Bayes continuo, independencia de variables aleatorias y distribuciones derivadas

Clase 7: *conditional PDFs*, *independence*, *Bayes' rule* y *derived distributions*

6.431x — Probability: The Science of Uncertainty and Data

(versión de clase: las soluciones se comparten al terminar)

B.Sc. Cristian Lazo Quispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Programa SDS (Statistics and Data Science, MITx) · **BREIT** (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca— en alianza con MIT IDSS.
Basado en MIT 6.431x / MIT OCW RES.6-012 (CC BY-NC-SA 4.0).

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Dónde quedamos (Clase 6)

Variables continuas, en tres líneas

La **densidad** f_X reparte la masa sobre la recta: las probabilidades son **áreas** y f_X puede valer más que 1. Condicionar en un evento es **recortar y renormalizar** ($f_X/\mathbb{P}(A)$). La conjunta $f_{X,Y}$ es una superficie; la marginal se obtiene **integrando la franja** con los límites de la región.

Lo de hoy: tres preguntas que la Clase 6 dejó abiertas

- ¿Qué significa condicionar en $Y = y$ si $\mathbb{P}(Y = y) = 0$? La **densidad condicional** $f_{X \mid Y}$: cortar la conjunta en una rebanada y renormalizar.
- ¿Cuándo X e Y son **independientes**? Cuando $f_{X,Y} = f_X f_Y$ en todo el plano. Con eso, la **regla de Bayes** funciona igual con densidades: inferir un parámetro continuo (el sesgo de una moneda) a partir de datos.
- Si conozco f_X, ¿cuál es la densidad de $Y = g(X)$? **Distribuciones derivadas**: el método de la CDF (siempre funciona) y la fórmula del caso monótono (cuando se puede).

Cómo trabajaremos hoy

Primero la teoría de cada sección con ejemplos simples; al final, **cinco problemas difíciles**: el enunciado (lo desarrollamos juntos en la pizarra) y después la solución escrita paso a paso.

¿Qué veremos hoy? (Unit 5 — Lecture 10 · Unit 6 — Lecture 11)

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

La densidad condicional $f_{X \mid Y}$: cortar la conjunta y renormalizar

Definición

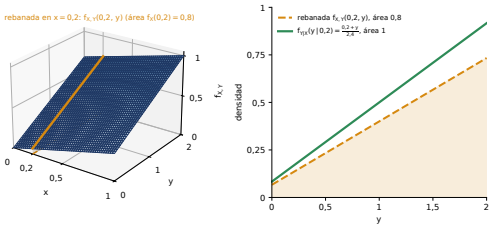
Para y con $f_Y(y) > 0$,

$$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Con y **fijo** es una densidad en x: $\int f_{X \mid Y}(x \mid y)\, dx = 1$. Es la **rebanada** de la superficie $f_{X,Y}$ en $Y = y$, dividida entre su área $f_Y(y)$.

Las mismas reglas de la Clase 5, con integrales

- **Multiplicación**: $f_{X,Y}(x,y) = f_Y(y)\, f_{X \mid Y}(x \mid y) = f_X(x)\, f_{Y \mid X}(y \mid x)$.
- **Probabilidad total**: $f_X(x) = \int f_Y(y)\, f_{X \mid Y}(x \mid y)\, dy$.
- **Valor esperado**: $\mathbb{E}[g(X) \mid Y = y] = \int g(x) f_{X \mid Y}(x \mid y)\, dx$.
- **Esperanza total**: $\mathbb{E}[X] = \int \mathbb{E}[X \mid Y = y]\, f_Y(y)\, dy$.



$f_{X,Y} = \frac{x+y}{3}$ en $[0,1] \times [0,2]$. La rebanada en $x = 0{,}2$ tiene área $f_X(0{,}2) = 0{,}8$; dividida entre 0,8 es $f_{Y \mid X}(y \mid 0{,}2) = \frac{0{,}2+y}{2{,}4}$, de área 1.

