)=\frac{p([Y\leq y]\cap[X\in S])}{p[X\in S]}=\frac{p[Y\leq y,X\in S]}{p[X\in S]}$$

Esperanza matemática de una función $g(X,Y)$:

$$E[g(X,Y)/Y\in S]=\int_R g(x,y).dp_c=\int_R g(x,y).dF_c(x/Y\in S)$$

$$E[g(X,Y)/X\in S]=\int_R g(x,y).dp'_c=\int_R g(x,y).dF_c(y/X\in S)$$

Media y Momentos iniciales:

$$M_1 = E[X / Y \in S] = \int_R x.dp_c = \int_R x.dF_c(x / Y \in S)$$
$$M_2 = E[Y / X \in S] = \int_R y.dp_c' = \int_R y.dF_c(y / X \in S)$$
$$M_1^r = E[X^r / Y \in S] = \int_R x^r.dp_c = \int_R x^r.dF_c(x / Y \in S)$$
$$M_2^r = E[Y^r / X \in S] = \int_R y^r.dp_c' = \int_R y^r.dF_c(y / X \in S)$$

Momentos centrales:

$$m_1^r = E[(X - M_1)^r / Y \in S] = \int_R (x - M_1)^r.dp_c = \int_R (x - M_1)^r.dF_c(x / Y \in S)$$
$$m_2^r = E[(Y - M_2)^r / X \in S] = \int_R (y - M_2)^r.dp'_c = \int_R (y - M_2)^r.dF_c(y / X \in S)$$

Varianza:

$$D_1^2 = E[(X - M_1)^2 / Y \in S] = \int_R (x - M_1)^2.dp_c = \int_R (x - M_1)^2.dF_c(x / Y \in S)$$
$$D_2^2 = E[(Y - M_2)^2 / X \in S] = \int_R (y - M_2)^2.dp'_c = \int_R (y - M_2)^2.dF_c(y / X \in S)$$

Función característica:

$$\varphi(t / X \in S) = E[e^{itY} / X \in S]$$
$$\varphi(t / Y \in S) = E[e^{itX} / Y \in S]$$

06.2. Un ejemplo de condición:

Consideremos el caso de que la condición restrictiva sea $X = x$, para la variable X, bien $Y = y$, para la variable Y. Se tendría:

- La distribución de la variable X condicionada a que $[Y = y]$ es el par $\{(U, \Phi, p_c), X\}$, donde $p_c : \Phi \to R$ está definida por

$$\forall A \in \Phi,\ p_c(A) = \lim_{h \to 0} \frac{p(A \cap [y - h < Y \le y + h])}{p[y - h < Y \le y + h]}$$

- La distribución de la variable Y condicionada a que $[X = x]$ es el par $\{(U, \Phi, p'_c), Y\}$, donde $p_c : \Phi \to R$ está definida por

$$\forall A \in \Phi,\ p'_c(A) = \lim_{h \to 0} \frac{p(A \cap [x - h < X \le x + h])}{p[x - h < X \le x + h]}$$

- Función de distribución condicional:

$$F_c(x / Y = y) = \frac{p([X \le x] \cap [Y = y])}{p[Y = y]} = \frac{p[X \le x, Y = y]}{p[Y = y]}$$

$$F_c(y / X = x) = \frac{p([Y \le y] \cap [X = x])}{p[X = x]} = \frac{p[Y \le y, X = x]}{p[X = x]}$$

- Esperanza matemática de una función $g(x, y)$:

$$E[g(X,Y)/Y=y]=\int_R g(x,y).dp_c=\int_R g(x,y).dF_c(x/Y=y)$$
$$E[g(X,Y)/X=x]=\int_R g(x,y).dp'_c=\int_R g(x,y).dF_c(y/X=x)$$

Media y Momentos iniciales:

$$M_1(y)=E[X/Y=y]=\int_R x.dp_c=\int_R x.dF_c(x/Y=y)$$
$$M_2(x)=E[y/X=x]=\int_R y.dp_c^{'}=\int_R y.dF_c(y/X=x)$$
$$M_1^r(y)=E[X^r/Y=y]=\int_R x^r.dp_c=\int_R x^r.dF_c(x/Y=y)$$
$$M_2^r(x)=E[Y^r/X=x]=\int_R y^r.dp_c^{'}=\int_R y^r.dF_c(y/X=x)$$

Momentos centrales:

$$m_1^r(y)=E[(X-M_1)^r/Y=y]=\int_R (x-M_1)^r.dp_c=\int_R (x-M_1)^r.dF_c(x/Y=y)$$
$$m_2^r(x)=E[(Y-M_2)^r/X=x]=\int_R (y-M_2)^r.dp_c^{'}=\int_R (y-M_2)^r.dF_c(y/X=x)$$

Varianza:

$$D_1^2(y)=E[(X-M_1)^2/Y=y]=\int_R (x-M_1)^2.dp_c=\int_R (x-M_1)^2.dF_c(x/Y=y)$$
$$D_2^2(x)=E[(Y-M_2)^2/X=x]=\int_R (y-M_2)^2.dp_c^{'}=\int_R (y-M_2)^2.dF_c(y/X=x)$$

Función característica:

$$\varphi(t/X=x)=E[e^{itY}/X=x]$$
$$\varphi(t/Y=y)=E[e^{itX}/Y=y]$$

06.3. En distribuciones discretas

Consideremos una distribución bidimensional discreta $\{(U,\Phi,p),(X,Y)\}$, tal que la masa de la distribución se concentra en los puntos $\{(x_1,y_1),(x_2,y_2),\ldots\}$, y llamemos $p[X=x_i]=p_{1i}$, $p[Y=y_k]=p_{2k}$, $p[X=x_i,Y=y_k]=p_{ik}$, $i=1,2,\ldots,\ k=1,2,\ldots$

