Segundo cuatrimestre 2016

Teoricas de Probabilidad y Estadistica M

Pablo A. Ferrari

Actualizadas 22 de febrero de 2017

Contenidos

Prefacio

Estas notas son la base de las teóricas del curso de probabilidad y estadística para Matemáticos en la licenciatura de Matemática de la Universidad de Buenos Aires, la última vez en 2015.

Se apoyan en notas no publicadas de Pablo Groisman y de Victor Yohai [13]. Usé también material del curso de Leonardo Rolla [10]. Parte de la sección sobre estadística la saqué de las notas de Ana Bianco y Elena Martínez [2]. Insoslayable es el libro de William Feller [5] del cual existe una versión en castellano que no encontré.

En relación al programa usual de la materia, le agregué una introducción a los procesos estocásticos, el proceso de Bernoulli, Binomial y el proceso de Poisson como límite de los procesos binomiales. Además hay material para dos clases de cadenas de Markov, con una demostración formal de la ley de grandes números.

Motivación

1. **Urna de Eherenfest** Paul y Tatiane Ehrenfest 1907, [4].

$2N$ bolas rojas numeradas. Dos urnas. Urna 0, Urna 1.

Elijo bola y cambio de urna.

$X_n = (X_n(1), \dots, X_n(2))$ vector de 0 y 1.

Despues de mucho tiempo cada bola tiene la misma probabilidad de estar en cada una de las urnas. Cada configuración $1/2^{2N}$.

La probabilidad de la mitad en cada urna

$$\binom{2N}{N}\frac{1}{2^N}$$

La probabilidad de todas en la urna 0:

$$\frac{1}{2^N}$$

Tiempo de retorno a una configuración determinada.

$$T = \frac{1}{\text{probabilidad de la config}}$$

Entonces el tiempo para volver a "todas en urna 0" es 2^{2N}.

Como $N = 2^{26}$, número de moleculas de un gas en una caja.

T= 10^{6000}. Se calcula la edad del universo en 10^9.

2. **Ranking de páginas de Internet**

Las páginas de internet forman un grafo orientado.

Vertices = páginas,

Aristas orientadas cuando una página tiene un link para otra.

Cómo ranquear las páginas? Páginas con más links llegando deberían tener mejor ranking.

Se propone una dinámica para recorrer el grafo. "Paseante aleatorio"

Cuando el paseante está en un nodo, elije una de las aristas que emanan con probabilidad $1/N$, N es el número de aristas emanantes.

Se lo deja paseando un tiempo "infinito" y se estudia la estadística de los sitios visitados.

Se ranquea de mayor a menor.

X_n es una cadena de Markov. No importa de donde vino para saber adonde va.

Se plantea una ecuacion de balance:

$$m(v) = \sum_w m(w)p(w,v)$$

Donde $p(w,v)$ etc.

La proporcion de veces que el paseante visita v tiende a la solución de esa ecuación. Teorema ergódico.

Modificación: con probabilidad $1-\epsilon$ hace lo de arriba y con probabilidad ϵ elige una página al azar entre todas las del grafo.

$$m(v) = \sum_w ((1-\epsilon)m(w)p(w,v) + \epsilon \frac{1}{D}$$

donde D es el número total de nodos.

1. Probabilidad. Definición y propiedades.

1.1. Espacio muestral

1. Dados. Interesa el número que sale. $\Omega = \{1, \dots, 6\}$.

2. Tirar el dado 10 veces $\Omega = \{1, \dots, 6\}^{10}$.

3. Si nos interesa el número de veces que sale cada cara:

$$\Omega = \{(k_1, \dots, k_6) : \sum_{i=1}^{6} k_i = 10\}$$

donde k_i representa número de veces que sale la cara i.

4. Lanzar una moneda n veces. Cara = 1, Ceca = 0. $\Omega = \{0,1\}^n$

5. Lanzar una moneda infinitas veces. $\Omega = \{0,1\}^{\mathbb{N}}$

6. Observar la duración de una batería de celular. $\Omega = [0, \infty)$ (continuo).

El **espacio muestral** es el *conjunto de resultados posibles de un experimento.*

Evento es un subconjunto del espacio muestral.

Operaciones con eventos

Unión, intersección, uniones e intersecciones numerables, complementos.

Ω es el evento cierto o seguro.

$\emptyset$ es el evento imposible.

$A \cup B$ Unión: Ocurre A ó B.

$A \cap B$ Intersección: Ocurre A y B.

A^c Complemento de A. No ocurre A.

$A \setminus B = A \cap B^c$. Diferencia: Ocurre A y no ocurre B.

Se dice que $A \subset B$ si cuando ocurre A, también ocurre B.

A y B son mutuamente excluyentes o disjuntos si $A \cap B = \emptyset$.

Propiedades:

Asociatividad: $A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$

$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$

Conmutatividad: $A \cup B = B \cup A, \quad A \cap B = B \cap A$

Distributividad: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

Leyes de **De Morgan**:

$$\Big(\bigcup_i A_i\Big)^c = \bigcap_i A_i^c, \qquad \Big(\bigcap_i A_i\Big)^c = \bigcup_i A_i^c$$

1.2. Probabilidad

Interpretación intuitiva Se repite n veces un mismo experimento aleatorio en forma independiente y bajo las mismas condiciones.

n_A: número de veces que ocurre A.

Frecuencia relativa de A:

$$\mathrm{fr}(A) = \frac{n_A}{n}$$

La evidencia empírica muestra que cuando n crece, $\mathrm{fr}(A)$ tiende a estabilizarse alrededor de un número $P(A)$.

Propiedades

1) $\mathrm{fr}(A)$ está entre 0 y 1

2) $\mathrm{fr}(\Omega) = 1$

3) Si $A \cap B = \emptyset$,

$$\mathrm{fr}(A \cup B) = \frac{n_{A\cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} = \mathrm{fr}(A) + \mathrm{fr}(B).$$

Axiomas de Probabilidad: Dado un experimento con espacio muestral Ω, consideramos $\mathcal{F} \subset \mathcal{P}(\Omega)$ un subconjunto de las partes, cuyos elementos se llaman *eventos*.

Una *Probabilidad* P es una función

$P : \mathcal{F} \to [0, 1]$

que satisface los siguientes **axiomas**:

A1. $P(A) \in [0, 1]$ para todo evento A.

A2. $P(\Omega) = 1$

A3. Si $A_1, A_2, \ldots$ mutuamente excluyentes (es decir si $A_i \cap A_j = \emptyset$, si $i \neq j$), entonces

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i)$$

Ejemplo: Moneda. $\Omega = \{cara, ceca\} = \{1, 0\}$. $P(\{1\}) = p$ y $P(\{0\}) = 1 - p$, $P(\{0, 1\}) = 1$, $P(\emptyset) = 0$, con $0 \leq p \leq 1$, satisface los axiomas.

1.3. Propiedades

1) $P(A^c) = 1 - P(A)$ para todo evento A

2) $P(\emptyset) = 0$

3) Si $A \subset B \Rightarrow P(A) \leq P(B)$ y $P(B \setminus A) = P(B) - P(A)$

Dem: Si $A \subset B \Rightarrow B = A \cup (B \setminus A)$ y éstos dos eventos son excluyentes. Por el axioma A3 $P(B) = P(A) + P(B \setminus A)$ Dado que, por el axioma A1, $P(B \setminus A) \geq 0$, resulta $P(B) \geq P(A)$ y, despejando, se obtiene la segunda afirmación.

4) Dados dos eventos cualesquiera A y B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dem: $A \cup B = A \cup (B \setminus A) = A \cup (B \cap A^c)$ y estos dos eventos son excluyentes, entonces, por el axioma A3,

$$P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \tag{1}$$

Por otra parte, $B = (B \cap A) \cup (B \cap A^c)$ y estos dos eventos son disjuntos, entonces

$$P(B) = P(B \cap A) + P(B \cap A^c) \Rightarrow P(B \cap A^c) = P(B) - P(B \cap A) \tag{2}$$

De (1) y (2) resulta que $P(A \cup B) = P(A) + P(B) - P(B \cap A)$ como queríamos demostrar.

5) Dados dos eventos cualesquiera A y B, $P(A \cup B) \leq P(A) + P(B)$.

Se deduce inmediatamente de 4) y de A1.

Ejercicios: a) Demostrar, usando la propiedad 4) que, dados tres eventos cualesquiera,

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_1 \cap A_2) + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

b) Probar, usando inducción que, dados $A_1, A_2, \ldots$ eventos cualesquiera,

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$$

6) Monotonía creciente:

Si $A_n \supset A_{n-1}$ y definimos $A := \cup_n A_n$, entonces $P(A) = \lim_n P(A_n)$

Dem: $A = A_1 \dot{\cup} \big(\dot{\cup}_{i=1}^{\infty} [A_{i+1} \setminus A_i]\big)$

$$\begin{aligned}
P(A) &= P(A_1) + \sum_{i=1}^{\infty} P(A_{i+1} \setminus A_i) = P(A_1) + \sum_{i=1}^{\infty} [P(A_{i+1}) - P(A_i)] \\
&= P(A_1) + \lim_n \sum_{i=1}^{n} [P(A_{i+1}) - P(A_i)] = \lim_n P(A_n)
\end{aligned}$$

porque $P(A_n) = P(A_1) + \sum_{i=1}^{n} [P(A_{i+1}) - P(A_i)]$, por suma telescópica.

7) Monotonía decreciente: Si $A_n \subset A_{n-1}$ entonces $P(\cap_n A_n) = \lim_n P(A_n)$. Queda como ejercicio: sigue de 6) y De Morgan.

1.4. Espacios discretos

Si Ω es finito o numerable, diremos que es *discreto*. En este caso $\mathcal{F} = \mathcal{P}(\Omega)$. Dada una *función de probabilidad puntual*

$$p : \Omega \to [0,1] \quad \text{tal que} \quad \sum_{\omega} p(\omega) = 1,$$

se define

$$P(A) := \sum_{\omega \in A} p(\omega)$$

P satisface los axiomas: $P(\Omega) = 1$ y para A_i disjuntos,

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{\omega : \omega \in \cup_i A_i} p(\omega) = \sum_i \sum_{\omega : \omega \in A_i} p(\omega) = \sum_{i=1}^{\infty} P(A_i).$$

Ejemplos: 1) Dado equilibrado. $\Omega = \{1,2,3,4,5,6\}$ con $p(\omega) = 1/6$ para $\omega = 1, \dots, 6$. Por ejemplo:

$$P(\text{el resultado es par}) = P(\{2,4,6\}) = p(2) + p(4) + p(6) = 1/2$$

2) Dado en el cual la probabilidad de las caras pares es el doble que la probabilidad de las caras impares:

$$p(1) = p(3) = p(5) = q, \;\; p(2) = p(4) = p(6) = 2q$$

Como $1 = P(\Omega) = 3q + 6q$, entonces $q = 1/9$.

3) Arrojamos una moneda equilibrada hasta obtener cara. Cuál es la probabilidad de que la cara sea obtenida en un número par de lanzamientos?

$$\Omega = \{(1), (0,1), (0,0,1), (0,0,0,1), \dots\} = \{1, 2, \dots\}$$

y le asignamos probabilidad $p(i) = \frac{1}{2^i}$.

El evento es $A=\{2,4,6,\dots\}$.

$$P(A)=\sum_{i\geq 1}p(2i)=\sum_{i\geq 1}1/2^{2i}=\frac{1}{1-\frac{1}{4}}-1=\frac{1}{3}.$$

Espacios de equiprobabilidad: Ω es finito y sea $n=\#\Omega$, donde $\#$ es el cardinal del conjunto.

El espacio es de equiprobabilidad si $p(i)=1/n$, para todo i.

$P(A):=\frac{\#A}{\#\Omega}$ favorables sobre posibles.

Ejemplos: 1) Urna contiene 5 bolillas numeradas de 1 a 5. Retiramos dos bolillas con reposición.

$\Omega=\{1,2,3,4,5\}\times\{1,2,3,4,5\}$; cardinal $\#\Omega=5\times 5=25$. Espacio equiprobable.

Bolillas 1 y 2 son blancas y las otras 3 rojas. Calcule la probabilidad de los siguientes eventos.

a) $A_1=$ extraer al menos una roja.

b) $A_2=$ extraer primero una roja y después una blanca.

El evento "ninguna roja" es $A_1^c=\{12,21,11,22\}$ que tiene 4 elementos. Así $P(A_1)=1-P(A_1^c)=21/25$.

b) A_2 tiene 3×2 elementos. Así $P(A_2)=6/25$.

Observe que el espacio "color de las dos bolas ordenado" $\{BB,BR,RB,RR\}$ no es equiprobable en este caso.

2) Sucesiones de n ceros y unos. Lanzamiento de n monedas.

Si la moneda es honesta Ω tiene 2^n elementos y todos tienen la misma probabilidad $1/2^n$.

3) Tablero de ajedrez con 32 fichas de dominó. elegimos al azar una de las posibles posiciones. Cual es la probabilidad que aparezcan todas orientadas para arriba?

1.5. Espacios no numerables

Álgebras y sigma-álgebras $\mathcal{P}(\Omega)$ *partes de* Ω.

Definición: **Álgebra** $\mathcal{A}\subset\mathcal{P}(\Omega)$ es una clase que contiene Ω y cerrada para complemento y uniones finitas.

Ejemplos: (1) $\{\Omega,\emptyset\}$

(2) $\{\Omega,\emptyset,A,A^c\}$

(3) $\mathcal{P}(\Omega)$

(4) $\Omega=\mathbb{R}$ y los elementos de $\mathcal{A}$ son uniones finitas de intervalos en el conjunto

$$\{(a,b]:-\infty\leq a\leq b\leq\infty\}$$

(incluye intervalos semi-infinitos).

Sigma-Álgebra Decimos que $\mathcal{F} \subset \mathcal{P}(\Omega)$ es una *sigma álgebra* si $\mathcal{F}$ es un álgebra cerrada para uniones numerables de eventos.

El ejemplo (4) no es una sigma-álgebra. Unión infinita de intervalos no es unión finita de intervalos.

Def. Definimos la *sigma álgebra generada* por una familia $\mathcal{C} \subset \mathcal{P}(\Omega)$ por

$$\sigma(\mathcal{C}) := \bigcap_{\mathcal{F} \supset \mathcal{C}\,:\,\mathcal{F} \text{ sigma álgebra}} \mathcal{F}$$

Intersección de todas las sigma algebras que contienen a $\mathcal{C}$.

Lema 1 *$\sigma(\mathcal{C})$ es sigma álgebra.*

Dem $\sigma(\mathcal{C}) \neq \emptyset$ porque $\mathcal{P}(\mathcal{C})$ contiene a $\mathcal{C}$ y es sigma álgebra.

a. Ω está en todas las $\mathcal{F}$, por lo tanto está en $\mathcal{C}$.

b. $A \in \mathcal{C}$ implica $A \in \mathcal{F}$ para toda $\mathcal{F}$ implica $A^c \in \mathcal{F}$ para toda $\mathcal{F}$ implica $A^c \in \mathcal{C}$

c. Idem para uniones numerables.

$\sigma(\mathcal{C})$ es la menor sigma álgebra que contiene a $\mathcal{C}$ en el sentido que si $\mathcal{F} \supset \mathcal{C}$ es sigma álgebra, entonces $\mathcal{F} \supset \sigma(\mathcal{C})$.

Teorema de extensión. (Ash [1], Theorem 1.3.6) *Una probabilidad en un álgebra $\mathcal{A}$ puede extenderse a $\sigma(\mathcal{A})$, la sigma álgebra generada por $\mathcal{A}$.*

Espacio de Probabilidad es una terna $(\Omega, \mathcal{F}, P)$, donde Ω es el espacio muestral $\mathcal{F} \subset \mathcal{P}(\Omega)$ es una sigma álgebra de eventos y $P : \mathcal{F} \to [0, 1]$ es una probabilidad.

Ejemplo: Sigma álgebra de Borel en $\mathbb{R}$ Sea $\Omega = \mathbb{R}$ y $\mathcal{A}$ el álgebra cuyos elementos son uniones finitas de intervalos abiertos a la izquierda y cerrados a la derecha, es decir, uniones finitas de elementos del conjunto

$$\big\{(a, b] : -\infty \leq a \leq b \leq \infty\big\}$$

$\mathcal{B} := \sigma(\mathcal{A})$ se denomina **sigma álgebra de Borel** y sus elementos se llaman *Borelianos.*

Dado un Boreliano $E \subset \mathbb{R}$, definimos $\mathcal{B}(E) = \{B \cap E : B \text{ Boreliano}\}$. Son los Borelianos de E.

Ejemplo: Sea $E = [0, 1]$ y definimos una P sobre los intervalos: $P((a, b]) = b - a$. Para un Boreliano $P(A)$ es la medida de Lebesgue del conjunto A.

Teorema 2 *Existe una unica probabilidad $P : \mathcal{B}([0, 1]) \to [0, 1]$ tal que $P((a, b]) = b - a$.*

Sigma álgebra de Borel en $\mathbb{R}^d$ $\quad \Omega = \mathbb{R}^d$,

$\mathcal{C} = \{$rectángulos semi abiertos$\} = \{(x_1, \dots, x_d) : a_i < x_i \le b_i\}$

$\mathcal{B} = \sigma(\mathcal{C})$ menor sigma álgebra que contiene $\mathcal{C}$: **sigma álgebra de Borel**. Borelianos.

Consideremos por ejemplo $E = [0,1]^d$. Definimos P sobre los rectángulos d-dimensionales:

$$P((a_1, b_1] \times \cdots \times (a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i)$$

que se extiende naturalmente al álgebra de uniones finitas de rectángulos. Para un Boreliano $P(A)$ es la medida de Lebesgue del conjunto A.

Teorema 3 *Existe una unica probabilidad* $P : \mathcal{B}([0,1]) \to [0,1]$ *tal que*

$$P((a_1, b_1] \times \cdots \times (a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i)$$

2. Probabilidad condicional e independencia

Un estudio sobre la eficiencia de la vacuna contra la gripe en una muestra de 100 personas observó:

2 enfermos y vacunados (ev)
13 enfermos y no vacunados (en)
75 sanos y vacunados (sv)
10 sanos y no vacunados (sn)

	e	s	
v	2	75	77
n	13	10	23
Total	15	85	100

Elijo una persona al azar de esa muestra y observo su estado.

El espacio muestral es $\Omega = \{ev, en, sv, sn\}$,

$P(\{ev\}) = 0{,}02$, $P(\{en\}) = 0{,}13$, $P(\{sv\}) = 0{,}75$, $P(\{sn\}) = 0{,}10$

(cálculos hechos con casos favorables sobre posibles)

Eventos $E = \{ev, en\}$ (enfermo),

$V = \{ev, sv\}$ (vacunado).

Cual es la probabilidad que una persona esté enferma?

$P(E) = P(\{ev, en\}) = 0{,}02 + 0{,}13 = 0{,}15$.

Cual es la probabilidad que una persona esté enferma y vacunada?

$P(EV) = P(\{ev\}) = 0{,}02$.

Probabilidad que una persona vacunada esté enferma?

Casos favorables 2, casos posibles $75 + 2$ (los vacunados)

Si sabemos que la persona elegida está vacunada, cual es la probabilidad que esté enferma?

Hay que restringir el espacio muestral a los vacunados:

$$P(\text{enfermo dado vacunado}) = \frac{2}{77} = \frac{P(EV)}{P(V)}.$$

2.1. Probabilidad condicional

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad , considere los eventos A, B con $P(B) > 0$ y defina

$$P(A|B) := \frac{P(AB)}{P(B)}.$$

como la *probabilidad condicional* de A dado que conocemos B.

Valen las siguientes afirmaciones

- $P(AB) = P(A|B)P(B)$
- $(B, \mathcal{F}_B, P_B)$ es un espacio de probabilidad.
- $(\Omega, \mathcal{F}_B, P_B)$ es un espacio de probabilidad.

Ejemplos

Un dado. Calcule la probabilidad de ver un 3 dado que el resultado es un número no mayor que 4.

Dos dados. Calcule la probabilidad de que haya salido un seis dado que la suma es mayor o igual a 9.

Monedas Lanzamos 3 monedas. Calcule la probabilidad que la tercera moneda sea cara dado que el número de caras es 2.

Familias de dos hijos

$\Omega = \{vv, vm, mv, mm\}$, espacio equiprobable.

1) Una familia tiene dos hijos. Sabemos que el primer hijo es varón. Cual es la probabilidad que el segundo hijo sea también varón?

$A = \{vv\}$ (dos hijos varones), $C = \{vv, vm\}$ (primer hijo varón),

Queremos calcular $P(A|C) = P(AC)/P(C) = \frac{1/4}{2/4} = 1/2$

2) Sabemos que una familia con dos hijos tiene por lo menos un hijo varón. Cual es la probabilidad que los dos sean varones?

Buscamos $P(A|C)$, con $A = \{vv\}$ (dos hijos varones), y $C = \{vv, vm, mv\}$ (por lo menos un varón).

Usando las fórmulas $P(A|C) = P(AC)/P(C) = \frac{1/4}{3/4} = 1/3$.

3) Supongamos que visitamos a la familia, tocamos el timbre y un chico varón abre la puerta. Cual es la probabilidad que el otro chico sea varón?

$\Omega = \{v^*v, vv^*, m^*v, mv^*, v^*m, vm^*, m^*m, mm^*\}$

donde * quiere decir "abrió la puerta". Por ejemplo mv^* es el evento que el primer hijo es mujer, el segundo hijo es varón y es él quien abre la puerta. Espacio equiprobable.

Buscamos $P(A|C)$, donde $A = \{v^*v, vv^*\}$ (los dos hijos son varones) y $C = \{v^*v, vv^*, mv^*, v^*m\}$ (abre la puerta un varón)

$P(A|C) = \frac{P(AC)}{P(C)} = \frac{2/8}{4/8} = 1/2$.

Regla de la multiplicación Cálculo de probabilidades usando árboles. Vale la siguiente fórmula

$$P(A_1 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_1\dots A_{n-1})$$

Dem: Por inducción. $P(A_1A_2) = P(A_1)P(A_2|A_1)$, por definición.

$$P(A_1 \dots A_n) = P(A_1 \dots A_{n-1})P(A_n|A_1\dots A_{n-1})$$

usando el caso $n = 2$. Culmine aplicando la hipótesis inductiva a $P(A_1 \dots A_{n-1})$. $\square$

Ejemplo Una urna tiene 4 bolas negras y 3 rojas.

Sacamos tres bolas sin reposición. Cual es la probabilidad que salga negra-negra-roja?

A_i = negra en la i-ésima bola, o sea A_i^c = roja en la $i$ésima bola.

$$P(A_1A_2A_3^c) = P(A_1)P(A_2/A_1)P(A_3^c/A_1A_2) = \frac{4}{7}\frac{3}{6}\frac{3}{5}$$

2.2. Probabilidad total y Bayes

Una *partición* de Ω es una familia de eventos $\{B_1, B_2, \dots\}$, disjuntos dos a dos tal que la unión es Ω:

$$\Omega = \dot\cup_i B_i, \quad B_i \cap B_j = \emptyset, \quad \text{para } i \neq j.$$

En ese caso $P(\Omega) = \sum_i P(B_i)$

Ejemplo. Dado. $\Omega = \{1, 2, 3, 4, 5, 6\}$.

$B_1 = \{1, 2\}$, $B_2 = \{3, 4, 5\}$, $B_3 = \{6\}$ es una partición de Ω.

Teorema de la Probabilidad total *Sea B_i una partición de Ω tal que $P(B_i) > 0$ para todo i. Sea A un evento. Entonces,*

$$P(A) = \sum_i P(A|B_i)P(B_i).$$

Dem $P(A) = P(\dot\cup_i(A \cap B_i)) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$.

Ejemplo Engripados y vacunados. 80 % de la población está vacunada. De los vacunados 2 % se enferman de gripe. De los no vacunados, 15 % se enferman.

Cual es la probabilidad que una persona tenga gripe?

A = engripado, $P(A) = ?$

B_0 = no vacunado

$B_1 =$ vacunado

Conocemos $P(B_0) = 0,2,\ P(B_1) = 0,8,\ P(A|B_0) = 0,15,\ P(A|B_1) = 0,02.$

Usando probabilidad total:

$$\begin{aligned} P(A) &= P(A|B_0)P(B_0) + P(A|B_1)P(B_1) \\ &= 0{,}15\ 0{,}2 + 0{,}02\ 0{,}8 = 0{,}19 \end{aligned}$$

Fórmula de Bayes *Sea B_i una partición de Ω tal que $P(B_i) > 0$ para todo i. Sea A un evento. Entonces,*

$$P(B_j|A) = \frac{P(B_jA)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

Se usa cuando sabemos calcular $P(A|B_i)$ y $P(B_i)$

Vacunas

Cual es la probabilidad que una persona con gripe haya sido vacunada?

Queremos calcular $P(B_1|A)$. Se aplica Bayes directo.

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{0{,}8\ 0{,}2}{0{,}19} = \dots$$

Juego de las 3 puertas. Monty Hall.

Tres puertas cerradas y un premio atras de una de las puertas. Elijo una puerta y el presentador abre una de las otras dos que no tiene premio. Me da la opcion de cambiar de puerta. Conviene cambiar?

$B_i =$ premio en puerta i. $P(B_i) = 1/3$

Jugador elige la puerta 1 (los otros casos son análogos).

$A_3 =$ presentador abre la puerta 3 (el otro caso es análogo).

Quiero comparar $P(B_1|A_3)$ con $P(B_2|A_3)$

El **protocolo** que usa el presentador cuando el premio está en la puerta 1 es abrir una de las otras dos puertas con la misma probabilidad. Es decir:

$$P(A_3|B_3) = 0, \quad P(A_3|B_2) = 1, \quad P(A_3|B_1) = 1/2 \tag{3}$$

Puede haber otros protocolos. Usando el protocolo (3) calculamos la probabilidad que el presentador abra la puerta 3:

$$\begin{aligned} P(A_3) &= P(A_3|B_1)P(B_1) + P(A_3|B_2)P(B_2) + P(A_3|B_3)P(B_3) \\ &= \frac{1}{2}\frac{1}{3} + 1\frac{1}{3} + 0\frac{1}{3} = \frac{1}{2} \end{aligned}$$

y las probabilidades que el premio esté en las puertas 1 y 2 dado que el presentador abrió la 3:

$$P(B_1|A_3) = \frac{P(A_3|B_1)P(B_1)}{P(A_3)} = \frac{1/6}{1/2} = 1/3$$

$$P(B_2|A_3) = \frac{P(A|B_2)P(B_2)}{P(A_3)} = \frac{1/3}{1/2} = 2/3.$$

O sea que $P(\text{ganar sin cambiar de puerta}) = 1/3$ y $P(\text{ganar cambiando de puerta}) = 2/3$

Simulación en R: ver Monty Hall

2.3. Independencia de eventos

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Los eventos A y B se dicen *independientes* si $P(AB) = P(A)P(B)$.

