Repaso de Probabilidad

Germán Capdehourat, Sergio Martínez, Pablo Musé
{gcapde, sematag, pmuse}@fing.edu.uy

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

4 de agosto de 2022

Desigualdades de Markov y Chebyshev

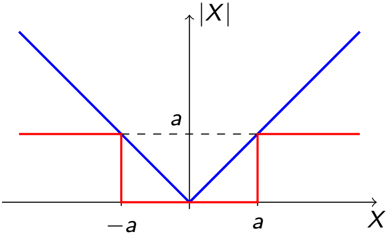
Desigualdad de Markov

- X VA tal que $\mathbb{E}\left[|X|\right] < \infty$, $a > 0$ constante
- Desigualdad de Markov: $\mathrm{P}\left(|X| \geq a\right) \leq \dfrac{\mathbb{E}(|X|)}{a}$

Demostración.

- $\mathbb{I}\left\{|X| \geq a\right\} = 1$ si $|X| \geq a$, 0 si no.
  Entonces (figura de la derecha)

$$a\,\mathbb{I}\left\{|X| \geq a\right\} \leq |X|$$



- Linealidad de la esperanza

$$a\,\mathbb{E}(\mathbb{I}\left\{|X| \geq a\right\}) \leq \mathbb{E}(|X|)$$

- Esperanza de función indicatriz = Probab. del suceso indicado

$$a\,\mathrm{P}\left(|X| \geq a\right) \leq \mathbb{E}(|X|)$$

$\square$

Desigualdad de Chebyshev

- X VA con $\mathbb{E}(X) = \mu$ y $\mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$, $k > 0$ constante
- Desigualdad de Chebyshev: $\mathrm{P}\left(|X-\mu| \geq k\right) \leq \dfrac{\sigma^2}{k^2}$

Demostración.

- Desigualdad de Markov para la VA $Z = (X-\mu)^2$ y $a = k^2$ constante

$$\mathrm{P}\left((X-\mu)^2 \geq k^2\right) = \mathrm{P}\left(|Z| \geq k^2\right) \leq \frac{\mathbb{E}\left[|Z|\right]}{k^2} = \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2}$$

- Notar que $(X-\mu)^2 \geq k^2$ si y solo si $|X-\mu| \geq k$, entonces

$$\mathrm{P}\left(|X-\mu| \geq k\right) \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2}$$

- La desigualdad de Chebyshev sigue de la definición de varianza

$\square$

Convergencia de variables aleatorias

Límites

- Sea la suceción de VAs $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$

  $\Rightarrow$ Distinguimos las VAs $X_{\mathbb{N}}$ de sus realizaciones $x_{\mathbb{N}}$

Q1) ¿Podemos decir algo sobre X_n para n grande? $\Rightarrow$ No claro, X_n es VA

Q2) ¿Podemos decir algo sobre x_n para n grande? $\Rightarrow$ Sí, $\lim_{n\to\infty} x_n$

Q3) ¿Podemos decir algo sobre $\mathrm{P}\left(X_n \in \mathcal{X}\right)$ para n grande? $\Rightarrow$ Sí, $\lim_{n\to\infty} \mathrm{P}\left(X_n \in \mathcal{X}\right)$

- Queremos traducir lo que sabemos sobre límites convencionales a definiciones para VAs
- Comenzamos por la convergencia de la sucesión: $\lim_{n\to\infty} x_n$

  $\Rightarrow$ Convergencia segura y casi segura
- Luego: Convergencia en probabilidad, en media cuadrática y en distribución

Convergencia de sucesiones y convergencia segura

- Denotemos como sucesión de números $x_{\mathbb{N}} = x_1, x_2, \ldots, x_n, \ldots$
- **Def:** La sucesión $x_{\mathbb{N}}$ converge al valor x si dado $\epsilon > 0$

  $\Rightarrow$ existe n_0 tal que para todo $n > n_0$, $|x_n - x| < \epsilon$
- La sucesión x_n se acerca arbitrariamente a su límite $\Rightarrow |x_n - x| < \epsilon$