- La distribución de la variable X condicionada a que $[Y=y_k]$ es el par $\{(U,\Phi,p_c),X\}$, donde $p_c:\Phi\to R$ está definida por

$$\forall A\in\Phi,\ p_c(A)=p[A/Y=y_k]=\frac{p(A\cap[Y=y_k])}{p[Y=y_k]}$$

- La distribución de la variable Y condicionada a que $[X=x_i]$ es el par $\{(U,\Phi,p_c^{'}),Y\}$, donde $p_c^{'}:\Phi\to R$ está definida por

$$\forall A\in\Phi,\ p_c^{'}(A)=p[A/X=x_i]=\frac{p(A\cap[X=x_i])}{p[X=x_i]}$$

- Función de distribución condicional:

$$F_c(x/y_k)=\frac{p\left(\left[X\le x\right]\cap\left[Y=y_k\right]\right)}{p\left[Y=y_k\right]}=\frac{p\left[X\le x,Y=y_k\right]}{p\left[Y=y_k\right]}=\frac{\sum\limits_{x_i\le x}p_{ik}}{p_{2k}}$$

$$F_c(y/x_i)=\frac{p\left(\left[Y\le y\right]\cap\left[X=x_i\right]\right)}{p\left[X=x_i\right]}=\frac{p\left[Y\le y,X=x_i\right]}{p\left[X=x_i\right]}=\frac{\sum\limits_{y_k\le y}p_{ik}}{p_{1i}}$$

- Esperanza matemática de una función $g(X,Y)$:

$$E\left[g(X,Y)/x_i\right]=\int_R g(x,y).dp'_c=\int_R g(x,y).dF_c(y/x_i)=\frac{\sum\limits_{k=1}^{\infty}g(x_i,y_k).p_{ik}}{p_{1i}}$$

$$E\left[g(X,Y)/y_k\right]=\int_R g(x,y).dp_c=\int_R g(x,y).dF_c(x/y_k)=\frac{\sum\limits_{i=1}^{\infty}g(x_i,y_k).p_{ik}}{p_{2k}}$$

- Media y Momentos iniciales:

$$M_1(y_k)=E\left[X/y_k\right]=\int_R x.dF_c(x/y_k)=\frac{\sum\limits_{i=1}^{\infty}x_i.p_{ik}}{p_{2k}}$$

$$M_2(x_i)=E\left[Y/x_i\right]=\int_R y.dF_c(y/x_i)=\frac{\sum\limits_{k=1}^{\infty}x_i.p_{ik}}{p_{1i}}$$

$$M_1^r(y_k)=E\left[X^r/y_k\right]=\int_R x^r.dF_c(x/y_k)=\frac{\sum\limits_{i=1}^{\infty}x_i^r.p_{ik}}{p_{2k}}$$

$$M_2^r(x_i)=E\left[Y^r/x_i\right]=\int_R y^r.dF_c(y/x_i)=\frac{\sum\limits_{k=1}^{\infty}y_k^r.p_{ik}}{p_{1i}}$$

- Momentos centrales:

$$m_1^r(y_k)=E\left[(X-M_1)^r/Y=y_k\right]=\int_R (x-M_1)^r.dF_c(x/y_k)=\frac{\sum\limits_{i=1}^{\infty}(x_i-M_1)^r.p_{ik}}{p_{2k}}$$

$$m_2^r(x_i)=E\left[(Y-M_2)^r/X=x_i\right]=\int_R (y-M_2)^r.dF_c(y/x_i)=\frac{\sum\limits_{k=1}^{\infty}(y_k-M_2)^r.p_{ik}}{p_{2i}}$$

- Varianza:

$$D_1^2(y_k)=E\left[(X-M_1)^2/Y=y_k\right]=\int_R (x-M_1)^2.dF(x/Y=y_k)=\frac{\sum\limits_{i=1}^{\infty}(x_i-M_1)^2.p_{ik}}{p_{2k}}$$

$$D_2^2(x_i) = E\left[(Y-M_2)^2 / X = x_i\right] = \int_R (y-M_2)^2.dF(y/X=x_i) = \frac{\sum_{k=1}^{\infty}(y_k - M_2)^2.p_{ik}}{p_{2i}}$$

- Función característica:

$$\varphi(t/x_i) = E\left[e^{itY} / X = x_i\right] = \frac{\sum_{k=1}^{\infty} e^{ity_k}.p_{ik}}{p_{1i}}$$

$$\varphi(t/y_k) = E\left[e^{itX} / Y = y_k\right] = \frac{\sum_{i=1}^{\infty} e^{itx_i}.p_{ik}}{p_{2k}}$$

06.4. En distribuciones absolutamente continuas

Consideremos una distribución bidimensional absolutamente continua, que representaremos por $\{(U,\Phi,p),(X,Y)\}$, y sea $f(x,y)$ su función de densidad.

- La distribución de la variable X condicionada a que $[Y=y]$ es el par $\{(U,\Phi,p_c),X\}$, donde $p_c : \Phi \to R$ está definida por

$$\forall A \in \Phi,\ p_c(A) = p[A/Y=y] = \lim_{h\to 0} \frac{p(A \cap [y-h<Y\leq y+h])}{p[y-h<Y\leq y+h]}$$

- La distribución de la variable Y condicionada a que $[X=x]$ es el par $\{(U,\Phi,p_c^{'}),Y\}$, donde $p_c^{'} : \Phi \to R$ está definida por

$$\forall A \in \Phi,\ p_c^{'}(A) = p[A/X=x] = \lim_{h\to 0} \frac{p(A \cap [x-h<X\leq x+h])}{p[x-h<X\leq x+h]}$$