A y B son independientes si y sólo si $P(A|B) = P(A)$. Demostración: Por definición de probabilidad condicional,

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A) \quad \text{si y sólo si} \quad P(AB) = P(A)P(B).$$

Es decir que independencia entre A y B puede interpretarse como "el conocimiento de B no modifica la probabilidad de A".

Si A y B son independientes, entonces A y B^c también lo son. Demostración: $P(AB^c) + P(AB) = P(A)$. Por lo tanto $P(AB^c) = P(A) - P(AB) = P(A)(1 - P(B)) = P(A)P(B^c)$.

Dos dados.

A_i = primer dado muestra la cara i, B_j segundo dado muestra la cara j. $i, j \in \{1, \dots, 6\}$.

A = suma 9; B = segundo dado es 5.

A = primer dado par. B = segundo dado impar.

Familia de eventos independientes

Se dice que los eventos $A_1, A_2, \dots$ son *independientes* si

$P(\cap_{i \in K} A_i) = \prod_{i \in K} P(A_i)$ para cualquier subconjunto finito $K \subset \{1, 2, \dots\}$

Ejemplo: descomposición binaria de un número real elegido al azar en $[0, 1]$.

A_i := i-ésimo digito en la descomposición binaria del número es 1. Los eventos A_i son independientes!

Ejemplo de conjuntos independientes dos a dos pero no independientes: 3 monedas

A_1 primera moneda cara.

A_2 segunda moneda cara.

A_3 las dos monedas son iguales.

Son independientes dos a dos pero no independientes.

3. Variables aleatorias

3.1. Definición y propiedades

$(\Omega, \mathcal{F}, \mathcal{P})$. Una variable aleatoria es una funcion

$$X : \Omega \to \mathbb{R}$$

tal que $X^{-1}(B) \in \mathcal{F}$ para todo Boreliano B.

Notacion:

$$\{X \in B\} = \{\omega : X(\omega) \in B\} = X^{-1}(B) \quad \text{es un } \textbf{evento}$$

Una variable aleatoria es una manera de **describir** eventos.

Proposición 4 *Sea $\mathcal{C}$ una familia de conjuntos que generan los Borelianos, $\sigma(\mathcal{C}) = \mathcal{B}$. $X : \Omega \to \mathbb{R}$ una función. Si $\{X \in C\} \in \mathcal{F}$ para todo $C \in \mathcal{C}$ entonces X es una variable aleatoria.*

Dem Sea $\mathcal{D} \subset \mathcal{B}$ el conjunto de Borelianos D para los que $\{X \in D\} \in \mathcal{F}$. Se ve que $\mathcal{D}$ es una sigma álgebra. Como $\mathcal{C} \subset \mathcal{D}$ tenemos que $\mathcal{D} = \mathcal{B}$.

Veamos que $\mathcal{D}$ es una sigma algebra:

$\{X \in D^c\} = \{X \in D\}^c \in \mathcal{F}$ implica $D^c \in \mathcal{D}$.

$\{X \in \cup_i D_i\} = \cup_i \{X \in D_i\} \in \mathcal{F}$ implica $\cup D_i \in \mathcal{F}$.

$\{X \in \mathbb{R}\} = \Omega \in \mathcal{F}$ implica $\mathbb{R} \in \mathcal{D}$. $\square$

Corolario 5 *$(\Omega, \mathcal{F})$, X es variable aleatoria* **sii** *$\{X \leq z\} \in \mathcal{F}$ para todo $z \in \mathbb{R}$.*

Ejemplos de variables aleatorias

Variable aleatoria constante $X(\omega) = c$ para todo $\omega \in \Omega$.

$$\{X \in B\} = \begin{cases} \Omega & \text{si } B \ni c \\ \emptyset & \text{si } B \not\ni c \end{cases}$$

es una variable aleatoria.

Función indicadora. Sea $A \in \mathcal{F}$, entonces la función definida por $\mathbf{1}_A = I_A$ dada por

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

es una variable aleatoria: Si $B \in \mathcal{B}$.

$$\{X \in B\} = \begin{cases} \emptyset & \text{si } B \not\ni 1,\ B \not\ni 0 \\ A & \text{si } B \ni 1,\ B \not\ni 0 \\ A^c & \text{si } B \not\ni 1,\ B \ni 0 \\ \Omega & \text{si } B \ni 1,\ B \ni 0 \end{cases}$$

Ejemplo de función que no es variable aleatoria Sea $\Omega = \{1,2,3,4\}$, $\mathcal{F} = \{\emptyset, \{1,2\}, \{3,4\}, \Omega\}$

$$X = I_{\{2,3\}} \text{ no es variable aleatoria.}$$

porque $\{X \in B\} = \{2,3\} \notin \mathcal{F}$ para $B \ni 1$.

Lema 6 *Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y X variable aleatoria. Entonces $(\mathbb{R}, \mathcal{B}, P_X)$ es espacio de probabilidad, donde $P_X(B) = P(X \in B)$.*

Dem: (ejercicio) Como $\mathcal{B}$ es una sigma álgebra, sólo falta ver que P_X es una probabilidad. $\square$

3.2. Función de distribución acumulada

Definición: $F_X(x) := P(X \leq x)$

Propiedades de la función de distribución acumulada: $F = F_X$

1) F es monótona no decreciente: $x \leq y$ implica $F(x) \leq F(y)$

2) F es continua a derecha, es decir $\lim_{h \to 0^+} F(x+h) = F(x)$

3) $\lim_{x\to\infty} F(x) = 1$ y $\lim_{x \to -\infty} F(x) = 0$

Lema 7 *Cualquier función que satisfaga 1, 2, 3 es la función de distribución de alguna variable aleatoria.*

Dem. Pospuesta. $\square$

Lema 8 *F_X caracteriza X. Es decir, si X e Y son variables aleatorias, entonces*

$$P_X = P_Y \quad \text{si y sólo si} \quad F_X = F_Y. \tag{4}$$

Es decir, $P(X \in A) = P(Y \in A)$ para todo evento A si y sólo si $P(X \leq x) = P(Y \leq x)$ para todo $x \in \mathbb{R}$.

Dem:

1 implica 2: $P(X \leq x) = P(X \in (-\infty, x]) = P_X((-\infty, x])$.

2 implica 1: $P_X((a,b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$. Con esto definimos la probabilidad en los intervalos. Por el teorema de extensión, queda definida sobre todo Boreliano. $\square$

Uso de la distribución acumulada Se usa para calcular la probabilidad P_X sobre los intervalos:

$$\begin{aligned}
P(a < X \leq b) &= F(b) - F(a)\\
P(a \leq X \leq b) &= F(b) - F(a-)\\
P(a \leq X < b) &= F(b-) - F(a)\\
P(a < X < b) &= F(b-) - F(a-)
\end{aligned}$$

3.3. Variables aleatorias discretas

Decimos que una variable aleatoria es **discreta** si asume un número finito o numerable de valores. La variable X induce la partición de Ω dada por:

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

donde $R_X =$ es el **rango** de X definido por

$$R_X := \{x \in \mathbb{R} : P(X = x) > 0\}.$$

Función de probabilidad puntual: Definimos $p_X(x) := P(X = x)$. En este caso discreto decimos que p_X es la **distribución** de X. Es una tabla.

Función acumulada de variables discretas: En el caso de variables discretas, la función acumulada es constante por pedazos con saltos en los x en el rango de X:

$$P(X = x) = F(x) - F(x-).$$

Observe que no hay salto cuando $x \notin R_X$.

Espacio de probabilidad inducido por variables discretas Una variable aleatoria discreta X induce naturalmente el espacio de probabilidad

$$(R_X, \mathcal{P}(R_X), P_X).$$

Diagrama de barras: gráfico de la función $x \mapsto P(X = x)$.

Histograma: A cada x del rango se le asigna un rectángulo cuyo área es igual a $P(X = x)$.

Ejemplos *(1) Dos monedas* $\Omega = \{00, 01, 10, 11\}$. Es un espacio equiprobale, $p(\omega) = 1/4$ para todo ω. Definimos la variable aleatoria $X =$ número de caras.

$$X(00) = 0, \quad X(01) = X(10) = 1, \quad X(11) = 2.$$

X induce la partición:

$$\{X = 0\} = \{00\}, \quad \{X = 1\} = \{01, 10\}, \quad \{X = 2\} = \{11\}$$

que permite calcular la distribución:

X	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(2) Uniforme en un conjunto finito B: Para $x \in B$,

$$P(X = x) = \frac{1}{|B|}$$

donde $|B|$ es el cardinal de $|B|$.

(3) Bernoulli(p). Jacob Bernoulli (1654-1705), p es el parámetro.

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p.$$

(4) Distribución geométrica de parámetro p. Informalmente, representa el número de ensayos de Bernoulli(p) hasta el primer éxito. La probabilidad puntual está dada por

$$p_X(k) = P(X = k) = (1-p)^{k-1}p, \qquad k \geq 1. \tag{5}$$

Deducimos que $P(X > k) =$ probabilidad de k fracasos $= (1-p)^k$. La función de distribución acumulada es

$$F(k) = P(X \leq k) = 1 - P(X > k) = 1 - (1-p)^k$$

Falta de memoria de la geometrica.

$$\begin{aligned} P(T > n + m | T > n) &= \frac{P(T > n + m)}{P(T > n)} && (6) \\ &= \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = P(T > m) && (7) \end{aligned}$$

(5) Distribución binomial. La variable S_n representa el número de éxitos en n ensayos independientes de Bernoulli(p): Sus probabilidades puntuales están dadas por

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, \dots, n$$

Decimos que S_n es Binomial(n, p) (parámetros).

Para ver que el número de éxitos en n ensayos es binomial hay que verificar que los ensayos son independientes y la probabilidad de éxito es siempre p. Por ejemplo

El experimento: se extraen 4 bolillas *con reposición* de urna con 5 blancas y 3 negras. $S_4 =$ número de bolillas blancas extraídas tiene distribución Binomial($4, 5/8$).

El experimento: se extraen 2 bolillas *sin reposición* de urna con 5 blancas y 3 negras. $S_2 =$ número de bolillas blancas extraídas, **no** tiene distribución binomial (falla la independencia).

(6) Distribución de Poisson. Simeon-Denis Poisson (1781-1840). Sea $\lambda > 0$ un parámetro real. Decimos que una variable X es Poisson(λ) si su distribución está dada por

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k \geq 0.$$

Para verificar que es una probabilidad, recordemos que por Taylor:

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

Esto implica que $\sum_{k \geq 0} P(X = k) = 1$.

Aproximación Poisson de la binomial Sea $S_n \sim$ Binomial($n, p(n)$) con $p(n) = \lambda/n$, $\lambda > 0$ es un parámetro.

Lema 9 *Vale*

$$\lim_{n \to \infty} P(S_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k \geq 0.$$

Dem:

$$P(S_n = k) = \binom{n}{k} p(n)^k (1-p(n))^{n-k} = \frac{n!}{k!(n-k)!}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k}$$
$$= \frac{\lambda^k}{k!}\Big(1-\frac{\lambda}{n}\Big)^n \frac{n!}{(n-k)!\,n^k}\Big(1-\frac{\lambda}{n}\Big)^{-k}$$

Esto demuestra el lema porque

$$\begin{aligned}
&\lim_{n\to\infty}\Big(1-\frac{\lambda}{n}\Big)^n = e^{-\lambda}\\
&\lim_{n\to\infty}\frac{n!}{(n-k)!\,n^k} = \lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{n^k} = 1\\
&\lim_{n\to\infty}\Big(1-\frac{\lambda}{n}\Big)^{-k} = 1. \qquad \square
\end{aligned}$$

(7) Binomial negativa o Pascal. Sea Y_k el número de ensayos de Bernoulli(p) necesarios para obtener k éxitos? Hay dos parametros, k y p.

$$P(Y_k = j) = \binom{j-1}{k-1} p^k (1-p)^{j-k}$$

Dualidad entre la Binomial y la Binomial negativa

Lema 10 *Sean $S_n \sim$ Binomial(n,p) y $Y_k \sim$ Binomial-negativa(k,p). Entonces vale*

$$P(S_n \geq k) = P(Y_k \leq n).$$

Demostración En realidad vale una identidad más fuerte: el instante del k-ésimo éxito es menor o igual a n si y sólo si hay por lo menos k éxitos en los primeros n ensayos:

$$\{S_n \geq k\} = \{Y_k \leq n\}. \qquad \square$$

3.4. Variables aleatorias continuas

Ejemplo: X_n: duración de una batería en unidades $1/n$.

$X_n \sim$ Uniforme en $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$.

Cuando n es grande X_n aproxima una variable aleatoria X "continua", el tiempo de duración $X \in [0,1]$.

Histogramas con área total igual a 1.

dias, horas, minutos, segundos, décimas de segundo, etc, como límite de los histogramas una curva suave.

Probabilidad de que la duración X esté entre los números reales a y b ($a < b$) estará dada por el área bajo la curva entre a y b.

$$P(X_n \in [a,b]) = (b-a) + O(1/n) \underset{n\to\infty}{\longrightarrow} b-a = P(a \leq X \leq b).$$

Definición: Una variable aleatoria X es **continua** si existe una función $f : \mathbb{R} \to \mathbb{R}^+ = [0, \infty)$ llamada **función de densidad** de X tal que

$$P(X \in A) = \int_A f(x)dx, \qquad A \subset \mathbb{R}$$

para cualquier Boreliano A. Cuando $A = [a, b]$ es un intervalo tenemos

$$P(a \le X \le b) = \int_a^b f(x)dx$$

La función de densidad $f(x)$ debe satisfacer

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$f(x)$ puede ser mayor que 1.

Ejemplo: $f(x) = ax^2\mathbf{1}\{x \in [1, 3]\}$.

Calcular $a = \left(\int_1^3 x^2 dx\right)^{-1} = \frac{3}{26}$.

Calcular $P(X \ge 2) = \frac{19}{26}$

Función de distribución acumulada En este caso se expresa como una integral:

$$F_X(x) := P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

Calcular la F de la variable X

Propiedades de la función de distribución acumulada de variables continuas Sea F la función de distribución acumulada de X, una variable aleatoria continua. Entonces F tiene las siguiente propiedades

1) $F(x)$ es monótona no decreciente.

2) $F(x)$ es **continua** en todo punto.

3) $\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = 1$

Lema 11 *Si X es continua y $a \le b$ reales, vale*

$$\begin{aligned} P(a < X < b) &= P(a \le X < b) = P(a < X \le b) \\ &= P(a \le X \le b) = F(b) - F(a) \end{aligned}$$

Dem: Basta ver que $P(X = a) = P(X = b) = 0$. $\square$

Lema 12 *Si X continua con densidad f y acumulada F, entonces en todo punto x donde $F(x)$ es derivable,*

$$f(x) = F'(x)$$

Dem: Resulta del Teorema Fundamental del Cálculo Integral, y de la definición de $F(x)$.
$\square$

Lema 13 *La función de distribución caracteriza la variable aleatoria: Si $F_X = F_Y$ entonces $P(X \in A) = P(Y \in A)$ para todo Boreliano A.*

Percentiles de una distribución continua: Sea X una variable aleatoria continua con densidad f y acumulada F y sea $0 < p < 1$. El percentil $(100p)$-ésimo de la distribución de X es el valor x_p tal que $F(x_p) = P(X < x_p) = p$, es decir, la solución x_p de

$$\int_{-\infty}^{x_p} f(x) = p$$

Por ejemplo para encontrar el percentil $p = 0,25$ de X con densidad

$$f(x) = \frac{19}{26} x^2 \mathbf{1}\{x \in [1,3]\}$$

calculamos primero la acumulada

$$F(x) = \frac{x^3 - 1}{26} \mathbf{1}\{x \in [1,3]\} + \mathbf{1}\{x \geq 3\}$$

y resolvemos la ecuación para $x_p \in [1,3]$:

$$F(x_{0,25}) = 0,25 \quad \Rightarrow \quad \frac{x^3 - 1}{26} = 0,25 \quad \Rightarrow \quad x_{0,25} = 1,96$$

Ejemplos clásicos de distribuciones continuas

Uniforme

X tiene distribución uniforme en el intervalo $[a,b]$, si su función de densidad es

$$f(x) = \frac{1}{b-a} \mathbf{1}\{x \in [a,b]\}$$

La distribución acumulada está dada por:

$$F(x) = \frac{x-a}{b-a} \mathbf{1}\{x \in [a,b]\} + \mathbf{1}\{x \geq b\}$$

Note que $f(x) = F'(x)$ para todo $x \notin \{a,b\}$.

Exponencial

Buscamos una variable aleatoria X en $\mathbb{R}^+$ *sin memoria*, es decir, que satisfaga

$$P(X > t + s | X > s) = P(X > t), \quad \text{para todo} \quad s,t \geq 0.$$

Si el colectivo no llegó hasta el instante s, cuanto más va a tardar en llegar? Eso es equivalente a

$$\frac{P(X > t+s)}{P(X > s)} = P(X > t)$$

Si llamamos $g(t) = P(X > t)$, g debe satisfacer:

$$g(t+s) = g(t)g(s), \text{ para todo } s, t > 0. \tag{8}$$

Si g está así definida, entonces $g(t) = e^{-\lambda t}$ para algún $\lambda > 0$ es solución de (8). De hecho son la únicas soluciónes. La función de distribución acumulada de X es

$$F(t) = 1 - g(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0. \end{cases}$$

y su densidad $f(t) = \lambda e^{-\lambda t}$.

Decimos que la variable aleatoria X con esa densidad tiene distribución *exponencial* de parámetro λ. Denotamos $X \sim \text{Exponencial}(\lambda)$.

Exponencial como límite de geométricas rescaladas

Sea Y_n una geometrica de parametro $p_n = \lambda/n$ y calculemos

$$P\Big(\frac{Y_n}{n} > t\Big) = P\big(Y_n > tn\big) = \Big(1 - \frac{\lambda}{n}\Big)^{\lfloor tn \rfloor} \underset{n\to\infty}{\longrightarrow} e^{-\lambda t}$$

"En intervalos de tiempo de longitud $\frac{1}{n}$ se realizan ensayos con probabilidad $\frac{\lambda}{n}$ de éxito; Y_n/n es el tiempo necesario para obtener el primer éxito". Cuando n tiende a infinito, la distribución de Y_n/n converge a la distribución Exponencial(λ).

Gama

Una variable X tiene distribución Gama con parámetros $\alpha > 0$ y $\lambda > 0$ si su densidad es

$$f(x) = \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha - 1} \mathbf{1}\{x \geq 0\}$$

donde $\Gamma(\alpha)$ está definida por

$$\Gamma(\alpha) := \int_0^\infty e^{-y} y^{\alpha-1} dy$$

Integrando por partes se demuestra que

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

por lo que para α entero no negativo $\Gamma(\alpha) = (\alpha - 1)!$.

Cuando $\alpha = n$ es entero, X es el tiempo necesario para que haya n eventos, cuando el tiempo entre dos eventos es Exponencial(λ). Esto lo veremos después.

Dualidad Gama-Poisson

Lema 14 *Sea $N(t)$ una variable Poisson(λt) y Y_n una variable Gama(n, λ). Entonces*

$$P(N(t) \geq n) = P(Y_n \leq t).$$

Dem Sea

$$F(t) := P(N(t) \geq n) = \sum_{j=n}^{\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!}.$$

Diferenciando en t,

$$f(t) = F'(t) = \sum_{j=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{j-1}\lambda}{(j-1)!} - \sum_{j=n}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{j}\lambda}{j!}$$

como es una suma telescópica, la suma es igual al primer término:

$$= \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$$

que es la densidad de la Gama(n, λ). Integrando en t tenemos que $F(t)$ es la distribución acumulada de la gamma calculada en t. $\square$

3.5. Distribución Normal.

Poincaré pregunta si hay una distribución uniforme en una bola de radio infinito de dimensión infinita. No se puede definir una probabilidad uniforme en ese conjunto pero se puede proponer un procedimiento límite.

Sea $\Omega_k = \{\omega \in \mathbb{R}^k : |\omega|^2 \leq vk\}$ la bola k-dimensional de radio $\sqrt{vk}$. Sea P_k la distribución uniforme en Ω_k. Así, si $A \subset \Omega_k$ boreliano, definimos

$$P_k(A) = \frac{\text{Volumen de } A}{\text{Volumen de } \Omega_k}$$

Defino $X_1(\omega) = \omega_1$ primera coordenada.

Ejercicio: Pruebe que

$$\lim_k P_k(X_1 \leq a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi v}} e^{-x^2/2v} dx$$

Hay que calcular el volumen de $\{\omega \in \mathbb{R}^k : |\omega|^2 \leq vk \text{ y } \omega_1 \leq a\}$, dividir por el volumen de Ω_k y tomar el límite $k \to \infty$. La cuenta está hecha en Georgii [6].

Distribución Normal. Se dice que X tiene distribución Normal de parámetros μ y σ^2 si su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Notación: $X \sim N(\mu, \sigma^2)$. El gráfico tiene forma de campana con eje de simetría en $x = \mu$ y puntos de inflexión en $x = \mu \pm \sigma$. Es simetrica en relacion a μ: $f(\mu + x) = f(\mu - x)$ y alcanza el maximo en $x = \mu$.

Para probar que es la distribución de una variable aleatoria hay que verificar que $\int f = 1$. El truco es escribir

$$\begin{aligned}\Big(\int_{-\infty}^{\infty} f(x)dx\Big)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y)dxdy \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy = 1,\end{aligned}$$

pasando por coordenadas polares.

Distribución normal standard. Decimos que Z tiene distribución normal standard si $Z \sim \text{Normal}(0,1)$. Esta distribución está tabulada. Por ejemplo, el percentil 99 de la distribución es 2.33.

A partir de la tabla de la normal standard se pueden calcular probabilidades para la normal usando las siguientes relaciones.

• Si $X \sim \text{Normal}(\mu, \sigma^2)$ entonces $Z = \frac{X-\mu}{\sigma} \sim \text{Normal}(0,1)$. Veamos:

$$F_Z(z) = P(Z \leq z) = P\Big(\frac{X-\mu}{\sigma} \leq z\Big) = P(X \leq \sigma z + \mu) = F_X(\sigma z + \mu)$$

Diferenciando obtenemos la densidad:

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X(\sigma z + \mu) = \sigma\, f_X(\sigma z + \mu) \\ &= \frac{\sigma}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\Big) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{z^2}{2}\Big) \end{aligned}$$

• Si $Z \sim \text{Normal}(0,1)$ y $X = \sigma Z + \mu$ entonces $X \sim \text{Normal}(\mu, \sigma)$.

$$P(X \leq x) = P(\sigma Z + \mu \leq x) = P\Big(Z < \frac{x-\mu}{\sigma}\Big)$$

De donde

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Z\Big(\frac{x-\mu}{\sigma}\Big) \\ &= \frac{1}{\sigma} f_Z\Big(\frac{x-\mu}{\sigma}\Big) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(\frac{(x-\mu)^2}{2\sigma^2}\Big) \end{aligned}$$

que es la densidad de una $\text{Normal}(\mu, \sigma^2)$.

• Si $Z \sim \text{Normal}(0,1)$ y $X = \sigma Z + \mu$. Entonces los percentiles satisfacen

$$\frac{x_p - \mu}{\sigma} = z_p \quad y \quad x_p = z_p \sigma + \mu$$

Aquí los percentiles están definidos por $p = P(Z \leq z_p) = P(X \leq x_p)$.

3.6. Funciones de variables aleatorias

Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Cómo tiene que ser $g : \mathbb{R} \to \mathbb{R}$ para que $Y := g(X)$ sea variable aleatoria? Necesitamos que $\{Y \in B\} \in \mathcal{F}$, $B \in \mathcal{B}$.

$\{Y \in B\} = \{g(X) \in B\} = \{X \in g^{-1}(B)\} \in \mathcal{F}$, si $g^{-1}(B) \in \mathcal{B}$ (porque X es variable aleatoria)

O sea g tiene que ser *medible Borel*: $B \in \mathcal{B}$ implica $g^{-1}(B) \in \mathcal{B}$.

En general consideremos espacios $(\Omega_i, \mathcal{F}_i)$ con $\mathcal{F}_i$ sigma álgebra de conjuntos de Ω_i. Diremos que $f : \Omega_1 \to \Omega_2$ es $\mathcal{F}_1/\mathcal{F}_2$-medible si $f^{-1}(A) \in \Omega_1$ para todo $A \in \Omega_2$.

Lema 15 *Si $\mathcal{C}$ es una clase de conjuntos que genera $\mathcal{F}_2$ y $f^{-1}(A) \in \mathcal{F}_1$ para cada $A \in \mathcal{C}$, entonces f es $\mathcal{F}_1/\mathcal{F}_2$-medible.*

Dem La clase $\{A \in \mathcal{F}_2 : f^{-1}(A) \in \mathcal{F}_1\}$ es una sigma álgebra en Ω_2. Como esta sigma álgebra contiene a $\mathcal{C}$, contiene también a $\mathcal{F}_2$. □

Lema 16 *$g : \mathbb{R} \to \mathbb{R}$ es medible si (i) es continua; (ii) es monótona; (iii) es la indicadora de un Boreliano*

(iv) g_n medibles implica $\inf g_n$ $\sup g_n$ $\liminf g_n$ y $\limsup g_n$ medibles.

Dem (i) Los conjuntos abiertos generan $\mathcal{B}$ y $g^{-1}(A)$ es abierto porque g es continua. Por lo tanto f es medible.

(ii) Si g es monótona, $g^{-1}((-\infty, a])$ es un intervalo y por lo tanto está en $\mathcal{B}$. Como los intervalos generan $\mathcal{B}$, g es medible.