  $\Rightarrow$ Además permanece cerca de su límite para todo $n > n_0$
- Proceso estocástico (sucesión de VAs) $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$

  $\Rightarrow$ Las realizaciones de $X_{\mathbb{N}}$ son sucesiones $x_{\mathbb{N}}$
- **Def:** Decimos que $X_{\mathbb{N}}$ tiene convergencia segura a la VA X si

  $\Rightarrow \lim_{n\to\infty} X_n = X$ para toda realización $x_{\mathbb{N}}$ of $X_{\mathbb{N}}$
- Dicho de otra manera, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ para todo $\omega \in \Omega$
- No es un criterio adecuado: aún realizaciones (sin importancia) que suceden con probabilidad infinitamente pequeña impiden la convergencia segura

Convergencia casi segura

- Sean la VA X y el proceso estocástico $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- **Def:** Decimos que $X_{\mathbb{N}}$ converge casi seguramente a la VA X si

$$\mathrm{P}\left(\lim_{n\to\infty} X_n = X\right) = \mathrm{P}\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

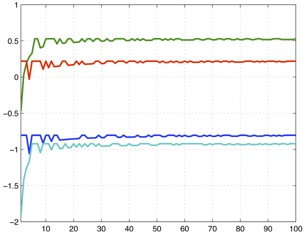
  $\Rightarrow$ Casi todas las sucesiones convergen, a excepción de un conjunto de probabilidad nula

- Denotamos la convergencia casi segura por $\Rightarrow \lim_{n\to\infty} X_n = X \quad$ c.s.

  $\Rightarrow$ El límite X es una VA

Ejemplo

- $X_0 \sim \mathcal{N}(0,1)$ (normal, media 0, varianza 1)
- Z_n sucesión de VAs de Bernoulli de parámetro p
- Definimos $\Rightarrow X_n = X_0 - \dfrac{Z_n}{n}$
- $\dfrac{Z_n}{n} \to 0$ por lo que $\lim_{n\to\infty} X_n = X_0$ c.s. (tambien seguramente)



Ejemplo de convergencia casi segura

- Consideremos $\Omega = [0, 1]$ y sea $\mathrm{P}(\cdot)$ la distribución uniforme

  $\Rightarrow \mathrm{P}([a, b]) = b - a,\ 0 \leq a \leq b \leq 1$

- Definimos las VAs $X_n(\omega) = \omega + \omega^n$ y $X(\omega) = \omega$

- Para todo $\omega \in [0, 1)$ $\Rightarrow \omega^n \to 0$ para $n \to \infty$, de donde $X_n(\omega) \to \omega = X(\omega)$

- Para $\omega = 1$ $\Rightarrow X_n(1) = 2$ para todo n, mientras que $X(1) = 1$

- La convergencia ocurre únicamente en el conjunto $[0, 1)$, y $\mathrm{P}([0, 1)) = 1$

  $\Rightarrow$ Decimos que $\lim_{n \to \infty} X_n = X$ c.s.

  $\Rightarrow$ Una vez más, nótese que el límite X es una VA

Convergencia en probabilidad

- **Def:** Decimos que $X_{\mathbb{N}}$ converge en probabilidad a la VA X si para todo $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left(|X_n - X| < \epsilon\right) = \lim_{n\to\infty} \mathrm{P}\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right) = 1$$

$\Rightarrow$ La probabilidad que $|X_n - X|$ se vuelva menor a ϵ tiende a 1

- La afirmación es sobre probabilidades, no sobre realizaciones

$\Rightarrow$ Si hay convergencia en probabilidad, las realizaciones $x_{\mathbb{N}}$ pueden o no converger

Teorema

La convergencia casi segura (c.s.) implica la convergencia en probabilidad

Demostración.