- Función de distribución condicional:

$$F_c(x/y) = \lim_{h\to 0} \frac{p([X\leq x] \cap [y-h<Y\leq y+h])}{p[y-h<Y\leq Y+h]} = \lim_{h\to 0} \frac{p[X\leq x, y-h<Y\leq y+h]}{p[y-h<Y\leq y+h]} =$$

$$= \lim_{h\to 0} \frac{\int_{u=-\infty}^{u=y}\int_{v=y-h}^{v=y+h} f(u,v).dudv}{\int_{y-h}^{y+h} f_2(v)dv}$$

$$F_c(y/x)=\lim_{h\to 0}\frac{p([Y\le y]\cap[x-h<X\le x+h])}{p[x-h<X\le x+h]}=\lim_{h\to 0}\frac{p[Y\le y, x-h<X\le x+h]}{p[x-h<X\le x+h]}=$$

$$=\lim_{h\to 0}\frac{\int_{v=-\infty}^{v=y}\int_{u=x-h}^{u=x+h} f(u,v).dudv}{\int_{x-h}^{x+h} f_1(u)du}$$

Para determinar las medias, momentos y función característica en estas distribuciones condicionales absolutamente continuas es necesario determinar previamente la función de densidad condicional $f(x/Y=y)$ (o bien $f(y/X=x)$), lo cual es posible mediante el siguiente teorema.

Teorema 10:
Sea (X,Y) una variable aleatoria bidimensional de distribución absolutamente continua, $\{(U,\Phi,p),(X,Y)\}$, con función de densidad $f(x,y)$. Si tal función de densidad es continua en el punto $x=x_0$ para casi todo y, y es $f_1(x_0)>0$ se tiene:

a) La función de Distribución de la variable aleatoria Y condicionada a $[X=x_0]$ es

$$F_c(y/x_0)\equiv F_c(y/X=x_0)=\frac{\int_{-\infty}^{y} f(x_0,v)dv}{f_1(x_0)}$$

b) La función de densidad de la variable Y condicionada a $[X=x_0]$ es

$$f(y/x_0)\equiv f(y/X=x_0)=\frac{f(x_0,y)}{f_1(x_0)}$$

(Siendo, como ya hemos visto en apartados anteriores, $f_1(x_0)=\int_{-\infty}^{+\infty} f(x_0,v)dv$)

Demostración:
a)

$$F_c(y/x)=\lim_{h\to 0}\frac{\int_{v=-\infty}^{v=y}\int_{u=x-h}^{u=x+h} f(u,v).dudv}{\int_{x-h}^{x+h} f_1(u)du}=\lim_{h\to 0}\frac{\frac{1}{2h}\int_{u=x-h}^{u=x+h}\int_{v=-\infty}^{v=y} f(u,v)dudv}{\frac{1}{2h}\int_{u=x-h}^{u=x+h}\int_{v=-\infty}^{v=+\infty} f(u,v)dudv}=\frac{\lim_{h\to 0}\frac{1}{2h}\int_{v=-\infty}^{v=y}\int_{u=x_0-h}^{u=x_0+h} f(u,v)dudv}{\lim_{h\to 0}\frac{1}{2h}\int_{v=-\infty}^{v=+\infty}\int_{u=x_0-h}^{u=x_0+h} f(u,v)dudv}$$

Calculemos la integral por separado: Si es $F(x_0)$ tal que $F'(x_0)=f(x_0,v)$, se tiene:

$$\lim_{h\to 0}\frac{1}{2h}\int_{v=-\infty}^{v=y}\int_{u=x_0-h}^{u=x_0+h} f(u,v)dudv=\int_{v=-\infty}^{v=y}\left[\lim_{h\to 0}\frac{1}{2h}\int_{u=x_0-h}^{u=x_0+h} f(u,v)du\right]dv=$$

$$=\int_{v=-\infty}^{v=y}\left[\lim_{h\to 0}\frac{1}{2h}\big(F(x_0+h)-F(x_0-h)\big)\right]dv=$$

$$= \int_{v=-\infty}^{v=y} \left[\lim_{h\to 0} \frac{1}{2} \frac{F(x_0+h)-F(x_0)}{h} + \lim_{h\to 0} \frac{1}{2} \frac{F(x_0)-F(x_0-h}{h} \right] dv = \int_{v=-\infty}^{v=y} \left[\frac{1}{2} F'(x_0) \frac{1}{2} F'(x_0) \right] dv =$$

$$= \int_{v=-\infty}^{v=y} F'(x_0) dv = \int_{v=-\infty}^{v=y} f(x_0, v) dv$$

por tanto, al sustituir:

$$F_c(y/x) = \frac{\lim_{h\to 0} \frac{1}{2h} \int_{v=-\infty}^{v=y} \int_{u=x_0-h}^{u=x_0+h} f(u,v) du dv}{\lim_{h\to 0} \frac{1}{2h} \int_{v=-\infty}^{v=+\infty} \int_{u=x_0-h}^{u=x_0+h} f(u,v) du dv} = \frac{\int_{v=-\infty}^{v=y} f(x_0,v) dv}{\int_{v=-\infty}^{v=+\infty} f(x_0,v) dv} = \frac{\int_{v=-\infty}^{v=y} f(x_0,v) dv}{f_1(x_0)}$$

b) $F_c(y/x) = \int_{-\infty}^{y} f(v/x_0).dv = \frac{\int_{v=-\infty}^{v=y} f(x_0,v) dv}{f_1(x_0)} = \int_{-\infty}^{y} \frac{f(x_0,v)}{f_1(x_0)} dv \to f(y/x_0) = \frac{f(x_0,y)}{f_1(x_0)}$

Las fórmulas análogas para la otra variable se tratarían del mismo modo:

$$F_c(x/y) \equiv F_c(x/Y=y_0) = \frac{\int_{-\infty}^{x} f(u,y_0) du}{f_2(y_0)} \quad \text{y} \quad f(x/y_0) \equiv f(x/Y=y_0) = \frac{f(x,y_0)}{f_2(y_0)}$$

