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

5 de agosto de 2021

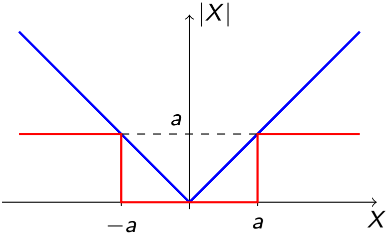
Desigualdad de Markov

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

Demostración.

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

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



- Linealidad de la esperanza

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

- Esperanza de función indicatriz = Probab. del suceso indicado

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

□

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

$\square$

Comentarios y observaciones

- Si el valor absoluto esperado es finito, i.e., $\mathbb{E}\left[|X|\right] < \infty$

  $\Rightarrow$ La cdf complementaria (ccdf) $\mathrm{P}\left(|X| \geq x\right)$ decae al menos como x^{-1} (Markov)

- Si la media $\mathbb{E}(X)$ y la varianza $\mathbb{E}\left[(X-\mu)^2\right]$ son finitas

  $\Rightarrow$ La ccdf decrece al menos como x^{-2} (Chebyshev)

- La mayor parte de las ccdfs decrecen exponencialmente (e.g. e^{-x^2} para normales)

  $\Rightarrow$ Las cotas $\propto x^{-\alpha}$ son débiles pero pueden ser útiles

- Muchas veces la desigualdad de Markov se utiliza para VAs no negativas $X \geq 0$

  $\Rightarrow$ Podemos dejar de lado el valor absoluto: $\mathrm{P}\left(X \geq a\right) \leq \frac{\mathbb{E}(X)}{a}$

  $\Rightarrow$ Cota general: $\mathrm{P}\left(X \geq a\right) \leq \frac{\mathbb{E}(X^r)}{a^r}$ para cualquier $r > 0$

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 $x_{\mathbb{N}}$ de $X_{\mathbb{N}}$
- 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

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

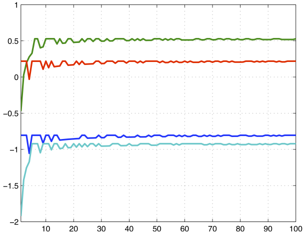
$\Rightarrow$ Casi todas las sucesiones convergen, a excepción de un conjunto de probabilidad nula

- 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 $\lim\limits_{n\to\infty} X_n = X$ 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 $\lim_{n\to\infty} \mathrm{P}\left(E_n\right) = \mathrm{P}\left(\cap_{n=1}^{\infty} E_n\right) = 0$. De aquí se obtiene que $\mathrm{P}\left(|X_n - X| \geq \epsilon\right) \leq \mathrm{P}\left(\cup_{k=n}^{\infty} |X_k - X| \geq \epsilon\right) = \mathrm{P}\left(E_n\right) \to 0$ si $n \to \infty$. $\square$

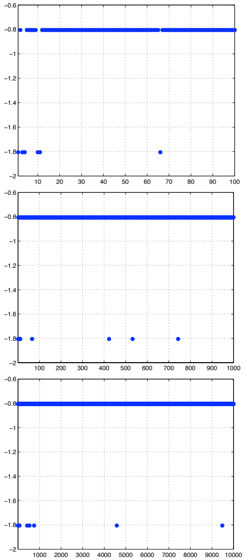
Ejemplo de convergencia en probabilidad

- $X_0 \sim \mathcal{N}(0,1)$
- 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
- La diferencia no es menor
  - Interpretemos Z_n como la tasa de cambios en los ahorros
  - Con la convergencia c.s. el riesgo se elimina
  - Con la convergencia en prob. el riesgo decrece pero no desaparece

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

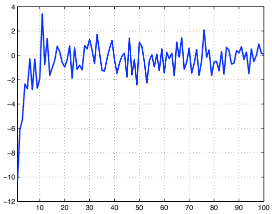
- No se afirma nada sobre las sucesiones individualmente, sólo sobre la cdf de X_n

  $\Rightarrow$ Es la convergencia más débil entre las que vimos

- Las convergencias c.s., en probabilidad y en r-media implican convergencia en distribución

