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

Eduardo I. Silva*

25 de abril de 2012

1. Introducción

Las notas siguientes corresponden al material que pretendía discutir en la clase del día 26 de abril del 2012. En ellas, se usa la notación, convenciones, y las ideas ya vistas en clases anteriores.

2. Distribuciones y esperanzas condicionales

Considere un experimento al cual se han asociado dos variables aleatorias, digamos X e Y. Para caracterizar la distribución de X, dado que Y tomó un valor específico, introducimos la siguiente noción:

Definición 1 *Dadas variables aleatorias conjuntamente distribuidas, X e Y, definimos la densidad de probabilidad condicional de X, dado que $Y = y$, como*

$$f_{X|Y=y}(x) \triangleq \frac{f_{XY}(x,y)}{f_Y(y)}, \tag{1}$$

eq:condicional

para todo y tal que $f_Y(y) \neq 0$. ■

Nota 1 *Si y es tal que $f_Y(y) = 0$, entonces no es posible que $Y = y$ y, por lo tanto, no tiene sentido pretender calcular $f_{X|Y=y}(x)$.* ■

Para cada realización y de Y, (1) define una densidad de probabilidad distinta para X. Dicha densidad caracteriza a la variable aleatoria X cuando $Y = y$. Note que, si la realización x de X está fija, entonces $f_{X|Y=y}(x)$ puede entenderse como una función de la variable aleatoria Y. En otras palabras, para cada x, $f_{X|Y}(x)$ es una variable aleatoria cuyas realizaciones se obtienen dando valores a Y. Dichas realizaciones son $f_{X|Y=y}(x)$.

Ejercicio 1 *Pruebe que $f_{X|Y=y}$ tiene las propiedades de una densidad de probabilidad, i.e., pruebe que $f_{X|Y=y}(x) \geq 0$ para todo x, y que*

$$\int_{\mathbb{R}^{n_x}} f_{X|Y=y}(x)dx = 1. \tag{2}$$

■

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

Ejemplo 1 (Teorema de Bayes) *La definición de distribución condicional permite concluir inmediatamente que*

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y) f_X(x)}{f_Y(y)}. \tag{3}$$

Esta expresión permite intercambiar condicionantes. ■

Ejemplo 2 (Probabilidades totales) *Dadas dos variables conjuntamente distribuidas, X e Y, se tiene que*

$$f_X(x) = \int_{\mathbb{R}^{n_y}} f_{X|Y=y}(x) f_Y(y) dy. \tag{4}$$

De hecho,

$$\int_{\mathbb{R}^{n_y}} f_{X|Y=y}(x) f_Y(y) dy = \int_{\mathbb{R}^{n_y}} \frac{f_{XY}(x,y)}{f_Y(y)} f_Y(y) dy = \int_{\mathbb{R}^{n_y}} f_{XY}(x,y) dy = f_X(x), \tag{5}$$

donde la primera igualdad es consecuencia de la definición de densidad condicional, y la tercera es consecuencia de la definición de densidad marginal.

El resultado anterior permite construir densidades marginales usando las densidades condicionales y las densidades de los condicionantes. ■

Ejemplo 3 *La definición de independencia permite concluir que las variables (conjuntamente distribuidas) X e Y son independientes si y sólo si $f_{X|Y=y} = f_X$ para todo y.* ■

Ejemplo 4 *En este ejemplo probaremos que $f_X(x) = \mathcal{E}\left\{f_{X|Y}(x)\right\}$. Para ello, notamos que lo único aleatorio en $f_{X|Y}(x)$ es Y (X está fijo, de hecho, $X = x$). Por lo tanto, la definición de esperanza permite concluir que*

$$\mathcal{E}\left\{f_{X|Y}(x)\right\} = \int_{\mathbb{R}^{n_y}} f_{X|Y=y}(x) f_Y(y) dy = \int_{\mathbb{R}_{n_y}} f_{XY}(x,y) dy = f_X(x). \tag{6}$$

En la primera igualdad hemos usado el que las realizaciones de la variable aleatoria $f_{X|Y}(x)$ están dadas por $f_{X|Y=y}(x)$ y el que lo único aleatorio en $f_{X|Y}(x)$ es Y. Las restantes igualdades son consecuencia de las definiciones de densidades condicionales y marginales. ■