- Esperanza matemática de una función $g(X,Y)$:

$$E[g(X,Y)/X=x] = \int_R g(x,y).dp_c = \int_R g(x,y).f(y/x)dy$$

$$E[g(X,Y)/Y=y] = \int_R g(x,y).dp_c = \int_R g(x,y)f(x/y).dx$$

- Media y Momentos iniciales:

$$M_1(y) = E[X/Y=y] = \int_R x.f(x/y).dx$$

$$M_2(x) = E[Y/X=x] = \int_R y.f(y/x).dy$$

$$M_1^r(y) = E[X^r/Y=y] = \int_R x^r.f(x/y).dx$$

$$M_2^r(x) = E[Y^r/X=x] = \int_R y^r.f(y/x).dy$$

- Momentos centrales:

$$m_1^r(y) = E[(X-M_1)^r/Y=y] = \int_R (x-M_1)^r.f(x/y).dx$$

$$m_2^r(x) = E[(Y-M_2)^r/X=x] = \int_R (y-M_2)^r.f(y/x).dy$$

- Varianza:

$$D_1^2(y) = E[(X-M_1)^2/Y=y] = \int_R (x-M_1)^2.f(x/y).dx$$

$$D_2^2(x) = E[(Y-M_2)^2/X=x] = \int_R (y-M_2)^2.f(y/x).dy$$

- Función característica:

$$\varphi(t/X=x)=E\left[e^{itY}/X=x\right]=\int_R e^{ity}.f(y/x).dy$$
$$\varphi(t/Y=y)=E\left[e^{itX}/Y=y\right]=\int_R e^{itx}.f(x/y).dx$$

Teorema 11:
Si las dos variables aleatorias, X e Y, son estocásticamente independientes, entonces las medias condicionales $M_1(y)$ y $M_2(x)$ coinciden, respectivamente, con las medias M_1 y M_2 de las correspondientes distribuciones marginales.
Demostración:
Puesto que $f(x/y)=\dfrac{f(x,y)}{f_2(y)}$ y $f(y/x)=\dfrac{f(x,y)}{f_1(x)}$, se tiene que

$$f(x,y)=f(x/y).f_2(y)=f(y/x).f_1(x)$$

y si son estocásticamente independientes sabemos, por el teorema 6, 2), que se verifica:

$$f(x,y)=f_1(x).f_2(y)$$

por lo que, al identificar términos, se tiene que

$$f_1(x)=f(x/y)$$
$$f_2(y)=f(y/x)$$

con lo que las medias serán:

$$M_1(y)=\int_R x.f(x/y).dx=\int_R x.f_1(x).dx=M_1$$
$$M_2(x)=\int_R y.f(y/x).dy=\int_R y.f_2(y).dx=M_2$$

07. Regresión

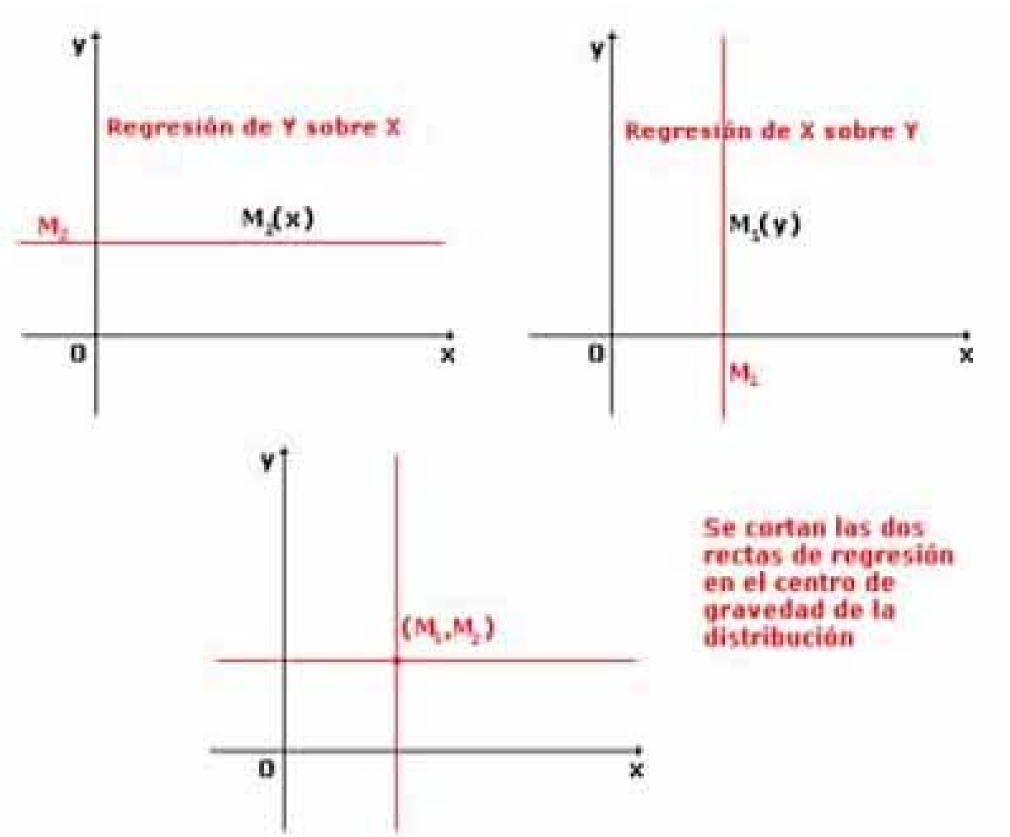
Sea (X,Y) una variable aleatoria bidimensional con distribución absolutamente continua. Consideremos la distribución de la variable aleatoria Y condicionada a que $[X=x]$ y sea $f_2(x)$ una medida sobre esta distribución condicional (media, moda, varianza, etc.). Es obvio que cuando varía x también varía la medida $f_2(x)$, es decir, el punto $(x,f_2(x))$ describe una cierta curva, que se denomina *curva de regresión de Y sobre X para* $f_2(x)$. Lo mismo se diría para $(f_1(y),y)$ como *curva de regresión de X sobre Y para* $f_1(y)$.