(iv) $(\inf g_n)^{-1}([a, \infty)) = \{x \in \mathbb{R} : \inf g_n(x) \geq a)\} = \cap_n\{x \in \mathbb{R} : g_n(x) \geq a)\} \in \mathcal{B}$. Como los intervalos semi-infinitos generan $\mathcal{B}$, listo. La misma demostración sirve para el sup. Como $\liminf f_n = \sup_n \inf_{k\geq n} f_k$, el $\liminf$ es un supremo de medibles y por lo tanto es medible.

Cambio de variable

Cálculo de la distribución de $g(X)$ a partir de la distribución de X.

Teorema 17 *Sea X una variable aleatoria con densidad $f_X(x)$ tal que $P(X \in (a, b)) = 1$. Sea $g : (a, b) \to \mathbb{R}$* ***estrictamente creciente*** *o bien* ***estrictamente decreciente****. Considere la nueva variable aleatoria $Y = g(X)$. Entonces, para y en $\{g(x) : x \in (a, b)\}$,*

$$f_Y(y) = f_X(g^{-1}(y)) \left|\left(g^{-1}(y)\right)'\right| .$$

Dem (a) g estrictamente creciente. Calculamos la distribución acumulada de Y

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

pero como la función es estrictamente creciente en el intervalo (a, b), podemos invertirla:

$$= P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Para obtener f_Y derivamos F_Y y obtenemos

$$f_Y(y) = f_X(g^{-1}(y)) \left|\left(g^{-1}(y)\right)'\right|$$

el valor absoluto no agrega nada porque la derivada es positiva.

(b) g estrictamente decreciente. Como g^{-1} es decreciente,

$$P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

y derivando,

$$f_Y(y) = -f_X(g^{-1}(y)) \left(g^{-1}(y)\right)' = f_X(g^{-1}(y)) \left|\left(g^{-1}(y)\right)'\right|$$

porque la derivada de g^{-1} es negativa.

Ejemplos (1) $X \sim$ Uniforme$[0,1]$ y $Y = X^2$. Entonces

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}}.$$

Caso no invertible A veces, aunque g no sea invertible, se puede calcular la función de densidad de $Y = g(X)$. Por ejemplo, consideremos $X \sim$ Uniforme$[-3,3]$ y $Y = X^2$. Como $X \in [-3,3]$, $Y \in [0,9]$. Calculemos F_Y y f_Y.

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(g(X) \le y) = P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 2F_X(\sqrt{y}) - 1 \end{aligned}$$

porque $F_X(x) = 1 - F_X(-x)$, por simetría de la f_X. Derivando,

$$f_Y(y) = f_X(\sqrt{y})/\sqrt{y} = \frac{1}{6}\sqrt{y}, \qquad y \in [0,9].$$

(2) $Z \sim$ Normal$(0,1)$ y $Y = Z^2$. Con el mismo razonamiento que en el caso anterior:

$$\begin{aligned} F_Y(y) &= 2F_X(\sqrt{y}) \\ f_Y(y) &= f_X(\sqrt{y})/\sqrt{y} \end{aligned}$$

3.7. Convergencia en distribución

Decimos que una sucesión de variables aleatorias $Y_1, Y_2, \dots$ converge en distribución a una variable Y si

$$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$$

para todo y donde $F_Y(y)$ es continua.

(contra) Ejemplo. Sea $F_n(x) = \mathbf{1}\{x \ge 1/n\}$, la distribución de la variable X_n concentrada en $1/n$: $P(X_n = 1/n) = 1$. Cuando $n \to \infty$, X_n converge en distribución a la variable X concentrada en cero: $P(X = 0) = 1$. De hecho $\lim_n F_n(x) = F(x)$ para todo $x \ne 0$, el único punto de discontinuidad de F, la distribución acumulada de X. Sin embargo, vemos que $F_n(0)$ no converge a $F(0)$: $F_n(0) = 0$ para todo n, pero $F(0) = 1$.

Ejemplos de convergencia en distribución

1) $U_n \sim$ Uniforme$(\{1/n, \dots, n/n\})$, $U \sim$ Uniforme$[0,1]$. $U_n \xrightarrow{D} U$.

2) $X_n \sim$ Binomial$(n, \lambda/n)$, $X \sim$ Poisson(λ). $X_n \xrightarrow{D} X$.

3) $Y_n \sim$ geometrica (λ/n), $X_n = Y_n/n$, $X \sim$ Exponencial(λ). $X_n \xrightarrow{D} X$.

$$P(X_n/n > t) = P(X_n > tn) = \Big(1 - \frac{\lambda}{n}\Big)^{tn-1} \underset{n\to\infty}{\to} e^{-\lambda t}$$

4) $X_n =$ primera coordenada de un vector aleatorio distribuido uniformemente en $B_n(0, \sqrt{n})$, la bola de dimensión n de radio $\sqrt{n}$. $Z \sim$ Normal$(0,1)$. $X_n \xrightarrow{D} Z$.

4. Vectores Aleatorios

4.1. Definición

Empezamos con un ejemplo. Pedimos a un estudiante que lance dos veces una moneda. El resultado es un vector (X_1, X_2). Hay dos tipos de estudiante, el que lanza la moneda dos veces, con resultados posibles $(0,0), (0,1), (1,0), (1,1)$ con probabilidad $1/4$ cada uno. Y hay otro estudiante que simplemente lanza una vez la moneda y repite el resultado: $(0,0)$, $(1,1)$; en este caso cada coordenada tiene la misma probabilidad: $P(X_1 = 0) = P(X_2 = 0) = 1/2$. Observando sólo X_1 o X_2 no podemos diferenciar entre los dos. Hay que mirar el resultado del vector (X_1, X_2).

Sea $(\Omega, \mathcal{F}, P)$ espacio de probabilidad. Un **vector aleatorio** $X = (X_1, \dots, X_d)$ es una función

$$X : \Omega \to \mathbb{R}^d$$

tal que cada coordenada X_i es una variable aleatoria.

Espacio de probabilidad inducido por el vector X

El conjunto de los *Borelianos* d dimensionales $\mathcal{B}^d$ es la sigma álgebra generada por los rectángulos semi-infinitos $\{x \in \mathbb{R}^d : x \leq a\}$ $a \in \mathbb{R}^d$, donde el orden de vectores es coordenada a coordenada:

$$x \leq a \text{ si y sólo si } x_i \leq a_i \text{ para todo } i = 1, \dots, d.$$

Un vector aleatorio induce el espacio de probabilidad

$$(\mathbb{R}^d, \mathcal{B}^d, P_X),$$

donde el espacio muestral es $\mathbb{R}^d$ los eventos son los Borelianos d-dimensionales y la probabilidad P_X se define

$$P_X(B) = P(X \in B).$$

Para que P_X esté bien definida necesitamos el siguiente lema.

Lema 18 *Un vector aleatorio satisface que para todo $B \in \mathcal{B}^d$,*

$$\{X \in B\} \in \mathcal{F}.$$

Dem Por definición, como las X_i son variables aleatorias, para $B_i \in \mathcal{B}$,

$$\{X \in B_1 \times \cdots \times B_d\} \ = \ \cap_{i=1}^d \{X_i \in B_i\} \ \in \ \mathcal{F}.$$

Por otro lado $\mathcal{C} = \{B : \{X \in B\} \in \mathcal{F}\}$ es sigma-álgebra. Como $\mathcal{C}$ contiene los rectangulos, contiene $\mathcal{B}^d$ la sigma álgebra de Borel d-dimensional, pues es la menor que contiene rectangulos. $\square$

4.2. Función de distribución acumulada

Para un vector aleatorio X definimos $F_X : \mathbb{R}^d \to [0,1]$:

$$F_X(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d) = P(X \leq x),$$

usando la notación $x = (x_1, \ldots, x_d)$ y el orden parcial $x \leq y$ si $x_i \leq y_i$ para cada coordenada i.

Ejemplo Lanzamos dos monedas $X_1 :=$ cantidad de caras; $X_2 := \mathbf{1}\{$los resultados son iguales$\}$. La Función de distribución en este caso se calcula a partir de las probabilidades puntuales:

Ω	P	X_1	X_2
0,0	1/4	0	1
0,1	1/4	1	0
1,0	1/4	0	1
1,1	1/4	2	1

	0	1	2	X_1
0	0	1/4	0	
1	1/2	0	1/4	
X_2				

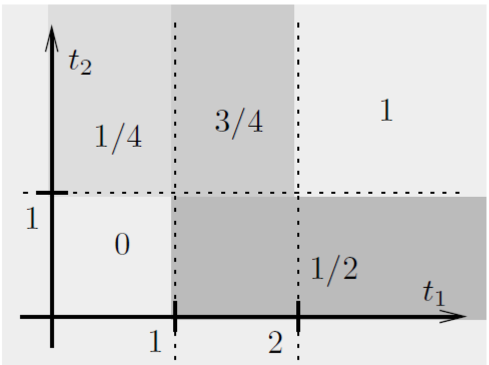


Figura 1: Función acumulada bi-dimensional

Propiedades de la función de distribución acumulada

1) $\lim_{x \to \infty} F_X(x) = 1$

2) $\lim_{x \to -\infty} F_X(x) = 0$

3) Continuidad a derecha: Si $x^n \searrow x \Rightarrow \lim_n F_X(x^n) = F_X(x)$

4) Monotonía. $x \leq y \Rightarrow F(x) \leq F(y)$.

Para que una F que satisface 1 a 4 sea de distribución hay que pedirle que la probabilidad inducida en rectangulos sea no negativa. Por ejemplo en dimensión 2, debe valer la condición

$$P(X \in (a,b] \times (c,d]) = F(b,d) - F(b,c) - F(a,d) + F(a,c) \geq 0. \tag{9}$$

Esto caracteriza la probabilidad en los rectangulos de $\mathbb{R}^2$ y por lo tanto caracteriza la probabilidad. En $\mathbb{R}^d$ hay una formula análoga a (9).

5) El valor inducido por F sobre rectangulos debe ser no negativo. En dimension 2, se debe satisfacer la desigualdad (9).

En general no es necesario chequear la propiedad (9) porque las probabilidades se definen en forma más directa en el caso discreto por la probabilidad puntual y en el caso continuo por una densidad.

4.3. Funciones de vectores aleatorios

$X \in \mathbb{R}^d$ vector aleatorio y $g : \mathbb{R}^d \to \mathbb{R}^k$. Defina $Y = g(X)$. Decimos que g es *medible Borel* si $g^{-1}(B) \in \mathcal{B}^d$ para todo $B \in \mathcal{B}^k$.

Lema 19 *Si g es medible Borel y $X \in \mathbb{R}^d$ es un vector aleatorio, entonces $Y \in \mathbb{R}^k$ es un vector aleatorio.*

Lema 20 *(i) g continua implica g medible. (ii) g limite de medibles Borel implica medible Borel.*

Por ejemplo, si $X_1, \dots, X_n$ son variables aleatorias definidas en $(\Omega, \mathcal{F}, P)$ entonces son vectores aleatorios $X_1 + \cdots + X_n$, $X_1^2 + e^{X_3+X_4}$, las funciones continuas de esas variables, etc.

4.4. Vectores aleatorios discretos

Un vector $X = (X_1, \dots, X_d)$ es *discreto* si cada coordenada es una variable aleatoria discreta. El rango $R_X := \{x \in \mathbb{R}^d : P(X = x) > 0\}$ de un vector aleatorio discreto está contenido en $R_{X_1} \times \cdots \times R_{X_d}$, el producto cartesiano de los rangos de las X_i, y por lo tanto es a lo sumo numerable.

En el ejemplo precedente el alumno aplicado tiene rango $R_X = \{(0,0), (0,1), (1,0), (1,1)\}$ y el fiaca $R_X = \{(0,0), (1,1)\}$ ambos contenidos en $R_{X_1} \times R_{X_2} = \{0,1\} \times \{0,1\}$.

Función de probabilidad conjunta puntual Sea $X = (X_1, \dots, X_d)$ un vector aleatorio discreto. Definimos la probabilidad puntual conjunta por

$$p(x_1, \dots, x_d) = P(X_1 = x_1, \dots, X_d = x_d)$$

y la probabilidad del evento $\{X \in A\}$ por

$$P((X_1, \dots, X_d) \in A) = \sum_{(x_1,\dots,x_d) \in A} p(x_1, \dots, x_d)$$

La probabilidad conjunta satisface $\sum_{x_1} \cdots \sum_{x_d} p(x_1, \dots, x_d) = 1$.

Distribuciones marginales

Lema 21 *Para el vector $X = (X_1, \dots, X_d)$, vale que*

$$P(X_1 = x_1) = \sum_{x_2} \cdots \sum_{x_d} p(x_1, \dots, x_d)$$

y similar para las otras coordenadas.

Las distribuciones de las coordenadas de un vector aleatorio se llaman *distribuciones marginales.*

Dem Esto sigue del teorema de la probabilidad total. La familia $\{\{X_2 = x_2, \dots, X_d = x_d\} : x_2, \dots, x_d \in R\}$ es una partición. $\square$

Ejemplo Sea (X_1, X_2) vector con $p(0,0) = 0{,}4$, $p(0,1) = 0{,}2$, $p(1,0) = 0{,}1$ y $p(1,1) = 0{,}3$.

Las marginales son

$P(X_1 = 0) = p(0,0) + p(0,1) = 0{,}6$

$P(X_1 = 1) = p(1,0) + p(1,1) = 0{,}4$

Toda la información se concentra en la tabla:

	0	1	X_1
0	0.4	0.2	0.6
1	0.1	0.3	0.4
X_2	0.5	0.5	1

4.5. Independencia

Si $(X_1, \dots, X_d)$ es un vector aleatorio, decimos que las variables $X_1, \dots, X_d$ son *independientes* si

$$P(X_1 \in B_1, \dots, X_d \in B_d) = P(X_1 \in B_1) \dots P(X_d \in B_d)$$

para todo $B_1, \dots, B_d$ Borelianos en $\mathcal{B}^1$.

Independencia en el caso general discreto

Lema 22 *Sea $(X_1, \dots, X_d)$ un vector discreto. Entonces son equivalentes*

i. $X_1, \dots, X_d$ son variables aleatorias independientes

ii. Las puntuales factorizan

$$P(X_1 = x_1, \dots, X_d = x_d) = P(X_1 = x_1) \dots P(X_d = x_d)$$

para todo $(x_1, \dots, x_d) \in R_X$.

iii. Las acumuladas factorizan:

$$F_X(x) = F_{X_1}(x_1) \dots F_{X_d}(x_d)$$

iv. Existen funciones $q_1, \dots, q_d$ tal que

$$P(X_1 = x_1, \dots, X_d = x_d) = q_1(x_1) \dots q_d(x_d) \tag{10}$$

v. Existen funciones $H_1, \dots, H_d$ tales que

$$F_X(x) = H_1(x_1) \dots H_d(x_d)$$

Dem $i \Rightarrow ii]$ Los puntos son Borelianos implica la factorización.

$ii \Rightarrow i]$

$$\begin{aligned}
P(X_i \in B_i,\ i = 1\dots,d) &= P((X_i \in B_i \cap R_{X_i},\ i = 1\dots,d) \\
&= \sum_{x:x_i \in B_i \cap R_{X_i} \forall i} P(X_1 = x_1,\dots,X_d = x_d)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{x:x_i \in B_i \cap R_{X_i} \forall i} P(X_1 = x_1)\dots P(X_d = x_d) \qquad \text{por hipotesis} \\
&= \sum_{x_1 \in B_1 \cap R_{X_1}} P(X_1 = x_1) \cdots \sum_{x_d \in B_d \cap R_{X_d}} P(X_d = x_d) \\
&= P(X_1 \in B_1)\dots P(X_d \in B_d)
\end{aligned}$$

$ii \Rightarrow iv]$ Inmediato poniendo $q_i = p_{X_i}$

$iv \Rightarrow ii]$ Sumando (10) sobre las coordenadas $j \neq i$ y usando iv:

$$\begin{aligned}
P(X_i = x_i) &= \sum_{j:j\neq i} \sum_{x_j} P(X_1 = x_1,\dots,X_d = x_d) \\
&= q_i(x_i) \sum_{j:j\neq i} \sum_{x_j} q_1(x_1)\dots q_d(x_d) = q_i c_i,
\end{aligned}$$

porque la suma no depende de x_i. Por lo tanto,

$$P(X_1 = x_1,\dots,X_d = x_d) = q_1 \dots q_d = \frac{P(X_1 = x_1)\dots P(X_d = x_d)}{c_1 \dots c_d}$$

pero sumando en $x_1,\dots,x_d$ obtenemos $\frac{1}{c_1\dots c_d} = 1$ y concluimos.

$ii \Rightarrow iii]$ Inmediata.

$iii \Rightarrow ii]$ Se hace una cuenta explícita, por ejemplo si X Y son discretas en $d = 2$, vimos que

$$p(x,y) = \big[F(x,y) - F(x,y-) - F(x-,y) + F(x-,y-)\big] \tag{11}$$

Si las acumuladas factorizan, también factorizan los limites $x-$ y $y-$:

$$= F_X(x)F_Y(y) - F_X(x)F_Y(y-) - F_X(x-)F_Y(y) + F_X(x-)F_Y(y-),$$

por hipótesis; ahora, sacando factor común:

$$\begin{aligned}
&= F_X(x)[F_Y(y) - F_Y(y-)] - F_X(x-)[F_Y(y) - F_Y(y-)] \\
&= [F_X(x) - F_X(x-)]\,[F_Y(y) - F_Y(y-)] = p_X(x)p_Y(y)
\end{aligned}$$

$iv \Rightarrow v]$ Inmediata.

$v \Rightarrow iv]$ Lo pruebo para $d = 2$, la prueba general es igual, usando la fórmula que relaciona las puntuales con las acumuladas. Usando (11) y (v),

$$\begin{aligned}
p(x,y) &= \big[H_1(x)H_2(y) - H_1(x)H_2(y-) - H_1(x-)H_2(y) + H_1(x-)H_2(y-)\big] \\
&= (H_1(x) - H_1(x-))(H_2(x) - H_2(x-)) =: q_1(x)q_2(x). \qquad \square
\end{aligned}$$

Ejemplos

(1) $P(X_1 = x_1, X_2 = x_2) = \frac{1}{Z}a^{x_1+x_2}$, $x_1, x_2 \geq 1$ enteros y $a > 0$. Como factorizan, son independientes.

(2) $P(X_1 > x_1, X_2 > x_2) = a^{x_1+x_2}$, si x_1, x_2 enteros positivos, mientras que $P(X_1 > x_1, X_2 > x_2) = 1$ para x_1, x_2 enteros no positivos. Por lo tanto las acumuladas factorizan. Poniendo $x_2 = 0$ obtenemos

$$P(X_1 > x_1, X_2 > 0) = P(X_1 > x_1) = a^{x_1}.$$

Analogamente $P(X_2 > x_2) = a^{x_2}$, por una fórmula análoga a la anterior:

$$\begin{aligned}
P(X_1 = x_1, X_2 = x_1) &= P(X_1 > x_1 - 1, X_2 > x_1 - 1) - P(X_1 > x_1 - 1, X_2 > x_1) \\
&\quad - P(X_1 > x_1, X_2 > x_1 - 1) + P(X_1 > x_1, X_2 > x_1) \\
&= a^{x_1+x_2-2} - a^{x_1+x_2-1} - a^{x_1+x_2-1} + a^{x_1+x_2} \\
&= a^{x_1+x_2-2}(1 - 2a + a^2) = a^{x_1+x_2-2}(1-a)^2
\end{aligned}$$

Se trata de dos geometricas independientes de parámetro $(1 - a)$.

4.6. Vectores aleatorios continuos

Decimos que el vector $X : \Omega \to \mathbb{R}^d$ es continuo si existe una función $f : \mathbb{R}^d \to \mathbb{R}^+$ llamada función de densidad conjunta con $\int_{\mathbb{R}^d} f(x)dx = 1$ tal que

$$P(X \in [a_1, b_1] \times \cdots \times [a_d, b_d] = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(x_1, \ldots, x_d)dx_1 \ldots dx_d.$$

Ya vimos que las probabilidades de rectángulos en R^d determinan probabilidad. Así, para todo Boreliano d dimensional B:

$$P(X \in B) = \int_B f(x)dx$$

Ejemplo Sea X Uniforme en un Boreliano A de volumen finito, $\int_A dx < \infty$. Sea

$$f(x) = \frac{\mathbf{1}\{x \in A\}}{\int_A dx}$$

Así, para $B \subset A$, Boreliano,

$$P(X \in B) = \int_B f(x)dx = \frac{\int_B dx}{\int_A dx} = \frac{\text{Volumen de } B}{\text{Volumen de } A}$$

Relación entre la acumulada y la densidad

Acumulada en el caso continuo:

$$F(x) := \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y_1, \ldots, y_d)dy_1 \ldots y_d$$

Si f es continua en $(x_1, \ldots, x_d)$, derivando respecto de cada coordenada queda

$$\frac{\partial F}{\partial x_1 \ldots \partial x_d} = f(x_1, \ldots, x_d)$$

Distribuciones marginales

Lema 23 *Si $\underline{X} = (X_1, \dots, X_d)$ es un vector aleatorio continuo con densidad f, entonces X_1 es continua con densidad*

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_2 \dots dx_d$$

Esta es la densidad *marginal* de la primera coordenada. El lema vale para las otras coordenadas con las modificaciones notacionales apropiadas.

Dem

$$\begin{aligned} P(X_1 \in (a,b)) &= P(\underline{X} \in [(a,b) \times \mathbb{R} \cdots \times \mathbb{R}]) \\ &= \int_a^b \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 dx_2 \dots dx_d \\ &= \int_a^b \Big(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_2 \dots dx_d \Big) dx_1 = \int_a^b f_{X_1}(x_1), \end{aligned}$$

usando Fubini. □

Ejemplo Dardo. (X, Y) uniforme en $B(0, 1)$. La densidad es

$$f(x, y) = \frac{1}{\pi} \mathbf{1}\{(x, y) \in B(0, 1)\}$$

Marginal de x:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy \ \mathbf{1}\{x \in (-1, 1)\} \\ &= \frac{2\sqrt{1 - x^2}}{\pi} \mathbf{1}\{x \in (-1, 1)\}. \end{aligned}$$

4.7. Independencia de variables continuas

Lema 24 *$(X_1, \dots, X_d)$ vector aleatorio continuo con densidad f. Son equivalentes*

i. $X_1, \dots, X_d$ son independientes

ii. $f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \dots f_{X_d}(x_d)$.

iii. $F_X(x_1, \dots, x_d) = F_{X_1}(x_1) \dots F_{X_d}(x_d)$

iv. Existen funciones $h_1, \dots, h_d : \mathbb{R} \to \mathbb{R}^+$ tal que $f_X(x_1, \dots, x_d) = h_1(x_1) \dots h_d(x_d)$

v. Existen funciones $H_1, \dots, H_d : \mathbb{R} \to [0, 1]$ tal que $F_X(x_1, \dots, x_d) = H_1(x_1) \dots H_d(x_d)$

Dem $i \Rightarrow iii]$ Fácil. $(-\infty, x]$ es Boreliano, entonces

$$P(X_1 \le x_1, \dots, X_d \le x_d) = P(X_1 \le x_1) \dots P(X_d \le x_d)$$

$ii \Rightarrow i]$ Facil:

$$\begin{aligned} P(X_i \in B_i, \ i = 1 \dots, d) &= \int_{B_1} \cdots \int_{B_d} f_X(y_1, \dots, y_d) dy_1 \dots y_d \\ &= \int_{B_1} f_{X_1}(y_1) dy_1 \cdots \int_{B_d} f_{X_d}(y_d) y_d = \prod_i P(X_i \in B_i) \end{aligned}$$

$ii \Rightarrow iii$] Misma que la anterior para los borelianos $(-\infty, x_i]$.

$iii \Rightarrow ii$] Facil. Como X es continua y las funciones acumuladas factorizan,

$$\begin{aligned} f(y_1, \ldots, y_d) &= \frac{\partial F}{\partial x_1 \ldots \partial x_d} = \frac{\partial^d (F_{X_1}(x_1) \ldots F_{X_d}(x_d))}{\partial x_1 \ldots \partial x_d} \\ &= \frac{\partial F_{X_1}(x_1)}{\partial x_1} \ldots \frac{\partial F_{X_d}(x_d)}{\partial x_d} = f_{X_1} \ldots f_{X_d} \end{aligned}$$

$iv \Rightarrow ii$] Calculemos la distribución marginal de X_i en función de las h_j. Sea B_i Boreliano en $\mathbb{R}$.

$$\begin{aligned} P(X_i \in B_i) &= \int \cdots \int f(x_1, \ldots, x_d)\, \mathbf{1}\{x_i \in B_i\} dx_1 \ldots dx_d \\ &= \int \cdots \int h_1(x_1) \ldots h_d(x_d)\, \mathbf{1}\{x_i \in B_i\} dx_1 \ldots dx_d \\ &= \int_{B_i} h_i(x_i) dx_i \prod_{j: j \neq i} \int h_j(x_j) dx_j = c_i \int_{B_i} h_i(x_i) dx_i \end{aligned}$$

porque el producto de las integrales no depende de x_i. Usando iv tenemos

$$\begin{aligned} P(X_1 \in B_1, \ldots, X_d \in B_d) &= \int \cdots \int f(x_1, \ldots, x_d) \Big(\prod_{i=1}^d \mathbf{1}\{X_i \in B_i\} \Big) dx_1 \ldots dx_d \\ &= \prod_i \int h_i(x_i) \mathbf{1}\{X_i \in B_i\} dx_i \\ &= \prod_i \frac{P(X_i \in B_i)}{c_i} \end{aligned}$$

Pero tomando $B_i = \mathbb{R}$ para todo i, obtenemos $\frac{1}{c_1 \ldots c_d} = 1$ y concluimos.

$ii \Rightarrow iv$] Basta tomar $h_i = f_{X_i}$.

$iii \Rightarrow v$] Basta tomar $H_i = F_{X_i}$.