Ejemplo

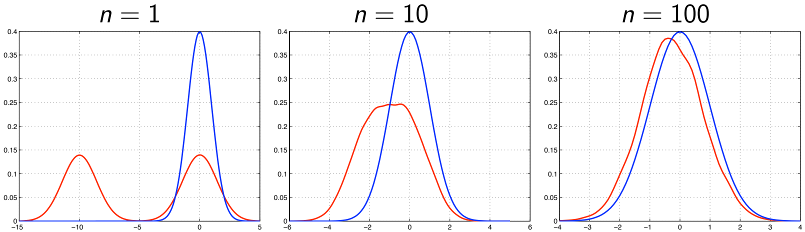
- $Y_n \sim \mathcal{N}(0,1)$
- Z_n Bernoulli de parámetro p
- Definimos $\Rightarrow X_n = Y_n - Z_n/n$
- $\dfrac{Z_n}{n} \to 0 \Rightarrow \lim_{n\to\infty} F_n(x)$ "$=$" $\mathcal{N}(0,1)$



Convergencia en distribución (cont.)

- Las sucesiones individuales x_n no convergen en ningún sentido

  $\Rightarrow$ Es la función de distribución quien converge

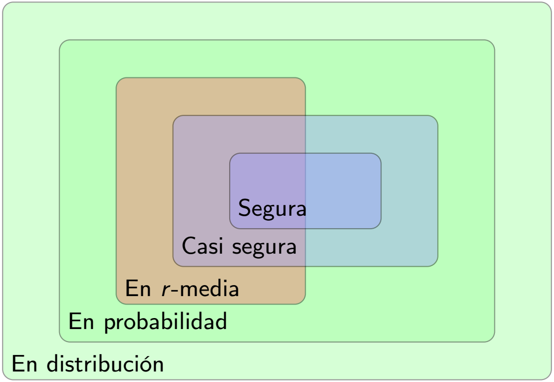


- A medida que el efecto de Z_n/n disminuye, la pdf de X_n converge a la pdf de Y_n

  $\Rightarrow$ Normal estándar $\mathcal{N}(0, 1)$

Relaciones de implicancia

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



Teoremas límite

Desigualdades de Markov y Chebyshev

Convergencia de variables aleatorias

Teoremas límite

Probabilidades condicionales

Esperanza condicional

Suma de VAs independientes e idénticamente distribuídas

- $X_1, X_2, \ldots, X_n, \ldots$ VAs independientes, idénticamente distribuídas (i.i.d.)
- Media $\mathbb{E}[X_n] = \mu$ y varianza $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ para tod n
- Q: ¿Qué sucede con la suma $S_N := \sum_{n=1}^{N} X_n$ cuando N crece?
- El valor esperado de la suma es $\mathbb{E}[S_N] = N\mu \;\Rightarrow$ Diverge si $\mu \neq 0$
- La varianza es $\mathbb{E}\left[(S_N - N\mu)^2\right] = N\sigma^2 \;\Rightarrow$ Diverge si $\sigma \neq 0$
- Una normalización interesantes es $\Rightarrow \bar{X}_N := (1/N) \sum_{n=1}^{N} X_n$
- Con esto $\mathbb{E}\left[\bar{X}_N\right] = \mu$ y var $\left[\bar{X}_N\right] = \sigma^2/N$

  $\Rightarrow$ Ley de los grandes números (débil y fuerte)
- Otra normalización interesante es $\Rightarrow Z_N := \dfrac{\sum_{n=1}^{N} X_n - N\mu}{\sigma\sqrt{N}}$
- Con esto $\mathbb{E}[Z_N] = 0$ and var $[Z_N] = 1$ para todo N

  $\Rightarrow$ Teorema central del 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

Prueba de la ley débil de los grandes números

- La ley débil de los grandes números es muy fácil de demostrar

Demostración.