¿Condicionar en un evento de probabilidad cero?

$\mathbb{P}(Y = y) = 0$, pero $f_{X \mid Y}(x \mid y)$ es el límite de condicionar en $\{y \le Y \le y + \delta\}$ cuando $\delta \to 0$. Si $f_Y(y) = 0$, **no está definida**.

Ejemplo: una densidad sobre un triángulo, rebanada a rebanada

$f_{X,Y}(x,y) = 8xy$ en $0 \le x \le y \le 1$

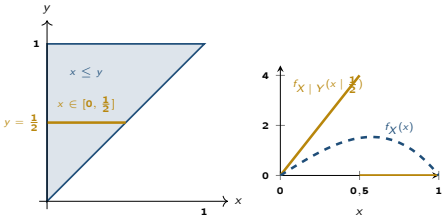
Marginal de Y (franja horizontal, $x \in [0, y]$):
$f_Y(y) = \int_0^y 8xy\,dx = 4y^3$.

$$f_{X \mid Y}(x \mid y) = \frac{8xy}{4y^3} = \frac{2x}{y^2}, \qquad 0 \le x \le y.$$

En $y = \frac{1}{2}$: $f_{X \mid Y}(x \mid \frac{1}{2}) = 8x$ en $[0, \frac{1}{2}]$ (área $8 \cdot \frac{1}{8} = 1$). ✓

Dos atajos que ahorran trabajo

- Como función de x, la condicional tiene **la misma forma** que la conjunta (aquí, lineal en x): condicionar solo **recorta el soporte y renormaliza**. Basta hallar la constante: $c \int_0^{1/2} x\,dx = 1 \Rightarrow c = 8$.
- $\mathbb{E}[X \mid Y = y] = \int_0^y x \cdot \frac{2x}{y^2}\,dx = \frac{2y}{3}$, y por esperanza total $\mathbb{E}[X] = \int_0^1 \frac{2y}{3} \cdot 4y^3 dy = \frac{8}{15}$ (sin calcular f_X).



Izquierda: la rebanada en $y = \frac{1}{2}$ recorre $x \in [0, \frac{1}{2}]$. Derecha: la condicional (naranja, hasta $\frac{1}{2}$) frente a la marginal $f_X(x) = 4x(1 - x^2)$ (azul).

Error clásico

$f_{X \mid Y}(2 \mid 0{,}5) = 0$ (fuera del soporte), pero $f_{X \mid Y}(0{,}5 \mid 2)$ **no existe**: $f_Y(2) = 0$.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

Independencia: la conjunta es el producto de las marginales

Definición

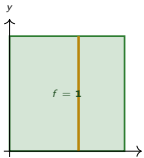
$X \perp\!\!\!\perp Y$ si $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$ para **todo** (x,y).
Equivalentemente, $f_{X \mid Y}(x \mid y) = f_X(x)$: saber Y no cambia la ley de X.

Consecuencias (las mismas que en la Clase 5)

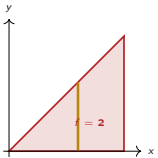
- $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$; en particular $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.
- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ (y al revés: derivando se recupera la conjunta).
- Funciones de variables independientes siguen siendo independientes.

El test más rápido: mira el soporte

Si la región donde $f_{X,Y} > 0$ **no es un rectángulo** (producto de intervalos), no pueden ser independientes. Si lo es, aún hay que ver si la fórmula factoriza: $x + y$ no; $6xy^2 = (2x)(3y^2)$ sí.



uniforme en el cuadrado: $X \perp\!\!\!\perp Y$



uniforme en el triángulo: no

En el cuadrado, la rebanada en $x = 0{,}6$ siempre recorre $[0,1]$: $f_{Y \mid X} = f_Y$.
En el triángulo, recorre $[0, 0{,}6]$: $f_{Y \mid X}(y \mid x) = \frac{1}{x}$ depende de x.

