Elements of Structural Modelling

Conditional Probability

Cursillo de Verano 2020, Facultad de Matemáticas, UC

Ernesto San Martín

Laboratorio Interdisciplinario de Estadística Social LIES, UC
The Economics School of Louvain, Université Catholique de Louvain, Belgium

January 2021

Probabilidad condicional

- Sea $(M, \mathcal{M}, P)$ un espacio de probabilidad.
- Sea $A, B \in \mathcal{M}$. La probabilidad de A condicionalmente a B se define como

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

si $P(B) \neq 0$.

Probabilidad condicional: Herramienta

- Novick (1979):

  Conditional independence is the key concept of structural modelling.

Probabilidad condicional: Herramienta

Laboratorio Interdisciplinario de Estadística Social

- Novick (1979):

  Conditional independence is the key concept of structural modelling.

- But $P(A \mid B)$ when $P(B) \neq 0$ is just a number: why is it a key tool for structural modelling?

Descomposición marginal-condicional

- El siguiente argumento ha sido tomado de Laplace (*Théorie Analytique des Probabilité*, cuarto principio).
- Sea p el número total de casos posibles. Supongamos además que:
  - De los p casos totales, hay p' casos favorables a la ocurrencia del evento E_1.
  - De estos p' casos, hay q casos favorables a la ocurrencia del evento E_2.
- Suponemos que $q < p'$. Entonces si el evento E_1 ocurre, **no necesariamente** ocurre el evento E_2. Sin embargo, si observamos que el evento E_2 ocurre, entonces **necesariamente** el evento E_1 ocurre.
- Por ejemplo, E_1 corresponde a "día nublado en un año determinado" y E_2 corresponde a "día lluvioso en un año determinado". Así, si llueve, entonces está nublado, pero si está nublado no necesariamente llueve.

Descomposición marginal-condicional

- Por lo tanto, la ocurrencia de los eventos E_1 y E_2 es dependiente la una de la otro.
- La **ocurrencia de eventos dependientes** no es necesariamente simétrica.

Descomposición marginal-condicional

Laboratorio Interdisciplinario de Estadística Social

- La probabilidad de que ocurra el evento E_1 está dada por

$$\frac{p'}{p};$$

en el contexto del ejemplo, esto corresponde a la proporción de días en un año que estaba nublado.

- La probabilidad de que ocurra el evento E_2 **bajo la condición de que el evento E_1 ocurrió** está dada por

$$\frac{q}{p'};$$

en el contexto del ejemplo, esto correspondería a la proporción de días lluviosos entre ls días nublados.

- La probabilidad de que ocurran **simultáneamente** los eventos E_1 y E_2 está dada por

$$\frac{q}{p};$$

en el contexto del ejemplo, esto corresponde a la proporción de días lluviosos (y por tanto nublados) en un año determinado.

Descomposición marginal-condicional

Laboratorio Interdisciplinario de Estadística Social

- Se sigue que

$$\frac{q}{p} = \frac{p'}{p} \times \frac{q}{p'};$$

- esto es, la probabilidad del evento compuesto E_1 y E_2 es igual a la probabilidad del evento E_1, multiplicada por la probabilidad del evento E_2, bajo el supuesto que E_1 ha ocurrido.

- En el contexto del ejemplo, esto significa que la proporción de días lluviososo (y por tanto nublados) en un año determinado, es igual a la proporción de días nublados, multiplicado por la proporción de dias lluviosos entre los días nublados.

Descomposición marginal-condicional

Laboratorio Interdisciplinario de Estadística Social

- La probabilidad del evento compuesto E_1 y E_2 denotémosla por $P(E_1 \text{ y } E_2)$.
- La probabilidad del evento E_1 denotémosla por $P(E_1)$.
- La probabilidad del evento E_2, bajo el supuesto que E_1 ha ocurrido, denotémosla por $P(E_2 \mid E_1)$.
- Por lo tanto, obtenemos la siguiente descomposición marginal-condicional:

$$P(E_1 \text{ y } E_2) = P(E_1) \times P(E_2 \mid E_1).$$

- $P(E_2 \mid E_1)$ se llama **probabilidad condicional de E_2 dado E_1**.
- La descomposición marginal-condicional indica en qué medida el conocimiento de E_1 es **informativo** para calcular la probabilidad de E_2.