- La varianza de $\bar{X}_N$ se decae a cero para N grande

$$\mathrm{var}\left[\bar{X}_N\right] = \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{var}\left[X_n\right] = \frac{\sigma^2}{N} \to 0$$

- Como

$$0 \leftarrow \frac{\sigma^2}{N} = \mathrm{var}\left[\bar{X}_N\right] = \mathbb{E}\left[(\bar{X}_N - \mu)^2\right]$$

tenemos que $\bar{X}_N$ converge a μ en media cuadrática

$\Rightarrow$ Entonces $\bar{X}_N$ converge a μ en probabilidad $\square$

- La ley fuerte es más difícil de demostrar (no lo haremos aquí)

Recapitulando

- Serie de experimentos $\Rightarrow$ Sucesión de VAs $X_1, X_2, \ldots, X_n, \ldots$

  $\Rightarrow$ Consideramos un suceso de interés $X \in E$. Ej: obtener "cara" al tirar una moneda
- La fracción de veces que $X \in E$ sucede en N experimentos es

$$\bar{X}_N = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}\left\{X_n \in E\right\}$$

- Como las indicatrices son también i.i.d., la ley fuerte asegura que

$$\lim_{N\to\infty} \bar{X}_N = \mathbb{E}\left[\mathbb{I}\left\{X_1 \in E\right\}\right] = \mathrm{P}\left(X_1 \in E\right) \quad a.s.$$

- La ley fuerte coincide con nuestra noción intuitiva de probabilidad

  $\Rightarrow$ Frecuencia relativa de ocurrencia de un suceso en varios experimentos

  $\Rightarrow$ Justifica los procedimientos de estimación basados en simulaciones (e.g. histogramas)

Teorema central del límite (TCL)

Teorema (Teorema central del límite)

Sea $X_1, X_2, \dots, X_n, \dots$ 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)

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

Ejemplo de pmf condicional

- Consideremos las VAs Bernoulli independientes Y y Z. Definimos $X = Y + Z$
- Q: ¿pmf condicional de X dado Y? Para $X = 0,\ Y = 0$

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{\mathrm{P}\left(X = 0, Y = 0\right)}{\mathrm{P}\left(Y = 0\right)} = \frac{(1-p)^2}{1-p} = 1 - p$$

- También la podemos obtener de las pmfs conjunta y marginal

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{p_{XY}(0,0)}{p_Y(0)} = \frac{(1-p)^2}{1-p} = 1 - p$$

- Podemos calcular el resto de forma análoga

$$\begin{aligned} &p_{X|Y}(0|0) = 1 - p, \quad && p_{X|Y}(1|0) = p, && p_{X|Y}(2|0) = 0 \\ &p_{X|Y}(0|1) = 0, && p_{X|Y}(1|1) = 1 - p, \quad && p_{X|Y}(2|1) = p \end{aligned}$$

Condicionamiento de una suma de VAs Poisson

- Sean Y, Z VAs de Poisson independientes, de parámetros λ_1 y λ_2
- Definimos $X = Y + Z$. Q: ¿pmf condicional de Y dado X?

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

- Usamos independencia de Y y Z. Ahora recordemos que X es Poisson de parámetro $\lambda = \lambda_1 + \lambda_2$

$$\begin{aligned} p_{Y|X}(Y = y \,|\, X = x) &= \frac{e^{-\lambda_1}\lambda_1^y}{y!}\frac{e^{-\lambda_2}\lambda_2^{x-y}}{(x-y)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^x}{x!}\right]^{-1} \\ &= \frac{x!}{y!(x-y)!}\frac{\lambda_1^y\lambda_2^{x-y}}{(\lambda_1+\lambda_2)^x} \\ &= \binom{x}{y}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^y\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x-y} \end{aligned}$$

$\Rightarrow$ Condicionada a $X = x$, Y es binomial $(x, \lambda_1/(\lambda_1+\lambda_2))$

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

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

  $\Rightarrow$ 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)

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

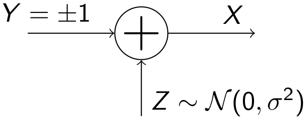
- Pero como Z es normal de media nula y varianza σ^2

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

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

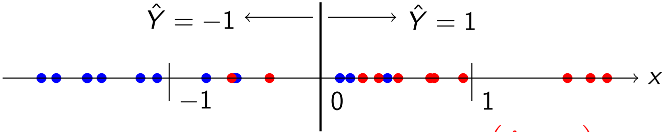
Ejemplo: modelo simple de canal de comunicación (cont.)