Ejemplo rápido

$X \sim \mathrm{Unif}(0,2)$, $Y \sim \mathrm{Exp}(1)$ independientes:
$f_{X,Y} = \frac{1}{2}e^{-y}$ en $[0,2] \times [0,\infty)$; $\mathbb{E}[XY] = 1$;
$\mathbb{P}(X \le 1, Y \le 1) = \frac{1}{2}(1 - e^{-1}) \approx 0{,}316$.

Normales independientes: una campana en dos dimensiones

Si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ son independientes

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{ -\frac{(x-\mu_X)^2}{2\sigma_X^2} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2} \right\}.$$

Las curvas de nivel son **elipses** con ejes paralelos a los de coordenadas (círculos si $\sigma_X = \sigma_Y$). Un término cruzado xy en el exponente las inclinaría: eso sería dependencia (Clase 8).

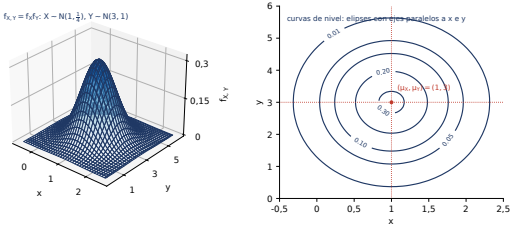
Leer una conjunta: completar cuadrados

$f_{X,Y} = c\, e^{-\frac{1}{2}(4x^2 - 8x + y^2 - 6y + 13)}$. Como $4x^2 - 8x = 4(x-1)^2 - 4$ y $y^2 - 6y = (y-3)^2 - 9$,

$$f_{X,Y} = c\, e^{-\frac{(x-1)^2}{2\cdot(1/4)}}\, e^{-\frac{(y-3)^2}{2\cdot 1}} :$$

$X \sim \mathcal{N}(1, \frac{1}{4})$, $Y \sim \mathcal{N}(3,1)$, independientes, $c = \frac{1}{2\pi\cdot\frac{1}{2}\cdot 1} = \frac{1}{\pi}$.

Entonces $\mathbb{P}(X > 1, Y < 3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ sin integrar nada.



La densidad conjunta de $\mathcal{N}(1, \frac{1}{4})$ y $\mathcal{N}(3,1)$ independientes y sus curvas de nivel, centradas en $(1,3)$ y más estrechas en x ($\sigma_X = \frac{1}{2}$) que en y ($\sigma_Y = 1$).

Lo que viene detrás

Si X, Y son normales independientes, $X + Y$ también es normal (Clase 8) y por eso la inferencia con ruido normal (Unit 7) es tan limpia.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

Una sola frase: posterior $\propto$ prior $\times$ verosimilitud

Las cuatro combinaciones (desconocido Θ, observación Y)

	Y discreta	Y continua
Θ discreto	$\dfrac{p_\Theta(\theta)\, p_{Y\mid\Theta}(y\mid\theta)}{\sum_{\theta'} p_\Theta(\theta')\, p_{Y\mid\Theta}(y\mid\theta')}$	$\dfrac{p_\Theta(\theta)\, f_{Y\mid\Theta}(y\mid\theta)}{\sum_{\theta'} p_\Theta(\theta')\, f_{Y\mid\Theta}(y\mid\theta')}$
Θ continuo	$\dfrac{f_\Theta(\theta)\, p_{Y\mid\Theta}(y\mid\theta)}{\int f_\Theta(\theta')\, p_{Y\mid\Theta}(y\mid\theta')\, d\theta'}$	$\dfrac{f_\Theta(\theta)\, f_{Y\mid\Theta}(y\mid\theta)}{\int f_\Theta(\theta')\, f_{Y\mid\Theta}(y\mid\theta')\, d\theta'}$

Se cambia PMF por densidad según la variable sea discreta o continua; el denominador es siempre la **probabilidad total** de lo observado, y no depende de θ.