$v \Rightarrow iii$] Por hipótesis v,

$$F(z_1, \ldots, z_d) = \prod_{j=1}^d H_j(z_j) \tag{12}$$

Sacando límites $z_j \to \infty$ para todo $j \neq i$ tenemos

$$F_{X_i}(z_i) = H_i(z_i) \prod_{j \neq i} \lim_{z_j \to \infty} H_j(z_j) = H_i(z_i) c_i. \tag{13}$$

(justifique que los límites a la derecha existen). Poniendo (13) en (12), tenemos

$$F(z_1, \ldots, z_d) = \prod_{j=1}^d \frac{F_{X_j}(z_j)}{c_j}$$

Sacando limite $z_i \to \infty$ para todo i vemos que el lado izquierdo y el numerador del lado derecho convergen a 1. Por lo tanto, $\frac{1}{c_1 \ldots c_d} = 1$ y concluimos. □

Ejemplo X Y con densidad conjunta $f(x, y) = e^{-x-y}$, $x, y > 0$. Entonces $f(x, y)$ se factoriza como $f(x, y) = e^{-x} e^{-y}$ y son independientes.

4.8. Distribución de funciones de vectores continuos

Ejemplo Si un dardo es lanzado a un punto uniformemente distribuído en la bola unitaria. Cual es la distribución de R, la distancia del dardo al origen? $X = (X_1, X_2) \sim$ Uniforme en $B(0,1) \subset \mathbb{R}^2$, es decir $f_X(x) = \frac{1}{\pi}\mathbf{1}\{x \in B(0,1)\}$.

La distancia al origen es $R := \sqrt{X_1^2 + X_2^2} \in [0,1]$. Para $r \in [0,1]$, calculamos

$$\begin{aligned} F_R(r) &= P(R \le r) = P(X_1^2 + X_2^2 \le r^2) = P((X_1, X_2) \in B(0,r)) \\ &= \frac{\text{area de } B(0,r)}{\text{area de } B(0,1)} = \frac{\pi r^2}{\pi} = r^2. \end{aligned}$$

$F_r(r) = 0$ para $r \le 0$ y $F_R(r) = 1$ para $r \ge 1$. Derivando en $r \notin \{0,1\}$, obtenemos la densidad:

$$f_R(r) = 2r\,\mathbf{1}\{0 \le r \le 1\}.$$

Caso general Sea (X,Y) vector aleatorio continuo y $g : \mathbb{R}^2 \to \mathbb{R}^2$ medible e **invertible**. Defino $(U,V) := g(X,Y)$. Sea $\mathrm{Im}(g)$ la imagen de g. Para un Boreliano $B \subset \mathbb{R}^2$,

$$\begin{aligned} P((U,V) \in B) &= \int_{g^{-1}(B)} f(x,y)dxdy \\ &= \int_B f(g^{-1}(u,v))\,|Jg^{-1}(u,v)|\,\mathbf{1}\{(u,v) \in \mathrm{Im}(g)\}\,dudv \end{aligned}$$

por el teorema de cambio de variable para g invertible continua con Jacobiano diferente de cero. El Jacobiano es

$$Jg^{-1} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \\ \frac{\partial x}{\partial v} & \frac{\partial x}{\partial v} \end{pmatrix}$$

donde $(x,y) = g^{-1}(u,v)$.

Ejemplo Vector (X,Y) con X, Y normales standard independientes:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

Cambio a coordenadas polares definiendo $g(x,y) := (u,v)$ donde

$$u := x^2 + y^2, \qquad v := \arctan(y/x)$$

Para calcular la distribución de $(U,V) := g(X,Y)$, usamos que la función

$$g : \mathbb{R}^2 \setminus \{(0,0)\} \to (0,\infty) \times [0, 2\pi)$$

es invertible con inversa $g^{-1}(u,v) = (x,y)$ dados por

$$x = \sqrt{u}\,\cos(v), \qquad y = \sqrt{u}\,\mathrm{sen}(v)$$

Cálculo del Jacobiano:

$$\begin{aligned} Jg^{-1}(u,v) &= \det \begin{pmatrix} \frac{1}{2\sqrt{u}}\cos(v) & -\sqrt{u}\,\mathrm{sen}(v) \\ \\ \frac{1}{2\sqrt{u}}\mathrm{sen}(v) & \sqrt{u}\cos(v) \end{pmatrix} \\ &= \frac{1}{2}\cos^2(v) + \frac{1}{2}\mathrm{sen}^2(v) = \frac{1}{2} \end{aligned}$$

Entonces

$$
\begin{aligned}
f_{U,V}(u,v) &= \frac{1}{2\pi}\, e^{-u/2}\, \frac{1}{2}\, \mathbf{1}_{(0,\infty)}(u)\, \mathbf{1}_{(0,2\pi)}(v) \\
&= \Big(\frac{1}{2} e^{-u/2}\, \mathbf{1}_{(0,\infty)}(u)\Big)\Big(\frac{1}{2\pi}\, \mathbf{1}_{(0,2\pi)}(v)\Big)
\end{aligned}
$$

Implica que U, V independientes, U exponencial, V Uniforme$[0, 2\pi]$. Para verlo calcule las marginales.

Suma de variables. Sea (X, Y) vector aleatorio con densidad $f_{X,Y}$ y $U = X + Y$.

$$F_U(u) = P(X + Y \le u) = \int\!\!\int_{A_u} f(x,y)dxdy$$

donde $A_z = \{(x, y) : x + y \le z\}$. No se puede aplicar el teorema de cambio de variable porque $g(x, y) = x+y$ no es invertible. Se agrega otra coordenada para hacer g invertible: $g(x, y) = (x+y, x)$ es invertible. Tenemos $g^{-1}(u, v) = (v, u-v)$ con Jacobiano 1. Entonces:

$$f_{U,V}(u,v) = f_{X,Y}(v, u-v)$$

Calculamos densidad de la marginal de $U = X + Y$:

$$f_U(u) = \int f_{X,Y}(v, u-v)dv$$

Cuando X e Y son independientes obtenemos la fórmula

$$f_U(u) = \int f_X(v) f_Y(u-v)dv := (f_X \star f_Y)(u)$$

que se llama la *convolución* de f_X y f_Y.

Ejemplo Suma de exponenciales independientes. $f_{X_i}(u) = \lambda e^{-\lambda u}\mathbf{1}\{u \ge 0\}$.

$$
\begin{aligned}
f_{X_1+X_2}(u) &= \int_0^\infty f_{X_1}(v) f_{X_2}(u-v)dv \\
&= \lambda^2 \int_0^\infty e^{-\lambda v} e^{-\lambda(u-v)}\mathbf{1}\{v \ge 0\}\mathbf{1}\{u - v \ge 0\}dv \\
&= \lambda^2 u e^{-\lambda u}
\end{aligned}
$$

Gama $(2, \lambda)$.

Ejercicio: Algo análogo vale para el caso discreto. Si X e Y son variables discretas independientes y $Z = X + Y$, entonces

$$p_Z(z) = \sum_x p_X(x) p_Y(z-y)$$

Funciones de vectores independientes son independientes

Lema 25 *Sean $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_m)$ independientes, es decir, para Borelianos B y C tenemos $P(X \in B, Y \in C) = P(X \in B)P(Y \in C)$. Sean $g : R^n \to \mathbb{R}^k$ y $h : R^m \to \mathbb{R}^j$ funciones medibles. Entonces $g(X)$ y $h(Y)$ son independientes.*

Dem: Como para $B \in \mathcal{B}^k$, $\{x \in \mathbb{R}^n : g(x) \in B\} = g^{-1}(B)$,

$$
\begin{aligned}
P(g(X) \in B, h(Y) \in C) &= P(X \in g^{-1}(B), Y \in h^{-1}(C)) \\
&= P(X \in g^{-1}(B))P(Y \in h^{-1}(C)) \quad \text{(indep de } X, Y) \\
&= P(g(X) \in B)P(h(Y) \in C). \quad \square
\end{aligned}
$$

5. Esperanza

5.1. Definición

Si X es una variable **discreta** definimos

$$EX := \sum_{x \in R_X} xP(X = x) = \sum_x xp(x)$$

si alguna de las series $\sum_{x \in \mathbb{R}_X, x>0} xp(x)$, $\sum_{x \in \mathbb{R}_X, x<0} xp(x)$ converge.

Interpretaciones. Centro de gravedad. Promedio asintótico de los resultados de n experimentos (esta es la ley de grandes números que veremos después).

Ejemplos: Bernoulli $EX = 1p + 0(1-p) = p$,

Binomial:

$$EX = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np$$

Poisson:

$$EX = \sum_{x \geq 0} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x \geq 1} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \sum_{x \geq 0} \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$$

(porque el primer término de la suma a la izquierda es cero).

Indicadoras:

A evento en Ω, definimos $X(\omega) = \mathbf{1}\{\omega \in A\}$. X es una variable discreta que asume dos valores: 1 con probabilidad $P(A)$ y 0 con probabilidad $1 - P(A)$.

$$EX = E(\mathbf{1}\{A\}) = 1\,P(A) + 0\,(1 - P(A)) = P(A).$$

Si X es **continua** con densidad f. Definimos.

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$

si las integrales positiva y negativa están bien definidas.

Ejemplos *Uniforme*$[0,1]$, $EX = 1/2$. *Uniforme*$[a,b]$, $EX = (b-a)/2$.

Normal(μ, σ^2) Sumar y restar μ, una parte da μ y la otra da 0 porque es la integral de una función impar. $EX = \mu$.

Esperanza de funciones de variables aleatorias

Proposición 26 *$X : \Omega \to \mathbb{R}^d$ vector aleatorio, $g : \mathbb{R}^d \to \mathbb{R}$. Entonces, en el caso discreto:*

$$Eg(X) = \sum_x g(x)P(X = x)$$

En el caso continuo, si X tiene densidad f:

$$Eg(X) = \int_{\mathbb{R}^d} g(x)\, f(x)\, dx$$

Dem Caso discreto:

$$
\begin{aligned}
Eg(X) &= \sum_y y\, P(g(X)=y) = \sum_y \sum_x yP(g(X)=y, X=x) \\
&= \sum_y \sum_{x:g(x)=y} yP(X=x) = \sum_y \sum_{x:g(x)=y} g(x)P(X=x) \\
&= \sum_x g(x)P(X=x)
\end{aligned}
$$

En la primera linea usamos probabilidad total. De la primera a la segunda usamos que el evento $\{g(X)=y, X=x\}$ es igual a vacío si $g(x)\neq y$ e igual a $\{X=x\}$ si $g(x)=y$.

Caso continuo: X vector continuo con densidad f. Si la imagen de g es discreta, se calcula como antes:

$$
\begin{aligned}
Eg(X) &= \sum_y yP(g(X)=y) = \sum_y y \int_{\{x:g(x)=y\}} f(x)\,dx \\
&= \sum_y \int_{\{x:g(x)=y\}} g(x)\,f(x)\,dx = \int g(x)f(x)dx.
\end{aligned}
$$

Si la imagen de g no es discreta, aproximamos g por funciones g_n con imagenes discretas. $\square$

5.2. Propiedades

Para X, Y variables aleatorias,

1. **Linealidad**

$$
E\Big(\sum_i a_i X_i\Big) = \sum_i a_i EX_i
$$

Dem: Caso discreto:

$$
\begin{aligned}
E\Big(\sum_i a_i X_i\Big) &= \sum_{x_1,\dots,x_n} p(x_1,\dots,x_n)(a_1x_1+\dots+a_nx_n) \\
&= \sum_{i=1}^n \sum_{x_i} a_i x_i \sum_{\{x_1,\dots,x_n\}\setminus\{x_i\}} p(x_1,\dots,x_n) \\
&= \sum_{i=1}^n a_i \sum_{x_i} x_i p_{X_i}(x_i) = \sum_i a_i EX_i
\end{aligned}
$$

El caso continuo sigue de la misma forma.

2. **Monotonía** Si $P(X\ge Y)=1$ entonces $EX\ge EY$.

Dem: Sea $V=X-Y$. Claramente $P(V\ge 0)=1$. Por lo tanto $EV=\sum_x xP(V=x)\ge 0$ porque los sumandos nunca son negativos. Por linealidad $0\le E(X-Y)=EX-EY$.

El mismo argumento se aplica al caso continuo.

3. **Esperanza de constantes** Si X es constante, es decir $P(X=c)=1$ para algún c, entonces $EX=c$.

Dem: es una variable discreta. $EX = cP(X = c) = c$.

4. **Módulos** $|EX| \le E|X|$

Dem: Veamos el caso discreto

$$|EX| = \Big|\sum_x xp(x)\Big| \le \sum_x |x|p(x) = E|X|.$$

Vale para sumas generales: a_i reales. Por definición de módulo, $-|a_i| \le a_i \le |a_i|$, que sumando en i da $-\sum |a_i| \le \sum a_i \le \sum |a_i|$. Por lo tanto $|\sum a_i| \le \sum |a_i|$.

Ejemplos

$S_n \sim$ *Binomial*(n, p). $S_n = X_1 + \cdots + X_n$ con X_i iid Bernoulli(p).

$$ES_n = E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n = np$$

$W_r \sim$ *Binomial negativa* (r, p). Número de ensayos Bernoulli(p) hasta el r-ésimo éxito. $W_r = Y_1 + \cdots + Y_r$, con $Y_j \sim$ Geometrica (p). Como $Y_j = \frac{1}{p}$, tenemos

$$EW_r = EY_1 + \cdots + EY_r = \frac{r}{p}.$$

Proposición 27 *Sea $X \ge 0$ una variable aleatoria no negativa con función de distribución acumulada F y esperanza finita $EX < \infty$. Entonces $EX = \int_0^\infty (1 - F(x))dx$*

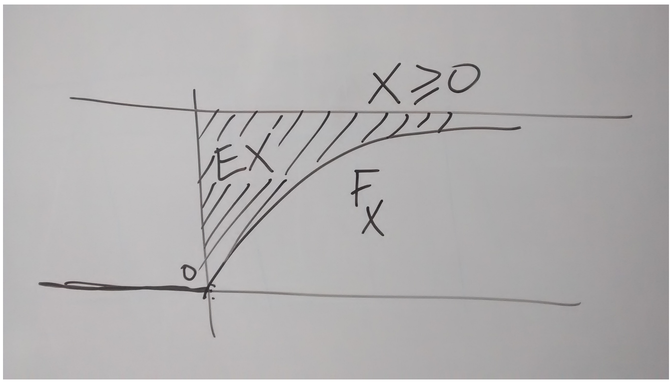
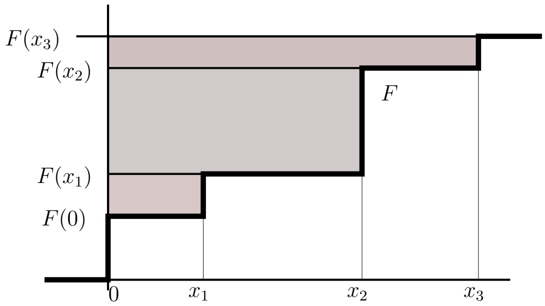


Figura 2: La esperanza de $X \ge 0$ es el área sobre la función acumulada.

Dem 1. Caso discreto. Sea X discreta con rango $R_X = \{x_0, x_1, x_2, \dots\}$ con $x_0 = 0$ y $x_{i-1} < x_i$ para todo $i \ge 1$. Como $p(x_i) = F(x_i) - F(x_{i-1})$, tenemos

$$\begin{aligned}
EX &= \sum_{i=1}^\infty x_i p(x_i) = \sum_{i=1}^\infty x_i (F(x_i) - F(x_{i-1})) \\
&= \sum_{i=1}^\infty \sum_{j=1}^{i} (x_j - x_{j-1})(F(x_i) - F(x_{i-1})) \\
&= \sum_{j=1}^\infty \sum_{i=j}^{\infty} (x_j - x_{j-1})(F(x_i) - F(x_{i-1})) \qquad \text{(Fubini)} \\
&= \sum_{j=1}^\infty (x_j - x_{j-1})(1 - F(x_{j-1})) = \int_0^\infty (1 - F(x))dx
\end{aligned}$$



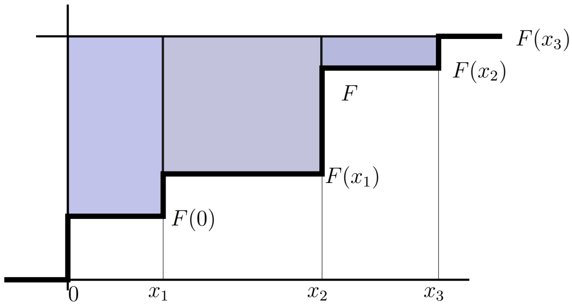


Figura 3: Las figuras muestran dos maneras de expresar $\int_0^\infty (1-F(x))dx$ en el caso discreto. A la izquierda $\sum_{i\geq 1} x_i(F(x_i)-F(x_{i-1}))$ y a la derecha $\sum_{j\geq 1}(x_j-x_{j-1})(1-F(x_{j-1}))$.

Esto es una integración por partes discreta. Se llama sumas de Abel. Si 0 no está en el rango, tendremos $F(0)=0$ y exactamente la misma cuenta vale.

Ejemplo: Geométrica(p).

$$EX = \sum_{k\geq 0} P(X>k) = \sum_{k\geq 0} (1-p)^k = \frac{1}{p}$$

2. Caso continuo. $X \geq 0$ con densidad f y función de distribución acumulada F. Escribimos $x = \int_0^x dy$ y obtenemos

$$\int_0^\infty xf(x)dx = \int_0^\infty \int_0^x dy f(x)dx = \int_0^\infty \int_y^\infty f(x)dxdy = \int_0^\infty (1-F(y))dy$$

Intercambio de integrales por Fubini porque todo es positivo. □

Si X toma valores negativos y positivos, $X = X^+ - X^-$, por lo tanto

$$EX = EX^+ - EX^- = \int_0^\infty (1-F_{X^+}(x))dx - \int_0^\infty (1-F_{X^-}(x))dx$$

Lema 28 *$X \geq 0$ y $EX = 0$ implica $P(X=0)=1$.*

Dem La siguiente desigualdad es inmediata:

$$X \geq \frac{1}{n}\mathbf{1}\{X > 1/n\}, \quad \text{para todo } n > 0.$$

De hecho, cuando $X \leq 1/n$, el segundo miembro es 0; cuando $X > 1/n$, el segundo miembro es $1/n$. Por hipótesis $EX = 0$ y por monotonía, usando la desigualdad de arriba,

$$0 = nEX \geq P(X > 1/n) \underset{n\to\infty}{\longrightarrow} P(X>0).$$

Esto implica que $P(X=0)=1$. □

5.3. Ejemplos

Decimos que un juego es honesto si la esperanza del lucro es cero.

Juego de cara o ceca Si te doy 15 pesos por cada vez que salga cara, cuanto apostarías?

$L =$ lucro. Si apuesto k la esperanza del lucro es

$EL = -k\frac{1}{2} + 15\frac{1}{2} = \frac{15-k}{2}$. Parece razonable (honesto) apostar 15.

Y si son 15 millones de pesos? Apostarías 15 millones de pesos?

Juego de San Petersburgo Se lanza una moneda hasta que sale cara. $N =$ número de veces que la moneda es lanzada hasta la primera cara. Geometrica 1/2.

Premio: 2^N.

Cuanto pagarías para participar de este juego? digamos K

$L =$ ganancia $= 2^N - K$. $EL = \infty$.

Pagarías $K = 1{,}000{,}000 \sim 2^{20}$ por jugar una única vez?

La probabilidad de ganar por lo menos lo mismo que aposté es 2^{-21} mmmmmm...

5.4. El espacio L^2

Sea

$$L^2 = L^2(\Omega, \mathcal{F}, P) = \{X : \Omega \to \mathbb{R} : X \text{ variable aleatoria y } EX^2 < \infty\}$$

Definimos $\langle X, Y\rangle := E(XY)$.

Lema 29 *$\langle \cdot, \cdot \rangle$ es un producto interno.*

Dem En primer lugar, veamos que está bien definido. Como $(X - Y)^2 \geq 0$, $2|XY| \leq X^2 + Y^2$. Esto implica que si $X, Y \in L^2$, entonces

$$2E|XY| \leq EX^2 + EY^2 < \infty. \tag{14}$$

La bilinealidad $\langle aX + bY, Z\rangle = a\langle X, Z\rangle + b\langle Y, Z\rangle$ sigue de la linealidad de la esperanza. La simetría es inmediata y es definido positivo porque $\langle X, X\rangle = 0$ implica $P(X = 0) = 1$. $\square$

Lema 30 *L^2 es un espacio vectorial.*

Dem Hay que ver que si X, $Y \in L^2$, $aX + bY$ tambien:

$$E(aX + bY)^2 = a^2EX^2 + b^2EY^2 + 2abE(XY) < \infty. \quad \square$$

La esperanza como la constante que mejor aproxima X

Sea X variable aleatoria con $EX^2 < \infty$. Encontrar la constante $c \in \mathbb{R}$ tal que $E(X - c)^2$ sea mínimo.

Para eso derivamos $EX^2 - 2cEX + c^2$ en c, igualamos a cero: $-2EX + 2c = 0$ y obtenemos que el mínimo se alcanza cuando $c = EX$.

Alternativamente: Si consideramos el subespacio $V \subset L^2$ de las constantes

$$V = \{W \in L^2 : P(W = c') = 1\}$$

(verifique que es un subespacio), el c que minimiza $E(X-c)^2$ es la proyección de X sobre V: es decir el c tal que $E((X-c)c') = 0$ para todo $c' \in \mathbb{R}$. Pero esto ocurre sólo si $E(X-c) = 0$, es decir $c = EX$.

5.5. Varianza y covarianza

Varianza Definimos la *Varianza* de X por

$$VX := E(X - EX)^2.$$

Es una medida de la dispersión de la distribución de X alrededor de la media.

Ejemplo: $X \sim$N(μ, σ^2). $VX = \sigma^2$. Como $EX = \mu$,

$$VX = \int (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \cdots = \sigma^2$$

Propiedades de la varianza

1. $VX \geq 0$; $VX = 0 \Leftrightarrow X = EX$

2. $V(X+b) = VX$

3. $V(aX) = a^2 VX$

Covarianza Si $X^2 < \infty$, $EY^2 < \infty$ definimos la *covarianza* de X e Y por

$$\text{Cov}(X,Y) := E((X-EX)(Y-EY))$$

Es una medida de la dependencia de las variables centradas.

Propiedades de la covarianza

0. $\text{Cov}(X,Y) := \text{Cov}((X-EX),(Y-EY))$, no depende de las medias.

1. $\text{Cov}(X,Y) = E(XY) - EX\,EY$

2. $\text{Cov}(X,X) = VX$.

3. $|\text{Cov}(X,Y)| \leq VX + VY$.

4. $|\text{Cov}(X,Y)|^2 \leq VX\,VY$ (Cauchy-Schwarz)

5. $\text{Cov}(X,Y) = \text{Cov}(Y,X)$ simetría

6. $\text{Cov}(aX+bY,Z) = a\text{Cov}(X,Z) + b\text{Cov}(Y,Z)$ bilineal

7. X, Y independientes implica $\text{Cov}(X,Y) = 0$. No vale la vuelta.

8. $V(X+Y) = VX + VY + 2\text{Cov}(X,Y)$.

9. X, Y independientes implica $V(X+Y) = VX + VY$.

Dem

3. Sigue de (14).

4. *Desigualdad de Cauchy-Schwarz. Si* $EX^2 < \infty$ *y* $EY^2 < \infty$, *entonces* $E|XY| < \infty$ *y*

$$(E(XY))^2 \leq EX^2 EY^2. \tag{15}$$

Dem Por 3, $E(XY)$ existe y es finita. Defina

$$g(t) := E(X - tY)^2 = EX^2 - 2tE(XY) + t^2 EY^2 \geq 0$$

Se trata de un polinomio de segundo grado en t con coeficientes $a = EY^2$, $b = -2E(XY)$, $c = EX^2$ y discriminante $b^2 - 4ac = 4(E(XY))^2 - 4EX^2EY^2 \leq 0$ porque el polinomio es no negativo. Esto demuestra (15). Para concluir la demostración de 4 substituya X e Y por $(X - EX)$ e $(Y - EY)$, respectivamente.

7. Usando 1, basta probar que $E(XY) = EX\, EY$. En el caso discreto tenemos

$$\begin{aligned} E(XY) &= \sum_{x,y} xy\, p(x,y) = \sum_{x,y} xy\, p_X(x)\, p_Y(y) \\ &= \sum_x x\, p_X(x) \sum_y y\, p_Y(y) = EX\, EY. \end{aligned}$$

por independencia. El caso continuo es análogo:

$$\begin{aligned} E(XY) &= \int\int xy\, f(x,y) dx\, dy = \int\int xy\, f_X(x)\, f_Y(y) dx dy \\ &= \int x f_X(x)\, dx \int y\, f_Y(y) dy = EX\, EY, \end{aligned}$$

por independencia. □

8. Cuentas

9. Consecuencia de 7 y 8.

Coeficiente de correlación Sean X Y variables aleatorias con segundo momento finito. Definimos *coeficiente de correlación de* X *e* Y por

$$\rho(X,Y) := \frac{\text{Cov}(X,Y)}{\sqrt{VX\, VY}}$$

Propiedades

1. $|\rho(X,Y)| \leq 1$.

2. $|\rho(X,Y)| = 1 \iff$ existen a, b tales que $Y = aX + b$.

3. Sean a, b, c y d números reales, $a \neq 0$, $c \neq 0$ y X e Y variables aleatorias con varianza positiva, entonces

$$\rho(aX + b, cY + d) = \text{sg}(ac)\, \rho(X,Y)$$

donde sg denota la función signo. Esto quiere decir que el coeficiente de correlación es invariante por cambio multiplicativo de escala y por traslaciones.

Dem 1. Cauchy-Schwarz demuestra $\rho^2 \leq 1$.

2. Supongamos que $\rho = 1$. Esto implica que el discriminante de $g(t)$ es cero y que g tiene una única raiz t_0. Es decir

$$E(X - t_0 Y)^2 = 0$$

Como $(X - t_0 Y)^2 \geq 0$, por el Lema 28, $P((X - t_0 Y)^2 = 0) = 1$, lo que implica $X = t_0 Y$ con probabilidad 1.

En el caso general, substituyendo

$$E(X - EX - t_0(Y - EY))^2 = 0$$

implica que $Y = \frac{1}{t_0} X + \frac{1}{t_0} EY - EX$.