- El condicionamiento es una herramienta usual para estimar probabilidades
- Mensaje 1 (con probabilidad p)
  $\Rightarrow$ Transmito $Y = 1$



- Message 2 (con probabilidad q)
  $\Rightarrow$ Transmito $Y = -1$
- Señal recibida $\Rightarrow X = Y + Z$
- Regla de decisión: $\Rightarrow \hat{Y} = 1$ si $X \geq 0$, $\hat{Y} = -1$ si $X < 0$

  $\Rightarrow$ **Errores:** ● a la izquierda de 0 y ● a la derecha de 0



- Q: ¿Cuál es la probabilidad de error, $P_e := \mathrm{P}\left(\hat{Y} \neq Y\right)$?

Modelo simple de canal de comunicación (cont.)

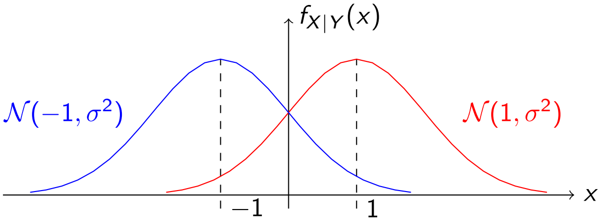
- pdf de la salida: sabemos que

  $\Rightarrow$ Si $Y = 1$ entonces $X \mid Y = 1 \sim \mathcal{N}(1, \sigma^2)$. La pdf condicional es

$$f_{X|Y}(x \mid 1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-1)^2/2\sigma^2}$$

  $\Rightarrow$ Si $Y = -1$ entonces $X \mid Y = -1 \sim \mathcal{N}(-1, \sigma^2)$. La pdf condicional es

$$f_{X|Y}(x \mid -1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x+1)^2/2\sigma^2}$$



Modelo simple de canal de comunicación (cont.)

- Probabilidad de error: la escribimos condicionando en $Y = \pm 1$ (prob. total)

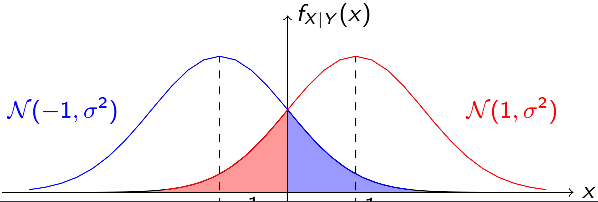
$$P_e = \mathrm{P}\left(\hat{Y} \neq Y \,\middle|\, Y = 1\right) \mathrm{P}\left(Y = 1\right) + \mathrm{P}\left(\hat{Y} \neq Y \,\middle|\, Y = -1\right) \mathrm{P}\left(Y = -1\right)$$
$$= \mathrm{P}\left(\hat{Y} = -1 \,\middle|\, Y = 1\right) \; p \qquad + \mathrm{P}\left(\hat{Y} = 1 \,\middle|\, Y = -1\right) \; q$$

- De acuerdo a la regla de decisión

$$P_e = \mathrm{P}\left(X < 0 \,\middle|\, Y = 1\right) p + \mathrm{P}\left(X \geq 0 \,\middle|\, Y = -1\right) q$$

- Como X dado Y es normal, tenemos

$$P_e = \frac{p}{\sqrt{2\pi}\sigma} \int_{-\infty}^{0} e^{-(x-1)^2/2\sigma^2} dx + \frac{q}{\sqrt{2\pi}\sigma} \int_{0}^{\infty} e^{-(x+1)^2/2\sigma^2} dx$$



Esperanza condicional

Definición de esperanza condicional

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

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

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

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

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

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

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

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

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

Ejemplo de esperanza condicional

- Sean Y y Z VAs de Bernoulli independientes, definimos $X = Y + Z$
- Q: ¿Cuánto vale $\mathbb{E}\left[X \mid Y = 0\right]$? Recordemos que encontramos la pmf condicional

$$
\begin{array}{lll}
p_{X|Y}(0|0) = 1 - p, & p_{X|Y}(1|0) = p, & p_{X|Y}(2|0) = 0 \\
p_{X|Y}(0|1) = 0, & p_{X|Y}(1|1) = 1 - p, & p_{X|Y}(2|1) = p
\end{array}
$$