Ejemplo (desconocido discreto, observación continua)

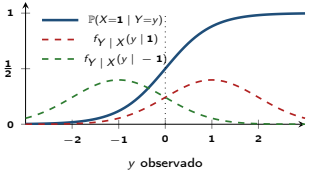
Una pieza viene de la máquina A o B (mitad y mitad). Su medida $Y \in [0,1]$ tiene densidad 1 si viene de A y $2y$ si viene de B. Observado $Y = 0{,}8$:

$$\mathbb{P}(B \mid Y = 0{,}8) = \frac{\frac{1}{2}\cdot 1{,}6}{\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1{,}6} = \frac{1{,}6}{2{,}6} = \frac{8}{13} \approx 0{,}615.$$

Ejemplo (desconocido discreto, ruido normal)

Se envía un bit $X = \pm 1$ (equiprobable) y se recibe $Y = X + W$, $W \sim \mathcal{N}(0,1)$. Las dos verosimilitudes son campanas centradas en ± 1:

$$\mathbb{P}(X = 1 \mid Y = y) = \frac{e^{-(y-1)^2/2}}{e^{-(y-1)^2/2} + e^{-(y+1)^2/2}} = \frac{1}{1 + e^{-2y}}.$$



La posterior es una **sigmoide**: cruza $\frac{1}{2}$ en $y = 0$ (donde las dos campanas se cortan) y se acerca a 0 o 1 lejos del centro. Es el Problema 3.

Desconocido continuo, observación discreta: el sesgo de una moneda

El modelo

Q = probabilidad de cara, desconocida; prior $Q \sim \text{Unif}(0,1)$. Dado $Q = q$, K = caras en n tiros es $\text{Bin}(n, q)$: $p_{K\,|\,Q}(k\,|\,q) = \binom{n}{k} q^k (1-q)^{n-k}$.

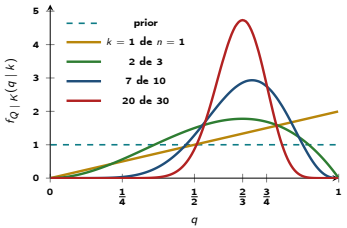
$$f_{Q\,|\,K}(q\,|\,k) = \frac{f_Q(q)\, p_{K\,|\,Q}(k\,|\,q)}{p_K(k)} \;\propto\; q^k (1-q)^{n-k}.$$

Un tiro, una cara ($n = k = 1$)

$p_K(1) = \int_0^1 q\, dq = \frac{1}{2}$, así que $f_{Q\,|\,K}(q\,|\,1) = \dfrac{1 \cdot q}{1/2} = 2q$: la creencia se inclina hacia sesgos altos. Media posterior $\frac{2}{3}$ (antes era $\frac{1}{2}$).

Tres números que resumen la posterior

Moda (MAP): k/n, la frecuencia observada. **Media**: $\frac{k+1}{n+2}$ (regla de Laplace), siempre más cerca de $\frac{1}{2}$ que la moda. **Probabilidad de la siguiente cara**: la media posterior. Con n grande la posterior se **concentra**.



Posteriores $\propto q^k(1-q)^{n-k}$ con la misma frecuencia $\approx \frac{2}{3}$ y cada vez más datos: la campana se estrecha alrededor de $\frac{2}{3}$. Densidades que valen 4 y 5: no son probabilidades.

Es el anticipo de la Unit 7

Inferencia bayesiana: prior → datos → posterior; estimadores MAP y de mínimo error cuadrático (la media posterior). Problema 2 de hoy.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

La densidad de $Y = g(X)$: el método de la CDF (el que siempre funciona)

Receta

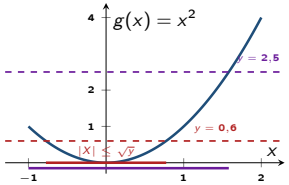
1. Identifica el **soporte** de Y (qué valores puede tomar).
2. Para cada y, traduce $\{g(X) \le y\}$ a un conjunto de valores de X y calcula $F_Y(y) = \mathbb{P}(g(X) \le y)$ con F_X o integrando f_X.
3. Deriva: $f_Y(y) = F_Y'(y)$. Chequea que integra 1.