Descomposición marginal-condicional

- La probabilidad condicional no está clásicamente definida como un cociente.
- Ayuda a representar la dependencia entre eventos.

Ejemplo: datos del SIMCE

Laboratorio Interdisciplinario de Estadística Social

- Aplicación SIMCE 4° básco 2012, prueba de lenguaje.
  - 216.523 estudiantes.
  - 7.683 escuelas.

Datos SIMCE

- Información oficial generada:
  - Puntaje SIMCE Lenguaje: 267.5 puntos de promedio, con una desviación estándar de 51.6 puntos.

Q1	Q2	Q3	Q4	Q5
195.02	221.20	241.76	258.13	272.21

Q6	Q7	Q8	Q9	Q10
285.09	298.20	312.91	332.81	381.64

Datos SIMCE

- Niveles de logro: inicial, intermedio, avanzado.
    - Nivel inicial (menos de 241 puntos en el SIMCE:

Estos alumnos y alumnas aún no han consolidado los aprendizajes del Nivel Intermedio, ya que en ocasiones demuestran logros en algunos de los aprendizajes descritos en ese nivel, pero con una menor frecuencia y de manera poco consistente. Aquí se agrupan desde estudiantes que están aprendiendo a leer frases breves, junto con estudiantes que, con un poco de ayuda podrían demostrar los aprendizajes del Nivel Intermedio.

Datos SIMCE

Laboratorio Interdisciplinario de Estadística Social

- Niveles de logro: inicial, intermedio, avanzado.
  - Nivel intermedio (entre 241 y 280 puntos en el SIMCE:

Los alumnos y alumnas alcanzan, en este nivel, una comprensión de los textos leídos que les permite extraer información explícita fácil de encontrar, realizar inferencias claramente sugeridas, reconocer algunos aspectos de la situación comunicativa y opinar sobre el contenido de textos familiares.
Los estudiantes que alcanzan este nivel son capaces, entre otras cosas, de:

- *Identificar información explícita que se visualiza fácilmente.*
- *Realizar inferencias a partir de información reiterada y/o destacada en el texto.*
- *Interpretar expresiones familiares en lenguaje figurado.*
- *Identificar tipo de texto.*
- *Identificar propósito, emisor y receptor cuando estos son evidentes.*
- *Reconocer de qué se trata un texto cuando es evidente.*
- *Expresar y fundamentar una opinión acerca de acciones de personajes o hechos descritos en un texto.*

Datos SIMCE

Laboratorio Interdisciplinario de Estadística Social

- Niveles de logro: inicial, intermedio, avanzado.
    - Nivel avanzado (más de 280 puntos en el SIMCE:

Los alumnos y alumnas alcanzan, en este nivel, una comprensión de los textos leídos que les permite relacionar e integrar diversas informaciones, tanto explícitas como implícitas (inferidas) y opinar sobre el contenido de textos poco familiares.
Los alumnos y alumnas que alcanzan este nivel son capaces, entre otras cosas, de:

- *Identificar información explícita que no se visualiza fácilmente o que está junto a información semejante.*
- *Realizar inferencias indirectamente sugeridas en el texto.*
- *Reconocer relaciones de causalidad en el texto.*
- *Interpretar expresiones no familiares en lenguaje figurado.*
- *Comprender el significado de una palabra a partir de diversas claves3 del texto.*
- *Expresar y fundamentar una opinión sobre informaciones o puntos de vista presentados en un texto.*

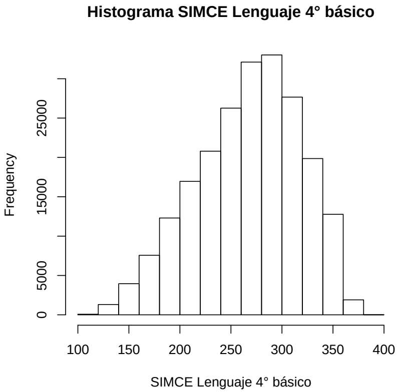
Datos SIMCE

- Niveles de logro:

Inicial	Internedio	Avanzado
30 %	30 %	40 %

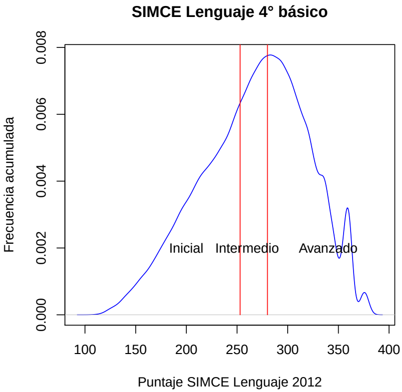
Datos del SIMCE

- Histograma de los puntajes SIMCE:



Datos del SIMCE

- Estimador no-paramétrico de la densidad de probabilidad:

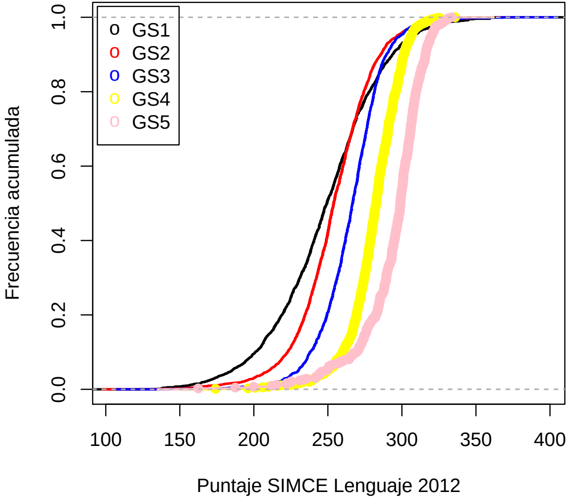


Datos del SIMCE

- Consideremos el desempeño de las escuelas según su nive socio-económico:

GSE	Q1	Mediana	Q3	Media	Desv. Est.	N
1	225.15	249.24	271.55	248.30	37.01	2373
2	238.58	253.83	270.24	253.90	27.32	2543
3	253.08	266.59	278.59	265.36	22.25	1601
4	272.20	282.69	293.11	281.20	18.18	746
5	285.64	299.01	307.46	294.11	22.34	420

SIMCE Lenguaje Promedio 4° básico



Distribución condicional

- Las distribuciones de puntajes para cada uno de los niveles socio-económicos corresponde a describir la variabilidad de dichos puntajes bajo una condición: el nivel socio-económico de la escuela.
- Así, la distribución del puntaje SIMCE Lenguaje Promedio 4° básico para el GSE 1 se denota como

$$P(Leng04_j \leqslant x \mid GSE = 1)$$

- De manera similar,

$$P(Leng04_j \leqslant x \mid GSE = g) \quad g = 2, 3, 4, 5.$$

Distribución condicional

- La distribución condicional $P(Leng04_j \leqslant x \mid GSE)$ describe la variabilidad del puntaje SIMCE Lenguaje 4° básico para cada una de los grupos socio-económicos.
- La distribución condicional $P(Leng04_j \leqslant x \mid GSE)$ es una función de la variable aleatoria GSE.
- Por lo tanta, la distribución condicional $P(Leng04_j \leqslant x \mid GSE)$ describe la dependencia que hay entre el puntaje lenguaje SIMCE 4° básico y el grupo socio-económico.

Distribución condicional

- Consideremos ahora la proporción de escuelas que tiene a lo más 240 puntos en el SIMCE Lenguaje 4° básico.

- Entonces

$$P(Leng04_j \leqslant 240) = \frac{1827}{7683} = 0{,}24.$$

- Consideremos esta misma información pero condicionando al GSE:

GSE 1	$P(Leng04_j \leqslant 240 \mid GSE = 1) = \frac{935}{2373} = 0{,}38$
GSE 2	$P(Leng04_j \leqslant 240 \mid GSE = 2) = \frac{686}{2543} = 0{,}27$
GSE 3	$P(Leng04_j \leqslant 240 \mid GSE = 3) = \frac{174}{1601} = 0{,}11$
GSE 4	$P(Leng04_j \leqslant 240 \mid GSE = 4) = \frac{20}{746} = 0{,}03$
GSE 5	$P(Leng04_j \leqslant 240 \mid GSE = 5) = \frac{12}{420} = 0{,}03$

Distribución condicional como función

- Para enfatizar que la probabilidad $P(Leng04_j \mid GSE)$ es una función del GSE, definamos la siguiente variable indicadora:

$$\mathbb{1}_{\{GSE=g\}} = \begin{cases} 1, & \text{si el GSE es } g; \\ 0, & \text{en caso contrario.} \end{cases}$$