- Usando la definición de esperanza condicional para variables discretas

$$
\begin{aligned}
\mathbb{E}\left[X \mid Y = 0\right] &= \sum_{x} x \, p_{X|Y}(x|0) \\
&= 0 \times (1 - p) + 1 \times p + 2 \times 0 = p
\end{aligned}
$$

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

Ejemplo de esperanzas iteradas

- Consideremos un curso de probabilidad en alguna universidad

  $\Rightarrow$ Avanzados aprueban con $A = 4$ (prob. 0,5), $B = 3$ (prob. 0,5)

  $\Rightarrow$ Principiantes con $B = 3$ (prob. 0,6), $C = 2$ (prob. 0,4)

  $\Rightarrow$ Un estudiante de intercambio es avanzado con prob. 0,7, y principiante con prob. 0,3

- Q: ¿Valor esperado de X = nota del estudiante de intercambio?
- Comenzamos por condicionar en grado de avance en la carrera

$$\begin{aligned}\mathbb{E}\left[X \,\middle|\, \text{Avanzado}\right] &= 0{,}5 \times 4 + 0{,}5 \times 3 = 3{,}5 \\ \mathbb{E}\left[X \,\middle|\, \text{Ppte}\right] &= 0{,}6 \times 3 + 0{,}4 \times 2 = 2{,}6\end{aligned}$$

- Sumamos sobre las probabilidades de grado de avance

$$\begin{aligned}\mathbb{E}\left[X\right] &= \mathbb{E}\left[X \,\middle|\, \text{Avanzado}\right] \mathrm{P}\left(\text{Avanzado}\right) + \mathbb{E}\left[X \,\middle|\, \text{Ppte}\right] \mathrm{P}\left(\text{Ppte}\right) \\ &= 3{,}5 \times 0{,}7 + 2{,}6 \times 0{,}3 = 3{,}23\end{aligned}$$

Condicionamiento en una suma de VAs de Poisson

- Y, Z VAs de Poisson independientes, de parámetros λ_1, λ_2
- Definimos $X = Y + Z$. ¿Cuánto vale $\mathbb{E}\left[Y \,\middle|\, X = x\right]$?

  $\Rightarrow$ Vimos que $Y \,\big|\, X = x$ es binomial $(x, \lambda_1/(\lambda_1 + \lambda_2))$, por lo que

$$\mathbb{E}\left[Y \,\middle|\, X = x\right] = \frac{x\lambda_1}{\lambda_1 + \lambda_2}$$

- Ahora usamos esperanzas iteradas para calcular $\mathbb{E}[Y]$

  $\Rightarrow$ Recordemos que X es Poisson con parámetro $\lambda = \lambda_1 + \lambda_2$

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{x=0}^{\infty} \mathbb{E}\left[Y \,\middle|\, X = x\right] p_X(x) = \sum_{x=0}^{\infty} \frac{x\lambda_1}{\lambda_1 + \lambda_2} p_X(x) \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \mathbb{E}[X] = \frac{\lambda_1}{\lambda_1 + \lambda_2} (\lambda_1 + \lambda_2) = \lambda_1 \end{aligned}$$

- Obviamente, ya que Y es Poisson con parámetro λ_1

Condicionamiento para cálculo de esperanzas

- Como para las probabilidades, condicionar es útil para calcular esperanzas

  $\Rightarrow$ Reduce la dificultad a problemas más simples

Ejemplo

- Un jugador de baseball anota X_i carreras por partido

  $\Rightarrow$ Número esperado de carreras: $\mathbb{E}[X_i] = \mathbb{E}[X]$, indep. del partido

- El jugador juega N partidos por temporada, N aleatorio

  $\Rightarrow$ Número esperado de partidos: $\mathbb{E}[N]$

- ¿Cuál es el número esperado de carreras en la temporada?

