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

3 de agosto de 2021

Probabilidad

- Un suceso es algo que ocurre
- Todo suceso aleatorio tiene un resultado incierto

  $\Rightarrow$ La probabilidad de un suceso mide cuán verosímil es que este ocurra

Ejemplo

- Escribo el número de cédula de identidad de un estudiante en un papel. Quién es?
- Suceso: La C.I. del estudiante x está escrita en el papel
- Probabilidad: $P(x)$ mide cuán verosímil es que el número de C.I. de x esté escrito

- La probabilidad es un instrumento de medida

  $\Rightarrow$ Lenguaje matemático para cuantificar la incertidumbre

Sigma-álgebra

- Dado un espacio muestral or universo Ω (también espacio de sucesos elementales)
  - Ej: Todos los estudiantes del curso, $\Omega = \{x_1, x_2, \ldots, x_N\}$ (x_n representa número de C.I.)
- **Def:** Un suceso elemental es un elemento o punto de Ω, e.g., x_3
- **Def:** Un suceso E es un subconjunto de Ω
  - Ej: $\{x_1\}$, estudiante con número de C.I. x_1
  - Ej: $\{x_1, x_4\}$, estudiantes con número de C.I. x_1 o x_4

  $\Rightarrow$ Suceso elemental x_3 y suceso $\{x_3\}$ son cosas distintas, éste último es un conjunto
- **Def:** Una sigma-álgebra $\mathcal{F}$ es un conjunto de sucesos $E \subseteq \Omega$ tal que
  (i) El conjunto vacío (*suceso imposible*) $\emptyset$ pertence a $\mathcal{F}$: $\emptyset \in \mathcal{F}$
  (ii) Es cerrado por complemento: Si $E \in \mathcal{F}$, entonces $E^c \in \mathcal{F}$
  (iii) Es cerrado por unión numerable: Si $E_1, E_2, \ldots \in \mathcal{F}$, entonces $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$
- $\mathcal{F}$ es un conjunto de conjuntos

Ejemplos de sigma-álgebras

Ejemplo

- Ningún estudiante y todos los estudiantes, i.e., $\mathcal{F}_0 := \{\emptyset, \Omega\}$

Ejemplo

- Conjunto vacío, mujeres, hombres, todos, i.e., $\mathcal{F}_1 := \{\emptyset, \text{Mujeres}, \text{Hombres}, \Omega\}$

Ejemplo

- $\mathcal{F}_2$ incluyendo el conjunto vacío $\emptyset$ más

  todos los sucesos (conjuntos) con un estudiante $\{x_1\}, \ldots, \{x_N\}$ más

  todos los sucesos con dos estudiantes $\{x_1, x_2\}$, $\{x_1, x_3\}$, ..., $\{x_1, x_N\}$,
  $\{x_2, x_3\}$, ..., $\{x_2, x_N\}$,
  ...
  $\{x_{N-1}, x_N\}$ más

  todos los sucesos con tres, cuatro, ..., N estudiantes

  $\Rightarrow$ $\mathcal{F}_2$ se conoce como conjunto potencia o conjunto de las partes de Ω, y se escribe 2^Ω ($|2^\Omega| = 2^{|\Omega|}$)

Axiomas de la teoría de la probabilidad

- Sea una función $P(E)$ de una sigma-álgebra $\mathcal{F}$ a los números reales
- $P(E)$ es una probabilidad si
  A1) Non-negatividad: $P(E) \geq 0$
  A2) Probabilidad del universo: $P(\Omega) = 1$
  A3) Aditividad: Dada un sucesión de sucesos disjuntos $E_1, E_2, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P\left(E_i\right)$$

$\Rightarrow$ Sucesos disjuntos (o incompatibles dos a dos) significa $E_i \cap E_j = \emptyset,\ i \neq j$

$\Rightarrow$ Unión de infinitos sucesos disjuntos numerables

- La terna $(\Omega, \mathcal{F}, P)$ se llama espacio de probabilidad

Consecuencias de los axiomas

- Implicancias de los axiomas A1)-A3)
  - $\Rightarrow$ Suceso imposible: $P(\emptyset) = 0$
  - $\Rightarrow$ Monotonía: $E_1 \subset E_2 \;\Rightarrow P(E_1) \leq P(E_2)$
  - $\Rightarrow$ Imagen: $0 \leq P(E) \leq 1$
  - $\Rightarrow$ Complemento: $P(E^c) = 1 - P(E)$
  - $\Rightarrow$ Union disjunta finita: Para suceso disjuntos $E_1, \ldots, E_N$

$$P\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} P\left(E_i\right)$$

  - $\Rightarrow$ Inclusión-exclusión: Para cualquier par de sucesos E_1 and E_2

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Ejemplo de espacio de probabilidad

- Construyamos un espacio de probabilidad para nuestro ejemplo
- Universo de todos los estudiantes del curso $\Omega = \{x_1, x_2, \ldots, x_N\}$
- Sigma-álgebra con todas las combinaciones de estudiantes, i.e., $\mathcal{F} = 2^{\Omega}$
- Los números de C.I. escritos son equiprobables $\Rightarrow P(\{x_n\}) = 1/N$, $n = 1, \ldots, N$

  $\Rightarrow$ Especifiquemos la probabilidad para todo $E \in \mathcal{F}$ $\Rightarrow$ definimos $P(E) = \frac{|E|}{|\Omega|}$
- Q: Es esta función una probabilidad?

  $\Rightarrow$ A1): $P(E) = \frac{|E|}{|S|} \geq 0 \checkmark$ $\Rightarrow$ A2): $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1 \checkmark$

  $\Rightarrow$ A3): $P\left(\bigcup_{i=1}^{N} E_i\right) = \frac{\left|\bigcup_{i=1}^{N} E_i\right|}{|\Omega|} = \frac{\sum_{i=1}^{N} |E_i|}{|\Omega|} = \sum_{i=1}^{N} P(E_i) \checkmark$
- La función $P(\cdot)$ así definida se llama distribución uniforme

Probabilidad condicional, probabilidad total, regla de Bayes

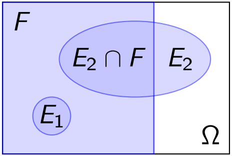
Probabilidad condicional

- Consideremos los sucesos E y F, y supongamos que sabemos que F ha ocurrido
- Q: Qué información nos da esto sobre la probabilidad de E?
- **Def:** La probabilidad condicional de E dado F es (necesitamos que $P(F) > 0$)

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

$\Rightarrow$ En general $P(E|F) \neq P(F|E)$

- Renormalizamos las probabilidades al conjunto F
  - Descartamos una parte de Ω
  - Eventualmente también puede descartarse una parte de E



- Dado F con $P(F) > 0$, $P(\cdot|F)$ satisface los axiomas de la probabilidad

Ejemplo de probabilidad condicional

- El nombre de C.I. escrito corresponde a un estudiante de sexo masculino. Cuál es la probabilidad de x_n?
- Supongamos que los hombres son $F = \{x_1, \ldots, x_M\}$ $\Rightarrow P(F) = \frac{M}{N}$
- Si el número de C.I. x_n es de un hombre, $x_n \in F$ y tenemos el suceso $E = \{x_n\}$

$$P(E \cap F) = P(\{x_n\}) = \frac{1}{N}$$

$\Rightarrow$ La probabilidad condicional es, como esperamos

$$P(E \,|\, F) = \frac{P(E \cap F)}{P(F)} = \frac{1/N}{M/N} = \frac{1}{M}$$

- Si el número de C.I. es de una mujer $x_n \notin F$, entonces $P(E \cap F) = P(\emptyset) = 0$

$\Rightarrow$ Tenemos entonces $P(E \,|\, F) = 0$

Ley de la probabilidad total

- Consideremos el suceso E y los sucesos F y F^c
  - F y F^c forman una partición de Ω ($F \cup F^c = S$, $F \cap F^c = \emptyset$)
- Como $F \cup F^c = \Omega$ cubre Ω, podemos escribir el conjunto E como

$$E = E \cap S = E \cap [F \cup F^c] = [E \cap F] \cup [E \cap F^c]$$

- Como $F \cap F^c = \emptyset$ son disjuntos, también lo son $[E \cap F] \cap [E \cap F^c] = \emptyset$

  $\Rightarrow P(E) = P([E \cap F] \cup [E \cap F^c]) = P(E \cap F) + P(E \cap F^c)$
- Usando la definición de probabilidad condicional

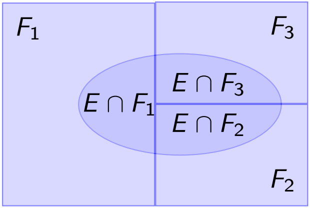
$$P(E) = P(E \mid F)P(F) + P(E \mid F^c)P(F^c)$$

- Esto traduce la información condicional $P(E \mid F)$ y $P(E \mid F^c)$

  $\Rightarrow$ En información incondicional $P(E)$

Ley de la probabilidad total (cont.)