Funciones lineales: $Y = aX + b$, $a \neq 0$

Si $a > 0$: $F_Y(y) = \mathbb{P}(X \le \frac{y-b}{a}) = F_X\big(\frac{y-b}{a}\big)$ y derivando,

$$f_Y(y) = \frac{1}{|a|} f_X\Big(\frac{y-b}{a}\Big) \quad \text{(vale también si } a < 0\text{)}.$$

- $X \sim \text{Unif}(0,1)$, $Y = 2X + 3$: $f_Y = \frac{1}{2}$ en $[3,5]$, es $\text{Unif}(3,5)$.
- $X \sim \mathcal{N}(\mu, \sigma^2)$: $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$; con $a = \frac{1}{\sigma}$, $b = -\frac{\mu}{\sigma}$ sale $Z \sim \mathcal{N}(0,1)$. Es la **demostración** de la estandarización.
- $X \sim \text{Exp}(\lambda)$, $Y = aX$ con $a > 0$: $\text{Exp}(\lambda/a)$. Pero $X + 1$ ya no es exponencial.



$X \sim \text{Unif}(-1,2)$, $Y = X^2$. El conjunto $\{g(X) \le y\}$ cambia de forma: para $y \le 1$ es $[-\sqrt{y}, \sqrt{y}]$; para $1 < y \le 4$ es $[-1, \sqrt{y}]$ (el brazo izquierdo ya se acabó). Por eso F_Y tiene **dos tramos**. Es el Problema 4.

Error clásico

"Sustituir": $f_Y(y) = f_X(g^{-1}(y))$ sin más. Falta el factor que mide cuánto **estira o comprime** g; y si g no es monótona, falta sumar los tramos.

Ejemplo completo: $X \sim \text{Unif}(0,1)$, $Y = X^2$ y $Y = -\ln X$

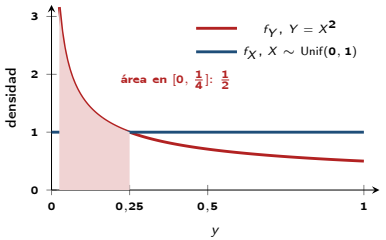
$Y = X^2$ (una densidad que explota)

Soporte $[0,1]$. $F_Y(y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(X \le \sqrt{y}) = \sqrt{y}$, luego

$$f_Y(y) = \frac{1}{2\sqrt{y}}, \qquad 0 < y \le 1.$$

Tiende a ∞ en 0 (el cuadrado aplasta los valores pequeños), pero $\int_0^1 \frac{dy}{2\sqrt{y}} = 1$.

$\mathbb{E}[Y] = \int_0^1 \frac{y}{2\sqrt{y}}\,dy = \frac{1}{3} = \mathbb{E}[X^2]$. ✓



La mitad de la masa de $Y = X^2$ cae en $[0, \frac{1}{4}]$ (porque $X \le \frac{1}{2}$ la mitad de las veces). La densidad de Y no se parece a la de X: transformar **deforma** la masa.

$Y = -\ln X$ (cómo se fabrican exponenciales)

Soporte $[0, \infty)$.
$F_Y(y) = \mathbb{P}(-\ln X \le y) = \mathbb{P}(X \ge e^{-y}) = 1 - e^{-y}$:
$Y \sim \text{Exp}(1)$. En general $-\frac{1}{\lambda}\ln U$ con U uniforme es $\text{Exp}(\lambda)$: así simula una computadora tiempos exponenciales (banco, Reto+).

Dos preguntas distintas

Si solo quieres $\mathbb{E}[g(X)]$, **no necesitas** f_Y: usa la regla del valor esperado $\int g(x) f_X(x)\,dx$. La densidad de Y hace falta para probabilidades de Y o para describirla por completo.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

La fórmula del caso monótono: $f_Y(y) = f_X(h(y))\,|h'(y)|$

Teorema

Si g es estrictamente monótona y derivable, con inversa $h = g^{-1}$, entonces

$$f_Y(y) = f_X(h(y))\left|\frac{dh}{dy}(y)\right|$$

en el soporte de Y. Demostración: g creciente $\Rightarrow F_Y(y) = F_X(h(y))$ y se deriva (regla de la cadena); g decreciente $\Rightarrow F_Y(y) = 1 - F_X(h(y))$, y el signo menos se come el $h' < 0$. Por eso el **valor absoluto**.