$$\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right]$$

- Tanto N como X_i son aleatorios. Aquí las suponemos independientes

  $\Rightarrow$ La suma $\sum_{i=1}^{N} X_i$ se conoce como VA compuesta

Suma de un número aleatorio de variables aleatorias

Paso 1: Condicionamiento por $N = n$

$$\left[\sum_{i=1}^{N} X_i \,\middle|\, N = n\right] = \sum_{i=1}^{n} X_i$$

Paso 2: Calcular el valor esperado con respecto a X_i, usando la independencia de N y X_i

$$\mathbb{E}_{X_i}\left[\sum_{i=1}^{N} X_i \,\middle|\, N = n\right] = \mathbb{E}_{X_i}\left[\sum_{i=1}^{n} X_i \,\middle|\, N = n\right] = \mathbb{E}_{X_i}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}\left[X\right]$$

$\Rightarrow$ La tercer igualdad es posible porque n es determinístico

Paso 3: Calcular el valor esperado con respecto a los valores n de N

$$\mathbb{E}_{N}\left[\mathbb{E}_{X_i}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right] = \mathbb{E}_{N}\left[N\mathbb{E}\left[X\right]\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$$

Se obtiene el resultado $\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$

Valor esperado de una VA geométrica

Ej: X VA geométrica de parámetro p (cuenta la cantidad de experimentos de Bernoulli hasta registrar el primer éxito)

- Calculamos $\mathbb{E}[X]$ condicionando por $Y = \mathbb{I}\{\text{“el primer experimento es exitoso”}\}$
  - $\Rightarrow$ Si $Y = 1$, entonces $\mathbb{E}\left[X \mid Y = 1\right] = 1$
  - $\Rightarrow$ Si $Y = 0$, por indep. de experimentos: $\mathbb{E}\left[X \mid Y = 0\right] = 1 + \mathbb{E}[X]$
- Usamos esperanzas iteradas

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X \mid Y = 1\right] \mathrm{P}(Y = 1) + \mathbb{E}\left[X \mid Y = 0\right] \mathrm{P}(Y = 0) \\ &= 1 \times p + (1 + \mathbb{E}[X]) \times (1 - p)\end{aligned}$$

- Despejando $\mathbb{E}[X]$ obtenemos $\mathbb{E}[X] = \frac{1}{p}$
- En este caso el cálculo directo es sencillo (serie geométrica, derivada)
  - $\Rightarrow$ A menudo las simplificaciones son mayores

Ejemplo: el minero atrapado

- Un minero está atrapado en una mina que contiene tres puertas
- Para todo tiempo $n \geq 1$ mientras que está atrapado
  - El minero escoge una puerta $P_n = j$, $j = 1, 2, 3$
  - La elección de una puerta P_n es independiente de las elecciones anteriores
  - La elección de la puerta es equiprobable, i.e., $\mathrm{P}\left(P_n = j\right) = 1/3$
- Cada puerta conduce a un túnel, pero solo un túnel va a la salida
  - Puerta 1: el minero alcanza la salida luego de dos horas de recorrido
  - Puerta 2: el minero retorna después de tres horas de recorrido
  - Puerta 3: el minero retorna después de cinco horas de recorrido
- Denotamos por X al tiempo total transcurrido hasta que el minero alcanza la salida
- Q: ¿Cuánto vale $\mathbb{E}\left[X\right]$?

- Calculamos $\mathbb{E}[X]$ condicionando a la primera puerta elegida, P_1
  - $\Rightarrow$ Si $P_1 = 1$, entonces sale a las 2 horas, i.e., $\mathbb{E}\left[X \mid P_1 = 1\right] = 2$
  - $\Rightarrow$ Si $P_1 = 2$, como elecciones son indep., $\mathbb{E}\left[X \mid P_1 = 2\right] = 3 + \mathbb{E}[X]$
  - $\Rightarrow$ De la misma forma, si $D_1 = 3$ tenemos $\mathbb{E}\left[X \mid P_1 = 3\right] = 5 + \mathbb{E}[X]$
- Usamos esperanzas iteradas

$$\mathbb{E}[X] = \sum_{j=1}^{3} \mathbb{E}\left[X \mid D_1 = j\right] \mathrm{P}\left(D_1 = j\right) = \frac{1}{3} \sum_{j=1}^{3} \mathbb{E}\left[X \mid D_1 = j\right]$$
$$= \frac{2 + 3 + \mathbb{E}[X] + 5 + \mathbb{E}[X]}{3} = \frac{10 + 2\mathbb{E}[X]}{3}$$