Muchas veces interesa calcular el valor esperado de una variable aleatoria, dado que otra variable ha tomado cierto valor. Esta noción puede capturarse usando el concepto de esperanza condicional:

Definición 2 *Considere dos variable alaeatorias conjuntamente distribuidas X e Y. Definimos, para "toda" función determinística $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \to \mathbb{R}^n$,*

$$\mathcal{E}\left\{g(X,Y)|Y = y\right\} \triangleq \int_{\mathbb{R}_{n_x}} g(x,y) f_{X|Y=y}(x) dx \tag{7}$$

eq:E-condicion

como la esperanza condicional de (la variable aleatoria) $g(X,Y)$, dado que $Y = y$. ■

Nota 2 *En (7), y tal como la notación sugiere, el valor que toma la variable Y está fijo (y es igual a y). Esto justifica el que se integre sobre x solamente.* ■

Nota 3 *Que g sea determinística implica que los valores que toma g dependen sólo de las realizaciones x e y de las variables aleatorias X e Y, y no de realizaciones de variables aleatorias adicionales. Por ejemplo, la función g definida a través de $g(X,Y) \triangleq XY$ es determinística. Sin embargo, la función g definida a través de $g(X,Y) \triangleq XY + Z$, donde Z es una variable aleatoria uniformemente distribuida en cierto intervalo, no es una función determinística. Muchas veces, funciones no determinísticas pueden* *re-escribirse como funciones determinísticas, haciendo explícita la dependencia de la función con respecto a las variables aleatorias adicionales. En el ejemplo anterior, se podría definir una función $\bar{g}$ a través de la relación $\bar{g}(X,Y,Z) = XY + Z$. Esta es una función determinística de X, Y y Z. Calcular la esperanza condicional de $\bar{g}(X,Y,Z)$, dado que, por ejemplo, $Y = y$, implica conocer la distribución condicional $f_{XZ|Y=y}(x,z)$.* ■

Es evidente que $\mathcal{E}\{g(X,Y)|Y=y\}$ es una función de la realización y de la variable aleatoria Y. Por lo tanto, podemos definir una función, digamos ϕ, a través de la relación

$$\phi(y) \triangleq \mathcal{E}\{g(X,Y)|Y=y\}\,. \tag{8}$$

eq:phi

La variable aleatoria que surge cuando, en (8), se reemplaza la realización y por la variable aleatoria Y (o sea, cuando el valor de Y se deja libre), i.e., la función definida a través de

$$\phi(Y) \triangleq \mathcal{E}\{g(X,Y)|Y\}\,, \tag{9}$$

eq:phi-va

se llama esperanza condicional de $g(X,Y)$, dado Y.

anza-iteradas

Teorema 1 (Esperanzas iteradas) *Considere dos variables conjuntamente distribuidas X e Y. Entonces, $\mathcal{E}\{\mathcal{E}\{X|Y\}\} = \mathcal{E}\{X\}$.*

Demostración: Usando la definición de la función ϕ en (8)-(9), con g tal que $g(X,Y) = X$, tenemos que

$$\begin{aligned}
\mathcal{E}\{\mathcal{E}\{X|Y\}\} &\overset{(a)}{=} \mathcal{E}\{\phi(Y)\} \\
&\overset{(b)}{=} \int_{\mathbb{R}^{n_y}} \phi(y) f_Y(y) dy \\
&\overset{(c)}{=} \int_{\mathbb{R}^{n_y}} \mathcal{E}\{X|Y=y\} f_Y(y) dy \\
&\overset{(d)}{=} \int_{\mathbb{R}^{n_y}} \left(\int_{\mathbb{R}^{n_x}} x f_{X|Y=y}(x) dx \right) f_Y(y) dy \\
&\overset{(e)}{=} \int_{\mathbb{R}^{n_y}} \left(\int_{\mathbb{R}^{n_x}} x f_{XY}(x,y) dx \right) dy \\
&\overset{(f)}{=} \int_{\mathbb{R}^{n_x}} x \left(\int_{\mathbb{R}^{n_y}} f_{XY}(x,y) dy \right) dx \\
&\overset{(g)}{=} \int_{\mathbb{R}^{n_x}} x f_X(x) dx \\
&\overset{(h)}{=} \mathcal{E}\{X\}\,,
\end{aligned} \tag{10}$$