Por qué aparece $|h'|$: conservación de masa

La masa $f_X(x)\,dx$ de un trocito dx se transporta al trocito $dy = |g'(x)|\,dx$. Para que $f_Y(y)\,dy = f_X(x)\,dx$, la densidad se divide entre $|g'(x)|$, es decir, se multiplica por $|h'(y)|$. Donde g **comprime**, la densidad **sube**; donde estira, baja.

Ejemplos en una línea

$X \sim \text{Unif}(0,1)$, $Y = e^X$: $h = \ln y$, $f_Y = \frac{1}{y}$ en $[1, e]$. $X \sim \text{Exp}(1)$, $Y = e^X$: $f_Y = \frac{e^{-\ln y}}{y} = \frac{1}{y^2}$ en $[1, \infty)$ (Pareto: media infinita).
$Y = \ln X$: $f_Y(y) = f_X(e^y)\,e^y$.

Cuando g NO es monótona: partir y sumar

Para $Y = X^2$: $F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ y

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, \qquad y > 0.$$

Cada tramo monótono de g aporta un sumando (su propia fórmula monótona). Con $X \sim \mathcal{N}(0,1)$: $f_Y(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}}$ (chi-cuadrado con 1 grado de libertad).



$X \sim \text{Unif}(-1,2)$, $Y = X^2$: $f_Y = \frac{1}{3\sqrt{y}}$ en $(0,1]$ y $\frac{1}{6\sqrt{y}}$ en $(1,4]$. Usar la fórmula monótona con un solo brazo (morado) deja fuera un tercio de la masa.

Funciones de varias variables: máximo, mínimo, cociente

La misma receta: traducir $\{Z \le z\}$ a una región y medirla

Con X, Y **independientes**:

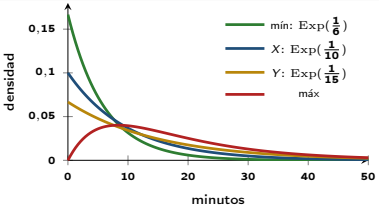
- **Máximo** $M = \max\{X, Y\}$: $\{M \le m\} = \{X \le m\} \cap \{Y \le m\}$, luego $F_M(m) = F_X(m)F_Y(m)$.
- **Mínimo** $L = \min\{X, Y\}$: $\{L > l\} = \{X > l\} \cap \{Y > l\}$, luego $\mathbb{P}(L > l) = \mathbb{P}(X > l)\mathbb{P}(Y > l)$.
- **Cociente, distancia, suma**: dibuja la región $\{(x, y) : g(x, y) \le z\}$ e integra la conjunta sobre ella (si es uniforme, es un **área**).

Dos uniformes independientes en $[0, 1]$

$F_M(m) = m^2$, $f_M = 2m$, $\mathbb{E}[M] = \frac{2}{3}$. $\mathbb{P}(L > l) = (1 - l)^2$, $f_L = 2(1 - l)$, $\mathbb{E}[L] = \frac{1}{3}$. Con $Z = X/Y$: $F_Z(z) = \frac{z}{2}$ si $z \le 1$ y $1 - \frac{1}{2z}$ si $z > 1$; $\mathbb{E}[Z] = \infty$. Con n uniformes: $F_M = m^n$, $\mathbb{E}[M] = \frac{n}{n+1}$.