Si $\lim_{n\to\infty} X_n = X$ c.s. entonces dado $\epsilon > 0$, para todo $\omega \in \Omega$ salvo un conjunto de probabilidad nula, existe $n = n(\epsilon, \omega)$ tal que $|X_k(\omega) - X(\omega)| < \epsilon$ para todo $k \geq n$. Es decir, $P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} |X_k(\omega) - X(\omega)| < \epsilon) = 1$. Tomando el complemento, $P(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} |X_k(\omega) - X(\omega)| \geq \epsilon) = 0 \Rightarrow$ como los $E_n = \{\cup_{k=n}^{\infty} |X_k(\omega) - X(\omega)| \geq \epsilon\}$ verifican $E_1 \supset E_2 \supset \ldots$, entonces $\lim_{n\to\infty} \mathrm{P}(E_n) = \mathrm{P}\left(\cap_{n=1}^{\infty} E_n\right) = 0$. De aquí se obtiene que $\mathrm{P}(|X_n - X| \geq \epsilon) \leq \mathrm{P}\left(\cup_{k=n}^{\infty} |X_k - X| \geq \epsilon\right) = \mathrm{P}(E_n) \to 0$ si $n \to \infty$. $\square$

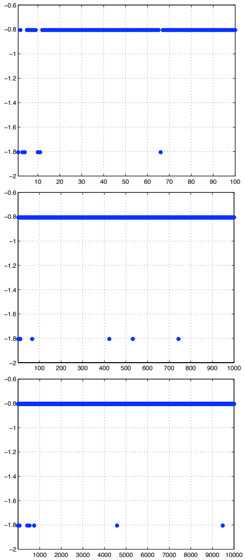
Ejemplo de convergencia en probabilidad

- $X_0 \sim \mathcal{N}(0,1)$
- Z_n sucesión de VAs Bernoulli, parámetro $1/n$
- Definimos $\Rightarrow X_n = X_0 - Z_n$
- X_n converge en probabilidad a X_0 ya que

$$\begin{aligned} \mathrm{P}\left(|X_n - X_0| < \epsilon\right) &= \mathrm{P}\left(|Z_n| < \epsilon\right) \\ &= 1 - \mathrm{P}\left(Z_n = 1\right) \\ &= 1 - \frac{1}{n} \to 1 \end{aligned}$$

- Gráficas: realizaciones de x_n hasta to $n = 10^2$, $n = 10^3$, $n = 10^4$

  $\Rightarrow Z_n = 1$ se vuelve raro pero igual sucede



Diferencia entre convergencia c.s. y en probabilidad

- La convergencia c.s. implica que casi todas las secuencias convergen
- La convergencia en probabilidad no implica la convergencia de las sucesiones
- Ejemplo anterior: $X_n = X_0 - Z_n$, Z_n Bernoulli de parámetro $1/n$

  $\Rightarrow$ Mostramos que converge en probabilidad

$$\mathrm{P}\left(|X_n - X_0| < \epsilon\right) = 1 - \frac{1}{n} \to 1$$

  $\Rightarrow$ Pero para casi todas las sucesiones, $\lim_{n\to\infty} x_n$ no existe

- Convergencia casi segura $\Rightarrow$ las perturbaciones dejan de suceder
- Convergencia en probabilidad $\Rightarrow$ las perturbaciones suceden, pero su frecuencia tiende a cero

Convergencia en r-media

Def: Supongamos que las $X_{\mathbb{N}}$ tienen momento de orden $r \geq 1$ finito. Decimos que $X_{\mathbb{N}}$ converge en r-media a la VA X si

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0$$

- $r = 1$: conv. en media; $r = 2$: conv. en media cuadrática (m.c.)
- A veces es (muy) fácil de comprobar
- Es fácil ver que si $r > s \geq 1$: conv. en s-media $\Rightarrow$ conv. en r-media

Teorema

La convergencia en r-media implica la convergencia en probabilidad

Demostración.