Así, refiriéndonos a la media condicional, será:

$(x,M_2(x))$: curva de regresión de Y sobre X para la media $M_2(x)$.

$(M_1(y),y)$: curva de regresión de X sobre Y para la media $M_1(y)$.

Teorema 12:
Si las variables X e Y son estocasticamente independientes, entonces las curvas de regresión para la media son líneas rectas paralelas a los ejes cartesianos y que se cortan en el centro de gravedad de la distribución.
Demostración:
Por teorema 11, es $M_2(x)=M_2$ y $M_1(y)=M_1$. Gráficamente:



Teorema 13:

Consideremos la variable aleatoria bidimensional (X,Y) con distribución absolutamente continua y sean $M_2(x)=E[y/X=x]$ y $M_1(y)=E[x/Y=y]$.

a) $E\left[(Y-g(x))^2\right]$ tiene el valor mínimo si $g(x)=M_2(x)$

b) $E\left[(X-h(y))^2\right]$ tiene el valor mínimo si $h(y)=M_1(y)$

Demostración:

a)Sabemos que

$$E\left[(Y-g(X))^2\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-g(x))^2 f(x,y)dxdy=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}(y-g(x))^2 f(y/x)dy\right].f_1(x).dx$$

Sea la media $M=M_2(x)=E[Y/X=x]$, y llamemos, por simplificar, $g(x)=c$

$$\int_{-\infty}^{\infty}(y-g(x))^2 f(y/x)dy=\int_{-\infty}^{\infty}(y-c)^2 f(y/x)dy=\int_{-\infty}^{\infty}((y-M)-(c-M))^2 f(y/x)dy=$$

$$=\int_{-\infty}^{\infty}((y-M)^2-2(y-M)(c-M)+(c-M)^2)f(y/x)dy=\int_{-\infty}^{\infty}(y-M)^2 f(y/x)dy-$$

$$-2(c-M)\int_{-\infty}^{\infty}(y-M).f(y/x).dy+(c-M)^2\int_{-\infty}^{\infty}f(y/x).dy=m_2^2-0+(c-M)^2\geq m_2^2$$

o sea: $\int_{-\infty}^{\infty}(y-g(x))^2 f(y/x)dy=m_2^2+(g(x)-M_2(x))^2$, por lo cual es

$$E\left[(Y-g(X))^2\right]=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}(y-g(x))^2 f(y/x)dy\right].f_1(x).dx=\int_{-\infty}^{\infty}\left[m_2^2+(g(x)-M_2(x))^2\right].f_1(x).dx=$$

$$=m_2^2\int_{-\infty}^{\infty}f_1(x).dx+\int_{-\infty}^{\infty}(g(x)-M_2(x))^2 f_1(x).dx=m_2^2+\int_{-\infty}^{\infty}(g(x)-M_2(x))^2 f_1(x).dx\geq m_2^2$$

Obviamente, el mínimo valor se obtiene para $g(x)=M_2(x)$:

$$E\left[(Y-g(X))^2\right]=E\left[(Y-M_2(X))^2\right]=m_2^2+\int_{-\infty}^{\infty}(M_2(x)-M_2(x))^2 f_1(x).dx=m_2^2$$

b) La demostración para el caso $E\left[(X-h(y))^2\right]$ es del todo análoga.

07.1. Curvas de regresión mínimo-cuadráticas:

Sabemos del anterior teorema que la curva $y=g(x)$ que minimiza la esperanza matemática del cuadrado $(y-g(x))^2$ es la definida por la media $y=M_2(x)$, obteniéndose

$$E\left[(Y-g(X))^2\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-M_2(x))^2 f(x,y)dxdy=\int_{-\infty}^{\infty}(y-M_2(x))^2 f(y/x)dy=m_2^2$$

Podemos intentar encontrar funciones de un determinado tipo que se ajusten a la media en el sentido de minimizar tal cuadrado. Se denominan a estas funciones curvas de ajuste mínimo-cuadráticas.

Así, pueden estudiarse curvas de ajuste $g(x)=\phi(x,a,b,c,...)$ en general, mínimo-cuadráticas polinómicas, de la forma $g(x)=a+b_1x+b_2x^2+...+b_nx^n$, curvas trigonométricas, etc. El caso más sencillo corresponde a la curvas de ajuste lineal, en donde es $g(x)=a+bx$ la función rectilínea con la que podemos ajustar la regresión de Y sobre X, o bien $g(y)=a'+b'y$ que seria la recta de ajuste de la regresión de X sobre Y

Estudiaremos a continuación el caso de ajuste mínimo-cuadrático lineal, esto es, mediante una línea recta.

07.2. Regresión lineal:

Se trata de encontrar una línea recta $g(x)=a+bx$ que se ajuste a la media indicada, es decir, tal que $E\left[(Y-a-bX)^2\right]$ sea mínimo, por lo que los coeficientes a y b pueden determinarse por las condiciones

$$\frac{\partial}{\partial a}E\left[(Y-a-bX)^2\right]=0$$
$$\frac{\partial}{\partial b}E\left[(Y-a-bX)^2\right]=0$$

Antes de aplicar esta condición de mínimo, expresemos la función a derivar en función de los momentos de la distribución:

$$E\left[(Y-a-bX)^2\right]=E\left[((Y-M_2)-b(X-M_1)+(M_2-a-bM_1))^2\right]=$$

$$= E\left[\left(Y - M_2 - b(X - M_1)\right)^2 + 2\left(Y - M_2 - b(X - M_1)\right)\left(M_2 - a - bM_1\right) + \left(M_2 - a - bM_1\right)^2\right] =$$
$$= E\left[\left(Y - M_2 - b(X - M_1)\right)^2\right] + E\left[\left(M_2 - a - bM_1\right)^2\right] + 2E\left[\left(Y - M_2 - b(X - M_1)\right)\left(M_2 - a - bM_1\right)\right] =$$
$$= E\left[\left(Y - M_2 - b(X - M_1)\right)^2\right] + \left(M_2 - a - bM_1\right)^2 + 2\left(M_2 - a - bM_1\right)E\left[Y - M_2 - b(X - M_1)\right]$$