Reciprocamente, si $Y = AX + B$ entonces $|\rho| = 1$. $-1 \leq \rho(x, y) \leq 1$

3. Cuentas.

6. Esperanza condicional

6.1. Caso discreto

Sean (X, Y) un vector discreto con $E|X| < \infty$. Como $\{Y = y\}$ es un evento, $P(X = x|Y = y)$ está bien definido y podemos definir la esperanza de la variable X condicionada a $Y = y$ por

$$E(X|Y = y) = \sum_x x\, P(X = x|Y = y)$$

Definimos $\phi(y) := E(X|Y = y)$. Como $\phi(y)$ es una función de y, la función de la variable Y dada por $E(X|Y) := \phi(Y)$ es una variable aleatoria.

Para X e Y variables aleatorias discretas:

$$\begin{aligned} E\phi(Y) &= \sum_y \phi(y) P(Y = y) = \sum_y E(X|Y = y) P(Y = y) \\ &= \sum_y \sum_x x\, P(X = x|Y = y) P(Y = y) \\ &= \sum_x x \sum_y P(X = x, Y = y) = \sum_x x P(X = x) = EX \end{aligned}$$

Pudimos intercambiar las sumas porque asumimos $E|X| < \infty$. Concluimos que

$$E(E(X|Y)) = EX.$$

Ejemplo $\Omega = \{1, \dots, 6\}$. X resultado de un dado, $Y = \mathbf{1}\{\text{dado par}\}$. Calcular la distribución $p_{X|Y=1}$

$$\begin{aligned} E(X|Y=1) &= \sum_x xP(X=x|Y=1) = \sum_x x\frac{P(X=x, Y=1)}{P(Y=1)} \\ &= \sum_x x\frac{P(X=x, X\text{ par})}{P(X\text{ par})} = 2\frac{1/6}{1/2} + 4\frac{1/6}{1/2} + 6\frac{1/6}{1/2} = 4 \end{aligned}$$

Analogamente $E(X|Y=0) = 3$. Entonces $E(X|Y)$ es una variable aleatoria que toma el valor 4 con probabilidad 1/2 y el valor 3 con probabilidad 1/2. Su esperanza es

$$E(E(X|Y)) = \frac{1}{2}3 + \frac{1}{2}4 = 3{,}5 = EX$$

Ejemplo Voy de pesca, saco N pescados, $N \sim \text{Poisson}(\lambda)$. Cada pescado es comestible con probabilidad p independientemente. Es decir que la distribución de X dado $N = n$ es Binomial(n, p):

$$P(X = k|N = n) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \le k \le n.$$

Por lo tanto, $E(X|N = n) = np$, la esperanza de la Binomial(n, p). Substituyendo tenemos $E(X|N) = Np$. Podemos calcular EX usando la fórmula:

$$EX = E(E(X|N)) = E(Np) = \lambda p.$$

6.2. Esperanza condicional. Caso continuo

(X, Y) vector aleatorio continuo con densidad $f_{X,Y}$. Supongamos $E|X| < \infty$. Definimos

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}, & \text{si } f_Y(y) > 0 \\ 0, & \text{si no.} \end{cases}$$

Si $f_Y(y) > 0$, $f_{X|Y=y}$ es una función de densidad. Para verlo, integre $f_{X|Y=y}$ en x, obtenga la marginal de Y simplifique y da 1. Además,

$$\int_{-\infty}^{\infty} f_{X|Y=y}(x) f_Y(y) dy = f_X(x)$$

Para verlo, substituyo por la definición, cancelo las marginales en y e integro en y.

Def Para cada y tal que $f_Y(y) > 0$ definimos

$$\begin{aligned} \phi(y) := E(X|Y=y) &= \int_{-\infty}^{\infty} x\, f_{X|Y=y}(x) d\,x \\ E(X|Y) := \phi(Y) \end{aligned}$$

la *esperanza condicional de X dado Y*.

$E(X|Y)$ es una variable aleatoria con esperanza EX:

$$\begin{aligned} E(E(X|Y)) &= \int E(X|Y=y) f(y) dy = \int \Big(\int x f_{X|Y=y}(x) \Big) f(y) dy \\ &= \int x \Big(\int f(x,y) dy \Big) dx = \int x f_X(x) dx = EX. \end{aligned}$$

(se cambian las integrales porque $E|X| < \infty$).

Ejemplo Sea (X, Y) un vector continuo en $\mathbb{R}^2$ con densidad

$$f(x, y) = \lambda^2 e^{-\lambda y} \mathbf{1}\{0 \le x \le y\} \tag{16}$$

Calculemos primero la marginal de Y:

$$f_Y(y) = \int f(x, y)dx = \int_0^\infty \lambda^2 e^{-\lambda y} \mathbf{1}\{0 \le x \le y\}dx = \lambda^2 y e^{-\lambda y} \mathbf{1}\{y > 0\},$$

es decir que $Y \sim$ Gama$(2, \lambda)$. La condicionada de X dado $Y = y$ es

$$f_{X|Y=y}(x) = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} \mathbf{1}\{0 < x < y\} = \frac{1}{y} \mathbf{1}\{0 < x < y\}$$

es decir, Uniforme$[0, y]$. La marginal de X es

$$\begin{aligned} f_X(x) &= \int_{-\infty}^\infty f(x, y)dy = \int_0^\infty \lambda^2 e^{-\lambda y} \mathbf{1}\{0 \le x \le y\}dy \\ &= \int_x^\infty \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x} \mathbf{1}\{x \ge 0\} \end{aligned}$$

O sea, $X \sim$ Exponencial(λ). Finalmente, sea $Z = Y - X$. La distribución conjunta de (X, Z) está dada por

$$\begin{aligned} P(X > x, Z > z) &= P((X, Y) \in \{(x', y') : x' > x,\, y' > z + x'\}) \\ &= \int_x^\infty \int_{x'+z}^\infty \lambda^2 e^{-\lambda y'} dy' dx' = \lambda \int_x^\infty e^{-\lambda(x'+z)} dx' = e^{-\lambda x} e^{-\lambda z}. \end{aligned}$$

o sea, X y Z son Exponencial(λ) independientes.

Con esto en manos, se puede ver que la distribución de Y dado $X = x$ es la misma que la de $Z + x$ con $Z \sim$ Exponencial(λ). Por lo tanto las esperanzas condicionadas se calculan así.

$$E(Y|X = x) = \frac{1}{\lambda} + x \quad \Rightarrow \quad E(Y|X) = \frac{1}{\lambda} + X$$

Por otro lado,

$$E(X|Y = y) = \frac{y}{2} \quad \Rightarrow \quad E(X|Y) = \frac{Y}{2}$$

Controlando las esperanzas de las marginales da todo bien:

$$EY = E(E(Y|X)) = E\Big(\frac{1}{\lambda} + X\Big) = \frac{1}{\lambda} + EX = \frac{2}{\lambda}$$

$$EX = E(E(X|Y)) = E(Y/2) = \frac{1}{2} EY = \frac{1}{2}\frac{2}{\lambda} = \frac{1}{\lambda}.$$

porque Y era una Gamma$(2, \lambda)$ y X una Exponencial(λ).

Note finalmente que si $Y \sim$ Gama$(2, \lambda)$ y $U \sim$ Uniforme$[0, 1]$, entonces el vector (UY, Y) tiene distribución (16). Para verlo, observe que dado $Y = y$ la distribución de UY es uniforme en $[0, y]$. Por lo tanto la densidad conjunta de (UY, Y) es

$$\frac{1}{y} \lambda^2 y e^{-\lambda y}\, \mathbf{1}\{0 \le x \le y\}$$

y la densidad conjunta de $(X, Z) = (UY, (1 - U)Y)$ es

$$f_{X,Z}(x, z) = \lambda^2 e^{-\lambda x} e^{-\lambda z}\, \mathbf{1}\{x \ge 0, z \ge 0\}.$$

6.3. Esperanza condicional. Caso general

Motivación. Sea $h : \mathbb{R} \to \mathbb{R}$ medible acotada. Entonces, en el caso continuo,

$$\begin{aligned} E(E(X|Y)\, h(Y)) &= \int \Big(\int x \frac{f(x,y)}{f_Y(y)} dx \Big) h(y) f_Y(y) dy \\ &= \int \int x\, h(y)\, f(x,y) dx dy = E(X\, h(Y)) \end{aligned}$$

Y lo mismo vale en el caso discreto. Es decir que al integrar el producto de la esperanza condicional de X dado Y con cualquier función continua y acotada de Y, obtenemos la esperanza del producto $X\, h(Y)$.

Definición. Esperanza condicional.

Sea X variable aleatoria Y vector aleatorio $Y : \Omega \to \mathbb{R}^k$.

Z es la *esperanza condicional de X dado Y* si

1. existe $g : \mathbb{R}^k \to \mathbb{R}$ medible tal que $Z = g(Y)$

2. $E(Z\, h(Y)) = E(X\, h(Y))$ para toda h medible acotada.

Teorema. *Si $E|X| < \infty$, La esperanza condicional $E(X|Y)$ existe y es única.*

Dem Omitida.

Interpretación cuando hay momentos de orden 2.

Supongamos $E|X|^2 < \infty$ y $E|Y|^2 < \infty$, es decir $X, Y \in L^2$ definido en la Sección (5.4), un espacio vectorial con el producto interno $\langle X, Y\rangle = E(XY)$. El conjunto

$$V = \{\, h(Y) : h : \mathbb{R}^k \to \mathbb{R} \text{ medible }, E[h(Y)]^2 < \infty\}$$

es un subespacio de L^2. Buscamos un $Z \in V$ que minimice la distancia a X, es decir tal que

$$E(X - Z)^2 = \min_{h(Y)\in V} E(X - h(Y))^2$$

Se resuelve con la proyección ortogonal sobre V: la solución es el Z que verifica

$$\langle (X - Z), h(Y)\rangle = E((X - Z)\, h(Y)) = 0 \quad \forall\, h \text{ medible tq } E(h(Y))^2 < \infty$$

Pero esta condición es equivalente a

$$E(Z\, h(Y)) = E(X\, h(Y))$$

Como $Z \in V$, Z es función de Y.

Existe la proyección ortogonal? Si la dimensión de V es finita, no hay problema. En general, vale si el subespacio es cerrado para la convergencia en la norma del espacio.

6.4. Propiedades

Usando la definición abstracta, demuestre que

1. $E(E(X|Y)) = EX$.

Dem: Tome $h(Y) \equiv 1$

2. Si $X = f(Y)$, entonces $E(X|Y) = X = f(Y)$.

Dem: Tengo un candidato $Z = f(Y)$. Debo verificarlo

$$E(Z\,h(Y)) = E(f(Y)\,h(Y)) = E(X\,h(Y)) \qquad \forall h \text{ medible acotada.}$$

que implica que $f(Y) = E(X|Y)$.

Ejemplo Sea X una variable aleatoria continua. Sea $Y = e^X$. Por lo tanto $X = \log Y$. Usando la propiedad 2, $E(X|Y) = \log Y = X$. Concluímos que $E(X|e^X) = X$.

3. Si X, Y son independientes, $E(X|Y) = EX$.

Dem: EX es función (constante) de Y y $Z = EX$. Entonces

$$E(Z\,h(Y)) = E(EX\,h(Y)) = EX\,Eh(Y) = E(X\,h(Y))$$

por independencia.

4. Linealidad: Si $E|X_i| < \infty$ y a_i son constantes, entonces

$$E\Big(\sum_{i=1}^{k} a_i X_i \Big| Y\Big) = \sum_{i=1}^{k} a_i E(X_i|Y)$$

Dem: Propongo $Z = \sum_{i=1}^{k} a_i E(X_i|Y)$ y calculo

$$\begin{aligned} E(Zh(Y)) &= E\Big(\sum_{i=1}^{k} a_i E(X_i|Y) h(Y)\Big) = \sum_{i=1}^{k} a_i E\Big(E(X_i|Y)h(Y)\Big) \\ &= \sum_{i=1}^{k} a_i E(X_i h(Y)) = E\Big(\sum_{i=1}^{k} a_i X_i h(Y)\Big). \end{aligned}$$

5. Monotonía. $X \geq 0$ implica $E(X|Y) \geq 0$.

Dem: Sea $h(Y) = \mathbf{1}\{Z < 0\}$ es una función de Y y además $Xh(Y) \geq 0$ porque $X \geq 0$. Por otro lado $Zh(Y) \leq 0$

$$E(Zh(Y)) = E(Xh(Y)) \geq 0.$$

que implica $Zh(Y) = 0$ que implica $h(Y) = 0$.

6. $|E(X|Y)| \leq E(|X|\,|Y)$, que implica $E|E(X\,|Y)| \leq E|X|$

Dem:

$$\begin{aligned} E(|X|\,|Y) &= E(X^+ + X^-|Y) = E(X^+|Y) + E(X^-|Y) \\ &= |E(X^+|Y) + E(X^-|Y)| \geq |E(X^+|Y) - E(X^-|Y)| \\ &= |E(X|Y)|. \end{aligned}$$

7. Si $Z = E(X|Y)$ entonces

$$E(Z\,h(Y)) = E(X\,h(Y)) \quad \forall h \text{ medible tal que } E(X\,h(Y)) \text{ exista.}$$

(la diferencia con la definición es que aquí no estamos pidiendo que h sea acotada).

Dem: omitida.

8. f medible implica $E(f(Y)X|Y) = f(Y)E(X|Y)$.

Dem: Propongo $Z = f(Y)E(X|Y)$ (es una función de Y)

$$E(Z\,h(Y)) = E(f(Y)E(X|Y)\,h(Y)) = E(X\,f(Y)\,h(Y))$$

(tome h de la def como fh y usé 7), que implica $Z = E(X\,f(Y))|Y)$.

Casos particulares Vamos a ver que las definiciones de esperanza condicional que dimos al principio son consistentes con la definición abstracta.

(X,Y) **vector aleatorio discreto.** Si definimos

$$\phi(y) := E(X|Y=y) = \sum_x x\,P(X=x|Y=y)$$

entonces

$$E(X|Y) = \phi(Y)$$

En las mismas condiciones,

$$E(g(X,Y)|Y) = \phi(Y)$$

donde

$$\phi(y) := E(g(X,Y)|Y=y) = \sum_x g(x,y)\,P(X=x|Y=y)$$

Dem Vamos a demostrar la segunda identidad. La primera sale al tomar g como función constante en y.

$$\begin{aligned}
E(\phi(Y)h(Y)) &= \sum_y \phi(y)h(y)P(Y=y)\\
&= \sum_{x,y} g(x,y)\,h(y)\,P(X=x|Y=y)\,P(Y=y)\\
&= \sum_{x,y} g(x,y)\,h(y)\,P(X=x,Y=y) = E(g(X,Y)h(Y))
\end{aligned}$$

(X,Y) **vector aleatorio continuo.** Sea

$$\phi(y) = \int_{-\infty}^{\infty} x\,f_{X|Y=y}(x)dx$$

entonces

$$E(X|Y) = \phi(Y)$$

En las mismas condiciones,

$$E(g(X,Y)|Y) = \phi(Y)$$

donde

$$\phi(y) = \int_{-\infty}^{\infty} g(x,y)\, f_{X|Y=y}(x)dx$$

Dem Propongo $Z = \phi(Y)$.

$$\begin{aligned} E(Z\,h(Y)) = E(\phi(Y)\,h(Y)) &= \int \phi(y)h(y)f_Y(y)dy \\ &= \int h(y) \int x f_{X|Y=y}(x) f_Y(y)dxdy \\ &= \int\int h(y)x f_{XY}(x,y)dxdy = E(X\,h(Y)) \end{aligned}$$

Propongo $Z = \phi(Y)$. h medible acotada. $\phi(y)$ medible (ejercicio)

$$E(Z\,h(Y)) = E(\phi(Y)\,h(Y)) = \int \phi(y)h(y)f_Y(y)dy$$

Para el otro caso:

$$\begin{aligned} &\int h(y) \int g(x,y) f_{X|Y=y}(x) f_Y(y)dxdy \\ &\quad = \int\int h(y)g(x,y)f_{XY}(x,y)dxdy = E(g(X,Y)\,h(Y)). \quad \square \end{aligned}$$

Ejemplo Sea $X \sim$ Geométrica (p) y defina $Y = 1\{X = 1\}$. Sabemos que $EX = E(E(X|Y))$. Vamos a usar esto para calcular EX. Tenemos $E(X|Y = 1) = 1$ y

$$E(X|Y=0) = E(X|X>1) = \sum_{k\geq 1} kP(X=k|X>1) = \sum_{k\geq 2} kP(X=k-1) = 1 + EX,$$

donde la tercera igualdad es por la falta de memoria. Entonces $EX = 1p + (1+EX)(1-p)$ que implica que si $EX < \infty$ entonces se puede despejar y obtener $EX = 1/p$.

Ejemplo Sean $Z_i \sim$ Normal(μ, σ^2) independientes y $N \sim$ Geométrica(p), independiente de Z_i. Sea

$$X = \sum_{i=1}^{N} Z_i.$$

Calculemos su esperanza:

$$EX = E(E(X|N)) = \sum_n E(X|N=n)P(N=n)$$

$$E(X|N=n) = \frac{E(\sum_{i=1}^N Z_i \mathbf{1}\{N=n\})}{P(N=n)} = \frac{E(\sum_{i=1}^n Z_i \mathbf{1}\{N=n\})}{P(N=n)}$$

y por independencia entre Z_i y N,

$$= n\mu \frac{P(N=n)}{P(N=n)} = n\mu \quad \Rightarrow \quad E(X|N) = N\mu$$

De donde,

$$EX = E(N\mu) = \frac{\mu}{p}.$$

La misma demostración sirve en general:

Identidad de Wald. *Sean X_i identicamente distribuídas y N variable aleatoria entera, no negativa, independiente de los X_i. Si $EN < \infty$ y $E|X_i| < \infty$, entonces*

$$E\Big(\sum_{i=1}^{N} X_i\Big) \;=\; EN\, EX_1$$

Caso particular: $Y_N \sim$Binomial(N, p) con N aleatorio e independiente de X_i.

$$EY_N = EN\, EX_1 = EN\, p$$

Problema del minero Un minero está en el fondo de una mina y ve tres túneles: 1, 2 y 3. El tunel 1 lleva a la salida en una hora. El tunel 2 vuelve a la misma encrucijada en 2 horas y el tunel 3 vuelve a la encrucijada en 3 horas. Cada vez que el minero está en la encrucijada, elige uno de los túneles con probabilidad 1/3, independientemente de lo que eligió antes.

Sea T el tiempo que tarda en salir de la mina. T es finito con probabilidad 1:

$$P(T = \infty) \le \Big(\frac{2}{3}\Big)^n \qquad \text{para todo } n \ge 0.$$

Por otro lado $T \le T_1$ el tiempo que se tarda si todos los túneles demoran 3 horas. Pero $T_1 = 3G$, donde G es una geométrica de parámetro $1/3$. Así $ET \le ET_1 = 3/(1/3) = 9$. Esto es una cota.

Calculemos exactamente ET. Sea X el tunel que el minero elige en la primera tentativa. Entonces

$$E(T|X = 1) = 1, \quad E(T|X = 2) = 2 + ET, \quad E(T|X = 3) = 3 + ET$$

Usando esto, calculamos

$$\begin{aligned} ET = E(E(T|X)) &= \frac{1}{3}E(T|X = 1) + \frac{1}{3}E(T|X = 2) + \frac{1}{3}E(T|X = 3) \\ &= 1\frac{1}{3} + (2 + ET)\frac{1}{3} + (3 + ET)\frac{1}{3} = 2 + \frac{2}{3}ET. \end{aligned}$$

Despejando ET, obtenemos $ET = 6$. Usamos que $ET < \infty$ para poder pasar de término.

Se puede resolver tambien condicionando a $K :=$ número de intentos hasta encontrar el camino 1:

$$E(T|K = k) = 1 + (k - 1)2 + EY_{k-1}$$

donde $Y_k \sim$Binomial$(k, 1/2)$ representa el número de veces que el minero tomó el camino 3 sabiendo que solamente tomó los caminos 2 y 3 en los primeros k ensayos. Esto lo hicimos arriba con la identidad de Wald. Así

$$EY_K = EK/2$$

Calculando la esperanza :

$$ET = E(E(T|K)) = E(1 + (K-1) + EY_{K-1}) = 6,$$

usando que $EK = 3$.

Problema Y_1, Y_2 independientes $Y_i \sim$ Exponencial(λ). Calcular la distribucion condicional de $S = Y_1 + Y_2$ condicionado a $Y_1 = Y_2$.

Dos posibilidades.

a. $W = Y_1 - Y_2$ y probar que $f_{S|W=0}(y) = \lambda e^{-\lambda y}$; es la distribución Exponencial(λ).

En este caso estamos considerando la sucesión de variables aleatorias

$$P(S_n \le s) = P(Y_1 + Y_2 \le s ||Y_1 - Y_2| \le 1/n)$$

b. $V = Y_1/Y_2$ y probar que $f_{Y|V=1}(y) = \lambda^2 y e^{-\lambda y}$; Gama$(2, \lambda)$. En este caso

$$P(S_n \le s) = P\Big(Y_1 + Y_2 \le s \Big| \Big|\frac{Y_1}{Y_2} - 1\Big| \le 1/n\Big)$$

Los límites dan cosas diferentes.

Moraleja: OJO al condicionar a eventos de probabilidad 0.

Problema Sea (X, Y) vector continuo con densidad $f(x, y)$ y $f_Y(y) > 0$. Pruebe que

$$\lim_{h \to 0} P(X \in [x, x+h] \,|\, Y \in [y, y+h]) = f_{X|Y=y}(x)$$

y que

$$\lim_{h \to 0} P(X \le x \,|\, Y \in [y, y+h]) = \int_{-\infty}^{x} f_{X|Y=y}(u) du := F_{X|Y=y}(x)$$

7. Generadores de números aleatorios

Recomiendo fuertemente visitar la página www.random.org de donde saqué estas observaciones: PRNG son los generadores de números seudo aleatorios y TRNG los generadores de números verdaderamente aleatorios.

"TRNG extract randomness from physical phenomena and introduce it into a computer. You can imagine this as a dice connected to a computer, but typically people use a physical phenomenon that is easier to connect to a computer than a dice is. A suitable physical phenomenon is atmospheric noise, which is quite easy to pick up with a normal radio. This is the approach used by RANDOM.ORG.

The process of generating true random numbers involves identifying little, unpredictable changes in the data. For example, HotBits uses little variations in the delay between occurrences of radioactive decay, and RANDOM.ORG uses little variations in the amplitude of atmospheric noise.

The characteristics of TRNGs are quite different from PRNGs. First, TRNGs are generally rather inefficient compared to PRNGs, taking considerably longer time to produce numbers. They are also nondeterministic, meaning that a given sequence of numbers cannot be reproduced, although the same sequence may of course occur several times by chance. TRNGs have no period."

7.1. Generación de números seudo-aleatorios

Método de la congruencia Dados m, a, c y X_0,

$$X_{n+1} = (aX_n + c) \mod m, \qquad n \geq 0$$

X_{n+1} resto entero de dividir $aX_n + c$ por m $(0 \leq X_n \leq m-1)$.

Secuencia lineal congruente.

m es el módulo $m > 0$

a es el multiplicador $0 \leq a < m$

c es el incremento $0 \leq c < m$

X_0 es la semilla o valor inicial

Método multiplicativo secuencial: $c = 0$

Knuth [9]: $m = 2^{64}$, $a = 6364136223846793005$, $c = 1442695040888963407$

Ver Wikipedia: Linear congruential generator

Este generador es determinístico y periódico. Una vez repetido un número, se repite todo el ciclo. El generador va a recorrer todos los números si se satisfacen las condiciones del Teorema de Hull-Dobell:

1. m y c son coprimos

2. $a-1$ es divisible por los primos en la descomposición de m

3. $a-1$ es divisible por 4 si m es divisible por 4.

7.2. Inversa generalizada de la distribución acumulada

Sea F la función de distribución de una variable aleatoria X. Defina la *función inversa generalizada* por

$$\begin{aligned} F^{-1}(u) &:= \inf\{y \in \mathbb{R} : F(y) \geq u\} \\ &= \min\{y \in \mathbb{R} : F(y) \geq u\}, \end{aligned}$$

por la continuidad a la derecha de F.

Lema 31 *Sea F una distribución y F^{-1} su inversa generalizada. Entonces, para todo $x \in \mathbb{R}$,*

$$\{u \in (0,1) : F^{-1}(u) \leq x\} = \{u \in (0,1) : u \leq F(x)\}$$

Dem $\subset$] Si $\min\{y \in \mathbb{R} : F(y) \geq u\} \leq x$, como F es no decreciente, $F(x) \geq u$.

$\supset$] Si $u \leq F(x)$, entonces $x \in \{y \in \mathbb{R} : F(y) \geq u\}$. Por lo tanto $x \geq \min\{y \in \mathbb{R} : F(y) \geq u\} = F^{-1}(u)$. $\square$

Corolario 32 *Sea $U \sim$ Uniforme$[0,1]$ y F una función de distribución acumulada. Entonces $Y := F^{-1}(U)$ tiene distribución F.*

Dem F^{-1} monotona. Por lo tanto medible Borel. Por lo tanto $F^{-1}(U)$ es una variable aleatoria. Sea $x \in \mathbb{R}$, entonces

$$P(Y \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x). \qquad \square$$

A partir de una variable uniforme y de una distribución, encontramos una variable aleatoria con esa distribución.

Ejemplo Sea F la distribución acumulada de una variable Bernoulli(p). Es decir $F(x) = (1-p)\mathbf{1}\{0 \leq x < 1\} + \mathbf{1}\{x \geq 1\}$. De tal manera que $F^{-1}(u) = 0$ para $u \in [0, 1-p]$ y $F^{-1}(u) = 1$ para $u \geq 1-p$. Verifique que $F^{-1}(U) \sim$ Bernoulli(p), si $U \sim$ Uniforme$[0,1]$.

Existencia de variables aleatorias con una distribución dada Como consecuencia, dada una distribución acumulada F, existe una variable aleatoria X (definida como $F^{-1}(U)$, con U uniforme en $[0,1]$) cuya distribución $F_X = F$.

Acoplamiento Construcción simultanea de dos o más variables aleatorias. Un *acoplamiento* de dos variables aleatorias con distribuciones F y G es la definición de ambas en función de la misma variable U: $(X, Y) := (F^{-1}(U), G^{-1}(U))$.