- De la desigualdad de Markov

$$\mathrm{P}\left(|X_n - X| \geq \epsilon\right) = \mathrm{P}\left(|X_n - X|^r \geq \epsilon^r\right) \leq \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\epsilon^r}$$

- Si $X_n \to X$ en r-media, $\mathbb{E}\left[|X_n - X|^r\right]/\epsilon^r \to 0$ para todo $\epsilon > 0$

$\square$

Diferencia entre convergencia casi segura y en r-media

- Convergencia casi segura y en r-media: ninguna implica la otra
- Ejemplo (convergencia c.s y no en media):

$$X_n = \begin{cases} n^3 & \text{con probabilidad } 1/n^2, \\ 0 & \text{con probabilidad } 1 - 1/n^2. \end{cases}$$

- Ejemplo (convergencia en media y no c.s.): X_n Bernoulli de parámetro $1/n$.

Convergencia en distribución

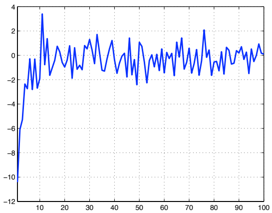
- Consideremos el proceso estocástico $X_{\mathbb{N}}$. La cdf de X_n es $F_n(x)$
- **Def:** decimos que $X_{\mathbb{N}}$ converge en distribución a la VA X con cdf $F_X(x)$ si

  $\Rightarrow \lim_{n\to\infty} F_n(x) = F_X(x)$ para todo x en el que $F_X(x)$ es continua
- No se afirma nada sobre las sucesiones individualmente, sólo sobre la cdf de X_n

  $\Rightarrow$ Es la convergencia más débil entre las que vimos
- Las convergencias c.s., en probabilidad y en r-media implican convergencia en distribución

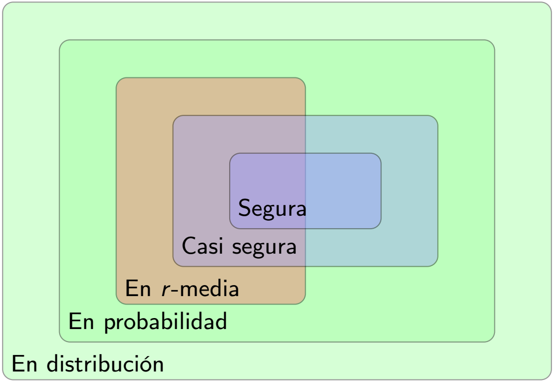
Ejemplo

- $Y_n \sim \mathcal{N}(0,1)$
- Z_n Bernoulli de parámetro p
- Definimos $\Rightarrow X_n = Y_n - Z_n/n$
- $\frac{Z_n}{n} \to 0 \Rightarrow \lim_{n\to\infty} F_n(x)$ "=" $\mathcal{N}(0,1)$



Relaciones de implicancia

- Segura $\Rightarrow$ casi segura $\Rightarrow$ en probabilidad $\Rightarrow$ en distribución
- En r-media $\Rightarrow$ en probabilidad $\Rightarrow$ en distribución



Teoremas límite

Ley de los grandes números

- $X_1, X_2, \ldots, X_n, \ldots$ sucesión de VAs i.i.d. de media μ
- Definimos la media muestral como $\bar{X}_N := (1/N) \sum_{n=1}^{N} X_n$

Teorema (Ley débil de los grandes números)

La media muestral $\bar{X}_N$ de la sucesión i.i.d. converge en prob. a $\mu = \mathbb{E}[X_n]$

$$\lim_{N\to\infty} \mathrm{P}\left(|\bar{X}_N - \mu| < \epsilon\right) = 1, \qquad \textit{for all } \epsilon > 0$$

Teorema (Ley fuerte de los grandes números)

La media muestral $\bar{X}_N$ de la sucesión i.i.d. converge c.s. a $\mu = \mathbb{E}[X_n]$

$$\mathrm{P}\left(\lim_{N\to\infty} \bar{X}_N = \mu\right) = 1$$

- La ley fuerte implica la ley débil

Teorema central del límite (TCL)

Teorema (Teorema central del límite)

Sea $X_1, X_2, \ldots, X_n, \ldots$ una sucesión de VAs i.i.d. de media $\mathbb{E}[X_n] = \mu$ y varianza $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ para todo n. Entonces

$$\lim_{N\to\infty} \mathrm{P}\left(\frac{\sum_{n=1}^{N} X_n - N\mu}{\sigma\sqrt{N}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

- Esto significa que para N suficientemente grande,

$$Z_N := \frac{\sum_{n=1}^{N} X_n - N\mu}{\sigma\sqrt{N}} \sim \mathcal{N}(0,1)$$

$\Rightarrow$ Z_N converge en distribución a una VA normal estándar

$\Rightarrow$ Universalidad notable! La distribución de los X_n es arbitraria

Teorema central del límite (TCL) (cont.)