- En general, sea la partición de Ω (posiblemente infinita), F_i, $i = 1, 2, \ldots$
- Los conjuntos son disjuntos $\Rightarrow F_i \cap F_j = \emptyset$ para $i \neq j$
- Los conjuntos cubren el espacio $\Rightarrow \cup_{i=1}^{\infty} F_i = \Omega$



- Igual que antes, como $\cup_{i=1}^{\infty} F_i = \Omega$ cubre el espacio, podemos escribir E como

$$E = E \cap S = E \cap \left[\bigcup_{i=1}^{\infty} F_i\right] = \bigcup_{i=1}^{\infty} [E \cap F_i]$$

- Como $F_i \cap F_j = \emptyset$ son disjuntos, también $[E \cap F_i] \cap [E \cap F_j] = \emptyset$. Luego

$$P(E) = P\left(\bigcup_{i=1}^{\infty} [E \cap F_i]\right) = \sum_{i=1}^{\infty} P(E \cap F_i) = \sum_{i=1}^{\infty} P(E \mid F_i) P(F_i)$$

Ejemplo de probabilidad total

- Consideremos un curso de probabilidad en alguna universidad

  $\Rightarrow$ Los estudiantes avanzados aprueban con A con probabilidad 0.9, los principiantes con probabilidad 0.8

  $\Rightarrow$ Un estudiante de intercambio es avanzado con probabilidad 0.7, y principiante con probabilidad 0.3
- Q: ¿Cuál es la probabilidad de que el estudiante de intercambio apruebe con A?
- Sea $E =$ "el estudiante de intercambio aprueba con A", a por avanzado, p por principiante

  $\Rightarrow$ Usamos la ley de la probabilidad total

$$\begin{aligned} P(E) &= P(E \mid a)P(a) + P(E \mid p)P(p) \\ &= 0{,}9 \times 0{,}7 + 0{,}8 \times 0{,}3 = 0{,}87 \end{aligned}$$

Regla de Bayes

- De la definición de probabilidad condicional: $P(E \mid F)P(F) = P(E \cap F)$
- De la misma forma: $P(F \mid E)P(E) = P(F \cap E)$
- Como ambas cantidades son iguales, obtenemos la regla de Bayes

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

- La regla de Bayes permite la inversión temporal o de causalidad. Si F (futuro) viene después de E (pasado),

  $\Rightarrow P(E \mid F)$, prob. que el pasado (E) haya visto el futuro (F)

  $\Rightarrow P(F \mid E)$, proba. que el futuro (F) corresponda al pasado (E)

- En general los modelos describen futuro | pasado (modelo directo). Usualmente interesa pasado | futuro (modelo inverso).

Ejemplo de regla de Bayes

- Consideremos la siguiente partición de mis emails

  $\Rightarrow$ E_1 =“spam” con probabilidad $P(E_1) = 0{,}7$

  $\Rightarrow$ E_2 =“baja prioridad” con probabilidad $P(E_2) = 0{,}2$

  $\Rightarrow$ E_3 =“alta prioridad” con probabilidad $P(E_3) = 0{,}1$

- Sea el suceso F=“el email contiene la palabra *gratis*”

  $\Rightarrow$ De la experiencia previa sabemos que $P(F \mid E_1) = 0{,}9$, $P(F \mid E_2) = P(F \mid E_3) = 0{,}01$

- Recibo un email con la palabra “gratis”. ¿Cuál es la probabilidad que sea spam?

- Aplicamos la regla de Bayes

$$P(E_1 \mid F) = \frac{P(F \mid E_1)P(E_1)}{P(F)} = \frac{P(F \mid E_1)P(E_1)}{\sum_{i=1}^{3} P(F \mid E_i)P(E_i)} = 0{,}995$$

  $\Rightarrow$ Ley de la probabilidad total: muy útil cuando se aplica la regla de Bayes

Independencia

Independencia

- **Def:** los suceso E y F son independientes si $P(E \cap F) = P(E)P(F)$
  - $\Rightarrow$ Sucesos que no son independientes son dependientes
- $P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E)$
  - $\Rightarrow$ Intuitivamente, conocer F no altera nuestra percepción sobre E
  - $\Rightarrow$ F no contiene información sobre E
  - $\Rightarrow$ El recíproco también es cierto: $P(F \mid E) = P(F)$
- El que E y F sean independientes depende fuertemente de $P(\cdot)$
- Atención: evitar confundir sucesos independientes con sucesos disjuntos, i.e. $E \cap F = \emptyset$
- Q: ¿Dos suceso disjuntos con probabilidades $P(E) > 0$, $P(F) > 0$ pueden ser independientes? No $\rightarrow$ Eventos disjuntos posibles son necesariamente dependientes
- Ej. 1: bolillero con los números $\{1, 2, 3, 4\}$. Saco una bolilla B. Los eventos "B es par" y "$B \leq 2$" son idependientes, pero no son disjuntos.
- Ej. 2: Ω universo, $E \subset \Omega$ evento. $P(\Omega \cap E) = P(E) = P(E)P(\Omega)$.

Ejemplo de independencia

- Escribo un número de C.I. de los estudiantes del curso, y le pido a un amigo que escriba otro (puede ser mismo)
- Espacio de probabilidad para el experimento: $(S, \mathcal{F}, P(\cdot))$

  $\Rightarrow$ S es el conjunto de todos los pares de números de C.I. de los estudiantes: $[x_n(1), x_n(2)]$, $|\Omega| = N^2$

  $\Rightarrow$ La sigma-álgebra es el producto cartesiano $\mathcal{F} = 2^\Omega \times 2^\Omega$

  $\Rightarrow$ Definimos $P(E) = \frac{|E|}{|\Omega|}$ como la ley uniforme
- Consideramos los sucesos $E_1 =$'Yo escribí x_1' y $E_2 =$'Mi amigo escribió x_2' Q: ¿Son independientes? Sí, ya que

$$P(E_1 \cap E_2) = P\big(\{(x_1, x_2)\}\big) = \frac{|\{(x_1, x_2)\}|}{|\Omega|} = \frac{1}{N^2} = P(E_1)P(E_2)$$

- Sucesos dependientes: $E_1 =$'Yo escribí x_1' y $E_3 =$'Ambos estudiantes son hombres'

Independencia para más de dos sucesos

- **Def:** Los sucesos E_i, $i = 1, 2, \ldots$ se dicen mutuamente independientes si

$$P\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} P(E_i)$$

  para todo subconjunto finito I de al menos dos enteros

- Ej: los sucesos E_1, E_2, y E_3 son mutuamente independientes si se cumple que

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$$
$$P(E_1 \cap E_2) = P(E_1)P(E_2)$$
$$P(E_1 \cap E_3) = P(E_1)P(E_3)$$
$$P(E_2 \cap E_3) = P(E_2)P(E_3)$$

- Si $P(E_i \cap E_j) = P(E_i)P(E_j)$ para todo (i, j), los E_i son independent dos a dos

  $\Rightarrow$ Independencia mutua $\rightarrow$ independencia dos a dos. No recíprocamente

Variables aleatorias

Definición de variable aleatorias (VA)

- **Def:** Una VA $X(\omega)$ es una función que le asigna un valor a un suceso elemental $\omega \in \Omega$