Orden estocástico En general, si F_i, $i = 1, \dots, n$, son funciones de distribución acumuladas que satisfacen $1 - F_i(y) \leq 1 - F_{i+1}(y)$ para todo y y si definimos $Y_i := F_i^{-1}(U)$, entonces por lo ya visto $Y_i \sim F_i$ para cada i y además las variables están ordenadas casi seguramente:

$$P(Y_1 \leq \cdots \leq Y_n) = 1.$$

Esto nos dá una noción de orden entre variables aleatorias que se llama *orden estocástico*.

7.3. Generación de variables aleatorias discretas

Vía particiones de $[0,1]$. Sea $p(x)$ una probabilidad puntual en un conjunto R discreto: $p(x) > 0$ y

$$\sum_{x \in R} p(x) = 1.$$

Sea U Uniforme$[0,1]$. Sea $(J(x) : x \in R)$ una partición del intervalo $[0,1]$ tal que $|J(x)| = p(x)$ para todo x con $p(x) > 0$. Defina

$$X = x \qquad \text{si } U \in J(x)$$

Equivalentemente:

$$X = \sum_x x\mathbf{1}\{U \in J(x)\}$$

X así definida es una variable con rango $R_X = R$ y probabilidad puntual $p(x)$. Veamos:

$$P(X = x) = P(U \in J(x)) = |J(x)| = p(x).$$

Vía función inversa generalizada. Los intervalos $J(x) = \{u : F^{-1}(u) = x\}$ forman una partición de $[0, 1]$ que satisface

$$|J(x)| = F(x) - F(x-) = p(x).$$

Ejemplo: Simule la variable con distribución

z	1	3	9
P(Z=z)	1/2	1/4	1/4

$$X = F^{-1}(U) = \begin{cases} 1, & \text{si } U < 1/2 \\ 3, & \text{si } 1/2 \le U < 3/4 \\ 9, & \text{si } U \ge 3/4 \end{cases}$$

Ejemplo: $X \sim \text{Bernoulli}(p)$

$$X = F^{-1}(U) = \begin{cases} 0, & \text{si } U < (1-p) \\ 1, & \text{si } U \ge (1-p) \end{cases}$$

Acoplamiento Considere variables $X_i \sim \text{Bernoulli}(p_i)$. $p_1 \le p_2$. Es más probable éxito para X_2 que para X_1. Se puede construir (X_1, X_2) de tal forma que $X_1 \le X_2$? Definimos $\hat{X}_i = F_i^{-1}(U)$, con la misma uniforme para las dos variables:

$$F_i^{-1}(U) = \begin{cases} 0, & \text{si } U < (1-p_i) \\ 1, & \text{si } U \ge (1-p_i) \end{cases}$$

Tenemos $\hat{X}_1 = \hat{X}_2 = 0$ si $U < (1 - p_2)$

$\hat{X}_1 = 0$, $\hat{X}_2 = 1$ si $1 - p_2 \le U < 1 - p_1$

$\hat{X}_1 = \hat{X}_2 = 1$ si $U \ge 1 - p_1$.

O sea: $P(\hat{X}_1 \le \hat{X}_2) = 1$.

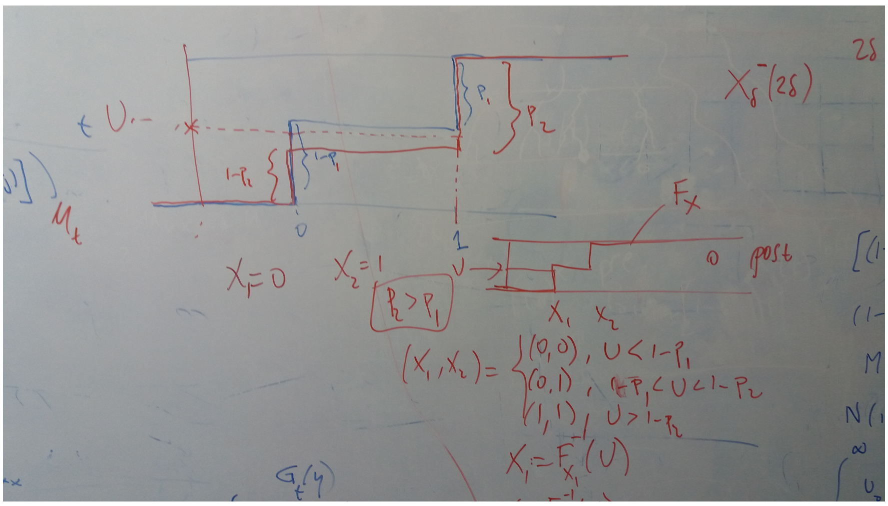


Figura 4: Construcción simultanea de dos variables Bernoulli.

7.4. Generación de variables aleatorias continuas

Método de inversión. Usaremos el Corolario 32.

Generación de una Exponencial(λ) En este caso $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$

$$F^{-1}(u) = \frac{-\log(1-u)}{\lambda}$$

Entonces la variable definida por

$$X = \frac{-\log(1-U)}{\lambda}$$

con $U \sim$ Uniforme$[0,1]$ es exponencial. Como $(1-U)$ tiene la misma distribución que U, la variable

$$X = \frac{-\log(U)}{\lambda}$$

tambien tiene distribución exponencial.

El método del rechazo Ross [11] página 443. Queremos generar una variable con densidad f. Sabemos como generar una variable con densidad g. Sabemos que existe una constante $c > 1$ tal que

$$f(x) \leq cg(x) \qquad \text{para todo } x$$

Algoritmo del rechazo

1. Simule Y con densidad g y $U \sim$ Uniforme$[0,1]$

2. Si $U \leq f(Y)/cg(Y)$, ponga $X = Y$ y termine. Si no, vaya a 1.

Lema 33 *La variable X así generada tiene densidad f.*

Demostración. Sean $(Y_1, U_1), (Y_2, U_2), \ldots$ una sucesion de vectores independientes con coordenadas independientes. $Y \sim g$ y $U \sim$ Uniforme$[0,1]$. Sea

$$A = \Big\{(x,u) : u \leq \frac{f(x)}{cg(x)}\Big\}$$

Sea $T = \min\{n : (Y_n, U_n) \in A\}$, una variable Geométrica de parámetro

$$p := P((Y_n, U_n) \in A).$$

Probamos más abajo que $p > 0$. Defina $X := Y_T$. Entonces

$$P(X \in B) = \sum_{n=1}^{\infty} P(Y_T \in B, T = n) = \sum_{n=1}^{\infty} P(Y_n \in B, T = n).$$

El sumando genérico vale (definiendo $\tilde{Y}_i := (Y_i, U_i)$):

$$\begin{aligned}
P(Y_n \in B, T = n) &= P(\tilde{Y}_1 \in A^c, \dots, \tilde{Y}_{n-1} \in A^c, \tilde{Y}_n \in A, Y_n \in B) \\
&= P(\tilde{Y}_1 \in A^c) \dots P(\tilde{Y}_{n-1} \in A^c) P(\tilde{Y}_n \in A, Y_n \in B) \\
&= (1-p)^{n-1} P(\tilde{Y}_n \in A, Y_n \in B) \qquad (17)
\end{aligned}$$

Como la distribución de $\tilde{Y}_n \sim (Y, U)$ no depende de n, la probabilidad en (17) es igual a

$$\begin{aligned}
P((Y,U) \in A, Y \in B) &= \int_{-\infty}^{\infty} \int_0^1 g(x) \mathbf{1}\{x \in B\} \, \mathbf{1}\{(x,u) \in A\} dx \, du \\
&= \int_{-\infty}^{\infty} g(x) \mathbf{1}\{x \in B\} \int_0^1 \mathbf{1}\Big\{u \le \frac{f(x)}{cg(x)}\Big\} du \, dx \\
&= \int_{-\infty}^{\infty} g(x) \mathbf{1}\{x \in B\} \, \frac{f(x)}{cg(x)} dx = \frac{1}{c} \int_B f(x) dx.
\end{aligned}$$

Usando $B = \mathbb{R}$ tenemos $p = P((Y,U) \in A) = 1/c$. Substituyendo en (17) y sumando (17) en n concluimos que

$$P(X \in B) = \int_B f(x) dx. \qquad \square$$

Generación de una variable normal standard Z No se puede usar el método de inversión. Empezamos por generar $X = |Z|$, que tiene densidad

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \ge 0$$

Proponemos $g(x) = e^{-x}$, $x \ge 0$.

$$\frac{f(x)}{g(x)} = \frac{2e^{-x^2/2}}{\sqrt{2\pi} e^{-x}} = \sqrt{\frac{2e}{\pi}} e^{-(x-1)^2/2} \le \sqrt{\frac{2e}{\pi}}$$

de donde $c = \sqrt{\frac{2e}{\pi}}$ y

$$\frac{f(x)}{cg(x)} = \exp\Big(\frac{-(x-1)^2}{2}\Big)$$

El **algoritmo** para generar $X = |Z|$ con $Z \sim$ Normal$(0,1)$ queda:

1. Genere Y exponencial de parametro 1, U Uniforme$[0,1]$

2. Si $U \le \exp(-(Y-1)^2/2)$, ponga $X = Y$. Si no, vaya a (1).

Ahora defina $Z = WX - (1-W)X$, con W Bernoulli$(1/2)$. Z es Normal$(0,1)$.

Simplificación En el paso (2), Y es aceptada si $U \le \exp(-(Y-1)^2/2)$, que es equivalente a $-\log U \ge -(Y-1)^2/2$. Como $Y_2 = -\log U$ es Exponencial(1), podemos usar el algoritmo:

1. Genere Y_1, $Y_2 \sim$ Exponencial(1)

2. Si $Y_2 \ge -(Y_1-1)^2/2$, ponga $X = Y_1$. Si no, vaya a (1).

8. Convergencia de variables aleatorias

8.1. Lema de Borel Cantelli

Sean A_n eventos en $(\Omega, \mathcal{F}, P)$. Definimos:

$$A^* = \{A_n \text{ infinitas veces }\} := \bigcap_{k\geq 1} \bigcup_{n\geq k} A_n$$

O sea si $B_k = \cup_{n\geq k} A_n$,

$$A^* = \bigcap_{k\geq 1} B_k$$

Teorema 34 (Lema de Borel Cantelli)

$$\begin{aligned} \sum_n P(A_n) < \infty \quad &\Longrightarrow \quad P(A^*) = 0 \\ \sum_n P(A_n) = \infty \ y \ A_n \ \textit{independientes} \quad &\Longrightarrow \quad P(A^*) = 1 \end{aligned}$$

Dem Primera parte, llamado *fácil*:

$$P(B_k) = P(\cup_{n\geq k} A_n) \leq \sum_{n\geq k} P(A_n) \underset{k\to\infty}{\longrightarrow} 0$$

porque la suma es finita. Así, como los B_k son decrecientes,

$$P(A^*) = P\Big(\bigcap_{k\geq 1} B_k\Big) = \lim_{k\to\infty} P(B_k) = 0$$

Segunda parte: por independencia tenemos

$$\begin{aligned} P(B_k^c) &= P(\cap_{n\geq k} A_n^c) = \lim_{K\to\infty} P(\cap_{n=k}^K A_n^c) \\ &= \lim_{K\to\infty} \prod_{n=k}^{K} (1 - P(A_n)) = \prod_{n=k}^{\infty} (1 - P(A_n)) \\ &\leq \prod_{n\geq k} e^{-P(A_n)} = e^{-\sum_{n\geq k} P(A_n)} = 0 \end{aligned}$$

porque por hipótesis la suma diverge. Ojo, hay que demostrar como ejercicio la segunda identidad. Entonces

$$P((A^*)^c) = P(\cup_{k\geq 0} B_k^c) = 0$$

porque es unión numerable de eventos de probabilidad 0. □

Ejemplo Ω sucesiones de 0 y 1.

X_n Bernoulli parametro p_n independientes.

$A_n = \{X_n = 1\}$

Si suma de p_n converge hay un número finito de unos.

Si suma de p_n diverge, hay un número infinito de unos.

8.2. Convergencia de variables aleatorias

$(X_n, n \geq 0)$ sucesión de variables aleatorias Queremos definir $\lim_{n\to\infty} X_n$.

Convergencia en Distribución $X_n \xrightarrow{D} X$ si $F_n(x) \to F(x)$ para todo $x \in \mathbb{R}$ punto de continuidad de F. Es una convergencia de las *funciones* de distribución acumulada, no hace falta que las variables estén definidas en el mismo espacio.

Convergencia en probabilidad $X_n \xrightarrow{P} X$ si todas las X_n y X estan definidas en el mismo espacio $(\Omega, \mathcal{F}, P)$ y para todo $\varepsilon > 0$,

$$\lim_{n\to\infty} P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon) = 0$$

Ejemplo: Sea U ~Uniforme$[0,1]$ y $I_n \subset [0,1]$ con medida de Lebesgue $|I_n| = 1/n$. Defina $X_n = \mathbf{1}\{U \in I_n\}$. Tenemos que X_n es Bernoulli con $P(X_n = 1) = 1/n$. Por lo tanto X_n converge a la constante 0: $P(X_n > \varepsilon) = 0$ para todo $n > 1/\varepsilon$.

Convergencia casi segura $X_n \xrightarrow{c.s.} X$ si todas las X_n y X estan definidas en el mismo espacio $(\Omega, \mathcal{F}, P)$ y

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

Ejemplo: U Uniforme$[0,1]$, $p_n \to p$ sucesión de números en $[0,1]$.

Definimos $X_n := \mathbf{1}\{U \leq p_n\}$ y $X := \mathbf{1}\{U \leq p\}$

Si $U = u < p$ entonces existe un n_0 que depende de u tal que si $n \geq n_0$ entonces $u < p_n$ etc.

Lema 35 $X_n \xrightarrow{c.s.} X$ *si y sólo si*

$$\lim_{k\to\infty} P\Big(\bigcup_{n\geq k}\Big\{|X_n - X| \geq \frac{1}{\ell}\Big\}\Big) = 0 \qquad \textit{para todo } \ell > 0 \tag{18}$$

Dem

$$\{X_n \to X\} = \bigcap_{\ell\geq 1}\bigcup_{k\geq 1}\bigcap_{n\geq k}\Big\{|X_n - X| < \frac{1}{\ell}\Big\}$$

$$\{X_n \to X\}^c = \bigcup_{\ell\geq 1}\bigcap_{k\geq 1}\bigcup_{n\geq k}\Big\{|X_n - X| \geq \frac{1}{\ell}\Big\}$$

$$X_n \xrightarrow{c.s.} X \quad \Leftrightarrow \quad P\Big(\bigcup_{\ell\geq 1}\bigcap_{k\geq 1}\bigcup_{n\geq k}\Big\{|X_n - X| \geq \frac{1}{\ell}\Big\}\Big) = 0$$

$$\Leftrightarrow \quad P\Big(\bigcap_{k\geq 1}\bigcup_{n\geq k}\Big\{|X_n - X| \geq \frac{1}{\ell}\Big\}\Big) = 0 \qquad \text{para todo } \ell > 0$$

que es equivalente a (43) porque son eventos decrecientes. □

Corolario 36 $X_n \overset{c.s.}{\to} X \implies X_n \overset{P}{\to} X$ *(si una sucesión converge casi seguramente a X, entonces también converge en probabilidad a X.)*

Dem Como la probabilidad de la unión va a cero, la probabilidad de cada conjunto también:

$$\lim_{k\to\infty} P\Big(|X_k - X| \geq \frac{1}{\ell}\Big) \leq \lim_{k\to\infty} P\Big(\bigcup_{n\geq k}\Big\{|X_n - X| \geq \frac{1}{\ell}\Big\}\Big) = 0$$

para todo ℓ. □

Corolario 37

$$\sum_{n\geq 1} P(|X_n - X| \geq \varepsilon) < \infty \quad \textit{para todo } \varepsilon > 0 \quad \implies \quad X_n \overset{c.s.}{\to} X$$

Dem La probabilidad de la unión en (43) está acotada por la suma de las probabilidades. Si la serie es convergente, sus colas convergen a 0. □

Obs Tambien se puede usar Borel-Cantelli:

$$X_n \overset{c.s.}{\to} X \Leftrightarrow P(|X_n - X| > \varepsilon, \text{ infinitas veces }) = 0, \quad \forall\varepsilon > 0.$$

Contraejemplo Convergencia en Probabilidad no implica convergencia casi segura. Considere los $I_n \subset [0, 1]$ del ejemplo anterior construidos de forma tal que todo $u \in [0, 1]$ pertenece a infinitos I_n. Esto puede hacerse sólo si $\sum_n \ell(I_n) = \infty$. Entonces

$$P(X_n = 1, \text{ infinitas veces}) = P(U \in I_n \text{ infinitas veces}) = 1$$

Así X_n **no** converge c.s. a 0. Y ya vimos que converge en probabilidad.

Lema 38 *Si $X_n \overset{P}{\to} X$, entonces $X_n \overset{D}{\to} X$*

Dem Sea x punto de continuidad de F.

$$\begin{aligned}
F_n(x) &= P(X_n \leq x) \\
&= P(X_n \leq x, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\
&\leq P(X \leq x + \varepsilon, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\
&\leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon)
\end{aligned}$$

Sacando limites, el segundo termino va a 0 por convergencia en probabilidad:

$$\limsup_n F_n(x) \leq P(X \leq x + \varepsilon) = F(x + \varepsilon) \underset{\varepsilon\to 0}{\to} F(x)$$

por la continuidad de la F en x.

Para ver la otra desigualdad escriba

$$\begin{aligned}
F_n(x) &= P(X_n \leq x, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\
&\geq P(X \leq x - \varepsilon, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon)
\end{aligned}$$

Tomo liminf en ambos lados. El primer término converge a $F(x-\varepsilon)$ porque $P(|X_n - X| < \varepsilon) \to_n 1$, por convergencia en probabilidad. El segundo término va a cero por convergencia en probabilidad. Queda

$$\liminf_{n\to\infty} F_n(x) \geq F(x-\varepsilon) \underset{\varepsilon\to 0}{\to} F(x),$$

por la continuidad de la F en x. $\square$

Contraejemplos (1) Sean X_i iid Bernoulli(p).

Es claro que X_i converge en distribución a $Y \sim$ Bernoulli(p).

Si $p \in (0,1)$, X_i no converge en probabilidad. Supongamos por absurdo que Y está en el mismo espacio que las X_i y que $X_i \to Y$ en probabilidad. Entonces, para $n \neq m$ y $0 < \varepsilon < 1$, como $\{|X_n - X_m| > \varepsilon\} = \{X_m = 0, X_n = 1\} \dot\cup \{X_m = 1, X_n = 0\}$,

$$\begin{aligned}&2p(1-p) = P(|X_n - X_m| > \varepsilon) = P(|X_n - Y + Y - X_m| > \varepsilon)\\ &\leq P(|X_n - Y| + |X_m - Y| > \varepsilon) \leq P(|X_n - Y| > \varepsilon/2) + P(|X_m - Y| > \varepsilon/2)\end{aligned}$$

pero por la convergencia en probabilidad, esto converge a cero cuando m y n van a infinito. Absurdo.

(2) Sea $U \sim$ Uniforme$(0,1)$ y

$$X_{2n} = \mathbf{1}\{U \leq \frac{1}{2}\}, \qquad X_{2n+1} = \mathbf{1}\{U > \frac{1}{2}\}.$$

Son todas variables Bernoulli($\frac{1}{2}$). Por lo tanto $X_n \overset{D}{\to}$ Bernoulli($\frac{1}{2}$). Sin embargo $X_n = 1 - X_{n+1}$ para todo n. Como las variables asumen valores en $\{0,1\}$, no puede haber convergencia. Más precisamente

$$\begin{aligned}1 &= P(|X_{2n} - X_{2n+1}| > \varepsilon) = P(|X_{2n} - X + X - X_{2n+1}| > \varepsilon)\\ &\leq P(|X_{2n} - X| + |X - X_{2n+1}| > \varepsilon)\\ &\leq P(|X_{2n} - X| > \varepsilon/2) + P(|X - X_{2n+1}| > \varepsilon/2)\end{aligned}$$

Pero si X es el límite en probabilidad de X_n, ambos términos tendrían que converger a 0. Una contradicción.

Corolario 39 $X_n \overset{c.s.}{\to} X \quad \Longrightarrow \quad X_n \overset{P}{\to} X \quad \Longrightarrow \quad X_n \overset{D}{\to} X$

Proposición 40 *Si $X_n \overset{P}{\to} X$, entonces existe una subsucesión X_{n_k} tal que $X_{n_k} \overset{c.s.}{\to} X$.*

Dem Para cada $k > 0$ sea n_k tal que $P(|X_{n_k} - X| > 1/k) < 2^{-k}$. Entonces, para cada $\varepsilon > 0$,

$$\sum_{k>0} P(|X_{n_k} - X| > \varepsilon) \leq \frac{1}{\varepsilon} + \sum_{k>1/\varepsilon} 2^{-k} < \infty$$

(los primeros $1/\varepsilon$ términos están acotados por 1). Por el Corolario 37, la sumabilidad de la serie implica convergencia c.s.. $\square$

Lema 41 *Si $F : \mathbb{R} \to [0,1]$ es una función no decreciente, entonces el conjunto de puntos de discontinuidad de F es finito o numerable.*

Dem Si F no es continuo en x, entonces $F(x-) < F(x+)$. Por lo tanto en el intervalo $[F(x-), F(x+))$ hay un número racional. Como la función es no decreciente, los intervalos $\{[F(x-), F(x+)),\ x \text{ punto de discontinuidad}\}$ son disjuntos. Por lo tanto establecimos una relación uno a uno entre los ptos de discontinuidad y un subconjunto de los racionales. Como los racionales son numerables, ese conjunto es finito o numerable. $\square$

8.3. Desigualdades de Markov y Chevichev

Desigualdad de Markov *Sea $X \geq 0$ una variable aleatoria no negativa con esperanza finita y $\varepsilon > 0$. Entonces*

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon}$$

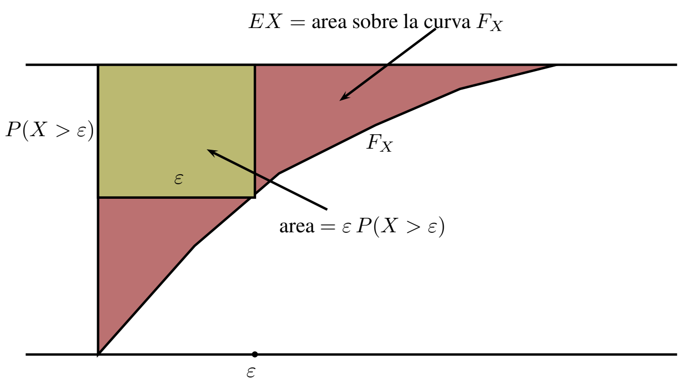


Figura 5: Desigualdad de Markov

Dem

$$X = X\,\mathbf{1}\{X < \varepsilon\} + X\,\mathbf{1}\{X \geq \varepsilon\} \ \geq\ \varepsilon\mathbf{1}\{X > \varepsilon\}$$

porque como $X \geq 0$ y $\mathbf{1}\{X < \varepsilon\} \geq 0$, el primer término es mayor o igual a 0. Por la monotonía de la esperanza:

$$EX \geq \varepsilon\, E\big(\mathbf{1}\{X \geq \varepsilon\}\big) = \varepsilon\, P(X > \varepsilon). \quad \square$$

Corolario 42 *Si $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ es no decreciente, entonces*

$$P(X \geq \varepsilon) \leq \frac{E\varphi(X)}{\varphi(\varepsilon)}$$

Casos particulares Sea X una variable aleatoria cualquiera,

$$P(|X| \geq \varepsilon) \leq \frac{E\varphi(|X|)}{\varphi(\varepsilon)}$$

Desigualdad de Chevichev

$$P(|X - EX| \geq \varepsilon) \leq \frac{VX}{\varepsilon^2}$$

Dem Usamos $\varphi(x) = x^2$ para obtener

$$P(|X - EX| \geq \varepsilon) = P(|X - EX|^2 \geq \varepsilon^2) \leq \frac{E(X - EX)^2}{\varepsilon^2}.$$

por la desigualdad de Markov.

8.4. Leyes de grandes números

Teorema 43 (Ley debil de los grandes números) *Sean X_n variable aleatoria independientes con $EX = \mu$, $VX = \sigma^2$. Sea*

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Entonces $\bar{X}_n \xrightarrow{P} \mu$.

Dem Primero calculamos la esperanza y la varianza de $\bar{X}_n$. Por la linealidad de la esperanza,

$$E\bar{X}_n = \frac{1}{n}E(X_1 + \cdots + X_n) = \frac{n}{n}\mu = \mu.$$

Usando que la varianza de la suma de variable aleatoria independientes es la suma de las varianzas,

$$V\bar{X}_n = \frac{1}{n^2}V(X_1 + \cdots + X_n) = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

Usando Chevichev:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad \text{para todo } \varepsilon > 0.$$

Es decir, $\bar{X}_n \xrightarrow{P} \mu$. □

kkkk **Aplicación** $X_i \sim \text{Bernoulli}(p)$. $EX_i = p$, $VX_i = p(1-p)$.

$$\bar{X}_n = \frac{\#\text{éxitos}}{\#\text{ensayos}} \xrightarrow{P} p$$

Teorema 44 (Ley fuerte de los grandes números) *Sean X_n variables aleatorias independientes con $EX = \mu$, $VX = \sigma^2$. Entonces $\bar{X}_n \xrightarrow{c.s.} \mu$.*

Dem Podemos asumir que $\mu = 0$. Por Chevichev,

$$P(|\bar{X}_{n^2}| > \varepsilon) \leq \frac{\sigma^2}{n^2\varepsilon^2},$$

que es sumable. Concluimos que $\bar{X}_{n^2} \overset{\text{c.s.}}{\to} 0$.

Sea $n_k = \text{máx}\{n : k \geq n^2\}$; así, $k \in [n_k^2, (n_k+1)^2)$. Los primeros valores son

$k:$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...	24	25	...
$n_k^2:$	1	1	1	4	4	4	4	4	9	9	9	9	9	9	9	16	...	16	25	...