- Equivale a decir que $\Rightarrow \sum_{n=1}^{N} X_n \sim \mathcal{N}(N\mu, N\sigma^2)$
- La suma de un gran número de VAs i.i.d. tiene distribución normal
  - $\Rightarrow$ Aquí tomar límite no tiene sentido
  - $\Rightarrow$ Pero intuitivamente, esto es lo que dice el TCL

Ejemplo

- X VA binomial de parámetros (n, p)
- Escribimos $X = \sum_{i=1}^{n} X_i$ con X_i VAs Bernoulli de parámetro p, i.i.d.
- Media $\mathbb{E}[X_i] = p$, varianza $\mathrm{var}[X_i] = p(1-p)$
  - $\Rightarrow$ Para n suficientemente grande $\Rightarrow X \sim \mathcal{N}(np, np(1-p))$

Probabilidades condicionales

pmf y cdf condicionales para VAs discretas

- Recordemos la definición de probabilidad condicional para sucesos E y F

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

- **Def:** La pmf condicional de la VA X dada Y, ambas discretas, es

$$p_{X|Y}(x \mid y) = \mathrm{P}\left(X = x \mid Y = y\right) = \frac{\mathrm{P}\left(X = x, Y = y\right)}{\mathrm{P}\left(Y = y\right)}$$

- Se puede reescribir como

$$p_{X|Y}(x \mid y) = \frac{\mathrm{P}\left(X = x, Y = y\right)}{\mathrm{P}\left(Y = y\right)} = \frac{p_{XY}(x, y)}{p_Y(y)}$$

$\Rightarrow$ pmf de la VA X, dado el parámetro y ("Y no es más aleatoria", tomó el valor y)

- **Def:** La cdf condicional es (un rango de X condicionado a un valor of Y)

$$F_{X|Y}(x \mid y) = \mathrm{P}\left(X \leq x \mid Y = y\right) = \sum_{z \leq x} p_{X|Y}(z \mid y)$$

pdf y cdf condicionales para VAs continuas

- **Def:** La pdf condicional de X dado Y, ambas continuas, es

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

- A modo de motivación, definamos los intervalos $\Delta x = [x, x+dx]$ y $\Delta y = [y, y+dy]$

  $\Rightarrow$ Aproximamos la prob. condicional $\mathrm{P}\left(X \in \Delta x \,\middle|\, Y \in \Delta y\right)$:

$$\mathrm{P}\left(X \in \Delta x \,\middle|\, Y \in \Delta y\right) = \frac{\mathrm{P}\left(X \in \Delta x, Y \in \Delta y\right)}{\mathrm{P}\left(Y \in \Delta y\right)} \approx \frac{f_{XY}(x,y)dxdy}{f_Y(y)dy}$$

- De la definición de pdf condicional surge que

$$\mathrm{P}\left(X \in \Delta x \,\middle|\, Y \in \Delta y\right) \approx f_{X|Y}(x \,|\, y)dx$$

  $\Rightarrow$ Es lo que esperamos de una densidad

- **Def:** La cdf condicional es $\Rightarrow F_{X|Y}(x) = \int_{-\infty}^{x} f_{X|Y}(u \,|\, y)du$

Ejemplo: modelo simple de canal de comunicación

- Mensaje aleatorio (VA) Y, transmitimos la señal y (realización de Y)
- La señal recibida es $x = y + z$ (z realización de ruido aleatorio)

  $\Rightarrow$ Modelamos un sistema de comunicación como la relación entre VAs

$$X = Y + Z$$

  $\Rightarrow$ Modelamos el ruido aditivo como $Z \sim \mathcal{N}(0, \sigma^2)$, independiente de Y

- Q: ¿Cómo es la pdf condicional de X dado Y? Probemos con la definición

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

  $\Rightarrow$ El problema es que no conocemos $f_{XY}(x, y)$. Debemos calcularla

- En general es más fácil calcular o estimar probabilidades condicionales que conjuntas

Ejemplo: modelo simple de canal de comunicación (cont.)

