Material para el Ramo de Probabilidad y Procesos Aleatorios (IPD-431)

Eduardo I. Silva*

24 de abril de 2012

1. Introducción

Las notas siguientes corresponden al material que pretendía discutir en la clase del día 25 de abril del 2012. En ellas, se usa la notación, convenciones, y las ideas ya vistas en clases anteriores.

2. Más sobre segundos momentos

Definition 1 *Suponga que X es una variable aleatoria continua n-dimensional con densidad de probabilidad f_X.*[1] *Suponga que $g : \mathbb{R}^n \to \mathbb{R}^m$ es una función medible. Entonces,*

$$\mathcal{E}\left\{g(X)\right\} \triangleq \int_{\mathbb{R}^n} g(x) f_X(x) dx \tag{1}$$

se denomina esperanza de (la variable aleatoria) $g(X)$. ■

Note que la integral en (1) puede no existir. En dicho caso, $\mathcal{E}\left\{g(X)\right\}$ tampoco existe. El operador esperanza, i.e., $\mathcal{E}\left\{\cdot\right\}$ en (1), es un operador lineal.

Definition 2 *Suponga que X es una variable aleatoria continua. Entonces,*

$$\mu_X \triangleq \mathcal{E}\left\{X\right\}, \quad P_X \triangleq \mathcal{E}\left\{(X-\mu_X)(X-\mu_X)^T\right\}, \quad Q_X \triangleq \mathcal{E}\left\{XX^T\right\} \tag{2}$$

se denominan esperanza (o valor esperado o medio) de la variable X, (matriz de co-) varianza de la variable X, y segundo momento de la variable X, respectivamente. ■

Se puede probar que tanto Q_X como P_X son matrices semi-positivas definidas (i.e., $y^T Q_X y \geq 0$ para todo vector real y de dimensiones apropiadas [2]).

Definition 3 *Suponga que X e Y son variables aleatorias conjuntamente distribuidas. Entonces,*

$$P_{XY} \triangleq \mathcal{E}\left\{(X-\mu_X)(Y-\mu_Y)^T\right\} \tag{3}$$

se denomina (matriz de co-) varianza cruzada entre X e Y. ■

*Departamento de Electrónica, Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile. Email: eduardo.silva@usm.cl.

[1] Definida sobre un espacio de probabilidad apropiado. En lo que sigue supondremos que todas las variables aleatorias son continuas y admiten una densidad de probabilidad.

Note que para que la definición de P_{XY} tenga sentido, no es necesario que X e Y tengan la misma dimensión. Basta con que ambas variables tengan la misma dimensión por columnas. Evidentemente, $P_{XX} = P_X$.

Definition 4 *Dos variables aleatorias conjuntamente distribuidas se dicen no correlacionadas si su varianza cruzada es cero.* ■

Definition 5 *Dadas dos variables aleatorias escalares*[2] *conjuntamente distribuidas, X e Y, definimos el coeficiente de correlación entre ellas como*

$$\rho_{XY} \triangleq \frac{P_{XY}}{\sqrt{P_X P_Y}}, \tag{4}$$

donde se ha supuesto que P_X y P_Y son no nulas. ■

El coeficiente de correlación entre dos variables aleatorias admite una interpretación interesante en términos de la dependencia existente entre ellas. Para poder presentar esta interpretación necesitamos el siguiente resultado:

Lemma 1 (Desigualdad de Cauchy-Schwartz) *Suponga que X e Y son variables aleatorias escalares conjuntamente distribuidas. Entonces,*

$$\left| \mathcal{E}\left\{XY\right\}^2 \right| \leq \mathcal{E}\left\{|XY|\right\}^2 \leq \mathcal{E}\left\{X^2\right\} \mathcal{E}\left\{Y^2\right\}. \tag{5}$$

Proof: La primera desigualdad es una consecuencia directa de que $\left|\int f(x)dx\right| \leq \int |f(x)|\, dx$. Para probar la segunda desigualdad, note que, para todo $a \in \mathbb{R}$,

$$0 \leq \mathcal{E}\left\{(a\,|X| - |Y|)^2\right\} = a^2 \mathcal{E}\left\{X^2\right\} - 2a\mathcal{E}\left\{|XY|\right\} + \mathcal{E}\left\{Y^2\right\}. \tag{6}$$