Exponenciales: el mínimo suma tasas

$X \sim \mathrm{Exp}(\lambda_1)$, $Y \sim \mathrm{Exp}(\lambda_2)$ independientes: $\mathbb{P}(L > l) = e^{-\lambda_1 l} e^{-\lambda_2 l}$, así que $L \sim \mathrm{Exp}(\lambda_1 + \lambda_2)$ y $\mathbb{P}(X < Y) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. El máximo **no** es exponencial. Es el Problema 5.



Dos combis con medias 10 y 15 min. El mínimo (verde) es exponencial con media 6; el máximo (rojo) arranca en 0, sube y tiene media 19: ya no tiene falta de memoria.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

Cómo trabajaremos los cinco problemas

El formato

Cada problema ocupa un slide casi en blanco: lo **desarrollamos juntos en la pizarra**. La solución escrita, paso a paso, se comparte al **final de la clase** (versión completa de estos slides en OneDrive): hoy, atención a la pizarra.

Los cinco problemas y la sección de teoría que usan

1. Romper un palo: **densidad condicional, marginal y esperanza total** (sección 1).
2. El sesgo de una moneda: **Bayes con desconocido continuo** y observación discreta (sección 3).
3. Un bit entre ruido: **Bayes con desconocido discreto** y observación normal; decidir y equivocarse (secciones 2 y 3).
4. La densidad de $Y = g(X)$: **método de la CDF, fórmula monótona y caso no monótono** (secciones 4 y 5).
5. Dos combis: **mínimo y máximo de exponenciales independientes** (secciones 2 y 5).

Antes de cada uno

Dibuja el soporte (o la gráfica de g), decide **qué es la conjunta y qué es la rebanada**, y escribe la pregunta como una probabilidad de un conjunto antes de integrar nada.

Problema 1: romper un palo (densidad condicional)

Adaptado de "Stick-breaking" (Lecture 10, MIT 6.431x). Un palo de longitud 1 se rompe en un punto $X \sim \mathrm{Unif}(0,1)$. El trozo izquierdo, de largo X, se vuelve a romper en un punto Y uniforme en $(0, X)$.

(a) Escribe $f_{X,Y}(x,y)$ con su soporte, halla la marginal $f_Y(y)$ y calcula $\mathbb{E}[Y]$ de dos maneras.

(b) Halla $f_{X \mid Y}(x \mid y)$ y $\mathbb{E}[X \mid Y = y]$. ¿Dónde estuvo, en promedio, el primer corte si el segundo cayó en $y = \frac{2}{3}$?

Problema 2: el sesgo de una moneda (Bayes con desconocido continuo)

Adaptado de "Inference of the bias of a coin" (Lecture 10, MIT 6.431x). La probabilidad de cara Q de una moneda es desconocida; a priori $Q \sim \text{Unif}(0,1)$. Dado $Q = q$, los tiros son independientes con probabilidad q de cara. Se lanza $n = 3$ veces y salen $k = 2$ caras.

a. Halla la posterior $f_{Q \mid K}(q \mid 2)$ y su moda (el estimador MAP).

b. Calcula la media posterior $\mathbb{E}[Q \mid K = 2]$ y la probabilidad de que el **cuarto** tiro sea cara. Compara con la frecuencia observada $\frac{2}{3}$.

Problema 3: un bit entre ruido (Bayes con desconocido discreto y observación continua)

Se transmite un bit $X \in \{+1, -1\}$ con $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$. El receptor observa $Y = X + W$, donde el ruido $W \sim \mathcal{N}(0, 1)$ es independiente de X.

a. Halla $\mathbb{P}(X = 1 \mid Y = y)$ y simplifícala. Evalúa en $y = 0{,}5$ y en $y = 2$.

b. El receptor declara "1" cuando $\mathbb{P}(X = 1 \mid Y = y) > \frac{1}{2}$. ¿Qué regla resulta? Calcula la probabilidad de error. ¿Cómo cambia si $\sigma_W = \frac{1}{2}$?