donde (a) es consecuencia de la definición de ϕ, (b) es consecuencia de la definición de esperanza, de que las realizaciones de $\phi(Y)$ están dadas por $\phi(y)$, y de que lo único aleatorio en $\phi(Y)$ es Y, (c) es consecuencia de la definición de ϕ, (d) es consecuencia de la definición de esperanza condicional, (e) es consecuencia de la definición de densidad condicional, en (f) se cambió el orden de las integrales, (g) es consecuencia de la definición de distribución marginal, y (h) es consecuencia de la definición de esperanza. ■

-condicionales

Ejercicio 2 *Pruebe las siguientes afirmaciones:*

1. *El operador $\mathcal{E}\{\#|\star\}$ es lineal en $\#$.*
2. *Si g es una función determinística, entonces $\mathcal{E}\{\mathcal{E}\{g(X)|Y\}\} = \mathcal{E}\{g(X)\}$.*
3. *Si g y h son funciones determinística, entonces $\mathcal{E}\{g(Y)h(X)|Y=y\} = g(y)\mathcal{E}\{h(X)|Y=y\}$.*
4. *Si g y h son funciones determinística, entonces $\mathcal{E}\{\mathcal{E}\{g(Y)h(X)|Y\}\} = \mathcal{E}\{g(Y)\mathcal{E}\{h(X)|Y\}\}$.*
5. *Si X e Y son independientes, entonces $\mathcal{E}\{X|Y=y\} = \mathcal{E}\{X\}$.* ■

Finalizamos esta sección presentando una propiedad fundamental de la esperanza condicional:

ejor-estimador **Teorema 2** *Considere el funcional (i.e., una función cuyo argumento es una función)*

$$J(F) \triangleq \mathcal{E}\left\{(X - F(Y))(X - F(Y))^T\right\}, \tag{11}$$

eq:def-J

donde $F : \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$ *es una función determinística, y* X *e* Y *son variables aleatorias conjuntamente distribuidas. Entonces,*

$$J(F) \geq \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(X - \mathcal{E}\{X|Y\})^T\right\} \tag{12}$$

con igualdad si $F(Y) = \mathcal{E}\{X|Y\}$.

Demostración: Primero notamos que

$$\begin{aligned} J(F) &= \mathcal{E}\left\{(X - F(Y))(X - F(Y))^T\right\} \\ &= \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\} + \mathcal{E}\{X|Y\} - F(Y))(X - \mathcal{E}\{X|Y\} + \mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &= \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(X - \mathcal{E}\{X|Y\})^T\right\} + \mathcal{E}\left\{(\mathcal{E}\{X|Y\} - F(Y))(\mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &\quad + M + M^T, \end{aligned} \tag{13}$$

eq:J-extendido

donde se ha usado la linealidad del operador esperanza y

$$M \triangleq \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(\mathcal{E}\{X|Y\} - F(Y))^T\right\}. \tag{14}$$

Usando el Teorema 1 se puede escribir

$$\begin{aligned} M &= \mathcal{E}\left\{\mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(\mathcal{E}\{X|Y\} - F(Y))^T|Y\right\}\right\} \\ &\stackrel{(a)}{=} \mathcal{E}\left\{\mathcal{E}\{X - \mathcal{E}\{X|Y\}|Y\}(\mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &\stackrel{(b)}{=} \mathcal{E}\left\{(\mathcal{E}\{X|Y\} - \mathcal{E}\{\mathcal{E}\{X|Y\}|Y\})(\mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &\stackrel{(c)}{=} \mathcal{E}\left\{(\mathcal{E}\{X|Y\} - \mathcal{E}\{X|Y\})(\mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &= 0, \end{aligned} \tag{15}$$

eq:M-es-0

donde (a) es consecuencia de la cuarta propiedad que se pide probar en el Ejercicio 2 (haga $g(Y) = \mathcal{E}\{X|Y\} - F(Y)$ y $h(X,Y) = X - \mathcal{E}\{X|Y\}$), (b) es consecuencia de la linealidad de la esperanza condicional, y (c) es consecuencia de que, claramente, $\mathcal{E}\{\mathcal{E}\{X|Y\}|Y\} = \mathcal{E}\{X|Y\}$ (complete Ud. los detalles!). Ahora, si se reemplaza (15) en (13) se puede concluir que