La última expresión es un polinomio de segundo orden en a que, para todo a, es positivo. Por lo tanto, el discriminante correspondiente (i.e., el término $b^2 - 4ac$ relacionado con las raíces de la ecuación $ax^2 + bx + c = 0$) debe ser negativo o cero. Por lo tanto,

$$4\mathcal{E}\left\{|XY|\right\}^2 - 4\mathcal{E}\left\{X^2\right\} \mathcal{E}\left\{Y^2\right\} \leq 0 \tag{7}$$

y el resultado es inmediato. ■

Note que la desigualdad de Cauchy-Schwartz aparece también en otros contextos, e.g., al estudiar espacios vectoriales finito dimensionales. Esto no es casualidad. De hecho, la desigualdad de Cauchy-Schwartz aparece cada vez que se considera un espacio vectorial (abstracto) equipado con un producto interno. Se puede probar que las variables aleatorias escalares de segundo orden son un espacio vectorial (en que los "vectores" son las variables aleatorias), el producto interno asociado es la varianza cruzada entre dos variables, y la idea de vectores ortogonales ha de reemplazarse por la noción de variables aleatorias no correlacionadas. Si quiere saber más, vea [3, Capítulo 4]. Lo anterior se puede extender incluso al caso de variables aleatorias multivariables en que el "producto interno" P_{XY} no es escalar, sino que toma valores matriciales [1, Capítulo 2].

Corollary 1 *Suponga que X e Y son variables aleatorias escalares conjuntamente distribuidas tales que $\mathcal{E}\left\{X^2\right\} < \infty$ y $\mathcal{E}\left\{Y^2\right\} < \infty$. Entonces:*

1. $|\mathcal{E}\left\{X\right\}| < \infty$, $|\mathcal{E}\left\{Y\right\}| < \infty$.
2. $P_X < \infty$, $P_Y < \infty$.
3. $|\mathcal{E}\left\{XY\right\}| < \infty$, $\mathcal{E}\left\{|XY|\right\} < \infty$, $|P_{XY}| < \infty$.

[2] i.e., variables aleatorias que toman valores en $\mathbb{R}$.

4. $\mathcal{E}\left\{|X+Y|^2\right\} < \infty$.

Proof:

1. Basta hacer $X \equiv 1$ ó $Y \equiv 1$ en (5).
2. Dado que el operador esperanza es lineal, $P_X = \mathcal{E}\left\{(X-\mu_X)^2\right\} = \mathcal{E}\left\{X^2\right\} - \mu_x^2$. Por lo tanto, el resultado es inmediato de la parte anterior.
3. Ejercicio. (Use Cauchy-Schwartz.)
4. Ejercicio. ■

El resultado anterior muestra que la existencia de segundos momentos garantiza la existencia de primeros momentos, varianzas, varianzas cruzadas, etc. Lo anterior puede extenderse a variable aleatorias multivariables.

Presentamos a continuación la interpretación del coeficiente de correlación anticipada anteriormente:

Theorem 1 *Considere dos variables aleatorias escalares, X e Y, de segundo orden y conjuntamente distribuidas. Entonces, $|\rho_{XY}| \leq 1$ con igualdad si y sólo si existen a y b reales tales que $Y = aX + b$.*

Proof:

1. Si usamos la desigualdad de Cauchy-Schwartz con $X - \mu_X$ en lugar de X e $Y - \mu_Y$ en lugar de Y, notamos que $\left|P_{XY}^2\right| \leq P_X P_Y$. Dado que tanto X como Y son de segundo orden, podemos dividir la expresión anterior por $P_X P_Y$ y sacar raíz cuadrada en ambos lados. El primer resultado es entonces evidente.
2. Probaremos nuestra segunda afirmación por partes:
   - ($\Leftarrow$) Si $Y = aX + b$, entonces
   $$P_Y = \mathcal{E}\left\{(Y-\mu_Y)^2\right\} = \mathcal{E}\left\{(aX+b-a\mu_X-b)^2\right\} = a^2\mathcal{E}\left\{(X-\mu_X)^2\right\} = a^2 P_X. \tag{8}$$
   Asimismo,
   $$P_{XY} = \mathcal{E}\left\{(X-\mu_X)(Y-\mu_Y)\right\} = aP_x. \tag{9}$$
   Por lo tanto,
   $$|\rho_{XY}| = \left|\frac{aP_x}{\sqrt{a^2 P_x \times P_x}}\right| = 1. \tag{10}$$
   - ($\Rightarrow$) Defina
   $$a \triangleq \frac{P_{XY}}{P_X}, \quad Z \triangleq aX - Y. \tag{11}$$
   Para probar el resultado basta con verificar que $P_Z = 0$ (¿por qué?). De hecho,
   $$\begin{aligned} P_Z = \mathcal{E}\left\{(aX-Y-a\mu_X-\mu_Y)(aX-Y-a\mu_X-\mu_Y)\right\} &= a^2 P_X + P_Y - 2aP_{XY} \\ &= \frac{P_{XY}^2}{P_X} + P_Y - \frac{2P_{XY}^2}{P_X} = P_Y - \frac{P_{XY}^2}{P_X} = 0, \end{aligned} \tag{12}$$
   donde hemos usado la linealidad de la esperanza, la definición de a, y donde la última igualdad es consecuencia de que $|\rho_{XY}| = 1$. ■

El resultado anterior muestra que si el coeficiente de correlación entre dos variables aleatorias es cercano a la unidad, entonces existe una relación afín entre ellas. (Note que la relación $Y = aX + b$ es lineal si y sólo si $b = 0$.) Es decir, existe una manera sencilla de explicar a Y como función de X. Si, por el contrario, dicho coeficiente es cercano a cero, entonces no existe relación afín que explique a Y como función de X.

Example 1 *Suponga que X es una variable aleatoria uniformemente distribuida en el intervalo $[-1/2, 1/2]$, i.e., $f_X(x) = 1$ si $x \in [-1/2, 1/2]$ y $f_X(x) = 0$ en otro caso. Defina la variable aleatoria $Y \triangleq X^2$. Claramente Y está íntimamente relacionada con la variable aleatoria X. Sin embargo, (complete Ud. los detalles)*

$$P_{XY} = \mathcal{E}\left\{(X - \mu_X)(Y - \mu_Y)\right\} = \mathcal{E}\left\{X(X^2 - P_X)\right\} = 0 \tag{13}$$

y, en consecuencia, $\rho_{XY} = 0$. (Note que tanto la varianza de X como la de Y es finita.) Lo anterior no contradice la discusión presentada después del Teorema 1 pues la relación existente entre X e Y no es afín. ■

3. Independencia

Hay ocasiones en que, dadas dos o más variables conjuntamente distribuidas, los valores que toman algunas de ellas no influyen en los valores de las otras. Esta idea se puede formalizar usando el concepto de independencia:

Definition 6

1. *Dos variables aleatorias conjuntamente distribuidas se dicen independientes si y sólo si su densidad conjunta es el producto de las marginales correspondientes. Es decir, las variables X e Y son independientes si y sólo si*[3]

$$f_{XY}(x, y) \triangleq f_{\begin{bmatrix} X \\ Y \end{bmatrix}}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = f_X(x) f_Y(y), \quad \forall x, y. \tag{14}$$

2. *Las variables aleatorias $X_1, \ldots, X_n$ se dicen mutuamente independientes si y sólo si*[4]

$$f_X(x) = \prod_{i=1}^{n} f_{X_i}(x_i), \tag{15}$$

donde $X \triangleq \begin{bmatrix} X_1^T & \ldots & X_n^T \end{bmatrix}^T$ y $x \triangleq \begin{bmatrix} x_1^T & \ldots & x_n^T \end{bmatrix}^T$.