Veamos por separado los sumandos primero y tercero:
- El primero:
$$E\left[\left(Y - M_2 - b(X - M_1)\right)^2\right] = E\left[(Y - M_2)^2\right] + b^2 E\left[(X - M_1)^2\right] - 2bE\left[(Y - M_2)(X - M_1)\right] =$$
$$= m_{02} + b^2 m_{20} - 2bm_{11}$$
- El tercero:
$$2\left(M_2 - a - bM_1\right)E\left[Y - M_2 - b(X - M_1)\right] = 2\left(M_2 - a - bM_1\right)\left(E\left[Y - M_2\right] - bE\left[X - M_1\right]\right) =$$
$$= 2\left(M_2 - a - bM_1\right)\left(0 - b.0\right) = 0$$
En definitiva, queda:
$$E\left[\left(Y - a - bX\right)^2\right] = m_{02} + b^2 m_{20} - 2bm_{11} + \left(M_2 - a - bM_1\right)^2$$
Apliquemos ahora la condición de mínimo:
$$\frac{\partial}{\partial a} E\left[\left(Y - a - bX\right)^2\right] = 0 \rightarrow 2(M_2 - a - bM_1).(-1) = 0 \rightarrow M_2 - a - bM_1 = 0 \rightarrow$$
$$\rightarrow a = M_2 - bM_1$$
$$\frac{\partial}{\partial b} E\left[\left(Y - a - bX\right)^2\right] = 0 \rightarrow 2bm_{20} - 2m_{11} + 2(M_2 - a - bM_1).(-b) = 0 \rightarrow$$
$$\rightarrow 2b(m_{20} - M_2 + a + bM_1) - 2m_{11} = 0$$
Si sustituimos $a = M_2 - bM_1$ en esta última igualdad, obtenemos el valor de b:
$$2bm_{20} - 2m_{11} = 0 \rightarrow b = \frac{m_{11}}{m_{20}}$$
Por tanto, al sustituir en la ecuación de la recta los valores de a y b, resulta:
$$y = a + bx = M_2 - bM_1 + bx = M_2 + b(x - M_1) = M_2 + \frac{m_{11}}{m_{20}}(x - M_1)$$

Y se tiene: $b = \frac{m_{11}}{m_{20}} = \frac{m_{11}\sqrt{m_{02}}}{\sqrt{m_{20}m_{20}}\sqrt{m_{02}}} = \left(\frac{m_{11}}{\sqrt{m_{02}m_{20}}}\right)\frac{\sqrt{m_{02}}}{\sqrt{m_{20}}} = \rho\frac{D_2}{D_1}$

Donde $D_1^2 = m_{20}$ y $D_2^2 = m_{02}$ son las varianzas respectivas de X e Y, y el término $\rho = \frac{m_{11}}{\sqrt{m_{02}m_{20}}} = \frac{m_{11}}{D_1 D_2}$ es el llamado *coeficiente de correlación.*

Y la recta de regresión lineal mínimo cuadrática de Y sobre X queda de esta forma:

$$\boxed{y - M_2 = \rho\frac{D_2}{D_1}(x - M_1)}$$

Recta que pasa, obviamente, por el centro de gravedad (M_1, M_2) de la distribución.

La cantidad $b = \frac{m_{11}}{m_{20}} = \rho\frac{D_2}{D_1}$ se denomina *coeficiente de regresión de Y sobre X.*

Análogamente, realizando el mismo proceso se puede encontrar una línea recta $g'(y) = a' + b'y$ tal que $E\left[\left(X - a' - b'Y\right)^2\right]$ sea mínimo, determinándose igualmente los coeficientes a' y b' del mismo modo. Se tiene así que la recta de regresión lineal mínimo cuadrática de X sobre Y es:

$$\boxed{x - M_1 = \rho \frac{D_1}{D_2}(y - M_2)}$$

La cantidad $b' = \dfrac{m_{11}}{m_{02}} = \rho \dfrac{D_1}{D_2}$ se denomina *coeficiente de regresión de X sobre Y.*

Teorema 14:
Se verifica que $E_{\min}\left[(Y - a - bX)^2\right] = D_2^2\left(1-\rho^2\right)$.
Demostración:
De la expresión obtenida anteriormente:

$$E\left[(Y-a-bX)^2\right] = m_{02} + b^2 m_{20} - 2bm_{11} + (M_2 - a - bM_1)^2$$

y al sustituir los valores de a y b que dan el mínimo $(a = M_2 - bM_1,\ b = m_{11}/m_{20})$:

$$E_{\min}\left[(Y-a-bX)^2\right] = m_{02} + \left(\frac{m_{11}}{m_{20}}\right)^2 m_{20} - 2\frac{m_{11}}{m_{20}} m_{11} + 0 = \frac{m_{02}m_{20} - m_{11}^2}{m_{20}} =$$

$$= m_{02} - \frac{m_{11}^2}{m_{20}m_{02}} m_{02} = m_{02}\left(1 - \frac{m_{11}^2}{m_{20}m_{02}}\right) = D_2^2\left(1-\rho^2\right)$$

Análogamente, para la regresión de X sobre Y: $E_{\min}\left[(X - a' - b'Y)^2\right] = D_1^2\left(1-\rho^2\right)$

Teorema 15:
La recta de regresión mínimo cuadrática de Y sobre X

$$y = M_2 + \rho \frac{D_2}{D_1}(x - M_1)$$

es la curva de más estricto ajuste a la curva de regresión $y = M_2(x)$. En el caso de que $y = M_2(x)$ sea una recta, ésta recta coincidiría con la recta de regresión mínimo cuadrática de Y sobre X.
Demostración:
De la obtención de la recta de regresión lineal mínimo cuadrática de Y sobre X, sabemos que para los valores de a y b obtenidos la varianza $E_{\min}\left[(Y-a-bX)^2\right]$ es mínima.