- Entonces

$$\begin{aligned} P(Leng04_j \leqslant 240 \mid GSE) &= 0{,}30 \times \mathbb{1}_{\{GSE=1\}} + 0{,}27 \times \mathbb{1}_{\{GSE=2\}} \\ &+ 0{,}11 \times \mathbb{1}_{\{GSE=3\}} + 0{,}03 \times \mathbb{1}_{\{GSE=4\}} \\ &+ 0{,}03 \times \mathbb{1}_{\{GSE=5\}}. \end{aligned}$$

- Así, por ejemplo,

$$\begin{aligned} P(Leng04_j \leqslant 240 \mid GSE)(2) &= 0{,}27 \times \mathbb{1}_{\{GSE=2\}}(2) \\ &= 0{,}27 \\ &\doteq P(Leng04_j \leqslant 240 \mid GSE = 2). \end{aligned}$$

Distribución condicional como función

Laboratorio Interdisciplinario de Estadística Social

- Así, una probabilidad condicional es una función de aquella variable aleatoria sobre la cual se condiciona (en nuestro ejemplo, el GSE).
- Por lo tanto, una probabilidad condicional es una variable aleatoria la cual puede ser representada por medio de una probabilidad.

Distribución condicional como función

Laboratorio Interdisciplinario de Estadística Social

- Así, una probabilidad condicional es una función de aquella variable aleatoria sobre la cual se condiciona (en nuestro ejemplo, el GSE).
- Por lo tanto, una probabilidad condicional es una variable aleatoria la cual puede ser representada por medio de una probabilidad.

Teorema de Probabilidades Totales

- Pero notemos que el numerador de

$$P(Leng04_j \leqslant 240) = \frac{1827}{7683} = 0{,}24$$

corresponde a la suma de los numeradores de

$$\begin{array}{ll}
\text{GSE 1} & P(Leng04_j \leqslant 240 \mid GSE = 1) = \frac{935}{2373} = 0{,}38 \\
\text{GSE 2} & P(Leng04_j \leqslant 240 \mid GSE = 2) = \frac{686}{2543} = 0{,}27 \\
\text{GSE 3} & P(Leng04_j \leqslant 240 \mid GSE = 3) = \frac{174}{1601} = 0{,}11 \\
\text{GSE 4} & P(Leng04_j \leqslant 240 \mid GSE = 4) = \frac{20}{746} = 0{,}03 \\
\text{GSE 5} & P(Leng04_j \leqslant 240 \mid GSE = 5) = \frac{12}{420} = 0{,}03
\end{array}$$

Teorema de Probabilidades Totales

- Similarmente con el denominador de

$$P(Leng04_j \leqslant 240) = \frac{1827}{7683} = 0{,}24$$

corresponde a la suma de los numeradores de

GSE 1 $\quad P(Leng04_j \leqslant 240 \mid GSE = 1) = \frac{935}{2373} = 0{,}38$

GSE 2 $\quad P(Leng04_j \leqslant 240 \mid GSE = 2) = \frac{686}{2543} = 0{,}27$

GSE 3 $\quad P(Leng04_j \leqslant 240 \mid GSE = 3) = \frac{174}{1601} = 0{,}11$

GSE 4 $\quad P(Leng04_j \leqslant 240 \mid GSE = 4) = \frac{20}{746} = 0{,}03$

GSE 5 $\quad P(Leng04_j \leqslant 240 \mid GSE = 5) = \frac{12}{420} = 0{,}03$

Teorema de Probabilidades Totales

- Por lo tanto, podemos relacionar ambas probabilidades (marginales y condicionales) de la siguiente manera:

$$\begin{aligned}\frac{1827}{7683} &= \frac{935}{2373} \times \frac{2373}{7683} \\ &+ \frac{686}{2543} \times \frac{2543}{7683} \\ &+ \frac{174}{1601} \times \frac{1601}{7683} \\ &+ \frac{20}{746} \times \frac{746}{7683} \\ &+ \frac{12}{420} \times \frac{420}{7683}\end{aligned}$$