3. *Las variable aleatorias $X_1, \ldots, X_n$ se dicen conjuntamente independientes de $Y_1, \ldots, Y_m$ si y sólo si $X \triangleq \begin{bmatrix} X_1^T & \ldots & X_n^T \end{bmatrix}^T$ e $Y \triangleq \begin{bmatrix} Y_1^T & \ldots & Y_m^T \end{bmatrix}^T$ son independientes.* ■

Example 2 *Si X e Y son independientes, entonces X e Y no están correlacionadas, i.e., $P_{XY} = 0$. De hecho,*

$$\begin{aligned} P_{XY} &= \mathcal{E}\left\{(X - \mu_X)(Y - \mu_Y)^T\right\} \\ &= \int_{\mathbb{R}^{n_x + n_y}} (X - \mu_X)(Y - \mu_Y)^T f_{XY}(x, y) dx dy \\ &= \int_{\mathbb{R}^{n_x}} (X - \mu_X) f_X(x) dx \int_{\mathbb{R}^{n_y}} (Y - \mu_Y)^T f_Y(y) dy \\ &= 0, \end{aligned} \tag{16}$$

[3] Note que la notación f_{XY} no ha sido definida con anterioridad. Esto justifica la igualdad intermedia en (14).
[4] $(\cdot)^T$ denota transpuesto.

donde la primera igualdad corresponde a la definición de P_{XY}, la segunda corresponde a la definición de esperanza, la tercera es consecuencia de que X e Y son independientes, y la última igualdad es trivial. Note que la afirmación contraria es falsa: Si dos variables aleatorias no están correlacionadas, entonces, en general, nada se puede decir acerca de la estructura de su distribución conjunta (vea también el Ejemplo 1, donde X e Y distan de ser independientes, pero $P_{XY} = 0$). ■

Example 3 *Si X e Y son independientes o no están correlacionadas, entonces, $P_{X+Y} = P_X + P_Y$. Para verificar lo anterior, basta con notar que*

$$\begin{aligned} P_{X+Y} &= \mathcal{E}\left\{(X+Y-\mu_X-\mu_Y)(X+Y-\mu_X-\mu_Y)^T\right\} \\ &= P_X + P_Y + \mathcal{E}\left\{(X-\mu_X)(Y-\mu_Y)^T\right\} + \mathcal{E}\left\{(Y-\mu_Y)(X-\mu_X)^T\right\} \\ &= P_X + P_Y, \end{aligned} \tag{17}$$

donde la primera igualdad corresponde a la definición de varianza, la segunda resulta de explotar la linealidad del operador esperanza, y la tercera es consecuencia de que X e Y no están correlacionadas (o son independientes). ■

Problem 1 *Pruebe que las afirmaciones siguientes son equivalentes:*

1. *X e Y son independientes.*
2. *$F_{XY}(x,y) = F_X(x)F_Y(y)$, para todo par (x,y).*
3. *$\mathcal{E}\{f(X)g(Y)\} = \mathcal{E}\{f(X)\}\mathcal{E}\{g(Y)\}$ para todas funciones f y g tales que las esperanzas involucradas existen.*

Note que la conclusión obtenida en el Ejemplo 2 es un caso particular de la condición en la Parte 3 (para ver esto defina $f(X) \triangleq X-\mu_X$ y $g(Y) \triangleq Y-\mu_Y$. Esto ilustra por qué $P_{XY} = 0$ no implica que X e Y sean independientes: La Parte 3 indica que independencia entre X e Y es equivalente a que $\mathcal{E}\{f(X)g(Y)\} = \mathcal{E}\{f(X)\}\mathcal{E}\{g(Y)\}$ para todas funciones f y g, y no sólo para las funciones afines $X-\mu_X$ e $Y-\mu_Y$. ■

Referencias

[1] B. Hassibi, A.H. Sayed, and T. Kailath. *Indefinite-quadratic estimation and control: a unified approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ theories.* SIAM, 1999.

[2] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press (9th reprint), 1999.

[3] D. Luenberger. *Optimization by Vector Space Methods.* John Wiley and Sons, Inc., London, 1969.

ESV - 24 de abril de 2012