Se trata ahora de probar que la distancia vertical entre la curva $y = M_2(x)$ y la recta de regresión mínimo cuadrática obtenida, $y = a + bx$, es mínima.
Es decir, se trata de probar que la varianza de $M_2 - a - bX$, $E\left[(M_2 - a - bX)^2\right]$, es mínima.
Sean, pues, a y b los parámetros que minimizan $E\left[(Y-a-bX)^2\right]$. Se tiene:

$$E\left[(Y-a-bX)^2\right] = E\left[(Y - M_2 + M_2 - a - bX)^2\right] = E\left[(Y-M_2)^2\right] + E\left[(M_2 - a - bX)^2\right] +$$
$$+ 2E\left[(Y - M_2)(M_2 - a - bX)\right]$$

Veamos que el tercero de estos sumandos es nulo:
Para ello, hemos de considerar que para una función $g(X,Y)$ integrable en las condiciones necesarias se verifica que

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(y/x).f_1(x)dxdy =$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y)f(y/x)dy\right]f_1(x).dx = \int_{-\infty}^{\infty} E[g(X,Y)/X = x].f_1(x).dx$$

En nuestro caso, para la función $(Y - M_2)(M_2 - a - bX)$ será:

$$E[(Y - M_2)(M_2 - a - bX)] = \int_{-\infty}^{\infty} E[(Y - M_2)(M_2 - a - bX)/X = x].f_1(x).dx$$

Ahora bien:

$$E[(Y - M_2)(M_2 - a - bX)/X = x] = \int_{-\infty}^{\infty} (y - M_2)(M_2 - a - bx).f(y/x).dy =$$

$$= (M_2 - a - bx)\int_{-\infty}^{\infty}(y - M_2).f(y/x).dy = (M_2 - a - bx)\left[\int_{-\infty}^{\infty} y.f(y/x).dx - M_2\int_{-\infty}^{\infty}.f(y/x).dx\right] =$$

$$= (M_2 - a - bx)[M_2 - M_2] = 0$$

En definitiva, la expresión queda así:

$$E\left[(Y - a - bX)^2\right] = E\left[(Y - M_2)^2\right] + E\left[(M_2 - a - bX)^2\right]$$

y puesto que $E\left[(Y - M_2)^2\right]$ no depende de los parámetros a y b, se tiene que la expresión $E\left[(Y - a - bX)^2\right]$ es mínima solo si $E\left[(M_2 - a - bX)^2\right]$ es mínima, con lo que queda probada la proposición.

Si la curva de regresión $y = M_2(x)$ fuera una recta: $M_2(x) = a + bx$, entonces los coeficientes a y b minimizan la expresión antedicha

$E\left[(Y - a - bX)^2\right] = E\left[(Y - M_2)^2\right] + E\left[(M_2 - a - bX)^2\right]$, pues el segundo sumando se anula: $E\left[(M_2 - a - bX)^2\right] = 0$

07.3. El coeficiente de correlación:

En el anterior apartado hemos llamado

$b = \frac{m_{11}}{m_{20}} = \rho\frac{D_2}{D_1}$, coeficiente de regresión de Y sobre X.

$b' = \frac{m_{11}}{m_{02}} = \rho\frac{D_1}{D_2}$, coeficiente de regresión de X sobre Y

$\rho = \frac{m_{11}}{\sqrt{m_{02}m_{20}}} = \frac{m_{11}}{D_1D_2}$, coeficiente de correlación entre ambas variables.

Veamos a continuación cómo el valor del coeficiente de correlación, válido tanto en la regresión de Y sobre X como en la regresión de X sobre Y, tiene ciertas propiedades que condicionan la forma de la distribución.

Por el teorema 14, sabemos que $E_{\min}\left[(Y-a-bX)^2\right]=D_2^2\left(1-\rho^2\right)$, de lo cual deducimos que para un valor mínimo dado, la varianza D_2^2 aumenta o disminuye en sentido contrario a la variación del término $\left(1-\rho^2\right)$, y lo mismo ocurre para la expresión $E_{\min}\left[(X-a'-b'Y)^2\right]=D_1^2\left(1-\rho^2\right)$.

Diremos que las variables aleatorias X e Y están *correlacionadas* sii $\rho \neq 0$. Caso contrario se dicen *incorrelacionadas* $(\rho = 0)$.

Teorema 16:
a) El coeficiente de correlación está comprendido en el intervalo real $[-1,1]$

$$-1 \leq \rho \leq 1$$

b) Si $\rho = 0$ las curvas de regresión son dos rectas paralelas a los ejes X e Y.
c) Si $\rho = \pm 1$ toda la masa de la distribución se encuentra en una línea recta.
Demostración:
a) Sea m la matriz de los momentos centrales

$$m = \begin{pmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{pmatrix}$$

$$m_{02} > 0, m_{20} > 0 \rightarrow m_{02}.m_{20} - m_{11}^2 \geq 0 \rightarrow \frac{m_{11}^2}{m_{02}m_{20}} \leq 1 \rightarrow \rho^2 \leq 1 \rightarrow -1 \leq \rho \leq 1$$

b) Si $\rho = 0$, entonces $m_{11}/D_1D_2 = 0 \rightarrow m_{11} = 0 \rightarrow m_{11} = M_{11} - M_1M_2 = 0 \rightarrow M_{11} =$ $= M_1.M_2 \rightarrow E[X.Y] = E[X].E[Y] \rightarrow f(x,y) = f(x).f(y) \rightarrow X,Y$ est. Independientes, con lo que las dos rectas de regresión se reducen a dos paralelas a los ejes:

$$y = M_2, \quad x = M_1$$

c) Si $\rho = \pm 1$, entonces $\rho = m_{11}\big/\sqrt{m_{02}m_{20}} = \pm 1 \rightarrow \rho^2 = m_{11}^2\big/m_{02}m_{20} = 1 \rightarrow$ $\rightarrow m_{02}m_{20} = m_{11}^2 \rightarrow rang(m) = 1 \rightarrow$ toda la masa de la distribución se encuentra en una línea recta. Es decir, las dos rectas de regresión, de Y sobre X y de X sobre Y, coinciden con la recta sobre la cual se halla toda la masa de la distribución.