$$\begin{aligned} J(F) &= \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(X - \mathcal{E}\{X|Y\})^T\right\} + \mathcal{E}\left\{(\mathcal{E}\{X|Y\} - F(Y))(\mathcal{E}\{X|Y\} - F(Y))^T\right\} \\ &\geq \mathcal{E}\left\{(X - \mathcal{E}\{X|Y\})(X - \mathcal{E}\{X|Y\})^T\right\}, \end{aligned} \tag{16}$$

con igualdad si $\mathcal{E}\{X|Y\} - F(Y) = 0$. (Note que la desigualdad es consecuencia de que los segundos momentos son semi-positivos definidos.) ■

Para interpretar el resultado anterior, suponga que en cierto experimento E se han identificado dos variables aleatorias, X e Y. Suponga que sólo Y es medible y que interesa utilizar dichas mediciones (i.e., las realizaciones de Y) para estimar los valores que toma X. Se define el error de estimación $\tilde{X}$ como $\tilde{X} \triangleq X - F(Y)$, donde F es un estimador genérico, i.e., una función que, dadas mediciones de Y, entrega la estimación $F(y)$ de la realización x de X (vea la Figura 1). Suponga que la calidad de la estimación se mide usando el segundo momento del error de estimación, i.e., usando el funcional J definido en (11). En este contexto, el mejor estimador corresponde a la función F que minimiza a J. El Teorema 2 muestra que dicho estimador óptimo no es más que la esperanza condicional de X dado Y. Este resultado es fundamental y está detrás de la solución de muchos problemas fundamentales en control, estimación, comunicaciones, etc. (vea, e.g., [1–6]). Sin embargo, en la práctica, puede resultar muy difícil calcular $\mathcal{E}\{X|Y\}$, salvo en algunos pocos casos particulares.

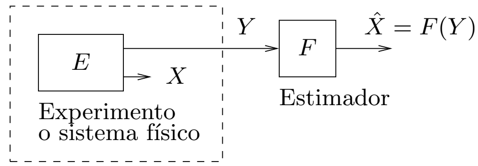

Figura 1: Esquema conceptual de un problema de estimación: El estimador F entrega la estimación $\hat{x} = F(y)$ de la variable no medible X en base a la realización y de la variable medible Y. (Note que la salida $\hat{X}$ del estimador es una también una variable aleatoria. Las realizaciones de esta variable se llaman estimaciones.) fig:estimador

Nota 4 *Dado que $\mathcal{E}\left\{\mathcal{E}\left\{X|Y\right\}\right\} = \mathcal{E}\left\{X\right\}$, concluimos que $\mathcal{E}\left\{E\right\} = 0$. Por lo tanto, el que J sea pequeño implica que las estimaciones son buenas. De hecho, si $J \approx 0$ entonces es al menos informalmente claro que $x \approx \mathcal{E}\left\{X|Y = y\right\}$ para todo x e y, i.e., las estimaciones son casi perfectas. (Recuerde la desigualdad de Chebyshev.)* ■

Nota 5 *Se puede probar que la esperanza condicional es el estimador óptimo de una variable aleatoria, dadas las mediciones de otra, cuando se usan criterios más generales que aquél considerado en el Teorema 2 (vea [6, Capítulo 5]).* ■

Referencias

andmoo79 [1] B.D.O. Anderson and J.B. Moore. *Optimal filtering.* Prentice Hall, Englewood Cliffs, N.J., 1979.

astrom70 [2] K.J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, New York, 1970.

baliki01 [3] Y. Bar-Shalom, X.R. Li, and T Kirubarajan. *Estimation with applications to tracking and navigation.* Wiley, 2001.

covtho06 [4] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* John Wiley and Sons, Inc., 2nd edition, 2006.

hasaka99 [5] 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.

soders02 [6] T. Söderström. *Discrete-time stochastic systems.* Springer, second edition, 2002.

ESV - 25 de abril de 2012