Clase 10: Covarianza y Correlación

Universidad Nacional de Colombia - Sede Medellín

Esperanza

Al igual que en el caso univariado se puede definir el valor esperado de funciones de variables aleatorias X e Y.

Definición

Sean X e Y variables aleatorias (Discretas o continuas) y sea $g(X, Y)$ una función de X e Y. El valor esperado de $g(X, Y)$ se define como:

$$E(g(X, Y)) = \begin{cases} \displaystyle\sum_{x} \sum_{y} g(x, y)\, p(x, y) & ; \quad \text{caso discreto} \\ \displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)\, f(x, y)\, dy\, dx & ; \quad \text{caso continuo} \end{cases}$$

Esperanza condicional

Sea Y una variable aleatoria y $g(Y)$ una función de Y, entonces la esperanza condicional de $g(Y)$ dado que $X = x$, está dada por:

$$E\left(g(Y)\,|X = x\right) = \begin{cases} \displaystyle\sum_{y \in \mathcal{A}_Y} g(y)\, p(y|x) & \text{Si } X, Y \text{ son v.a. discretas} \\ \displaystyle\int_{-\infty}^{\infty} g(y)\, f(y|x)\, dy & \text{Si } X, Y \text{ son v.a. continuas} \end{cases}$$

- En la definición anterior si $g(Y) = Y$, se tiene la media condicional de Y dado que $X = x$, $E(Y|X = x)$.
- Si $g(Y) = (Y - E(Y|X = x))^2$, se tiene la varianza condicional de Y dado que $X = x$.
- Recordemos que la varianza se puede obtener como $Var(Y) = E(Y - E(Y))$ o como $Var(Y) = E(Y^2) - (E(Y))^2$.

Ejemplo

Sean X e Y variables aleatorias discretas con f.m.p. conjunta dada por:

x	0	0	1	1	2	2
y	0	1	0	1	0	1
$p(x, y)$	$1/18$	$3/18$	$4/18$	$3/18$	$6/18$	$1/18$

Calcule $E(Y|X=0)$ y $Var(Y|X=0)$.

Solución

En la siguiente tabla tenemos tres distribuciones, la distribución conjunta de X y Y, la distribución marginal Y está en el renglón de abajo y la marginal de X está en la última fila.

		Y		
	$p(x,y)$	0	1	$p_X(x)$
X	0	$1/18$	$3/18$	$4/18$
	1	$4/18$	$3/18$	$7/18$
	2	$6/18$	$1/18$	$7/18$
	$p_Y(y)$	$11/18$	$7/18$	1

Como nos interesa $E(Y|X=0)$ y $Var(Y|X=0)$, debemos tener la distribución condicional de Y dado $X=0$. A continuación esa distribución.

y	0	1
$p_{Y\|0}(y)$	$1/4$	$3/4$

y	0	1
$p_{Y\|0}(y)$	$1/4$	$3/4$

Usando la distribución distribución condicional de Y dado $X = 0$ vamos a calcular lo solicitado.

$E(Y|X=0) = 0 \times 1/4 + 1 \times 3/4 = 3/4$.

Recordemos que $Var(X) = E(X^2) - (E(X))^2$. Así que calculemos primero $E(Y^2|X=0)$.

$E(Y^2|X=0) = 0^2 \times 1/4 + 1^2 \times 3/4 = 3/4$

Ahora ya podemos calcular $Var(Y|X=0)$ así:

$$\begin{aligned} Var(Y|X=0) &= E(Y^2|X=0) - (E(Y|X=0))^2 \\ &= 3/4 - (3/4)^2 = 3/4 - 9/16 = 3/16 \end{aligned}$$

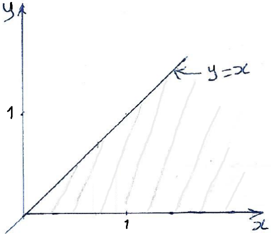
NOTA: en este ejemplo es una casualidad que $E(Y|X=0)$ y $E(Y^2|X=0)$ coinciden.