- Aquí

$$\frac{2373}{7683} = P(GSE = 1), \qquad \text{etc.}$$

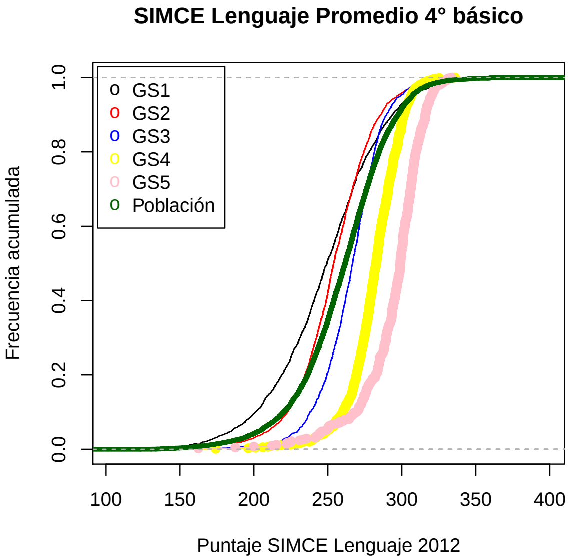
Teorema de Probabilidades TotalesE

Laboratorio Interdisciplinario de Estadística Social

- Por lo tanto,

$$\begin{aligned} P(Leng04_j \leqslant 240) &= P(Leng04_j \leqslant 240 \mid GSE = 1)\, P(GSE = 1) \\ &+ P(Leng04_j \leqslant 240 \mid GSE = 2)\, P(GSE = 2) \\ &+ P(Leng04_j \leqslant 240 \mid GSE = 3)\, P(GSE = 3) \\ &+ P(Leng04_j \leqslant 240 \mid GSE = 4)\, P(GSE = 4) \\ &+ P(Leng04_j \leqslant 240 \mid GSE = 5)\, P(GSE = 5). \end{aligned}$$

- Esto corresponde al Teorema de Probabilidades Totales.


SIMCE Lenguaje Promedio 4° básico
GS1
GS2
GS3
GS4
GS5
Población
Frecuencia acumulada
0.0
0.2
0.4
0.6
0.8
1.0
100
150
200
250
300
350
400
Puntaje SIMCE Lenguaje 2012

Definición de Probabilidad Condicional

- Sea $(M, \mathcal{M}, P)$ un espacio de probabilidad.
- Sea $\mathcal{C} \doteq \{C_1, C_2, C_3\}$ una partición medible –es decir, $C_1, C_2, C_3 \in \mathcal{M}$. Entonces para $B \in \mathcal{M}$, definimos la función

$$P(B \mid \mathcal{C}) \quad = \quad e_1 \mathbb{1}_{\{C_1\}} + e_2 \mathbb{1}_{\{C_2\}} + e_3 \mathbb{1}_{\{C_3\}},$$

donde e_1, e_2, e_3 son números reales distintos.

Definición de Probabilidad Condicional

Laboratorio Interdisciplinario de Estadística Social

- Sea $(M, \mathcal{M}, P)$ un espacio de probabilidad.
- Sea $\mathcal{C} \doteq \{C_1, C_2, C_3\}$ una partición medible –es decir, $C_1, C_2, C_3 \in \mathcal{M}$. Entonces para $B \in \mathcal{M}$, definimos la función

$$P(B \mid \mathcal{C}) \quad = \quad e_1 \mathbb{1}_{\{C_1\}} + e_2 \mathbb{1}_{\{C_2\}} + e_3 \mathbb{1}_{\{C_3\}},$$

  donde e_1, e_2, e_3 son números reales distintos.

- Entonces

$$P(B \mid \mathcal{C})(\omega) = e_1 \mathbb{1}_{\{C_1\}}(\omega) + e_2 \mathbb{1}_{\{C_2\}}(\omega) + e_3 \mathbb{1}_{\{C_3\}}(\omega)$$

  Si, por ejemplo, $\omega \in C_1$ entonces

$$P(B \mid \mathcal{C})(\omega) = e_1.$$

Definición de Probabilidad Condicional

Laboratorio Interdisciplinario de Estadística Social

- Recordemos que
$$\mathbb{1}_{\varnothing}(\omega) = 0.$$

- Luego, si un evento C es tal que $\{\omega \in M : \omega \in C\} = \varnothing$ y por tanto $P(C) = 0$.

- Luego, el número denotado por $P(B \mid C)$ se define arbitrariamente (con tal que sea finito) pues cualquier número multiplicado por 0 es igual a 0.