Denotemos $S_n := \sum_{i=1}^n X_i$.

$$|\bar{X}_k| = \frac{|S_k|}{k} \leq \frac{|S_k - S_{n_k^2} + S_{n_k^2}|}{n_k^2} \leq \frac{|S_k - S_{n_k^2}|}{n_k^2} + |\bar{X}_{n_k^2}| \tag{19}$$

Sea $n^2+1 \leq k \leq (n+1)^2$. Sea $A_k := \left\{\frac{|S_k - S_{n^2}|}{n^2} > \varepsilon\right\}$. Calculemos

$$\begin{aligned} V|S_k - S_{n^2}| &= V \sum_{i=n^2+1}^{k} X_i = (k-n^2)\sigma^2 \\ &\leq ((n+1)^2 - n^2)\sigma^2 = (2n+1)\sigma^2. \end{aligned}$$

Por Chevichev,

$$P(A_k) \leq \frac{(2n+1)\sigma^2}{n^4\varepsilon^2}$$

Veamos que $\sum_k P(A_k)$ es sumable. De hecho,

$$\sum_{k=1}^{\infty} P(A_k) = \sum_{n=0}^{\infty} \sum_{k=n^2+1}^{(n+1)^2} P(A_k) \leq \frac{\sigma^2}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{(2n+1)^2}{n^4} < \infty. \tag{20}$$

De donde sigue que

$$\frac{|S_k - S_{n_k^2}|}{n_k^2} \underset{k}{\overset{\text{c.s.}}{\to}} 0 \tag{21}$$

Por otro lado, como $\bar{X}_{n^2} \to 0$ y X_{n_k} es una sucesión que toma valores constantes entre n^2 y $(n+1)^2 - 1$, tenemos

$$\bar{X}_{n_k} \underset{k}{\overset{\text{c.s.}}{\to}} 0 \tag{22}$$

Sacando límites en (19), podemos concluir usando (21) y (22). □

Teorema 45 (Convergencia acotada) *Si $X_n \overset{P}{\to} X$ y existe una constante M tal que $|X_n| \leq M$, $|X| \leq M$ entonces*

$$EX_n \to EX$$

Dem

$$\begin{aligned} |EX_n - EX| &\leq E|X_n - X| \\ &= E(|X_n - X|\,\mathbf{1}\{|X_n - X| > \varepsilon\}) + E(|X_n - X|\,\mathbf{1}\{|X_n - X| \leq \varepsilon\}) \\ &\leq 2M\,P(|X_n - X| > \varepsilon) + \varepsilon \to \varepsilon \end{aligned}$$

Por la convergencia en probabilidad. Como vale para todo $\varepsilon > 0$, estamos. □

Teorema 46 (de Weierstrass) *Los polinomios son densos en* $\mathcal{C}[0,1] = \{f : [0,1] \to \mathbb{R},\ f$ *continua*$\}$.

Dem Dada una f continua, busco una sucesión de polinomios P_n que converge uniformemente a f.

Sea $X_i \sim$ Bernoulli(x), $x \in [0,1]$. Por la ley fuerte de grandes números, $\bar{X}_i \overset{\text{c.s.}}{\to} x$. Esto implica $f(\bar{X}_i) \overset{\text{c.s.}}{\to} f(x)$.

Como f es continua y f es acotada, por el teorema de convergencia acotada, tenemos la convergencia de las esperanzas $Ef(\bar{X}_i) \to_n f(x)$. Como $S_n \sim$ Binomial(n, x),

$$B_n(f,x) := Ef(\bar{X}_n) = Ef\Big(\frac{S_n}{n}\Big) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big)\binom{n}{k} x^k (1-x)^{n-k}$$

O sea que cuando $n \to \infty$, el polinomio $B_n(f,x)$ converge puntualmente a f.

Veamos que $B_n(f,x)$ converge uniformemente a f. Como f es acotada, f es uniformemente continua. Fije $\varepsilon > 0$ y sea $\delta(\varepsilon)$ tal que para todo $\delta < \delta(\varepsilon)$, $|f(x+\delta) - f(x)| < \varepsilon$. Escriba

$$\begin{aligned} |B_n(f,x) - f(x)| &= |Ef(\bar{X}_n) - f(x)| \le E|f(\bar{X}_n) - f(x)| \\ &\le E(|f(\bar{X}_n) - f(x)|\mathbf{1}_A) + E(|f(\bar{X}_n) - f(x)|\mathbf{1}_{A^c}) \end{aligned}$$

donde $A = \{|\bar{X}_n - x| < \delta(\varepsilon)\}$.

$$\le \varepsilon P(A) + 2MP(|\bar{X}_n - x| \ge \delta(\varepsilon))$$

donde $M = \sup_{x\in[0,1]} f(x)$. Por Chevichev,

$$\le \varepsilon + 2M\frac{V\bar{X}_n}{\delta(\varepsilon)^2} \le \varepsilon + 2M\frac{x(1-x)}{n\delta(\varepsilon)^2} \underset{n\to\infty}{\to} \varepsilon,$$

uniformemente en x. Como esto vale para todo $\varepsilon > 0$, concluímos que $B_n(f,\cdot)$ converge a f uniformemente. □

Observación: Los polinomios

$$b_{k,n}(x) := \binom{n}{k} x^k (1-x)^{n-k}$$

son llamados *polinomios de Bernstein.* Note que para cada $n \ge 1$ y $x \in [0,1]$, $(b_{k,n}(x),\ k = 0,\dots,n)$ es la probabilidad de la binomial.

Proposición 47 $X_n \overset{D}{\to} c$ *y todas definidas en el mismo espacio, entonces* $X_n \overset{P}{\to} c$.

Dem

$$\begin{aligned} P(|X_n - c| \ge \varepsilon) &\le P(X_n \ge c + \varepsilon) + P(X_n \le c - \varepsilon) \\ &= 1 - P(X_n < c + \varepsilon) + P(X_n \le c - \varepsilon) \\ &\le 1 - P(X_n \le c + \frac{\varepsilon}{2}) + P(X_n \le c - \varepsilon) \\ &\to_n 1 - F_c(c + \varepsilon/2) + F_c(c - \varepsilon) = 0 \end{aligned}$$

porque $F_c(x) = 0$ si $x < c$ y $F_c(x) = 1$ si $x \ge c$. □

8.5. Teorema de Skorohod

Propiedades de la inversa de la distribución acumulada Recordemos la definición de inversa generalizada de una función de distribución acumulada F:

$$F^{-1}(u) = \text{ínf}\{x \in \mathbb{R} : F(x) \geq u\} \quad u \in [0,1]$$

Las siguientes propiedades son consecuencia de la definición:

$$F^{-1} \text{ es no decreciente} \tag{23}$$

$$F(F^{-1}(u)-) \leq u \leq F(F^{-1}(u)), \quad u \in [0,1] \tag{24}$$

$$\left.\begin{array}{l} F^{-1} \text{ continua en } u \\ F(x-) \leq u \leq F(x) \end{array}\right\} \Longrightarrow F^{-1}(u) = x \tag{25}$$

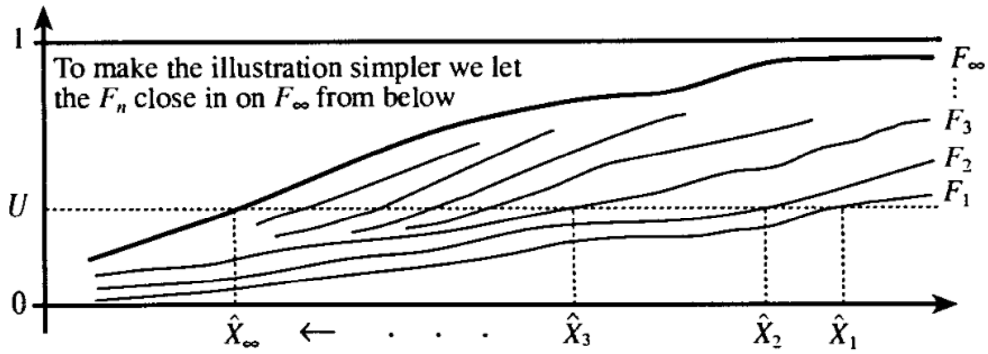


Figura 6: Teorema de Skorohod. Figura del libro de Thorisson [12])

Teorema 48 (de Skorohod) *Sean F_n, $n \geq 1$ y F funciones de distribución tales que*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

en los puntos x de continuidad de F. Entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, una sucesión de variables aleatorias X_n, $n \geq 0$ y X, todas definidas en Ω con $X_n \sim F_n$ y $X \sim F$ tales que $X_n \stackrel{c.s.}{\to} X$.

Dem Fije $u \in [0,1]$ y defina

$$x = \liminf_{n\to\infty} F_n^{-1}(u)$$

Como F es no decreciente, tiene a lo sumo numerables puntos de discontinuidad y se puede elegir ε arbitrariamente chico tal que F es continua en $x - \varepsilon$ y en $x + \varepsilon$.

Sea n_k una subsucesión tal que

$$x - \varepsilon < F_{n_k}^{-1}(u) \leq x + \varepsilon \quad \text{para todo } k \geq 1$$

Por la monotonicidad de F_{n_k},

$$F_{n_k}(x-\varepsilon) \le F_{n_k}(F_{n_k}^{-1}(u)-) \le F_{n_k}(F_{n_k}^{-1}(u)) \le F_{n_k}(x+\varepsilon)$$

Aplicando (24),

$$F_{n_k}(x-\varepsilon) \le u \le F_{n_k}(x+\varepsilon) \quad \text{para todo } k \ge 1$$

Como $x \pm \varepsilon$ son puntos de continuidad de F, mandando $k \to \infty$,

$$F(x-\varepsilon) \le u \le F(x+\varepsilon) \quad \text{para todo } k \ge 1$$

De donde, por la continuidad a derecha de F,

$$F(x-) \le u \le F(x)$$

Reemplacemos x por

$$y = \limsup_{n\to\infty} F_n^{-1}(u)$$

para obtener

$$F(y-) \le u \le F(y)$$

Así, si F^{-1} es continuo en u,

$$F^{-1}(u) = x = \liminf_{n\to\infty} F_n^{-1}(u)$$
$$F^{-1}(u) = y = \limsup_{n\to\infty} F_n^{-1}(u)$$

O sea que para todo punto de continuidad de F^{-1},

$$\lim_{n\to\infty} F_n^{-1}(u) = F^{-1}(u)$$

Sea $U \sim$ Uniforme$[0,1]$, definida por ejemplo en $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, $P =$ Lebesgue. $U(\omega) = \omega$. Defino

$$X_n := F_n^{-1}(U), \qquad X := F^{-1}(U)$$

Vimos que con esa definición, $X_n \sim F_n$ y $X \sim F$. Como el conjunto de puntos de discontinuidad de F^{-1} es a lo sumo numerable, $P(F^{-1}(U)-) < F^{-1}(U))) = 0$ y

$$\lim_{n\to\infty} F_n^{-1}(U) = F^{-1}(U)\square$$

Aplicación $X_n \stackrel{D}{\to} X$, $g : \mathbb{R} \to \mathbb{R}$ continua $\Longrightarrow g(X_n) \stackrel{D}{\to} g(X)$.

Dem Construimos $\hat{X}_n$ con el teo de Skorohod; $\hat{X}_n \sim X_n$, $\hat{X} \sim X$. $\hat{X}_n \stackrel{c.s.}{\to} X \Rightarrow g(\hat{X}_n) \stackrel{c.s.}{\to} g(X)$ porque g continua. Pero esto implica $g(X_n) \stackrel{D}{\to} g(X)$. $\square$

Teorema 49 $X_n \stackrel{D}{\to} X$ *<u>si y sólo si</u>* $Ef(X_n) \to Ef(X)$ *para toda f continua y acotada.*

Dem $\Rightarrow$] Usando el teorema de Skorohod, podemos suponer que $X_n \overset{c.s.}{\to} X$. Como f es continua,

$$f(X_n) \overset{c.s.}{\to} f(X)$$

Como f es acotada, por el Teorema de convergencia acotada,

$$Ef(X_n) \to Ef(X)$$

$\Leftarrow$] Quiero probar que $P(X_n \leq x) \to P(X \leq x)$ para x punto de continuidad de F_X. Definiendo

$$f_0(y) := \mathbf{1}\{y \leq x\},$$

tenemos

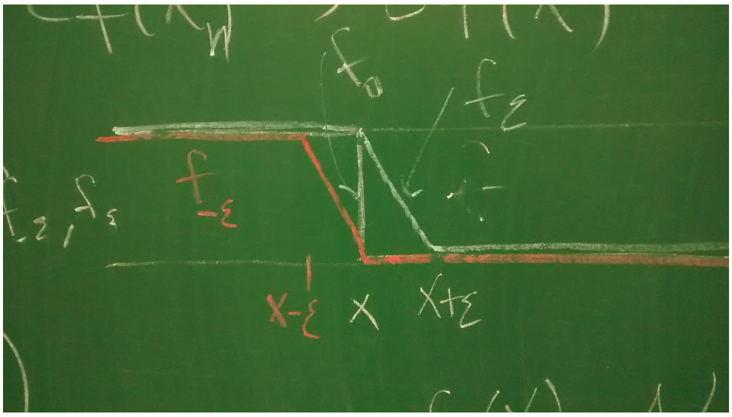
$$P(W \leq x) = Ef_0(W)$$

para cualquier variable aleatoria W. Por lo tanto queremos probar

$$\lim_{n\to\infty} Ef_0(X_n) = Ef_0(X).$$

El problema es que f_0 no es continua. Así que vamos a dominar f_0 por arriba y por abajo por funciones continuas f_ε y $f_{-\varepsilon}$ para las que podemos aplicar la hipótesis y sacar límites. Para $\varepsilon > 0$, definimos

$$f_\varepsilon(y) = \begin{cases} 1, & y \leq x \\ \text{interpolación}, & x \leq y \leq x+\varepsilon \\ 0, & y \geq x+\varepsilon \end{cases} \qquad f_{-\varepsilon}(y) = \begin{cases} 1, & y \leq x-\varepsilon \\ \text{interpolación}, & x-\varepsilon \leq y \leq x \\ 0, & y \geq x \end{cases}$$



Para cualquier y tenemos

$$f_{-\varepsilon}(y) \leq f_0(y) \leq f_\varepsilon(y) \tag{26}$$

y además

$$\mathbf{1}\{y \leq x-\varepsilon\} \leq f_{-\varepsilon}(y), \qquad f_\varepsilon(y) \leq \mathbf{1}\{y \leq x+\varepsilon\} \tag{27}$$

Usando la desigualdad (26) obtenemos $Ef_{-\varepsilon}(X_n) \leq Ef_0(X_n) \leq Ef_\varepsilon(X_n)$ y

$$\lim_n Ef_{-\varepsilon}(X_n) \leq \varliminf_n Ef_0(X_n) \leq \varlimsup_n Ef_0(X_n) \leq \lim_n Ef_\varepsilon(X_n),$$

porque como $f_{-\varepsilon}$ y f_ε son continuas y acotadas los límites existen y tenemos

$$Ef_{-\varepsilon}(X) \le \underline{\lim} Ef_0(X_n) \le \overline{\lim} Ef_0(X_n) \le Ef_\varepsilon(X).$$

Usando la desigualdad (27), obtenemos $F_X(x-\varepsilon) \le Ef_{-\varepsilon}(X_n)$ y $Ef_\varepsilon(X_n) \le F_X(x+\varepsilon)$, de donde sigue que

$$F_X(x-\varepsilon) \le \underline{\lim} Ef_0(X_n) \le \overline{\lim} Ef_0(X_n) \le F_X(x+\varepsilon).$$

Pero como x es un punto de continuidad de F, al hacer tender $\varepsilon \to 0$ tenemos $Ef_0(X_n) \to F_X(x) = Ef_0(X)$. □

8.6. Teorema de Slutsky

Teorema 50 (de Slutsky) *Si X_n converge en distribución a X y Y_n converge en probabilidad a una constante c entonces*

$$\begin{aligned} X_n + Y_n &\stackrel{D}{\to} X + c \\ X_n Y_n &\stackrel{D}{\to} Xc \\ X_n / Y_n &\stackrel{D}{\to} X/c, \qquad \text{si } c \neq 0. \end{aligned}$$

Dem. g continua acotada por $|g(x)| \le K$ para todo x.

$$\begin{aligned} &|Eg(X_n + Y_n) - Eg(X + c)| \\ &\quad \le |E[g(X_n + Y_n) - g(X_n + c)]| + |Eg(X_n + c) - Eg(X + c)| \\ &\quad = A_n + B_n \end{aligned}$$

B_n va a cero porque la función $g(\cdot + c)$ es continua y acotada más la convergencia en distribución de X_n.

Vamos a acotar A_n. Para todo $\varepsilon > 0$ podemos elegir $M(\varepsilon)$ y $-M(\varepsilon)$ puntos de continuidad de F_X tales que $P(|X| > M(\varepsilon)) < \varepsilon$. Escribimos

$$\begin{aligned} A_n &= \big|E\big[g(X_n + Y_n) - g(X_n + c)\big] \times \big[\mathbf{1}\{|X_n| > M(\varepsilon)\} + \mathbf{1}\{|X_n| \le M(\varepsilon)\}\big] \\ &\qquad \times \big[\mathbf{1}\{|Y_n - c| > \delta\} + \mathbf{1}\{|Y_n - c| \le \delta\}\big]\big| \\ &\le E\big|\big[g(X_n + Y_n) - g(X_n + c)\big] \times \big[\mathbf{1}\{|X_n| \le M(\varepsilon)\}\mathbf{1}\{|Y_n - c| \le \delta\}\big]\big| \\ &\qquad + 2K\, P(|X_n| > M(\varepsilon)) + 2K\, P(|Y_n - c| > \delta) \end{aligned}$$

El primer término está dominado por

$$\sup\{|g(x+y) - g(x+c)| : |x| < M(\varepsilon),\, |y - c| < \delta\} := h(\delta, \varepsilon) \to_\delta 0$$

para cada $\varepsilon > 0$ fijo, porque como g es continua y x es acotado, tenemos continuidad uniforme en el dominio $|x| \le M(\varepsilon) + c + 1$.

El segundo término converge a algo dominado por $2K\varepsilon$:

$$P(|X_n| > M(\varepsilon)) \to_n P(|X| > M(\varepsilon)) \le \varepsilon$$

por la convergencia de F_{X_n} a F_X en $\pm M(\varepsilon)$, puntos de continuidad de F_X. La dominación sigue de la definición de $M(\varepsilon)$.

El tercer término converge a 0:

$$P(|Y_n - c| > \delta) \to_n 0$$

por la convergencia en probabilidad de $Y_n \to c$. Así,

$$\lim_{\delta} \limsup_{n} A_n \leq \lim_{\delta}(h(\delta, \varepsilon) + 2K\varepsilon) \leq 2K\varepsilon \to_{\varepsilon} 0$$

Esto demuestra la primera convergencia del teorema.

Ligeras variaciones demuestran las otras dos convergencias. □

Aplicación X_i iid Bernoulli(p). $\bar{X}_n$ media muestral.

Por LGN

$$\bar{X}_n \xrightarrow{P} p$$

De donde (ejercicio),

$$\sqrt{\bar{X}_n(1-\bar{X}_n)} \xrightarrow{P} \sqrt{p(1-p)}$$

que equivale a

$$\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{p(1-p)}} \xrightarrow{P} 1$$

Como por el TCL

$$Z_n = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \to Z \sim \text{ Normal}(0,1)$$

tenemos

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\frac{\sqrt{p(1-p)}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}$$

que converge en distribución a Z.

9. Funciones características

9.1. Variables complejas

1. Variables aleatorias complejas.

$Z : \Omega \to \mathbf{C}$ variable aleatoria compleja si $Z = X + iY$ con X, $Y : \Omega \to \mathbb{R}$ variables aleatorias. Recordemos que $i^2 = -1$, el conjugado está definido por $\bar{Z} := X - iY$ y el módulo por $|Z| = \sqrt{Z\bar{Z}} = \sqrt{X^2 + Y^2}$.

2. $t \in \mathbb{R}$, $e^{it} = \cos(t) + i\,\mathrm{sen}(t)$, $|e^{it}| = 1$

3. $z \in \mathbf{C}$, $e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$

4. $f : \mathbb{R} \to \mathbf{C}$, $f = f_1 + if_2$ entonces $\int f = \int f_1 + i \int f_2$

Z variable aleatoria compleja, $Z = X + iY$, $EZ = EX + iEY$.

5. $|EZ| \leq E|Z|$.

Dem: Sea $Z = X + iY$. $EZ = EX + iEY$. Note que $EZ \in \mathbb{R}$ implica $EZ = E\mathrm{Re}Z$

$EZ = re^{i\theta}$, $r = |EZ|$, $\theta = \mathrm{ang}(EZ)$

$E(e^{-i\theta}Z) = e^{-i\theta}EZ = r \in \mathbb{R}$.

Entonces $r = E(\mathrm{Re}(e^{-i\theta}Z)) \leq E|e^{-i\theta}Z| = |e^{-i\theta}|\, E|Z|$

Entonces $|EZ| = r \leq E|Z|$.

9.2. Función característica

Def Dado $X : \Omega \to \mathbb{R}$, definimos la **Función característica de** X:

$$\varphi_X(t) := Ee^{itX} = E\cos(tX) + iE\,\mathrm{sen}(tX)$$

Observaciones 1. $\varphi_X(t)$ está bien definido para todo $t \in \mathbb{R}$ para toda variable X.

2. Si X es una variable aleatoria continua,

$$\varphi_X(t) = \int e^{itx} f_X(x)dx$$

es la *transformada de Fourier* de f_X.

3. Si X es discreta:

$$\varphi_X(t) = \sum_x e^{itx} P(X = x)$$

Ejemplos

1. Si $X \equiv a$ es constante,

$$\varphi_X(t) = e^{ita}$$

2. $X \sim$ Bernoulli(p). $\varphi_X(t) = e^{it0}(1-p) + e^{it1}p = p(e^{it} - 1) + 1$

3. $X \sim$ Poisson(λ).

$$\varphi_X(t) = \sum_k e^{itk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{(\lambda e^{it})} = e^{\lambda(e^{it}-1)}$$

4. $X \sim$ Binomial(n, p).

$$\varphi_X(t) = \cdots = (p(e^{it} - 1) + 1)^n$$

Propiedades X, Y variables aleatorias.

1. $|\varphi_X(t)| \leq 1$ porque $|\varphi_X(t)| \leq E|e^{itX}| = E1 = 1$.

2. $|\varphi_X(0)| = 1$. Inmediata.

3. $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$.

Si $\varphi_X(t)$ es real, entonces X y $-X$ tienen la misma distribución, es decir X simétrica.

$\overline{e^{it}} = \cos t - i \operatorname{sen} t = \cos(-t) + i \operatorname{sen}(-t) = e^{-it}$

4. φ_X es continua.

$$\begin{aligned}
\varphi_X(t) - \varphi_X(s) &= Ee^{itX} - Ee^{isX} = E(e^{itX} - e^{isX}) \\
&= E(e^{itX}(1 - e^{i(s-t)X})).
\end{aligned}$$

Entonces,

$$|\varphi_X(t) - \varphi_X(s)| \leq E(|e^{itX}|\,|1 - e^{i(s-t)X}|) \underset{s \to t}{\to} 0$$

donde usamos convergencia acotada porque $|e^{itX}| \leq 1$ y $|1 - e^{i(s-t)X}| \leq 2$.

5. Sean $X_1, \dots, X_n$ independientes, entonces

$$\varphi_{X_1 + \cdots + X_n}(t) = \varphi_{X_1}(t) \dots \varphi_{x_n}(t)$$

Probemos el caso $n = 2$:

$$\begin{aligned}
\varphi_{X+Y}(t) &= E(e^{itX} e^{isY}) \\
&= E[(cos(tX) + i \operatorname{sen}(tX))(cos(tY) + i \operatorname{sen}(tY))] \\
&= \cdots = \varphi_X(t)\varphi_Y(t),
\end{aligned}$$

usando independencia.

6. $\varphi_{aX+b}(t) \;=\; \varphi_X(at)e^{itb}$

Dem: $\varphi_{aX+b}(t) = E(e^{itaX} e^{itb}) = e^{itb}\varphi_X(at)$

7. **Fórmula de inversión.** Si $x < y$ son puntos de continuidad de F_X, entonces

$$F_X(y) - F_X(x) = \frac{1}{2\pi} \lim_n \int_{-n}^{n} \frac{e^{-itx} - e^{-ity}}{it} \varphi_X(t) dt$$

En particular, si X toma valores naturales, es decir, si $X \in \mathbb{N}$

$$P(X = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_X(t)\, dt$$

Demostración en el caso $\mathbb{N}$. Observe que

$$\int_{-\pi}^{\pi} e^{itk} dt = 2\pi\, \mathbf{1}\{k = 0\}$$

Entonces,

$$\begin{aligned}
P(X = k) &= \sum_j \mathbf{1}\{j - k = 0\} P(X = j) \\
&= \sum_{j=0}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itj - itk} P(X = j)\, dt \\
&= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{j=0}^{\infty} e^{itj} P(X = j) e^{-itk}\, dt.
\end{aligned}$$

porque $\sum_{j=0}^{\infty} e^{itj} P(X = j) = \varphi_X(t)$. Las sumas e integrales se pueden intercambiar porque

$$|e^{itj-itk} P(X = j)| \leq P(X = j)$$

que es sumable. □

8. Las **derivadas de la función característica** en el origen generan los **momentos de la distribución**. Si las siguientes derivadas existen, entonces

$$\varphi_X'(0) = \frac{d}{dt} E e^{itX}\Big|_{t=0} = E(iXe^{itX})\Big|_{t=0} = iEX$$

$$\varphi_X^{(2)}(0) = \frac{d^2}{dt^2} E e^{itX}\Big|_{t=0} = E(i^2X^2e^{itX})\Big|_{t=0} = i^2 EX^2$$

y en general si $E|X|^k < \infty$, entonces $\varphi_X \in C^n(\mathbb{R})$ y vale

$$EX^k = \frac{1}{i^k} \frac{d^k}{dt^k} \varphi_X(0) = (-1)^k i^k \varphi_X^{(k)}(0)$$

Demostración omitida.