  $\Rightarrow$ Podemos pensar en VAs como medidas asociadas a un experimento

Ejemplo

- Tiramos una pelota en un cuadrado de $1m \times 1m$. Nos interesa la posición de la pelota
- Una realización incierta es el lugar ω donde cae la pelota
- Las variables aleatorias son $X(\omega)$ e $Y(\omega)$, coordenadas espaciales

- Las probabilidades de las VAs se infieren de los sucesos elementales subyacentes

$$P(X(\omega) = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

$$P(X(\omega) \in (-\infty, x]) = P(\{\omega \in \Omega : X(\omega) \in (-\infty, x]\})$$

Ejemplo 1

- Se tira una moneda, cara (H) o cruz (T). La moneda no está cargada $P(H) = 1/2$, $P(T) = 1/2$. Se paga \$1 por H, se cobra \$1 por T. Ganancia?
- Las realizaciones posibles son H y T
- Para medir la ganancia se define la VA X con valores

$$X(H) = 1, \qquad X(T) = -1$$

- Las probabilidades de la VA son

$$\begin{aligned} P(X = 1) \quad &= P(H) = 1/2, \\ P(X = -1) &= P(T) = 1/2 \end{aligned}$$

$\Rightarrow$ Se tiene $P(X = x) = 0$ para todo $x \neq \pm 1$

Ejemplo 2

- Se tiran 2 monedas. Se paga \$1 por cada H, se cobra \$1 por cada T. Ganancia?
- Ahora las realizaciones posibles son HH, HT, TH, and TT
- Para medir la ganancia se define la VA Y con valores

$$Y(HH) = 2, \quad Y(HT) = 0, \quad Y(TH) = 0, \quad Y(TT) = -2$$

- Las probabilidades par la VA son

$$\begin{aligned} P(Y = 2) &= P(HH) &&= 1/4, \\ P(Y = 0) &= P(HT) + P(TH) &&= 1/2, \\ P(Y = -2) &= P(TT) &&= 1/4 \end{aligned}$$

Sobre los ejemplos 1 y 2

- Las VAs son más fáciles de manejar que los sucesos
- Sea $\omega_1 \in \{H, T\}$ una realización de la moneda 1 y $\omega_2 \in \{H, T\}$ de la moneda 2

  $\Rightarrow$ Se puede vincular Y y X como

$$Y(\omega_1, \omega_2) = X_1(\omega_1) + X_2(\omega_2)$$

- Se tiran N monedas. Ganancia? La enumeración se vuelve engorrosa
- Altermativamente, sea $\omega_n \in \{H, T\}$ la realización de la n-ésima tirada. Se define

$$Y(\omega_1, \omega_2, \ldots, \omega_N) = \sum_{n=1}^{N} X_n(\omega_n)$$

  $\Rightarrow$ En general se utiliza el abuso de notación $Y = \sum_{n=1}^{N} X_n$

Ejemplo 3

- Se tira una moneda hasta que sale H por primera vez. $P(H) = p$
- ¿Cuántas tiradas se hacen hasta que sale la primera H?
- Las realizaciones posibles son H, TH, TTH, $TTTH$, $\ldots$ Obs: $|S| = \infty$

  $\Rightarrow$ Se deja de tirar luego de obtener la primera H (e.g. THT no es una realización posible)
- Sea N la VA que cuenta la cantidad de tiradas

  $\Rightarrow$ $N = n$ si sale T en las primeras $n-1$ tiradas, y H en la n-ésima

$$
\begin{aligned}
P(N=1) &= P(H) && = p \\
P(N=2) &= P(TH) && = (1-p)p \\
&\vdots \\
P(N=n) &= P(\underbrace{TT\ldots T}_{n-1 \text{ veces}}H) && = (1-p)^{n-1}p
\end{aligned}
$$

Example 3 (continued)

- De A2) deberíamos tener $P(S) = \sum_{n=1}^{\infty} P(N = n) = 1$
- Cierto porque $\sum_{n=1}^{\infty} (1-p)^{n-1}$ es una serie geométrica

$$\sum_{n=1}^{\infty} (1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-(1-p)} = \frac{1}{p}$$

- Sustituyendo la suma de la serie geométrica en la expresión de $P(S)$

$$\sum_{n=1}^{\infty} P(N = n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = p \times \frac{1}{p} = 1 \checkmark$$

Función indicatriz

- La función indicatriz de un suceso es una VA
- Sea $\omega \in \Omega$ una realización, y $E \subset \Omega$ un suceso

$$\mathbb{I}\{E\}(\omega) = \begin{cases} 1, & \text{if } \omega \in E \\ 0, & \text{if } \omega \notin E \end{cases}$$

$\Rightarrow$ Indica que la realización ω pertenece al conjunto E, tomando el valor 1

Ejemplo

- Cantidad N de tiradas de monedas hasta la primera H. Nos interesa cuándo N supera N_0

$\Rightarrow$ El suceso es $\{N : N > N_0\}$. Las realizaciones posibles son $N = 1, 2, \ldots$

$\Rightarrow$ Llamamos a la función indicatriz $\mathbb{I}_{N_0} = \mathbb{I}\{N : N > N_0\}$

- La probabilidad $P(\mathbb{I}_{N_0} = 1) = P(N > N_0) = (1 - p)^{N_0}$

$\Rightarrow$ Para que N supere N_0 se necesitan N_0 consecutivas T

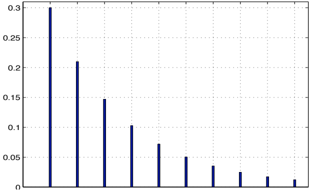
$\Rightarrow$ No importa lo que suceda luego

Variables aleatorias discretas

Funciones de masa de probabilidad y de distribución

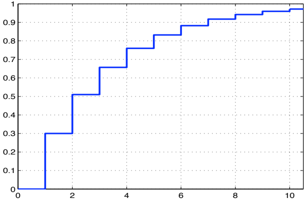
- Las VA discretas toman, a lo sumo, un conjunto numerable de valores
- Función de (masa de) probabilidad o probability mass function (pmf) $p_X(x) = P(X = x)$
  - Si la VA es clara del contexto, simplemente escribimos $p_X(x) = p(x)$

- Si X tiene soporte $\{x_1, x_2, \ldots\}$, su pmf satisface
  (i) $p(x_i) > 0$ for $i = 1, 2, \ldots$
  (ii) $p(x) = 0$ for all other $x \neq x_i$
  (iii) $\sum_{i=1}^{\infty} p(x_i) = 1$
  - pmf para "cantidad de tiradas hasta primera cara" ($p = 0{,}3$)



- Función de distribución (acumulada) (cdf) $F_X(x) = P(X \leq x) = \sum_{i:x_i \leq x} p(x_i)$

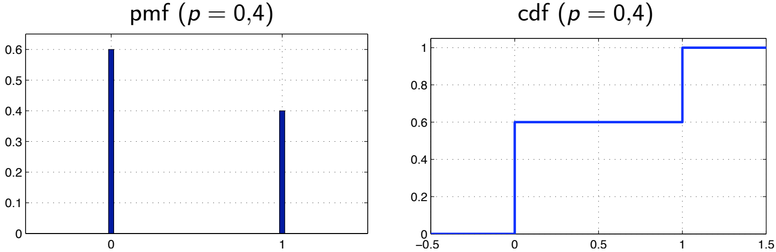
  $\Rightarrow$ Función creciente con saltos en x_i
  - cdf para "cantidad de tiradas hasta la primer cara" (con $p = 0{,}3$)



Distribución de Bernoulli

- Un experimento es exitoso con probabilidad p or fallido con probabilidad $q := 1 - p$
  - $\Rightarrow$ Ex: Tirar una moneda, cualquier indicatriz de un suceso
- Bernoulli X toma valores 0 o 1. Su pmf es $p(x) = p^x q^{1-x}$
- Su cdf es

$$F(x) = \begin{cases} 0, & x < 0 \\ q, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