Problema 4: la densidad de $Y = g(X)$ (monótono y no monótono)

a Sea $X \sim \text{Unif}(0, 1)$ e $Y = \dfrac{X}{1 - X}$. Halla f_Y por el método de la CDF y vuelve a hallarla con la fórmula del caso monótono. Calcula la mediana de Y y $\mathbb{E}[Y]$.

b Sea $X \sim \text{Unif}(-1, 2)$ e $Y = X^2$. Halla F_Y y f_Y (cuidado: g no es monótona y hay dos tramos). Comprueba que f_Y integra 1 y que $\mathbb{E}[Y] = \mathbb{E}[X^2]$.

Problema 5: dos combis (mínimo y máximo de exponenciales independientes)

En un paradero pasan dos líneas de combi que te sirven igual. El tiempo hasta la siguiente de la línea 1 es $X \sim \mathrm{Exp}(\lambda_1)$ con media 10 min; el de la línea 2 es $Y \sim \mathrm{Exp}(\lambda_2)$ con media 15 min; X e Y son independientes.

a. Sea $L = \min\{X, Y\}$ tu tiempo de espera. Halla $\mathbb{P}(L > t)$, identifica la distribución de L y calcula $\mathbb{E}[L]$. Calcula $\mathbb{P}(X < Y)$: la probabilidad de que la primera en pasar sea la de la línea 1.

b. Sea $M = \max\{X, Y\}$ el instante en que ya pasaron ambas. Halla F_M y f_M, calcula $\mathbb{P}(M \leq 10)$ y $\mathbb{E}[M]$ por el camino más corto.

Contenido

1. Condicionar en una variable aleatoria
2. Independencia de variables aleatorias continuas
3. La regla de Bayes en cuatro versiones
4. Distribuciones derivadas: funciones lineales y el método de la CDF
5. El caso monótono, el no monótono y funciones de varias variables
6. Cinco problemas para resolver juntos
7. Cierre

Resumen de la Clase 7

Lo que hay que saber hacer

1. **Densidad condicional** $f_{X\mid Y} = f_{X,Y}/f_Y$: cortar y renormalizar; regla de multiplicación, probabilidad total y **esperanza total** con integrales.
2. **Independencia**: $f_{X,Y} = f_X f_Y$ en todo el plano; mirar primero el soporte; consecuencias ($\mathbb{E}[XY]$, varianzas, CDFs); normales independientes (completar cuadrados).
3. **Bayes en cuatro versiones**: posterior $\propto$ prior $\times$ verosimilitud, con PMFs o densidades; el denominador es la probabilidad total. El sesgo de la moneda ($\propto q^k(1-q)^{n-k}$, MAP $\frac{k}{n}$, media $\frac{k+1}{n+2}$) y el bit entre ruido (sigmoide, umbral, error con Φ).
4. **Distribuciones derivadas**: funciones lineales ($\frac{1}{|a|} f_X(\frac{v-b}{a})$), el **método de la CDF** (siempre), la fórmula del **caso monótono** $f_X(h(y))|h'(y)|$ y partir en tramos cuando g no es monótona.
5. **Varias variables**: máximo $\to$ producto de CDFs; mínimo $\to$ producto de colas; cociente y distancia $\to$ regiones. Mínimo de exponenciales = exponencial con la suma de tasas.

Errores clásicos de esta unidad

Confundir la condicional con la conjunta · declarar independencia sin mirar todo el soporte · usar la verosimilitud como si fuera la posterior · "sustituir" sin el jacobiano · olvidar el valor absoluto · aplicar la fórmula monótona a x^2 · promediar medias para el mínimo.

Siguiente clase (Clase 8)

Unit 6, L12 y L13: suma de independientes (convolución), covarianza y correlación, y la esperanza condicional como variable aleatoria (ley de la varianza total). Con eso se cierra el material del **Midterm 2**.

¡Gracias!

Clase 7 — Bayes continuo, independencia y distribuciones derivadas (6.431x, Units 5 y 6)

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca · en alianza con MIT IDSS

Basado en MIT 6.431x / MIT OCW RES.6-012 (Bertsekas & Tsitsiklis), CC BY-NC-SA 4.0.