Definición de Probabilidad Condicional

- Ahora bien, podemos observar que $P(B \mid \mathcal{C})$ es una función –llamada función simple.
- Es decir, es una variable aleatoria definida sobre $(M, \sigma(\mathcal{C}))$, donde

$$\sigma(\mathcal{C}) = \{\emptyset, C_1, C_2, C_3, C_1 \cup C_2, C_1 \cup C_3, C_2 \cup C_3, M\}.$$

Definición de Probabilidad Condicional

- Notemos que

$$
\begin{aligned}
E[\,P(B \mid \mathcal{C})\,] &= E\left[e_1 1\!\!1_{\{C_1\}} + e_2 1\!\!1_{\{C_2\}} + e_3 1\!\!1_{\{C_3\}}\right] \\
&= e_1\, E[1\!\!1_{\{C_1\}}] + e_2 E[1\!\!1_{\{C_2\}}] + e_3 E 1\!\!1_{\{C_3\}}] \\
&= e_1 P(C_1) + e_2 P(C_2) + e_3 P(C_3).
\end{aligned}
$$

Definición de Probabilidad Condicional

- Kolmogorov (1933) define una probabilidad condicional como una variable aleatoria.

Definición de Probabilidad Condicional

- Kolmogorov (1933) define una probabilidad condicional como una variable aleatoria.
- Paso 1: se define una partición de M:

$$M = C_1 + \cdots + C_n.$$

- Paso 2: se define x una función real tal que a todo evento C_q le hace corresponder una constante a_q.
- Paso 2: La suma

$$E(x) = \sum_q a_q P(C_q)$$

se le llama esperanza matemática de x.

Definición de Probabilidad Condicional

Laboratorio Interdisciplinario de Estadística Social

- Una variable aleatoria para la cual el evento C_q asume el valor

$$P(B \mid C_q)$$

se llamará probabilidad condicional del evento B dado el experimento $\mathfrak{U}$ y se designará por $P(B \mid \mathfrak{U})$.

Definición de Probabilidad Condicional

- Volvamos entonces a

$$P(B \mid \mathcal{C}) = e_1 \mathbb{1}_{\{C_1\}} + e_2 \mathbb{1}_{\{C_2\}} + e_3 \mathbb{1}_{\{C_3\}}$$

donde

$$e_i \doteq P(B \mid C_i), \qquad i = 1, 2, 3.$$

Definición de Probabilidad Condicional

- Volvamos entonces a

$$P(B \mid \mathcal{C}) = e_1 \mathbb{1}_{\{C_1\}} + e_2 \mathbb{1}_{\{C_2\}} + e_3 \mathbb{1}_{\{C_3\}}$$

donde

$$e_i \doteq P(B \mid C_i), \qquad i = 1, 2, 3.$$

- Entonces

$$\begin{aligned}
E[\, P(B \mid \mathcal{C})\,] &= P(B \mid C_1)\, E(\mathbb{1}_{C_1}) + P(B \mid C_2)\, E(\mathbb{1}_{C_2}) + P(B \mid C_3)\, E(\mathbb{1}_{C_3}) \\
&= P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2) + P(B \mid C_3)P(C_3) \\
&= P(B \cap C_1) + P(B \cap C_2) + P(B \cap C_3) \\
&= P(B) \\
&= E(\mathbb{1}_B).
\end{aligned}$$

Definición de Probabilidad Condicional

- Es decir,

$$\begin{aligned} E[\,P(B \mid \mathcal{C})\,] &= E\,[E(\mathbb{1}_B \mid \mathcal{C})] \\ &= P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2) + P(B \mid C_3)P(C_3) \\ &= \\ &= E(\mathbb{1}_B). \end{aligned}$$

Definición de Probabilidad Condicional

- Es decir,

$$
\begin{aligned}
E[\,P(B \mid \mathcal{C})\,] &= E\,[E(\mathbb{1}_B \mid \mathcal{C})] \\
&= P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2) + P(B \mid C_3)P(C_3) \\
&= \\
&= E(\mathbb{1}_B).
\end{aligned}
$$

- El Teorema de Probabiliaddes Totales es lo mismo que la Definición kolmogoriana de Probabilidad Condicional, que a su vez es lo mismo que Marginalizar una probabilidad conjunta, que a su vez es ls mismo que la propiedad de una esperanza iterada.