Ejemplo

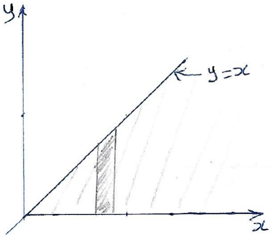
Sean X e Y variables aleatorias continuas con f.d.p conjunta dada por:

$$f(x, y) = e^{-x} \quad ; \quad 0 \leq y < x \, .$$

Calcule $E(Y|X = 2)$ y $Var(Y|X = 2)$.



Solución



Para obtener la distribución condicional debemos calcular primero la distribución marginal de X así:

$$f_X(x) = \int_0^x e^{-x}\, dy = \left(y\, e^{-x}\right|_0^x = x\, e^{-x} \quad , \quad x > 0\,.$$

Como la distribución condicional es el cociente entre la conjunta y la marginal, entonces:

$$f_{Y|X=2}(y) = \frac{f(2\,,\,y)}{f_X(2)} = \frac{e^{-2}}{2\,e^{-2}} = \frac{1}{2} \quad ; \quad 0 < y < 2\,.$$

Del resultado anterior vemos que $f_{Y|X=2}(y)$ no depende de y, es decir que la distribución es uniforme en el valor 1/2, en otras palabras:

$$f_{Y|X=2} \sim U(0,2)$$

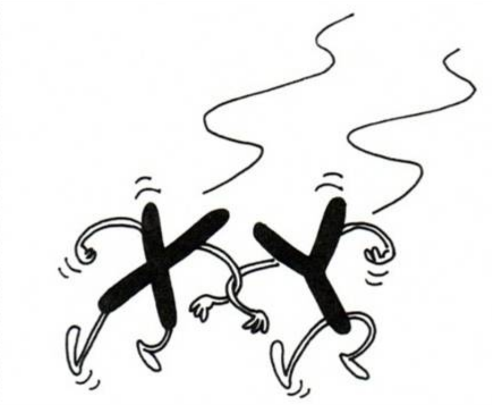
Eso significa que lo solicitado se puede calcular fácilmente así:

$$E(Y|X=2) = \frac{0+2}{2} = 1$$

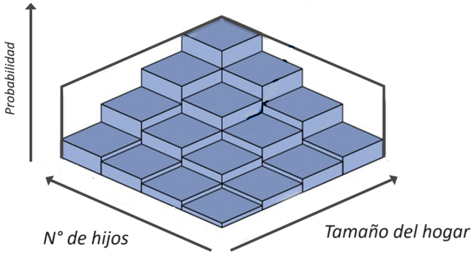
$$Var(Y|X=2) = \frac{(2-0)^2}{12} = \frac{4}{12} = \frac{1}{3}$$

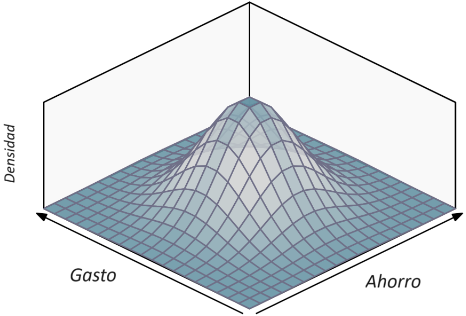
NOTA: si la distribución condicional no hubiese sido una distribución conocida, tendríamos que usar integración para obtener $E()$ y $Var()$.

Covarianza y Correlación



Covarianza y Correlación





Covarianza y Correlación

En el análisis de las distribuciones bivariadas para dos variables aleatorias X e Y, se pueden obtener medidas numéricas asociadas al grado o intensidad de la relación existente entre ambas variables, dos de estas medidas son la Covarianza y la Correlación. La Correlación es una medida de relación **lineal** entre las 2 variables. La Covarianza es una medida de relación entre las 2 variables.