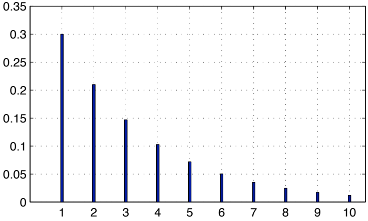


Distribución geométrica

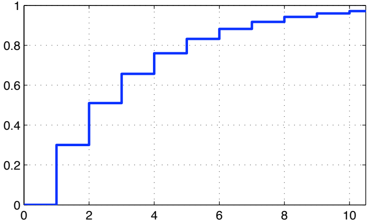
- Cuenta la cantidad de experimentos de Bernoulli hasta registrar el primer éxito

  $\Rightarrow$ Cada experimento tiene éxito con probabilidad p
- La cantidad de experimentos hasta éxito, X, sigue una distribución geométrica con parámetro p
- pmf: $p(x) = p(1-p)^{x-1}$
  - Un éxito luego de $x-1$ fallas. Los experimentos (las VAs de Bernoulli) son independientes
- cdf: $F(x) = 1 - (1-p)^{x}$
  - Recordar que $\mathrm{P}\left(X > x\right) = (1-p)^{x}$; sino sumar la serie geométrica

pmf ($p = 0{,}3$)



cdf ($p = 0{,}3$)



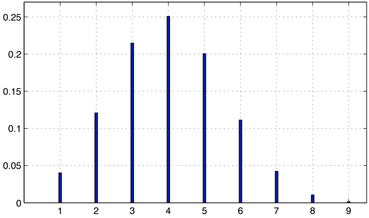
Distribución binomial

- Cuenta la cantidad de éxitos entre n experimentos de Bernoulli
  $\Rightarrow$ Cada experimento tiene éxito con probabilidad p
- La cantidad de éxitos X sigue una distribución binomial con parámetros (n, p). Su pmf es

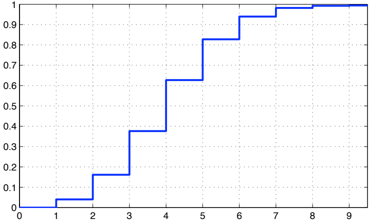
$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

$\Rightarrow X = x$ for x éxitos (p^x) and $n - x$ fallas ($(1-p)^{n-x}$).
$\Rightarrow \binom{n}{x}$ formas de escojer x éxitos entre n experimentos

pmf $(n = 9,\ p = 0{,}4)$



cdf $(n = 9,\ p = 0{,}4)$



Distribución binomial (cont.)

- Sean Y_i, $i = 1, \dots n$ VAs de Bernoulli con parámetro p
  - $\Rightarrow$ Las Y_i están asociadas a sucesos independientes
- Podemos escribir la binomial X como la con parametros (n, p) como

$$\Rightarrow X = \sum_{i=1}^{n} Y_i$$

Ejemplo

- Considere las binomiales Y and Z con parámetros (n_Y, p) y (n_Z, p)
  - $\Rightarrow$ Q: ¿Distribución de probabilidad de $X = Y + Z$?
- Escribimos $Y = \sum_{i=1}^{n_Y} Y_i$ and $Z = \sum_{i=1}^{n_Z} Z_i$, luego

$$X = \sum_{i=1}^{n_Y} Y_i + \sum_{i=1}^{n_Z} Z_i$$

$\Rightarrow$ X binomial con párametro $(n_Y + n_Z, p)$

Distribución de Poisson

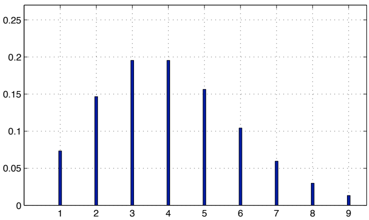
- Cuenta eventos raros (decay radioactivo, arribo paquetes, accidentes)
- Usualmente se modela como Poisson con parámetro λ y pmf

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

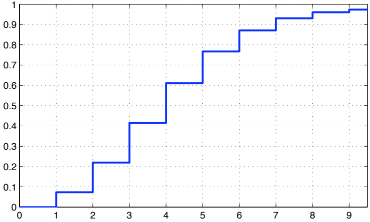
- Q: ¿La pmf está bien definida? Sí
- Desarrollo de Taylor de $e^x = 1 + x + x^2/2 + \ldots + x^i/i! + \ldots$.

$$\text{Entonces} \quad P(\Omega) = \sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1 \checkmark$$

pmf ($\lambda = 4$)

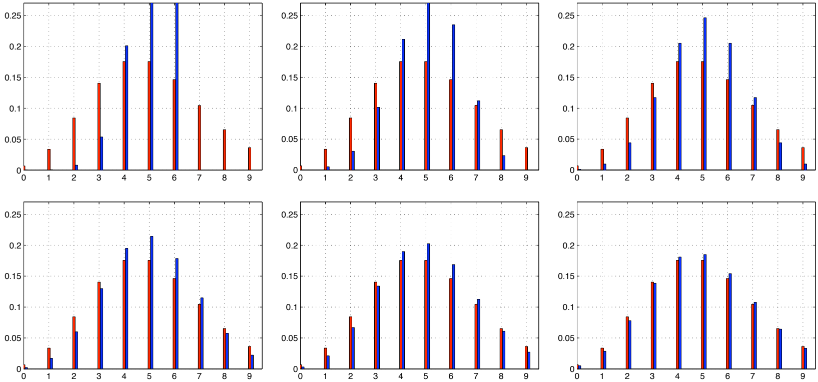


cdf ($\lambda = 4$)



Approximación de la binomial como Poisson

- X binomial con parámetros (n, p)
- Sea $n \to \infty$ manteniendo el producto $np = \lambda$ constante
  - (Si solo hacemos $n \to \infty$ la cantidad de éxitos diverge).
- Comparemos con la distribución de Poisson de parámetro λ
  - $\lambda = 5,\ n = 6, 8, 10, 15, 20, 50$



Poisson y binomial (cont.)

- Esta es, de hecho, la motivación para definir una VA de Poisson
- Sustituyendo $p = \lambda/n$ en la pmf de una VA binomial

$$\begin{aligned} p_n(x) &= \frac{n!}{(n-x)!x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n(n-1)\dots(n-x+1)}{n^x}\frac{\lambda^x}{x!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^x} \end{aligned}$$

$\Rightarrow$ (usamos la def. de factorial y $(1-\lambda/n)^{n-x} = \frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}$, y reordenamos)

- En el límite, el término rojo vale $\lim_{n\to\infty}(1-\lambda/n)^n = e^{-\lambda}$
- Los términos negro y azul convergen a 1. Luego,

$$\lim_{n\to\infty} p_n(x) = 1\frac{\lambda^x}{x!}\frac{e^{-\lambda}}{1} = e^{-\lambda}\frac{\lambda^x}{x!}$$

$\Rightarrow$ El límite es la pmf de una VA de Poisson

Conclusión

- La distribución binomial está motivada por el conteo de aciertos
- La distribución de Poisson es una aproximación para grandes números de experimentos n

  $\Rightarrow$ La distribución de Poisson es más manejable (comparar pmfs)
- A veces se la refiere como la "ley de eventos raros"
  - Los eventos individuales (aciertos) suceden con probabilidad pequeña $p = \lambda/n$
  - Los eventos agregados (cantidad de aciertos), sin embargo, no tienen por qué ser raros
- Notar que las cuatro VAs vistas hasta ahora refieren a "tiradas de monedas"

Variables aleatorias continuas

VAs continuas, funciones de densidad de probabilidad

- Los valores posibles que toman las VAs continuas X forman un subconjunto denso $\mathcal{X} \subseteq \mathbb{R}$

  $\Rightarrow$ Cantidad infinita de valores no numerables

- La función de densidad de probabilidad (pdf) $f_X(x)$ es tal que para cualquier subconjunto $\mathcal{X} \subseteq \mathbb{R}$
  (A la derecha, pdf normal)

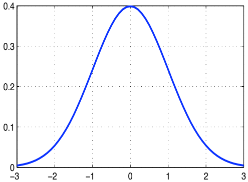
$$P(X \in \mathcal{X}) = \int_{\mathcal{X}} f_X(x)dx$$

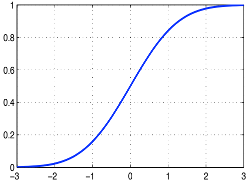
$\Rightarrow$ Tenemos $P(X = x) = 0$ para todo $x \in \mathcal{X}$