Definición de Probabilidad Condicional

- Lo que se tiene es
$$E\left[E(\mathbb{1}_B \mid \mathcal{C})\right] = E(\mathbb{1}_B),$$
donde
$$E(\mathbb{1}_B \mid \mathcal{C}) = P(B \mid C_1)\mathbb{1}_{\{C_1\}} + P(B \mid C_2)\mathbb{1}_{\{C_2\}} + P(B \mid C_3)\mathbb{1}_{\{C_3\}}$$
- La esperanza condicional es una función de la partición $\mathcal{C}$, la que es partición de M. Se dice que $E(\mathbb{1}_B \mid \mathcal{C})$ es medible con respecto a $\mathcal{C}$: es lo que dice efectivamente, que se puede medir pues no es otra cosa que parte de $\mathcal{M}$.

Nota para la enseñanza de la probabilidad

- Enseñamos como temáticas disconexas conceptualmente (**no** deductivamente) una serie de capítulos de la Teoría de Probabilidad.
- Estos capítulo disconexos solo ebfatizan reglas de cálculo, pero no el sentido de una probabilidad condicional como una representación de un objeto marginal a partir de la marginalización de una partición.
- Hay muchísimas particiones que dan sentido a una marginalización. Como se trata de explicitar dependencias, la partición elegida depende del contexto de la aplicación.
- Y no olvidemos que una variable aleatoria induce y está inducida por una partición.

Nota para la enseñanza de la probabilidad

- Enseñamos como temáticas disconexas conceptualmente (**no** deductivamente) una serie de capítulos de la Teoría de Probabilidad.
- Estos capítulo disconexos solo ebfatizan reglas de cálculo, pero no el sentido de una probabilidad condicional como una representación de un objeto marginal a partir de la marginalización de una partición.
- Hay muchísimas particiones que dan sentido a una marginalización. Como se trata de explicitar dependencias, la partición elegida depende del contexto de la aplicación.
- Y no olvidemos que una variable aleatoria induce y está inducida por una partición.
- ¿Por qué esta definición?

Conditioning with densities

Laboratorio Interdisciplinario de Estadística Social

- Let X, Y be random variables defined on $(M, \mathcal{M}, P)$ having an absolutely continuous (joint) distribution function $F_{X,Y}$ with density $f_{X,Y}$ defined as

$$P(X < x, Y < y) = F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dvdu.$$

- If F_X and F_Y are r¿the marginal distributions of X and Y, and f_X and f_Y are their respective densities, then

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u, v) dv \right) du.$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(v) dv = \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} f_{X,Y}(u, v) du \right) dv.$$

Conditioning with densities

- The conditional probability $P(X \in A \mid Y \in B)$, with $A, B \in \mathcal{M}$, becomes

$$
\begin{aligned}
P(X \in A \mid Y \in B) &= \frac{P(\{X \in A\} \cap \{Y \in B\})}{P(Y \in B)} \\
&= \frac{\int_A \int_B f_{X,Y}(u,v)dudv}{\int_B f_Y(v)dv}.
\end{aligned}
$$

- If $A = [a_1, a_2)$ and $B = [b, b+h)$ with $h > 0$, then

$$
P(a_1 \leqslant X < a_2 \mid b \leqslant Y < b+h) = \int_{a_1}^{a_2} \left(\frac{\int_b^{b+h} f_{X,Y}(u,v)dv}{\int_b^{b+h} f_Y(v)dv} \right) du \quad (1)
$$

Conditioning with densities

Laboratorio Interdisciplinario de Estadística Social

- Since both

$$P(b \leqslant Y < b + h), \qquad P(a_1 \leqslant X < a_2,\, b \leqslant Y < b + h)$$

tend to zero as $h \downarrow 0$. Therefore, the left side of the equality (1) results in an indeterminate form.

- These two probabilities become

$$P(Y = b), \qquad P(a_1 \leqslant X < a_2, Y = b)$$

which are positive only if Y has a discrete part in its distribution at b.

- Since these zero in the present case, one can apply a form of the L'Hôpital's rule and evaluate at the right hand of the expression (1) and define it as the value of the left side.