Definición

La covarianza entre dos variables aleatorias X e Y está dada por:

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y)$$

El coeficiente de correlación entre dos variables aleatorias X e Y está dada por:

$$\rho = \rho_{XY} = Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho_{XY} \leq 1.$$

Ilustración de la correlación

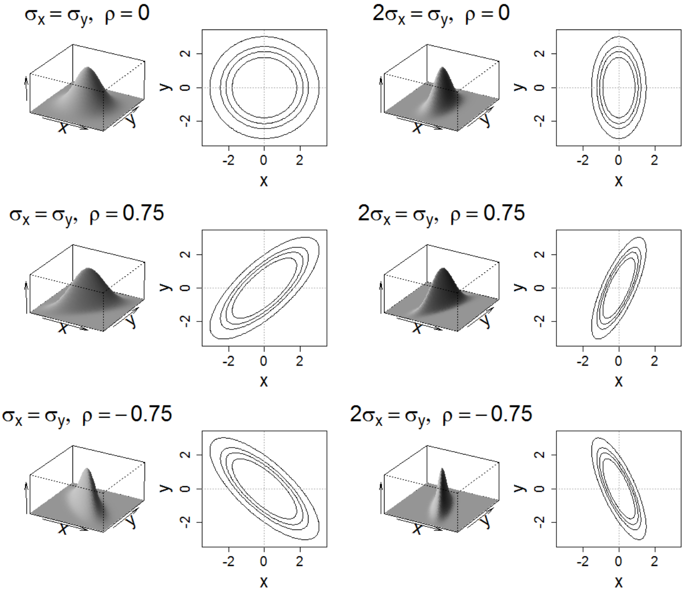


Ilustración de la correlación

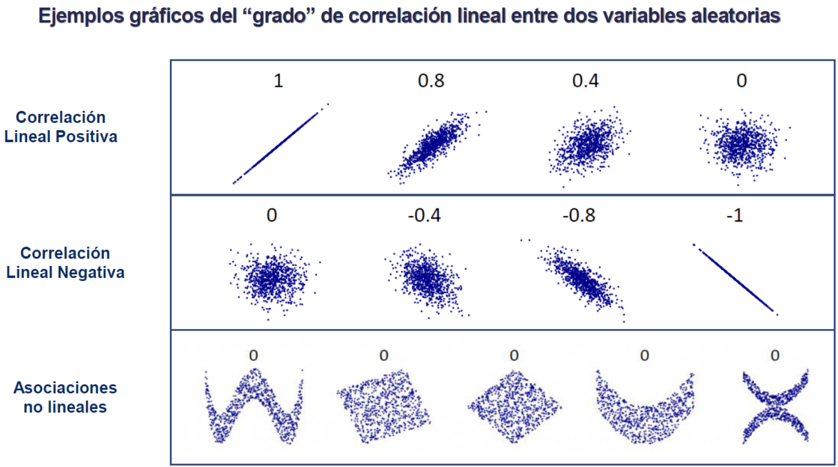
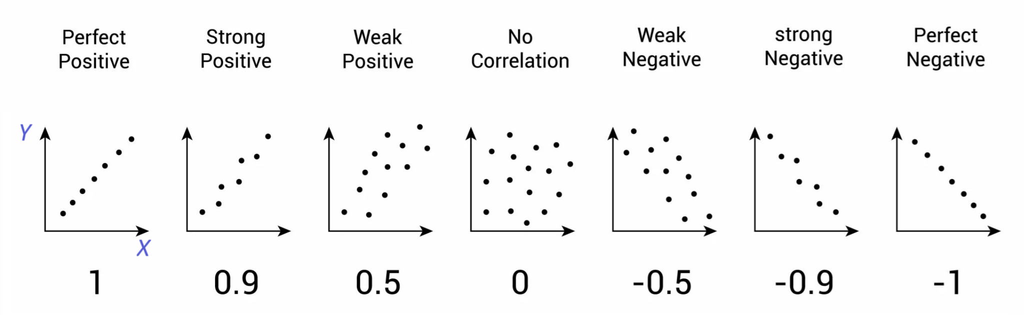