- La cdf definida como antes y relacionada a la (A la derecha, cdf normal)

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du$$

$$\Rightarrow P(X \leq \infty) = F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$$





Más sobre cdfs y pdfs

- Cuando el conjunto $\mathcal{X} = [a, b]$ es un intervalo de $\mathbb{R}$

$$\mathrm{P}\left(X \in [a, b]\right) = \mathrm{P}\left(X \leq b\right) - \mathrm{P}\left(X \leq a\right) = F_X(b) - F_X(a)$$

- En términos de la pdf se puede escribir como

$$\mathrm{P}\left(X \in [a, b]\right) = \int_a^b f_X(x)\, dx$$

- Para intervalos pequeños $[x_0, x_0 + \delta x]$, en particular

$$\mathrm{P}\left(X \in [x_0, x_0 + \delta x]\right) = \int_{x_0}^{x_0 + \delta x} f_X(x)\, dx \approx f_X(x_0)\delta x$$

$\Rightarrow$ La probabilidad es el "área bajo la pdf" (de dónde el nombre "densidad")

- Otra relación entre pdf y cdf es $\Rightarrow \dfrac{\partial F_X(x)}{\partial x} = f_X(x)$

  $\Rightarrow$ Teorema fundamental del cálculo

Distribución uniforme

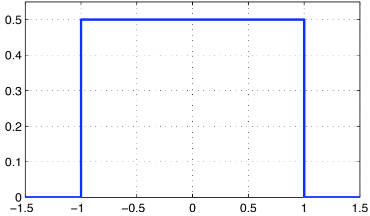
- Modela problemas con probabilidades equiprobables en $[a, b]$
- La pdf de una VA uniforme es $f(x) = 0$ fuera de $[a, b]$, y

$$f(x) = \frac{1}{b-a}, \quad \text{para } a \leq x \leq b$$

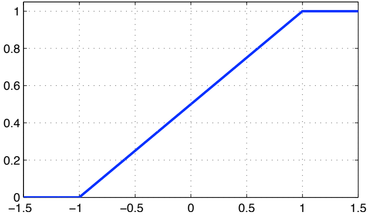
- La cdf es $F(x) = (x - a)/(b - a)$ en el intervalo $[a, b]$ (0 antes, 1 después)
- La prob. del intervalo $[\alpha, \beta] \subseteq [a, b]$ es $\int_\alpha^\beta f(x)dx = (\beta - \alpha)/(b - a)$

  $\Rightarrow$ Solo depende del ancho $\beta - \alpha$ del intervalo, no de su posición

pdf $(a = -1,\ b = 1)$



cdf $(a = -1,\ b = 1)$



Distribución exponential

- Duración de llamadas de teléfono, vida de componentes electrónicos
- La pdf de una VA exponencial es

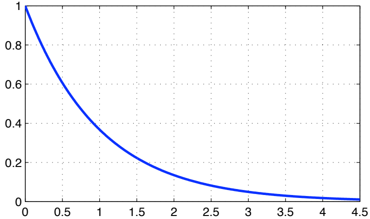
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$\Rightarrow$ Cuando λ crece, al "altura" aumenta y el "ancho" disminuye

- La cdf se obtiene integrando la pdf

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{0}^{x} \lambda e^{-\lambda u}\,du = -e^{-\lambda u}\Big|_0^x = 1 - e^{-\lambda x}$$

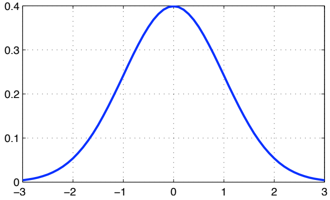
pdf ($\lambda = 1$) cdf ($\lambda = 1$)



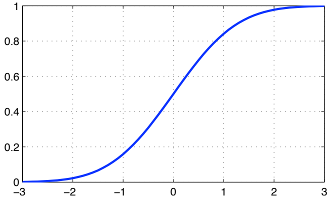
Distribución Normal o Gaussiana

- Modela la aleatoriedad que surge de un gran número de efectos aleatorios
- La pdf de una VA normal es $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
  - $\Rightarrow$ μ es la media (centro), σ^2 es la varianza (ancho)
  - $\Rightarrow$ 0,68 prob. between $\mu \pm \sigma$, 0,997 prob. in $\mu \pm 3\sigma$
  - $\Rightarrow$ Una VA normal estándar tiene $\mu = 0$ y $\sigma^2 = 1$
- La cdf $F(x)$ no se puede expresar en términos de funciones elementales

pdf ($\mu = 0$, $\sigma^2 = 1$)



cdf ($\mu = 0$, $\sigma^2 = 1$)



Valores esperados

Valores esperados

- Se nos pide que resumamos la información sobre una VA en un único valor,

  $\Rightarrow$ ¿cuál debería ser ese valor?

- Si se nos pidiera una descripción con unos pocos valores,

  $\Rightarrow$ ¿cuáles deberían ser esos valores?

- Los valores esperados (medias) son respuestas convenientes a estas preguntas

- **Atención:** Las esperanzas son descripciones condensadas

  $\Rightarrow$ No capturan todos los aspectos del fenómeno aleatorio

  $\Rightarrow$ La historia completa es contada por la distribución de probabilidad (cdf)

Definición para VAs discretas

- La VA discreta X toma valores x_i, $i = 1, 2, \ldots$ con pmf $p(x)$
- **Def:** El valor esperado de la VA discreta X es

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p(x_i) = \sum_{x: p(x)>0} x p(x)$$

- Es el promedio ponderado sobre los valores posibles x_i. Las probabilidades son pesos
- Es el promedio común si la VA toma valores x_i, $i = 1, \ldots, N$ equiprobables

$$\mathbb{E}[X] = \sum_{i=1}^{N} x_i p(x_i) = \sum_{i=1}^{N} x_i \frac{1}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Valor esperado de VAs de Bernoulli y geométrica

Ej: Para una VA Bernoulli $p(x) = p^x q^{1-x}$, for $x \in \{0, 1\}$

$$\mathbb{E}[X] = 1 \times p + 0 \times q = p$$

Ej: Para una VA geométrica $p(x) = p(1-p)^{x-1} = pq^{x-1}$, para $x \geq 1$

- Notar que $\partial q^x / \partial q = xq^{x-1}$ y que las derivadas son operadores lineales

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} \frac{\partial q^x}{\partial q} = p \frac{\partial}{\partial q} \left(\sum_{x=1}^{\infty} q^x \right)$$

- La suma dentro de la derivada es una geométrica. Vale $q/(1-q)$, de dónde

$$\mathbb{E}[X] = p \frac{\partial}{\partial q} \left(\frac{q}{1-q} \right) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

- El tiempo hasta el primer acierto es el inverso de la probabilidad de acierto. Razonable

Valor esperado de una VA Poisson

Ej: Para una VA Poisson $p(x) = e^{-\lambda}(\lambda^x/x!)$, for $x \geq 0$

- El primer sumando en la definición es 0; luego sacamos λ para afuera y usamos $\frac{x}{x!} = \frac{1}{(x-1)!}$

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

- La suma es la serie de Taylor de $e^{\lambda} = 1 + \lambda + \lambda^2/2! + \ldots + \lambda^x/x!$

$$\mathbb{E}[X] = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

- Poisson es el límite de la binomial para grandes números n, con $\lambda = np$

  $\Rightarrow$ Cuenta cantidad de aciertos en n experimentos, con probabilidad p

- El número esperado de aciertos es $\lambda = np$

  $\Rightarrow$ Cantidad de experimentos $\times$ probabilidad de acierto individual. Razonable

Definición para VAs continuas

- VA continua X toma valores en $\mathbb{R}$ con pdf $f(x)$
- **Def:** El valor esperado de la VA continua X es

$$\mathbb{E}\left[X\right] := \int_{-\infty}^{\infty} x f(x)\, dx$$

- Comparar con $\mathbb{E}\left[X\right] := \sum_{x:p(x)>0} x p(x)$ para el caso discreto
- Notar que la integral y la suma se suponen definidas