Conditioning with densities

- Thus one has

$$\begin{aligned}
P(a_1 \leqslant X \leqslant a_2 \mid Y = b) &= \lim_{h \downarrow 0} P(a_1 \leqslant X \leqslant a_2 \mid b \leqslant Y \leqslant b + h) \\
&= \lim_{h \downarrow 0} \int_{a_1}^{a_2} \left(\frac{\frac{1}{h} \int_b^{b+h} f_{X,Y}(u, v) dv}{\frac{1}{h} \int_b^{b+h} f_Y(v) dv} \right) du \\
&= \int_{a_1}^{a_2} \frac{f_{X,Y}(u, b)}{f_Y(b)} du.
\end{aligned}$$

Here we have taken the limits inside the integral and used a form of the fundamental theorem of integral calculus.

Conditioning with densities

- Now set

$$f_{X|Y}(u \mid b) = \begin{cases} f_{X,Y}(u,b)/f_Y(b), & \text{if } f_Y(b) > 0, \\ \alpha, & f_Y(b) = 0, \end{cases}$$

  where $\alpha \geqslant 0$ is an arbitrary number.
  Then

$$P(X \in A \mid Y = b) = \int_A f_{X|Y}(u \mid b) du.$$

The Borel-Kolmogorov Type Example

- Let $X, Y \sim \exp(1)$ such that $X \perp\!\!\!\perp Y$; so

$$P(X < x) = P(Y < x) = (1 - e^{-x})\mathbb{1}_{\{x>0\}}.$$

- Let $a > 0$ and

$$Z = \frac{X - a}{Y}.$$

- Z is a random variable, and if $\alpha \in \mathbb{R}$,

$$B = \{Z = \alpha\}$$

  then $P(B) = 0$.

- If $A = \{Y < y\}$, it is desired to find $P(A \mid B)$.

The Borel-Kolmogorov Type Example

- We need to find the joint distribution of (Y, Z). Using a change of variable technique one can get $f_{X,Z}$ from $f_{X,Y}$ by setting $y = y$ and $x = a + zy$, so that the Jacobian is y and then

$$f_{Y,Z}(y,z) = \begin{cases} y \exp[-(yz + a) - y], & y > 0, \quad yz > -a; \\ 0, & \text{otherwise.} \end{cases}$$

- Hence we obtain f_Z and finally

$$f_{Y|Z}(y \mid \alpha) = \begin{cases} y(1+\alpha)^{-2} \exp[-y(1+\alpha)], & y > 0; \\ 0, & \text{otherwise;} \end{cases}$$

and therefore, for $x > 0$,

$$P(A \mid B) = \int_0^x f_{Y|Z}(y \mid \alpha) dy = (1+\alpha)^4 [1 - e^{-x(1+\alpha)}\{1 + x(1+\alpha)\}].$$

The Borel-Kolmogorov Type Example

- But the desired conditional probability may also be obtained as follows:

$$B = \{Z = \alpha\} = \{X - Y\alpha = a\}.$$

We compute $P(A \mid B)$ by letting

$$U = X - Y\alpha,$$

and considering the conditional density of X given $U = a$:

$$f_{Y,U}(y, u) = \begin{cases} \exp[-y(1+\alpha) - u], & y > 0, \quad \alpha y + u > 0; \\ 0, & \text{otherwise}; \end{cases}$$

- Then

$$f_{Y|U}(y \mid a) = \begin{cases} (1+\alpha)\exp[-y(1+\alpha)], & y > 0; \\ 0, & \text{otherwise}; \end{cases}$$

Therefore

$$P(A \mid B) = 1 - e^{-x(1+\alpha)}, \quad x > 0.$$

The Borel-Kolmogorov Type Example

- Kolmogorov explains the Borel paradox:

  This show that the concept of a conditional probability with regard to an isolated given hypothesis whose probability equals 0 is inadmissible.

The Borel-Kolmogorov Type Example

- Kolmogorov explains the Borel paradox:

  This show that the concept of a conditional probability with regard to an isolated given hypothesis whose probability equals 0 is inadmissible.

- Kolmogorov's object was specially to define a conditional probability:

  I wish to call attention to those points of the present exposition which are outside the above-mentioned range of ideas familiar to the specialist. They are the following: Probability distributions in infinite-dimensional spaces (Chapter III, § 4); differentiation and integration of mathematical expectations with respect to a parameter (Chapter IV, § 5); and especially the theory of conditional probabilities and conditional expectations (Chapter V). It should be emphasizes that these new problems arose, of necessity, from some perfectly concrete physical problems.

Acknowledgements

Laboratorio Interdisciplinario de Estadística Social

- Partially financed by the FONDECYT Project Number 1181261.