- Si $Y = y$ está dado, entonces "Y no es más aleatoria"

  $\Rightarrow$ En realidad sigue siendo aleatoria, pero la pensamos como dada

- Si Y no fuese aleatoria, digamos $Y = y$ con y dada por $X = y + Z$

  $\Rightarrow$ La cdf de X dado $Y = y$ ahora es fácil (usamos independencia de Y y Z)

$$\mathrm{P}\left(X \leq x \,\middle|\, Y = y\right) = \mathrm{P}\left(y + Z \leq x | Y = y\right) = \mathrm{P}\left(Z \leq x - y\right)$$

- Pero como Z es normal de media nula y varianza σ^2

$$\begin{aligned}
\mathrm{P}\left(X \leq x \,\middle|\, Y = y\right) &= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x-y} e^{-z^2/2\sigma^2}\, dz \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x} e^{-(z-y)^2/2\sigma^2}\, dz
\end{aligned}$$

$\Rightarrow$ $[X$ dado $Y = y]$ es normal de media y y varianza σ^2

Esperanza condicional

Definición de esperanza condicional

- **Def:** Para las VAs continuas X e Y, la esperanza condicional es

$$\mathbb{E}\left[X \,\middle|\, Y = y\right] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

- **Def:** Para las VAs discretas X, Y, la esperanza condicional es

$$\mathbb{E}\left[X \,\middle|\, Y = y\right] = \sum_{x} x\, p_{X|Y}(x|y)$$

- Definida para un y dado $\Rightarrow \mathbb{E}\left[X \,\middle|\, Y = y\right]$ is un número

  $\Rightarrow$ Si consideramos que Y tomar valores y $\Rightarrow \mathbb{E}\left[X \,\middle|\, Y\right]$ es VA

- $\mathbb{E}\left[X \,\middle|\, Y\right]$ es una función de la VA Y, por lo tanto es una VA

  $\Rightarrow \mathbb{E}\left[X \,\middle|\, Y = y\right]$ valor asociado a la realización $Y = y$

- Si X e Y son independientes, entonces $\mathbb{E}\left[X \,\middle|\, Y\right] = \mathbb{E}\left[X\right]$

Esperanzas iteradas

- Como $\mathbb{E}\left[X \mid Y\right]$ es VA, podemos calcular su valor esperado $\mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right]$. Los subíndices especifican la variable sobre la que se calcula la esperanza
- Q: ¿Cuánto vale $\mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right]$? $\Rightarrow \mathbb{E}\left[X\right] = \mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right]$ (razonable)
- Lo demostramos para VAs discretas (VAs continuas, usar integrales)

$$
\begin{aligned}
\mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right] &= \sum_y \mathbb{E}_X\left[X \mid Y = y\right] p_Y(y) = \sum_y \left[\sum_x x\, p_{X|Y}(x|y)\right] p_Y(y) \\
&= \sum_x x \left[\sum_y p_{X|Y}(x|y) p_Y(y)\right] = \sum_x x \left[\sum_y p_{XY}(x,y)\right] \\
&= \sum_x x p_X(x) = \mathbb{E}\left[X\right]
\end{aligned}
$$

- Ofrece un método útil para calcular valores esperados

$\Rightarrow$ Condicionar por $Y = y$ $\qquad \Rightarrow X \mid Y = y$

$\Rightarrow$ Calcular valor esperado sobre X dado $Y = y$ $\qquad \Rightarrow \mathbb{E}_X\left[X \mid Y = y\right]$

$\Rightarrow$ Calcular valor esperado sobre todos los y de Y $\qquad \Rightarrow \mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right]$