  $\Rightarrow$ De no ser así decimos que la esperanza no existe

Valor esperado de una VA normal

Ej: Para una VA normal, sumamos y restamos μ, y separamos integrales

$$
\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x+\mu-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \mu\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx
\end{aligned}
$$

- La primera integral vale 1 ya que integra una pdf en todo $\mathbb{R}$
- La segunda integral es 0 por simetría. Resulta

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

- La media de una VA con pdf simétrica es el punto de simetría

Valores esperados de VAs uniforme y exponencial

Ej: Para una VA uniforme $f(x) = 1/(b-a)$, para todo $a \leq x \leq b$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{(a+b)}{2}$$

- Tiene sentido ya que la pdf es simétrica con respecto al punto medio $(a+b)/2$

Ex: Para una VA exponencial integramos por partes

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{\infty} x\lambda e^{-\lambda x}\, dx \\ &= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx \\ &= -xe^{-\lambda x}\Big|_0^{\infty} - \frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty} = \frac{1}{\lambda}\end{aligned}$$

Valor esperado de una función de una VA

- Consideremos una función $g(X)$ de una VA X. ¿Valor esperado de $g(X)$?
- g(X) es una VA, por lo que tiene una pmf $p_{g(X)}(g(x))$

$$\mathbb{E}\left[g(X)\right] = \sum_{g(x):p_{g(X)}(g(x))>0} g(x)p_{g(X)}(g(x))$$

$\Rightarrow$ Requiere calcular la pmf de $g(X)$. Hay una forma simple de hacerlo

Teorema

Sea $g(X)$ una función de una VA discreta X con pmf $p_X(x)$. Entonces

$$\mathbb{E}\left[g(X)\right] = \sum_{i=1}^{\infty} g(x_i)p_X(x_i)$$

- Suma ponderada de valores funcionales. No es necesario calcular la pmf de $g(X)$
- Lo mismo vale para X va continua

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$$

Valor esperado para una función lineal de una VA

- Consideremos la función afín $g(X) = aX + b$

$$\begin{aligned}
\mathbb{E}\left[aX + b\right] &= \sum_{i=1}^{\infty} (ax_i + b) p_X(x_i) \\
&= \sum_{i=1}^{\infty} a x_i p_X(x_i) + \sum_{i=1}^{\infty} b p_X(x_i) \\
&= a \sum_{i=1}^{\infty} x_i p_X(x_i) + b \sum_{i=1}^{\infty} p_X(x_i) \\
&= a\mathbb{E}\left[X\right] + b1
\end{aligned}$$

- Se puede intercambiar la esperanza con constantes aditivas o multiplicativas

$$\mathbb{E}\left[aX + b\right] = a\mathbb{E}\left[X\right] + b$$

$\Rightarrow$ Vale lo mismo para VAs continuas (linealidad de la integral)

Valor esperado de una función indicatriz

- Sean X una VA y $\mathcal{X}$ un conjunto

$$\mathbb{I}\{X \in \mathcal{X}\} = \begin{cases} 1, & \text{if } x \in \mathcal{X} \\ 0, & \text{if } x \notin \mathcal{X} \end{cases}$$

- El valor esperado de $\mathbb{I}\{X \in \mathcal{X}\}$ en el caso discreto es

$$\mathbb{E}[\mathbb{I}\{X \in \mathcal{X}\}] = \sum_{x: p_X(x)>0} \mathbb{I}\{x \in \mathcal{X}\} p_X(x) = \sum_{x \in \mathcal{X}} p_X(x) = \mathrm{P}(X \in \mathcal{X})$$

- De la misma forma, para el caso continuo

$$\mathbb{E}[\mathbb{I}\{X \in \mathcal{X}\}] = \int_{-\infty}^{\infty} \mathbb{I}\{x \in \mathcal{X}\} f_X(x) dx = \int_{x \in \mathcal{X}} f_X(x) dx = \mathrm{P}(X \in \mathcal{X})$$

- Valor esperado de la indicatriz = Probabilidad del suceso correspondiente

  $\Rightarrow$ Recordar que $\mathbb{E}[X] = p$ para una VA Bernoulli ("indica acierto")

Momentos, momentos centrados y varianza

- **Def:** El momento de orden n ($n \geq 0$) de una VA es

$$\mathbb{E}\left[X^n\right] = \sum_{i=1}^{\infty} x_i^n p(x_i)$$

- **Def:** El momento centrado de orden n corrige el sesgo de la media, esto es

$$\mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^n\right] = \sum_{i=1}^{\infty} \left(x_i - \mathbb{E}\left[X\right]\right)^n p(x_i)$$

- Momento de orden 0: $\mathbb{E}\left[X^0\right] = 1$; momento de orden 1: es la media $\mathbb{E}\left[X\right]$
- El momento centrado de orden 2 es la varianza. Mide el ancho de la pmf

$$\mathrm{var}\left[X\right] = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]$$

Ej: Para funciones afines

$$\mathrm{var}\left[aX + b\right] = a^2 \mathrm{var}\left[X\right]$$

Varianza de VAs Bernoulli y Poisson

Ex: Para una VA Bernoulli X de parámetro p, $\mathbb{E}[X] = \mathbb{E}\left[X^2\right] = p$

$$\Rightarrow \operatorname{var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] = p - p^2 = p(1-p)$$

Ej: Para una VA Poisson Y de parámetro λ, el segundo momento es

$$\begin{aligned}
\mathbb{E}\left[Y^2\right] &= \sum_{y=0}^{\infty} y^2 e^{-\lambda} \frac{\lambda^y}{y!} = \sum_{y=1}^{\infty} y \frac{e^{-\lambda}\lambda^y}{(y-1)!} \\
&= \sum_{y=1}^{\infty} (y-1) \frac{e^{-\lambda}\lambda^y}{(y-1)!} + \sum_{y=1}^{\infty} \frac{e^{-\lambda}\lambda^y}{(y-1)!} \\
&= e^{-\lambda}\lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} + e^{-\lambda}\lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} \\
&= e^{-\lambda}\lambda^2 e^{\lambda} + e^{-\lambda}\lambda e^{\lambda} = \lambda^2 + \lambda
\end{aligned}$$

$$\Rightarrow \operatorname{var}[Y] = \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Distribuciones de probabilidad conjunta

cdf conjunta

- Queremos estudiar problemas con más de una VA, e.g. X e Y
- Las distribuciones de probabilidad de X e Y no son suficientes

  $\Rightarrow$ La distribución de probabilidad conjunta (cdf) of (X, Y) se define como

$$F_{XY}(x, y) = \mathrm{P}\left(X \leq x, Y \leq y\right)$$

- If X, Y claras del contexto, omitimos el subíndice: $F_{XY}(x, y) = F(x, y)$
- Podemos recuperar $F_X(x)$ considerando todos los valores posibles de Y

$$F_X(x) = \mathrm{P}\left(X \leq x\right) = \mathrm{P}\left(X \leq x, Y \leq \infty\right) = F_{XY}(x, \infty)$$

  $\Rightarrow$ $F_X(x)$ y $F_Y(y) = F_{XY}(\infty, y)$ se llaman cdfs marginales

pmf conjunta

- Consideremos las VAs discretas X e Y
  X toma valores en $\mathcal{X} := \{x_1, x_2, \ldots\}$ e Y en $\mathcal{Y} := \{y_1, y_2, \ldots\}$
- La pmf conjunta de (X, Y) se define como

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

- Los valores posibles de (x, y) son elementos del producto Cartesiano $\mathcal{X} \times \mathcal{Y}$
  - (x_1, y_1), (x_1, y_2), $\ldots$, (x_2, y_1), (x_2, y_2), $\ldots$, (x_3, y_1), (x_3, y_2), $\ldots$
- La pmf marginal $p_X(x)$ se obtiene sumando sobre todos los valores posibles de Y

$$p_X(x) = \mathrm{P}\left(X = x\right) = \sum_{y \in \mathcal{Y}} \mathrm{P}\left(X = x, Y = y\right) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$$