07.4. La razón de Correlación:

La razón de correlación pretende ser una medida de la relación entre la dispersión estadística de las variables individuales y la dispersión de la población completa. Fue introducida por Karl Pearson, definiéndola del siguiente modo:

- Razón de correlación de Y en X:

$$\theta_{21}^2 = \frac{1}{D_2^2} E\left[(M_2(X) - M_2)^2\right]$$

- Razón de correlación de X en Y:

$$\theta_{12}^2 = \frac{1}{D_1^2} E\left[(M_1(Y) - M_1)^2\right]$$

Teorema 17: La razón de correlación de Pearson tiene las siguientes propiedades:

a) $0 \leq \theta_{21}^2 \leq 1$

b) $\theta_{21}^2 = \rho^2 + \frac{1}{D_2^2} E\left[(M_2(X) - a - bX)^2\right]$

c) Si $y = M_2(x)$ es una recta $\rightarrow \theta_{21}^2 = \rho^2$

d) $\theta_{21}^2 = 0 \leftrightarrow M_2(x) = y$ es una recta horizontal

e) $\theta_{21}^2 = 1 \leftrightarrow$ toda la masa está en $y = M_2(x)$

Demostración:

a) Por ser $D_2^2 = E\left[(Y - M_2)^2\right] = E\left[(Y - M_2(X) + M_2(X) - M_2)^2\right] =$

$$= E\left[(Y - M_2(X))^2 + (M_2(X) - M_2)^2 + 2(Y - M_2(X))(M_2(X) - M_2)\right] = E\left[(Y - M_2(X))^2\right] +$$

$$+ E\left[(M_2(X) - M_2)\right] + 2E\left[(Y - M_2(X))(M_2(X) - M_2)\right] \rightarrow 1 = \frac{1}{D_2^2} E\left[(Y - M_2(X))^2\right] + \theta_{21}^2 +$$

$$+ \frac{2}{D_2^2} E\left[(Y - M_2(X))(M_2(X) - M_2)\right] \rightarrow 1 - \theta_{21}^2 = \frac{1}{D_2^2} E\left[(Y - M_2(X))^2\right] \geq 0$$

O sea: $0 \leq \theta_{21}^2 \leq 1$

b) Puesto que $E\left[(Y - a - bX)^2\right] = E\left[(Y - M_2(X))^2\right] + E\left[(M_2(X) - a - bX)^2\right]$, se tiene:

$$D_2^2(1 - \rho^2) = D_2^2(1 - \theta_{21}^2) + E\left[(M_2(X) - a - bX)^2\right] \rightarrow$$

$$\rightarrow \theta_{21}^2 = \rho^2 + \frac{1}{D_2^2} E\left[(M_2(X) - a - bX)^2\right]$$

c) $y = M_2(x)$ recta entonces, por teorema 15 es $M_2(x) = a + bx$, por lo que teniendo en cuenta b):

$$\theta_{21}^2 = \rho^2 + \frac{1}{D_2^2} E\left[(M_2(X) - a - bX)^2\right] = \rho^2 + 0 = \rho^2$$

d) Si $\theta_{21}^2 = 0 \rightarrow \rho = 0 \rightarrow$ rectas de regresión paralelas a los ejes (por teorema 16, b))

Si $y = M_2(x)$ constante $\rightarrow \rho = b = 0 \rightarrow y = M_2(x) = a \rightarrow \theta_{21}^2 = 0$

e) Es obvio, por el resultado a).

08. Bibliografía

Chatterjee, S.- Price, B.; "Regression Analysis by Example". John Wiley & Sons, New York**,** 1977.
Domenech, J.M. y Riba, M.D. "Métodos estadísticos: modelo Lineal de Regresión". Herder,. Barcelona, 1985.
Pedhazur, E. J.; "Multiple Regression in Behavioral Research" (3rd ed.). Orlando, 1997
Cramer, H.; "Métodos matemáticos de la Estadística". Ediciones Aguilar.
Frechet, M.; "Recherches theoriques modernes sur la theorie des probabilities",
Gauthier-Villars, 10ª edic. 1950
Gndenko, B. ; "Teoría de probabilidades". Editorial Mir
Schweizer, B;Sklar, A.; "Probabilistic metric spaces", North Holland, N.York, 1983

Apéndice:

(*) Una serie $\sum_{k=0}^{\infty}\alpha_k$ es convergente si la sucesión $\{A_n\}_{n\geq 1}$ de sus sumas parciales es convergente: $A_0=\alpha_0,\, A_1=\alpha_0+\alpha_1,...,A_n=\sum_{k=0}^{n}\alpha_k,...$. Esto es: Si la sucesión $\{A_n\}_{n\geq 1}$ es convergente al límite l, entonces $\forall\varepsilon>0,\exists n_0\in N/\forall n>n_0,\left|l-A_n\right|<\varepsilon$. Si el límite al que tiende es $l=\sum_{k=0}^{\infty}\alpha_k$ se tendrá: $\left|\sum_{k=0}^{\infty}\alpha_k-\sum_{k=0}^{n}\alpha_k\right|=\sum_{k=n}^{\infty}\alpha_k<\varepsilon$

Por tanto, para la serie convergente $\sum_{k=0}^{\infty}p(M\cap[-(k+1)<Y\leq -k])$, se tiene que $\forall\varepsilon>0,\exists n_0\in N/\forall n>n_0,\sum_{k=n}^{\infty}p(M\cap[-(k+1)<Y\leq -k])<\varepsilon$