- Despejando $\mathbb{E}[X]$ obtenemos

$$\mathbb{E}[X] = 10$$

Fórmula de la varianza condicional

- **Def:** La varianza condicional de X dado $Y = y$ es

$$\begin{aligned}\operatorname{var}[X|Y=y] &= \mathbb{E}\left[(X - \mathbb{E}\left[X \mid Y=y\right])^2 \mid Y=y\right] \\ &= \mathbb{E}\left[X^2 \mid Y=y\right] - (\mathbb{E}\left[X \mid Y=y\right])^2\end{aligned}$$

$\Rightarrow$ $\operatorname{var}[X|Y]$ es una función de la VA Y, su valor para $Y = y$ es $\operatorname{var}[X|Y=y]$

- Calculo de $\operatorname{var}[X]$ por condicionamiento en $Y = y$. ¿Cuánto vale?

  $\Rightarrow \operatorname{var}[X] \neq \mathbb{E}_Y[\operatorname{var}_X(X \mid Y)]$

  $\Rightarrow \operatorname{var}[X] \neq \operatorname{var}_Y[\mathbb{E}_X(X \mid Y)]$

- Ninguno de los dos. fórmula de varianza condicional:

$$\operatorname{var}[X] = \mathbb{E}_Y[\operatorname{var}_X(X \mid Y)] + \operatorname{var}_Y[\mathbb{E}_X(X \mid Y)]$$

Fórmula de la varianza condicional (cont.)

Demostración.

- Primer sumando: usamos linealidad y esperanzas iteradas

$$\begin{aligned}
\mathbb{E}_Y[\mathrm{var}_X(X \mid Y)] &= \mathbb{E}_Y\left[\mathbb{E}_X(X^2 \mid Y) - (\mathbb{E}_X(X \mid Y))^2\right] \\
&= \mathbb{E}_Y\left[\mathbb{E}_X(X^2 \mid Y)\right] - \mathbb{E}_Y\left[(\mathbb{E}_X(X \mid Y))^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}_Y\left[(\mathbb{E}_X(X \mid Y))^2\right]
\end{aligned}$$

- Segundo sumando: usamos def. de varianza y esperanzas iteradas

$$\begin{aligned}
\mathrm{var}_Y[\mathbb{E}_X(X \mid Y)] &= \mathbb{E}_Y\left[(\mathbb{E}_X(X \mid Y))^2\right] - \left(\mathbb{E}_Y[\mathbb{E}_X(X \mid Y)]\right)^2 \\
&= \mathbb{E}_Y\left[(\mathbb{E}_X(X \mid Y))^2\right] - (\mathbb{E}[X])^2
\end{aligned}$$

- Sumando ambos términos (los términos en azul se cancelan):

$$\mathbb{E}_Y[\mathrm{var}_X(X \mid Y)] + \mathrm{var}_Y[\mathbb{E}_X(X \mid Y)] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \mathrm{var}[X]$$

$\square$

Varianza de una VA compuesta

- Sean $X_1, X_2, \ldots$ VAs i.i.d. con $\mathbb{E}[X_1] = \mu$ y $\text{var}[X_1] = \sigma^2$
- Sea N una VA a valores enteros no negativos independiente de los X_i
- Consideramos la VA compuesta $S = \sum_{i=1}^{N} X_i$. ¿Cuánto vale $\text{var}[S]$?
- La fórmula de varianza condicional resulta útil aquí
- Vimos anteriormente que $\mathbb{E}[S|N] = N\mu$. ¿Qué pasa con $\text{var}[S|N = n]$?

$$\text{var}\left[\sum_{i=1}^{N} X_i | N = n\right] = \text{var}\left[\sum_{i=1}^{n} X_i | N = n\right] = \text{var}\left[\sum_{i=1}^{n} X_i\right] = n\sigma^2$$

$\Rightarrow \text{var}[S|N] = N\sigma^2$. Aquí usamos la indep. de N y las VAs i.i.d. X_i

- La fórmula de varianza condicional es $\text{var}[S] = \mathbb{E}\left[N\sigma^2\right] + \text{var}[N\mu]$

Obtenemos el resultado $\Rightarrow \text{var}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}[N]\,\sigma^2 + \text{var}[N]\,\mu^2$