$\Rightarrow$ De la misma forma $p_Y(y) = \sum_{x \in \mathcal{X}} p_{XY}(x, y)$. Marginalizamos sumando

pdf conjunta

- Sean X, Y VAs continuas y $\mathcal{A} \subset \mathbb{R}^2$ un conjunto
- La pdf conjunta es una función $f_{XY}(x,y) : \mathbb{R}^2 \to \mathbb{R}^+$ tal que

$$\mathrm{P}\left((X,Y) \in \mathcal{A}\right) = \iint_{\mathcal{A}} f_{XY}(x,y)\, dxdy$$

- Marginalización. Hay dos formas de escribir $\mathrm{P}\left(X \in \mathcal{X}\right)$

$$\mathrm{P}\left(X \in \mathcal{X}\right) = \mathrm{P}\left(X \in \mathcal{X}, Y \in \mathbb{R}\right) = \int_{X \in \mathcal{X}} \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy\, dx$$

$\Rightarrow$ Definición de $f_X(x)$ $\Rightarrow \mathrm{P}\left(X \in \mathcal{X}\right) = \int_{X \in \mathcal{X}} f_X(x)\, dx$

- Tenemos entonces

$$\Rightarrow f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx$$

Ejemplo

- Sean B_1, B_2 VAs Bernoulli de igual parámetro p
  $\Rightarrow$ Definimos $X = B_1,\ Y = B_1 + B_2$
- La pmf de X es

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

- De la misma forma, la pmf de Y es

$$p_Y(0) = (1-p)^2, \quad p_Y(1) = 2p(1-p), \quad p_Y(2) = p^2$$

- La pmf conjunta de X e Y es

$$\begin{aligned} &p_{XY}(0,0) = (1-p)^2, && p_{XY}(0,1) = p(1-p), && p_{XY}(0,2) = 0 \\ &p_{XY}(1,0) = 0, && p_{XY}(1,1) = p(1-p), && p_{XY}(1,2) = p^2 \end{aligned}$$

Vectores aleatorios

- Por conveniencia es común agrupar las VAs en un vector

  $\Rightarrow$ La distribución de probabilidad de un vector es la distribución conjunta de sus elementos

- Sean por ejemplo dos VAs X e Y. El vector aleatorio es $\mathbf{X} = [X, Y]^T$
- Si X e Y son discretas, el vector $\mathbf{X}$ es discreto con pmf

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}\left([x, y]^T\right) = p_{XY}(x, y)$$

- Si X e Y son continuas, $\mathbf{X}$ es continuo con pdf

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}\left([x, y]^T\right) = f_{XY}(x, y)$$

- La cdf del vector es $\Rightarrow F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}\left([x, y]^T\right) = F_{XY}(x, y)$
- En general podemos definir VAs n-dimensionales $\mathbf{X} := [X_1, X_2, \ldots, X_n]^T$

  $\Rightarrow$ Es solo notación, las definiciones siguen del caso $n = 2$

Esperanzas conjuntas

Esperanzas conjuntas

- X e Y VAs, $g(X, Y)$ función es también una VA
- El valor esperado de $g(X, Y)$ para X, Y VAs discretas se escribe

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{x,y:p_{XY}(x,y)>0} g(x, y)p_{XY}(x, y)$$

- Si X, Y VAs continuas

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f_{XY}(x, y)\, dxdy$$

$\Rightarrow$ Podemos tener más de dos VAs y usar notación vectorial

Ej: Transformación lineal de un vector aleatorio $\mathbf{X} \in \mathbb{R}^n$: $g(\mathbf{X}) = \mathbf{a}^T\mathbf{X}$

$$\Rightarrow \mathbb{E}\left[\mathbf{a}^T\mathbf{X}\right] = \int_{\mathbb{R}^n} \mathbf{a}^T\mathbf{x}f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

Valor esperado de una suma de variables aleatorias

- Valor esperado de suma de dos VAs

$$\mathbb{E}[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{XY}(x,y)\, dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{XY}(x,y)\, dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{XY}(x,y)\, dxdy$$

- Luego,

$$\mathbb{E}[X+Y] = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy\, dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy$$
$$= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy$$
$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

$\Rightarrow$ Usamos expresiones marginales

- Esperanza $\leftrightarrow$ suma $\Rightarrow \mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i]$

La esperanza es un operador lineal

- Combinando $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ y $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$ probamos que

$$\mathbb{E}[a_x X + a_y Y + b] = a_x \mathbb{E}[X] + a_y \mathbb{E}[Y] + b$$

- En notación vectorial ($\mathbf{a} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^n$, b escalar)

$$\mathbb{E}\left[\mathbf{a}^T\mathbf{X} + b\right] = \mathbf{a}^T\mathbb{E}[\mathbf{X}] + b$$

- También, si $\mathbf{A}$ es una matriz $m \times n$ con filas $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ y $\mathbf{b} \in \mathbb{R}^m$ un vector con elementos $b_1, \ldots, b_m$, podemos escribir

$$\mathbb{E}[\mathbf{AX} + \mathbf{b}] = \begin{pmatrix} \mathbb{E}\left[\mathbf{a}_1^T\mathbf{X} + b_1\right] \\ \mathbb{E}\left[\mathbf{a}_2^T\mathbf{X} + b_2\right] \\ \vdots \\ \mathbb{E}\left[\mathbf{a}_m^T\mathbf{X} + b_m\right] \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^T\mathbb{E}[\mathbf{X}] + b_1 \\ \mathbf{a}_2^T\mathbb{E}[\mathbf{X}] + b_2 \\ \vdots \\ \mathbf{a}_m^T\mathbb{E}[\mathbf{X}] + b_m \end{pmatrix} = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}$$

- La esperanza conmuta con operadores lineales

Independencia de VAs

- Los suceso E y F son independientes si $\mathrm{P}(E \cap F) = \mathrm{P}(E)\,\mathrm{P}(F)$
- **Def:** Las VAs X e Y son independientes si los sucesos $\{X \leq x\}$ y $\{Y \leq y\}$ son independiente cualquiera sea x e y, i.e.

$$\mathrm{P}(X \leq x, Y \leq y) = \mathrm{P}(X \leq x)\,\mathrm{P}(Y \leq y)$$

$\Rightarrow$ Por definición, esto es equivalente a $F_{XY}(x, y) = F_X(x)F_Y(y)$

- Para VAs discretas, es equivalente a la relación análoga con pmfs

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$

- Para VAs continuas vale en análogo para pdfs

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

- Independencia $\Leftrightarrow$ La distribución conjunta se factoriza como producto de marginales

Suma de VAs de Poisson independientes

- X, Y VAs Poisson de parámetros λ_x y λ_y, Independientes
- Q: ¿Distribución de probabilidad de la VA $Z := X + Y$?
- $Z = n$ solo si $X = k$, $Y = n - k$ para algún $0 \leq k \leq n$

$$
\begin{aligned}
p_Z(n) &= \sum_{k=0}^{n} \mathrm{P}\left(X = k, Y = n-k\right) \quad = \sum_{k=0}^{n} \mathrm{P}\left(X = k\right) \mathrm{P}\left(Y = n-k\right) \\
&= \sum_{k=0}^{n} e^{-\lambda_x} \frac{\lambda_x^k}{k!} e^{-\lambda_y} \frac{\lambda_y^{n-k}}{(n-k)!} \quad = \frac{e^{-(\lambda_x+\lambda_y)}}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} \lambda_x^k \lambda_y^{n-k} \\
&= \frac{e^{-(\lambda_x+\lambda_y)}}{n!} (\lambda_x + \lambda_y)^n
\end{aligned}
$$

- Z es Poisson de parámetro $\lambda_z := \lambda_x + \lambda_y$

  $\Rightarrow$ La suma de VAs de Poisson independientes es Poisson (el parámetro es la suma de los parámetros originales)

Valor esperado de una VA binomial

- Las VAs binomiales cuentan cantidad de aciertos en n experimentos

Ej: Sean X_i, $i = 1, \ldots n$ VAs de Bernoulli independientes

- Podemos escribir la binomial como $X = \sum_{i=1}^{n} X_i$

$$\Rightarrow \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$$

- Cantidad esperada de aciertos = cantidad de experimentos $\times$ prob. individual de acierto
  - Misma interpretación que vimos para VAs de Poisson

Ej: VAs Bernoulli dependientes. $Y = \sum_{i=1}^{n} X_i$, pero los X_i no son independientes

- La cantidad esperada de aciertos sigue siendo $\mathbb{E}[Y] = np$
  - La linealidad de la esperanza no requiere independencia
  - Y no sigue una distribución binomial

Valor esperado de un producto de VAs independientes

Teorema

Para VAs X, Y independientes y funciones $g(X)$, $h(Y)$arbitrarias:

$$\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

El valor esperado del producto es el producto de los valores esperados

- Se demuestra que $g(X)$ y $h(Y)$ también son independientes. Intuitivo

Ej: Caso especial cuando $g(X) = X$ y $h(Y) = Y$ se tiene

$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$

- Esperanzas y productos se pueden intercambiar si las VAs son independientes
- Es distinto cuando las operaciones son lineales (siempre es posible conmutar)

Valor esperado de un producto de VAs independientes (cont.)