Ilustración de la correlación



Propiedades de la covarianza y la correlación

1. Si a, c, son constantes, ambas positivas o ambas negativas, entonces:
   $Cor(aX+b, cY+d) = Cor(X,Y)$
   Si a, c, son de signos opuestos, entonces:
   $Cor(aX+b, cY+d) = -Cor(X,Y)$
2. $-1 \leq Cor(X,Y) \leq 1$
3. $Cov(aX+b, cY+d) = acCov(X,Y)$
4. $Cov(X,X) = Var(X)$
5. $Var(aX+bY) = a^2\,Var(X) + b^2\,Var(Y) + 2abCov(X,Y)$
6. Si X, Y son independientes, entonces:
   $P(a < X < b, c < Y < d) = P(a < X < b)\,P(c < Y < d)$
   $Cov(X,Y) = Cor(X,Y) = 0$
   $E(XY) = E(X)\,E(Y)$

Nota: Si $Cov(X,Y) = 0$, no implica que X, Y sean independientes.

Ejemplo de covarianza y correlación

Sean X, Y dos v.a con f.m.p.c dada por:

	X	
Y	0	1
0	.38	.17
1	.14	.02
2	.24	.05

Chequee si X y Y son independientes, calcule $Cov(X, Y)$ y $Cor(X, Y)$.

Solución

A continuación se muestran la distribución conjunta y las dos marginales.

	X		
Y	0	1	$p_Y(y)$
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
$p_X(x)$	.76	.24	1.00

Para saber si las variables son independientes vamos a recorrer cada combinación de X y Y, y vamos a chequear si la conjunta se puede escribir como la multiplicación de las marginales.

- Iniciemos con $Y = 0$ y $X = 0$:
$p_Y(0) = 0.55$ y $p_X(0) = 0.76$, así que $p_Y(0)p_X(0) = 0.418$, pero $p(0,0) = 0.38$, como el producto no coincide a la probabidad conjunta, se concluye que Y y X NO son independientes.

	X		
Y	0	1	$p_Y(y)$
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
$p_X(x)$	.76	.24	1.00

Para calcular la $Cov(X, Y)$ necesitamos calcular tres cosas: $E(XY)$, $E(X)$ y $E(Y)$ así:

$$E(X) = 0 \times 0.76 + 1 \times 0.24 = 0.24$$
$$E(Y) = 0 \times 0.55 + 1 \times 0.16 + 2 \times 0.29 = 0.74$$
$$E(XY) = 0 \times 0 \times 0.38 + 0 \times 1 \times 0.17 + 1 \times 0 \times 0.14 + \ldots + 2 \times 1 \times 0.05 = 0.12$$

Así podemos calcular $Cov(X, Y) = E(XY) - E(X)E(Y) = 0.12 - 0.24 \times 0.74 = -0.0576$

	X		
Y	0	1	$p_Y(y)$
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
$p_X(x)$	.76	.24	1.00

Para calcular $Cor(X,Y)$ necesitamos calcular σ_X y σ_Y, las desviaciones de ambas variables.
Comencemos con X.

$$E(X^2) = 0^2 \times 0.76 + 1^2 \times 0.24 = 0.24$$
$$Var(X) = E(X^2) - (E(X))^2 = 0.24 - 0.24^2 = 0.1824$$
$$\sigma_X = \sqrt{Var(X)} = \sqrt{0.1824} = 0.427$$

Y	X		$p_Y(y)$
	0	1	
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
$p_X(x)$	.76	.24	1.00

Sigamos con Y.

$$E(Y^2) = 0^2 \times 0.55 + 1^2 \times 0.16 + 2^2 \times 0.29 = 1.32$$
$$Var(Y) = E(Y^2) - (E(Y))^2 = 1.32 - 0.74^2 = 0.7724$$
$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{0.7724} = 0.8789$$

Ahora si podemos obtener $Cor(X,Y)$ así:

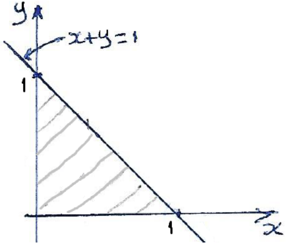
$$\begin{aligned} Cor(X,Y) &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\ &= \frac{-0.0576}{0.427 \times 0.8789} \\ &= -0.1535 \end{aligned}$$

Ejemplo

Sean X y Y dos variables aleatorias continuas con función de densidad de probabilidad conjunta dada por:

$$f(x,y) = \begin{cases} 2 & ; \quad \text{si} \quad x+y \leq 1, \quad x>0, \quad y>0 \\ 0 & ; \quad \text{en otro caso} \end{cases}$$

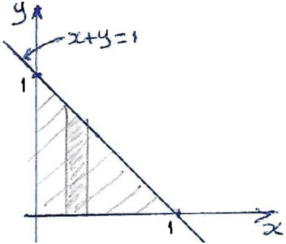
Encuentre $Cov(X,Y)$ y $Cor(X,Y)$.



Solución

Para hallar la covarianza se deben calcular los valores esperados de X e Y y el valor esperado $E(XY)$.

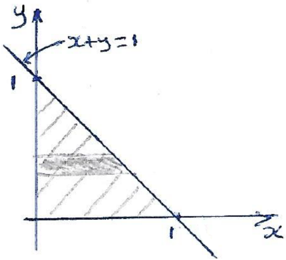
Para X.



$$f_X(x) = \int_0^{1-x} 2\,dy = 2y\Big|_0^{1-x} = 2(1-x) \quad \text{para } 0 \leq x \leq 1.$$

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \int_0^1 x \cdot 2(1-x)\,dx = \frac{1}{3}$$

Para Y.



$$f_Y(y) = \int_0^{1-y} 2\,dx = 2x\Big|_0^{1-y} = 2(1-y) \quad \text{para } 0 \le y \le 1.$$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_0^1 y \cdot 2(1-y)\, dy = \frac{1}{3}$$

El valor esperado del producto XY se obtiene así:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\, f(x,y)\, dy\, dx = \int_{0}^{1}\int_{0}^{1-x} x\, y\, 2\, dy\, dx = \int_{0}^{1} x\,(1-x)^2\, dx = \frac{1}{12}$$

Ahora que ya tenemos $E(X)$, $E(Y)$ y $E(XY)$ podemos calcular la covarianza así:

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{12} - \frac{1}{3} \times \frac{1}{3} = -\frac{1}{36}$$

Para obtener la correlación necesitamos las desviaciones estándar de X y Y.

Iniciemos con X.

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \, f_X(x) \, dx = \int_0^1 x^2 \, 2\,(1-x) \, dx = \frac{1}{6}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

$$\sigma_X = \sqrt{Var(X)} = \sqrt{1/18}$$

Sigamos con Y.

$$E(Y^2) = \int_{-\infty}^{\infty} y^2\, f_Y(y)\, dy = \int_0^1 y^2\, 2\,(1-y)\, dx = \frac{1}{6}$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{1/18}$$

Ya podemos calcular el coeficiente de correlación así:

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-\frac{1}{36}}{\sqrt{\frac{1}{18}}\sqrt{\frac{1}{18}}} = \frac{-\frac{1}{36}}{\frac{1}{18}} = -\frac{1}{2}$$

Resumiendo, la correlación y la covarianza entre X y Y son:

$$\boxed{Cor(X,Y) = -\frac{1}{2}}$$

$$\boxed{Cov(X,Y) = -\frac{1}{36}}$$

Ejercicio

Sean X y Y variables aleatorias continuas con f.d.p. conjunta dada por:

$$f(x, y) = 24\,x\,y \quad ; \quad 0 < x < 1\,,\, 0 < y < 1,\; x + y \leq 1$$

Hallar $Cor(X, Y)$.