Demostración.

- Supongamos X, Y VAs continuas. Usamos la definición de independencia

$$\begin{aligned}\mathbb{E}\left[g(X)h(Y)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dxdy\end{aligned}$$

- El integrando es el producto de una función de x y una función de y

$$\begin{aligned}\mathbb{E}\left[g(X)h(Y)\right] &= \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy \\ &= \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]\end{aligned}$$

□

Varianza de una suma de VAs independientes

- Sean X_n, $n = 1, \ldots N$ independientes, con $\mathbb{E}[X_n] = \mu_n$, $\text{var}[X_n] = \sigma_n^2$
- Q: ¿Varianza de la suma $X := \sum_{n=1}^{N} X_n$?
- Nótese que la media de X es $\mathbb{E}[X] = \sum_{n=1}^{N} \mu_n$. Luego

$$\text{var}[X] = \mathbb{E}\left[\left(\sum_{n=1}^{N} X_n - \sum_{n=1}^{N} \mu_n\right)^2\right] = \mathbb{E}\left[\left(\sum_{n=1}^{N} (X_n - \mu_n)\right)^2\right]$$

- Desarrollando e intercambiando suma y esperanza

$$\text{var}[X] = \sum_{n=1}^{N} \sum_{m=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)(X_m - \mu_m)\Big]$$

Varianza de una suma de VAs independientes (cont.)

- Separamos auto-productos y productos cruzados. Luego usamos independencia, y $\mathbb{E}(X_n - \mu_n) = 0$

$$\begin{aligned}\operatorname{var}[X] &= \sum_{n=1, n\neq m}^{N} \sum_{m=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)(X_m - \mu_m)\Big] + \sum_{n=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)^2\Big] \\ &= \sum_{n=1, n\neq m}^{N} \sum_{m=1}^{N} \mathbb{E}(X_n - \mu_n)\mathbb{E}(X_m - \mu_m) + \sum_{n=1}^{N} \sigma_n^2 = \sum_{n=1}^{N} \sigma_n^2\end{aligned}$$

- Si las VAs son independientes $\Rightarrow$ La varianza de la suma es la suma de las varianzas
- Hay resultados más generales para VAs independientes X_i, $i = 1, \ldots, n$

$$\operatorname{var}\left[\sum_i (a_i X_i + b_i)\right] = \sum_i a_i^2 \operatorname{var}[X_i]$$

Varianza de una VA binomial y media muestral

Ej: Sean X_i, $i = 1, \ldots n$ VAs Bernoulli independientes de parámetro p

$\Rightarrow$ Recordemos que $\mathbb{E}\left[X_i\right] = p$ y $\operatorname{var}\left[X_i\right] = p(1-p)$

- Escribimos la VA binomial X de parámetros (n, p) como: $X = \sum_{i=1}^{n} X_i$
- La varianza de la binomial es entonces

$$\Rightarrow \operatorname{var}\left[X\right] = \sum_{i=1}^{n} \operatorname{var}\left[X_i\right] = np(1-p)$$

Ej: Sean Y_i, $i = 1, \ldots n$ VAs independientes con $\mathbb{E}\left[Y_i\right] = \mu$, $\operatorname{var}\left[Y_i\right] = \sigma^2$

- La media muestral es $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. ¿Cuánto valen $\mathbb{E}\left[\bar{Y}\right]$ y $\operatorname{var}\left[\bar{Y}\right]$?
- Valor esperado $\Rightarrow \mathbb{E}\left[\bar{Y}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[Y_i\right] = \mu$
- Varianza $\Rightarrow \operatorname{var}\left[\bar{Y}\right] = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}\left[Y_i\right] = \frac{\sigma^2}{n}$ (independencia)

Covarianza

- **Def:** La covarianza de X and Y es (generaliza la varianza a pares de VAs)

$$\text{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

- Si $\text{cov}(X, Y) = 0$ las VAs X, Y se dicen no correlacionadas
- Si X, Y son independientes entonces $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ y $\text{cov}(X, Y) = 0$

  $\Rightarrow$ La independencia implica no correlación
- El inverso no vale, puede ser $\text{cov}(X, Y) = 0$ para X, Y dependientes
  - Ej: X uniforme en $[-a, a]$ y $Y = X^2$

  $\Rightarrow$ Pero si X, Y son normales, correlación nula implica independencia
- Si $\text{cov}(X, Y) > 0$ entonces X e Y tienden a moverse en la misma dirección

  $\Rightarrow$ Correlación positiva
- Si $\text{cov}(X, Y) < 0$ entonces X e Y tienden a moverse en direcciones opuestas

  $\Rightarrow$ Correlación negativa

Ejemplo de covarianza

- Sea X señal aleatoria de media nula y Z ruido de media nula

  $\Rightarrow$ Supongamos que la señal X y el ruido Z son independientes

- Consideremos las señales recibidas $Y_1 = X + Z$ e $Y_2 = -X + Z$

(I) Y_1 y X están correlacionadas positivamente (X, Y_1 se mueven en la misma dirección)

$$\begin{aligned} \operatorname{cov}(X, Y_1) &= \mathbb{E}[XY_1] - \mathbb{E}[X]\,\mathbb{E}[Y_1] \\ &= \mathbb{E}[X(X+Z)] - \mathbb{E}[X]\,\mathbb{E}[X+Z] \end{aligned}$$

- El segundo término es 0 ($\mathbb{E}[X] = 0$). Para el primer término usamos la independencia de X, Z

$$\mathbb{E}[X(X+Z)] = \mathbb{E}\left[X^2\right] + \mathbb{E}[X]\,\mathbb{E}[Z] = \mathbb{E}\left[X^2\right]$$

- Combinando observaciones $\Rightarrow \operatorname{cov}(X, Y_1) = \mathbb{E}\left[X^2\right] > 0$

Ejemplo de covarianza (cont.)

(II) Y_2 y X están correlacionadas negativamente (X, Y_2 se mueven en direcciones opuestas)

- Mismos cálculos $\Rightarrow \text{cov}(X, Y_2) = -\mathbb{E}\left[X^2\right] < 0$

(III) También podemos calcular la correlación entre Y_1 y Y_2

$$\begin{aligned}\text{cov}(Y_1, Y_2) &= \mathbb{E}\left[(X+Z)(-X+Z)\right] - \mathbb{E}\left[(X+Z)\right]\mathbb{E}\left[(-X+Z)\right] \\ &= -\mathbb{E}\left[X^2\right] + \mathbb{E}\left[Z^2\right]\end{aligned}$$

$\Rightarrow$ Correlación negativa si $\mathbb{E}\left[X^2\right] > \mathbb{E}\left[Z^2\right]$ (ruido débil)

$\Rightarrow$ Correlación positiva si $\mathbb{E}\left[X^2\right] < \mathbb{E}\left[Z^2\right]$ (ruido fuerte)

- Correlación entre X e Y_1 o entre X e Y_2 viene de la causalidad
- Correlación entre Y_1 e Y_2 no viene de la causalidad. Variables latentes X y Z

$\Rightarrow$ La correlación no implica causalidad

Es un posible modelo (simple) de canal de comunicación