9. **La función característica determina la distribución** (Corolario de 7)

$$F_X = F_Y \quad \Leftrightarrow \quad \varphi_X = \varphi_Y$$

Idea de la demostración. Ver Billingsley [3], Barry James [8]. Si $\varphi_X = \varphi_Y$, entonces

$$E\cos(tX) = E\cos(tY); \quad E\operatorname{sen}(tX) = E\operatorname{sen}(tY)$$

En particular, si a_k, b_k son sucesiones de reales,

$$\begin{aligned}&\sum_k E(a_k \cos(kX)) + i\sum_k E(b_k \operatorname{sen}(kX))\\ &\qquad = \sum_k E(a_k \cos(kY)) + i\sum_k E(b_k \operatorname{sen}(kY))\end{aligned}$$

Pero

$$\Big\{f : f(x) = \sum_k a_k \cos(kx) + \sum_k b_k \operatorname{sen}(kx)\Big\}$$

es denso en $\mathcal{C}(0, 2\pi)$. Esto "implica" $E(f(X)) = E(f(Y))$, $\forall f$ continua de soporte acotado. Que, a su turno, implica $X \sim Y$. Esto probaría (9).

10.

Teorema 51 (de continuidad de Paul Levy) *X, X_n variables aleatorias con*

$$\lim_{t\to\infty} \varphi_{X_n}(t) = \varphi_X(t), \qquad \textit{para todo } t \in \mathbb{R}.$$

Entonces $X_n \xrightarrow{D} X$.

Idea de la demostración Es un argumento análogo a la demostración de 9. $\varphi_{X_n}(t) \to \varphi_X(t)$ "implica" $Ef(X_n) \to Ef(X)$ para $f \in \mathcal{C}(\mathbb{R})$ con soporte acotado. Entonces $X_n \xrightarrow{D} X$.

Teorema de Taylor para variable compleja. Si $\varphi^{(k)}$ existe y k es par entonces $E|X^k|$ existe, si k es impar entonces $E|X^{k-1}|$ existe. Si $E|X^k|$ existe, entonces

$$\varphi(t) = \sum_{j=0}^{k} \frac{EX^j}{j!}(it)^j + o(t^k).$$

Y así, $\varphi^{(k)}(0) = i^k EX^k$.

Ley débil de grandes números asumiendo primer momento finito.

Teorema *Sean X_n i.i.d. con $EX_n = \mu < \infty$, Entonces*

$$\bar{X}_n \xrightarrow[n]{P} \mu$$

Dem Por independencia:

$$\begin{aligned}\varphi_{\bar{X}_n}(t) &= \prod_{i=1}^{n} \varphi_{X_i}(t/n) = \big(\varphi_{X_i}(t/n)\big)^n \\ &= \big(\varphi_{X_1}(0) + \varphi'_{X_1}(0)\frac{t}{n} + R(t/n)\big)^n\end{aligned}$$

por Taylor de orden 1, con $R(s)/s \to 0$ para $s \to 0$.

$$\begin{aligned}&= \Big(1 + \frac{i\mu\, t + R(t/n)n}{n}\Big)^n = \Big(1 + \frac{i\mu\, t + R(t/n)n}{n}\Big)^{\frac{n}{i\mu\, t + R(t/n)n}(i\mu\, t + R(t/n)n)} \\ &\to e^{i\mu t} = \varphi_\mu(y),\end{aligned}$$

la función característica de la constante μ. Esto implica que $\bar{X}_n \xrightarrow{D} \mu$ por el Teorema 51 de continuidad de Levy. Finalmente como se trata de convergencia en distribución a una constante, tenemos también convergencia en probabilidad $\bar{X}_n \xrightarrow{P} \mu$ por el Teorema 47. □

Cálculo de la función característica de la Normal

Si Z es Normal$(0, 1)$ entonces $\varphi_Z(t) = e^{-t^2/2}$:

$$\varphi_Z(s) = Ee^{itZ} = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} \exp(itz - (z^2/2))dz = e^{-t^2/2}$$

Vamos a calcular la *función generadora de momentos*:

$$\begin{aligned}M_Z(s) := Ee^{sZ} &= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} \exp(sz - (z^2/2))dz \\ &= e^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} \exp(-(s-z)^2/2))dz = e^{s^2/2},\end{aligned}$$

completando cuadrados. Ahora usamos el siguiente

Teorema 52 *Si* $|M(s)| < \infty$ *para* $|s| < a$*, entonces* $M(t)$ *tiene una extensión analítica en el conjunto* $|\mathrm{im}(t)| < a$ *y*

$$\varphi(t) = M(it), \qquad t \in \mathbb{R}.$$

Ver Grimmett-Stirzaker [7], página 184. Por lo tanto,

$$\varphi_Z(t) = e^{(it)^2/2} = e^{-t^2/2}.$$

Alternativamente, también se puede probar

$$\varphi_Z(t) = Ee^{itZ} = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-t^2/2}$$

usando residuos.

10. Teorema Central del Límite

Sabemos ahora, por la LGN, que la media muestral $\frac{S_n}{n}$ converge a la media μ. Nos gustaría saber a qué escala se produce esa convergencia, es decir, averiguar de qué orden es $\frac{S_n}{n} - \mu$. Como primer paso podemos calcularle la varianza. Obtenemos $\frac{\sigma^2}{n}$, por lo tanto el orden es $\frac{1}{\sqrt{n}}$. Corregimos por $\sqrt{n}/\sigma$ para evitar que tienda a 0, es decir consideramos

$$Z_n := \frac{\sqrt{n}}{\sigma} \big(\frac{S_n}{n} - \mu\big) = \frac{S_n - n\mu}{\sqrt{n}\sigma}, \tag{28}$$

que tiene media 0 y varianza 1. La siguiente pregunta es si existirá el límite de las variables (28) cuando $n \to \infty$, en algún sentido. La respuesta es afirmativa:

Teorema Central del Límite (TCL) *Sea* $(X_i,\, i \in \mathbb{N})$ *una sucesión de variables aleatorias i.i.d. con media* μ *y varianza* σ^2*,* $0 < \sigma < \infty$*. Sea* $(Z_n,\, n \in \mathbb{N})$ *la suma normalizada de* (28)*. Entonces* $Z_n \xrightarrow{d} N(0,1)$ *cuando* $n \to \infty$*, donde* $N(0,1)$ *es la distribución normal estándar.*

¿Por qué normal? Tratemos de justificar por qué el límite de las sumas (28) en caso de existir debe ser normal. Por simplicidad, podemos suponer que $\mu = 0$ y $\sigma = 1$ en (28). Observemos que si $Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{d} Y$, entonces tomando $n = 2k$

$$Z_n = \frac{X_1 + \cdots + X_k + X_{k+1} + \cdots + X_{2k}}{\sqrt{2k}} = \frac{Z_k + Z_k'}{\sqrt{2}} \xrightarrow{d} Y \tag{29}$$

también, donde $Z_k' = \frac{X_{k+1} + \cdots + X_{2k}}{\sqrt{n}} \sim Z_k$, y además Z_k y Z_k' son independientes, ambas propiedades consecuencia de que las variables $(X_i, i \in \mathbb{N})$ son i.i.d.. Tanto Z_k como Z_k' tienden en distribución a Y, y son independientes, luego cabe esperar que Y satisfaga la igualdad en distribución

$$\frac{Y' + Y''}{\sqrt{2}} \overset{d}{=} Y, \quad \text{donde } Y' \sim Y'' \sim Y \text{ son copias independientes de } Y. \tag{30}$$

Se puede probar que la distribución normál estándar es la única que cumple (30) y tiene varianza 1 (una propiedad que debiese heredar de las Z_n), y por lo tanto logramos identificar el límite.

10.1. Tercer momento finito

Pasemos ahora la demostración del teorema bajo la hipótesis del tercer momento finito. Esto permite hacer una demostración más intuitiva.

Teorema *Sea* $(X_i,\, i \in \mathbb{N})$ *una sucesión de variables aleatorias i.i.d.,* $X_i \sim X$*, con media* μ*, varianza* σ^2 *y* $E[|X|^3] < \infty$*. Sea* $(Z_n,\, n \in \mathbb{N})$ *la suma normalizada de* (28)*. Entonces* $Z_n \xrightarrow{d} N(0,1)$ *cuando* $n \to \infty$*.*

Demostración. Observemos primero que si el resultado vale para X entonces también valdrá para cualquier combinación $aX + b$ con $a, b \in \mathbb{R}$, $a \neq 0$. Podemos por lo tanto suponer que $\mu = 0$ y $\sigma = 1$, y en este caso (28) se reduce a $Z_n = S_n/\sqrt{n}$. Basta con probar que

$$\lim_{n\to\infty} E[f(Z_n)] = E[f(W)], \tag{31}$$

donde $W \sim N(0,1)$, y $f : \mathbb{R} \to \mathbb{R}$ es cualquier función de clase C^3 acotada, con derivadas acotadas hasta el tercer orden inclusive. Fijemos una tal función f.

Consideremos una sucesión $(Y_j,\, j \in \mathbb{N})$ de variables aleatorias i.i.d. $N(0,1)$, independientes de las variables $(X_i,\, i \in \mathbb{N})$ del enunciado, y sea

$$W_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \sim N(0,1). \tag{32}$$

Definamos las variables aleatorias

$$\begin{aligned}
&D_{n,i} = \frac{X_1 + \cdots + X_{i-1} + Y_{i+1} + \cdots + Y_n}{\sqrt{n}} \quad \text{si} \quad 2 \le i \le n-1 \\
&D_{n,1} = \frac{Y_2 + \cdots + Y_n}{\sqrt{n}}, \quad D_{n,n} = \frac{X_1 + \cdots + X_{n-1}}{\sqrt{n}}, \\
&\text{y} \quad Z_{n,i} = \frac{X_1 + \cdots + X_i + Y_{i+1} + \cdots + Y_n}{\sqrt{n}}, \quad 1 \le i \le n-1,\ Z_{n,0} = W_n,\ Z_{n,n} = Z_n.
\end{aligned}$$

Observemos que $Z_{n,i} = D_{n,i} + \frac{X_i}{\sqrt{n}}$ y $Z_{n,i-1} = D_{n,i} + \frac{Y_i}{\sqrt{n}}$, $1 \le i \le n$. Podemos escribir ahora $E[f(Z_n)] - E[f(W_n)]$ como suma telescópica

$$E[f(Z_n)] - E[f(W_n)] = \sum_{1 \le i \le n} E[f(Z_{n,i})] - E[f(Z_{n,i-1})]. \tag{33}$$

Consideremos la expansión de Taylor de segundo orden con la fórmula del resto en términos de la tercera derivada en un punto intermedio

$$f(Z_{n,i}) = f(D_{n,i}) + f'(D_{n,i})\frac{X_i}{\sqrt{n}} + \frac{1}{2}1f''(D_{n,i})\frac{X_i^2}{n} + \frac{1}{6}f'''(\theta_i)\frac{X_i^3}{n^{3/2}}$$

θ_i entre $D_{n,i}$ y $Z_{n,i}$, y similarmente

$$f(Z_{n,i-1}) = f(D_{n,i}) + f'(D_{n,i})\frac{Y_i}{\sqrt{n}} + \frac{1}{2}f''(D_{n,i})\frac{Y_i^2}{n} + \frac{1}{6}f'''(\theta_i')\frac{Y_i^3}{n^{3/2}},$$

para un valor θ_i' entre $D_{n,i}$ y $Z_{n,i-1}$. Recordemos que $E[X_i] = E[Y_i]$, $E[X_i^2] = E[Y_i^2]$, y que además tanto X_i como Y_i son independientes de $D_{n,i}$, $1 \le i \le n$. Por lo tanto, usando que la esperanza del producto de variables aleatorias independientes es igual al producto de las esperanzas, los primeros tres términos de los desarrollos de Taylor arriba se cancelan al tomar la diferencia $E[f(Z_{n,i})] - E[f(Z_{n,i-1})]$, y podemos acotar

$$\big|E[f(Z_{n,i})] - E[f(Z_{n,i-1})]\big| \leq \frac{C}{n^{3/2}}, \quad \text{donde} \quad C = \frac{1}{6}\sup_{x\in\mathbb{R}}|f'''(x)|\,\big(E[|X|^3] + E[|W|^3]\big) < \infty$$

es una constante positiva que no depende de n. De la identidad (33) deducimos que

$$|E[f(Z_n)] - E[f(W_n)]| \leq C/\sqrt{n}\,, \tag{34}$$

y como $W_n \sim N(0,1)$, obtenemos (31), como queríamos probar. □

Observación. La técnica que utilizamos en la demostración consiste en mostrar que, en el límite, $E[f(Z_n)]$ no depende de los distribución particular de la variable X, en particular coincide con $E[f(W_n)]$. Pero en este último caso el teorema es trivial, ya que la distribución $N(0,1)$ es un punto fijo: si $X_i \sim N(0,1), 1 \leq i \leq n$, son variables aleatorias i.i.d., entonces $W_n/\sqrt{n} \sim N(0,1)$ por (32) . Es decir, probamos que existe el límite y lo identificamos simultáneamente.

Velocidad de convergencia. La demostración del TCL que hicimos permite estimar las probabilidades acumuladas $P(Z_n \leq a)$, $a \in \mathbb{R}$. Mantenemos las hipótesis de la demostración, $\mu = 0$ y $\sigma = 1$, dejando al lector la tarea de adaptar la demostración al caso general.

Sea $\psi : \mathbb{R} \to \mathbb{R}$ de clase C^∞ tal que $\psi|_{(-\infty,0)} \equiv 1$, $\psi|_{(1,+\infty)} \equiv 0$, y para $\epsilon > 0$ definamos

$$g_{-\epsilon}(x) = \psi\big((x-a+\epsilon)/\epsilon\big), \quad g_{+\epsilon}(x) = \psi\big((x-a)/\epsilon\big).$$

Entonces $g_{-\epsilon}$, $g_{+\epsilon} \in C^\infty(\mathbb{R})$, sus derivadas de tercer orden están uniformemente acotadas por un múltiplo de $1/\epsilon^3$, y

$$\mathbf{1}(x \leq a-\epsilon) \leq g_{-\epsilon}(x) \leq \mathbf{1}(x \leq a) \leq g_{+\epsilon}(x) \leq \mathbf{1}(x \leq a+\epsilon), \quad x \in \mathbb{R}. \tag{35}$$

Tomando esperanzas en (35) y reemplazando en (34) obtenemos

$$P(Z_n \leq a) \leq E[g_{+\epsilon}(Z_n)] \leq E[g_{+\epsilon}(W)] + \frac{C'E[|X|^3]}{\sqrt{n}\,\epsilon^3} \leq P(W \leq a+\epsilon) + \frac{C'E[|X|^3]}{\sqrt{n}\,\epsilon^3} \tag{36}$$

$C' > 0$ independiente de n, ϵ y la distribución de X, y $W \sim N(0,1)$. Similarmente,

$$P(Z_n \leq a) \geq E[g_{-\epsilon}(Z_n)] \geq E[g_{-\epsilon}(W)] - \frac{C'E[|X|^3]}{\sqrt{n}\,\epsilon^3} \geq P(W \leq a-\epsilon) - \frac{C'E[|X|^3]}{\sqrt{n}\,\epsilon^3}. \tag{37}$$

Como la función de densidad de $W \sim N(0,1)$ está acotada, resulta

$$P(W \leq a+\epsilon) \leq P(W \leq a) + c\epsilon \quad \text{y también} \quad P(W \leq a-\epsilon) \geq P(W \leq a) - c\epsilon,$$

$c > 0$ independiente de a y ϵ. Sea ahora $\epsilon = n^{-1/8}$, tal que $\sqrt{n}\epsilon^3 = \epsilon^{-1}$. Por (36) y (37),

$$\sup_{a\in\mathbb{R}}\Big|P\big(Z_n \leq a\big) - P\big(W \leq a\big)\Big| \leq K\frac{1+E[|X|^3]}{n^{\frac{1}{8}}}. \tag{38}$$

Rastreando la composición de la constante a lo largo de la demostración resulta

$$K \leq \frac{1+\sup_x |\psi'''(x)|}{\sqrt{3}}.$$

Cota óptima. Andrew Berry y Carl Esseen obtuvieron en 1941 y 1942 las cotas óptimas de la velocidad de convergencia para el TCL, estableciendo que si $(X_i,\ i \in \mathbb{N})$ son v.a. i.i.d. $X_i \sim X$, con media μ, varianza σ^2 y $E[|X-\mu|^3] < \infty$, entonces existe una constante $0 < K' < 4$, que no depende de X ni de n, tal que si $W \sim N(0,1)$,

$$\sup_{a\in\mathbb{R}} \Big| P\big(Z_N \leq a\big) - P\big(W \leq a\big)\Big| \leq K' \frac{E[|X-\mu|^3]}{\sigma^3\sqrt{n}}. \tag{39}$$

Aplicación. Se realizan $n = 10^4$ ensayos Bernoulli en forma independiente, cada uno de ellos con probabilidad de éxito $0,6$. Estimar la probabilidad de que se produzcan entre 7,901 y 8,100 éxitos.

Sea $X_i = 1$ si el i-ésimo experimento es exitoso, 0 en otro caso. Entonces $X_i \sim \text{Ber}(0,8)$, la cantidad de éxitos es $S_n = \sum_{i=1} X_i \sim \text{Bi}(10^4; 0,8)$, y el problema consiste en estimar $P(7{,}900 < S_n \leq 8{,}100)$. Como $n = 10^4$ es un número bastante grande, intentemos aplicar el TCL y acotar el error que cometemos con (39), ya que $E[|X_i - 0,8|^3] < \infty$.

Escalando como en el TCL, con $\mu = 0,8$ y $\sigma^2 = 0,8(1-0,8) = 0,16$, obtenemos

$$\begin{aligned} P(7{,}900 < S_n \leq 8{,}100) &= P\Big(\frac{7{,}900 - 0,8n}{0,4\sqrt{n}} < \frac{S_n - 0,8n}{0,4\sqrt{n}} \leq \frac{8{,}100 - 0{,}8n}{0,4\sqrt{n}}\Big) \\ &= P\Big(-2,5 < \frac{S_n - 0,8n}{0,4\sqrt{n}} \leq 2,5\Big). \end{aligned}$$

Escribamos

$$P\Big(-2,5 < \frac{S_n - 0,8n}{0,4\sqrt{n}} \leq 2,5\Big) = P\Big(\frac{S_n - 0,8n}{0,4\sqrt{n}} \leq 2,5\Big) - P\Big(\frac{S_n - 0,8n}{0,4\sqrt{n}} \leq -2,5\Big)$$

y apliquemos (39) a cada una de estas dos probabilidades, para obtener

$$\Big| P\Big(-2,5 < \frac{S_n - 0,8n}{0,4\sqrt{n}} \leq 2,5\Big) - P\big(-2,5 < W \leq 2{,}5\big)\Big| \leq 8\frac{E[|X_1 - 0,8|^3]}{\sigma^3 10^2} \leq 0,13$$

Evaluamos $P(-2,5 < W \leq 2{,}5) = 0,9876$. Podemos estimar la probabilidad de que se produzcan entre 7901 y 8100 experimentos exitosos como $0,9876$, cometiendo un error inferior a $0,13$.

10.2. Teorema Central del Límite con segundo momento finito

Teorema Central del Límite *Sean $X_1, X_2, \dots$ i.i.d., $EX_i = \mu$, $VX_i = \sigma^2$. entonces*

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z. \qquad \text{donde } Z \sim \text{ Normal}(0,1).$$

Dem Queremos probar $\lim_{n\to\infty} \varphi_{Z_n}(t) = \varphi_Z(t) = e^{-t^2/2}$, $t \in \mathbb{R}$. Definiendo $Y_n := (X_n - \mu)/\sigma$ tenemos $EY_n = 0$, $VY_n = 1$,

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad Z_n = \sqrt{n}\,\bar{Y}_n = \sum_{i=1}^{n}(Y_i/\sqrt{n}).$$

Calculamos

$$\varphi_{Z_n}(t) = \varphi_{(\sum(Y_i/\sqrt{n})}(t) = \Big(\varphi_{Y_1}\Big(\frac{t}{\sqrt{n}}\Big)\Big)^n \tag{40}$$

$$= \Big(\varphi_{Y_1}(0) + \varphi'_{Y_1}(0)\frac{t}{\sqrt{n}} + \frac{1}{2}\varphi''_{Y_1}(0)\frac{t^2}{n} + R(t/\sqrt{n})\Big)^n \tag{41}$$

por Taylor segundo orden. Se pudo derivar dos veces porque la varianza es finita. El resto $R(t)/t^2 \to 0$ cuando $t \to 0$. Como la media de Y_i es cero y la varianza 1, (41) es igual a

$$= \Big(1 - \frac{t^2}{2n} + R(t/\sqrt{n})\Big)^n = \Big(1 + \frac{-\frac{t^2}{2} + nR(t/\sqrt{n})}{n}\Big)^n$$

que converge a $e^{-t^2/2}$. $\square$

10.3. Teorema de De Moivre Laplace

El Teorema de Demoivre Laplace es la versión del TCL para variables centradas que asumen valores -1 y 1. Vamos a demostrarlo directamente sin pasar por la función característica.

Teorema de De Moivre Laplace *Sean X_i, $i = 1, 2, \dots$ variables aleatorias independientes con distribución $P(X_i = -1) = P(X_i = 1) = 1/2$. Sea $S_n := \sum_{i=1}^n X_i$. Entonces para todo $x \geq 0$,*

$$\lim_{n\to\infty} P\Big(\frac{S_{2n}}{\sqrt{2n}} \leq x\Big) = \Phi(x),$$

donde Φ es la función de distribución acumulada de la Normal standard.

Dem Como $EX_i = 0$ y $VX_i = 1$, $ES_{2n} = 0$ y $VS_{2n} = 2n$. Por lo tanto $\frac{S_{2n}}{\sqrt{2n}}$ es centrada con varianza 1. Vamos a calcular explicitamente el límite. Defino

$$a_0 := P(S_{2n} = 0) = \binom{2n}{n}\Big(\frac{1}{2}\Big)^{2n} = \frac{(2n)!}{n!n!}\Big(\frac{1}{2}\Big)^{2n}.$$

que, usando la aproximación de Stirling, es aproximadamente igual a

$$\simeq \frac{(2n)^{2n}e^{-2n}\sqrt{2\pi 2n}}{\big(n^n e^{-n}\sqrt{2\pi n}\big)^2}\Big(\frac{1}{2}\Big)^{2n} = \frac{1}{\sqrt{\pi n}}$$

Vamos a calcular

$$a_k := P(S_{2n} = 2k) = \binom{2n}{n+k}\Big(\frac{1}{2}\Big)^{2n}$$
$$= a_0 \frac{n(n-1)\dots(n-k+1)}{(n+k)(n+k-1)\dots(n+1)}$$

Dividiendo cada término del denominador por el correspondiente del numerador,

$$= a_0\Big(\Big(1+\frac{k}{n}\Big)\Big(1+\frac{k}{n-1}\Big)\dots\Big(1+\frac{k}{n-k+1}\Big)\Big)^{-1}$$

Hay k factores en el producto. Sea b una constante positiva que no depende de n. Substituyendo $k = b\sqrt{2n}$ y simplificando,

$$= a_0\Big(\Big(1+\frac{b\sqrt{2}}{\sqrt{n}}\Big)\Big(1+\frac{b\sqrt{2}}{\sqrt{n}-\frac{1}{\sqrt{n}}}\Big)\dots\Big(1+\frac{b\sqrt{2}}{\sqrt{n}-\frac{b\sqrt{2}+1}{\sqrt{n}}}\Big)\Big)^{-1}$$

Como hay $\simeq b\sqrt{2}\sqrt{n}$ factores en el producto y los términos $\frac{1}{\sqrt{n}},\dots,\frac{b\sqrt{2}+1}{\sqrt{n}}$ se van a cero, los ignoramos y nos queda

$$\simeq a_0\Big(1+\frac{b\sqrt{2}}{\sqrt{n}}\Big)^{-b\sqrt{2}\sqrt{n}} \simeq \frac{1}{\sqrt{\pi n}}\,e^{-2b^2}$$

Consideremos $x > 0$ y calculemos

$$P\Big(0 \le \frac{S_{2n}}{\sqrt{2n}} \le x\Big) = \sum_{k=0}^{\sqrt{2n}x/2} a_k$$

Volviendo a escribir $k = b\sqrt{2n}$, para $b \in [0, x]$ tenemos $2b^2 = (k/\sqrt{n})^2$ y obtenemos

$$\simeq \sum_{k=0}^{\sqrt{n}x/\sqrt{2}} \frac{1}{\sqrt{\pi n}}\,e^{-(k/\sqrt{n})^2}$$

que es la suma de Riemann que aproxima

$$\underset{n\to\infty}{\longrightarrow} \int_0^{x/\sqrt{2}} \frac{1}{\sqrt{\pi}} e^{-z^2} dz = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

donde usamos el cambio de variable $y = \sqrt{2}z$. La última expresión es igual a

$$= \Phi(x) - \Phi(0).$$

Usando simetría, concluimos. $\square$

Formas alternativas del TCL:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z, \qquad\qquad \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z$$

10.4. TCL como punto fijo de un sistema dinámico

Particionemos en dos partes la suma que define Z_{2n}:

$$Z_{2n} = \frac{S_{2n}}{\sqrt{2n}} = \frac{S_n + S_{2n} - S_n}{\sqrt{2n}} = \frac{1}{\sqrt{2}}\Big(\frac{S_n}{\sqrt{n}} + \frac{S_n^*}{\sqrt{n}}\Big)$$

con $S_n^* \sim S_n$ y ambos independientes. O sea que el límite, si existe, tiene que satisfacer:

$$Z \sim \frac{Z + Z^*}{\sqrt{2}} \tag{42}$$

con $Z^* \sim Z$ y Z, Z^* independientes. En términos de la función característica, debería satisfacerse la ecuación $\varphi_Z(t) = (\varphi_Z(t/\sqrt{2}))^2$. Observamos que la normal satisface:

$$\varphi_Z(t) = e^{-t^2/2} = (e^{-t^2/4})^2 = (\varphi_Z(t/\sqrt{2}))^2$$

Para probar el TCL usando este enfoque habría que probar: (1) que el limite existe; (2) que la normal es la única que satisface la "ecuación" (42).

10.5. Observaciones, ejemplos y aplicaciones

Observación Qué significa n suficientemente grande? Cómo sabemos si la aproximación es buena? El tamaño de muestra requerido para que la aproximación sea razonable depende de la forma de la distribución de las X_i . Mientras más simétrica y acampanada sea, más rápidamente se obtiene una buena aproximación.

Ejemplo Al sumar números, una calculadora aproxima cada número al entero más próximo. Los errores de aproximación se suponen iid Uniforme$(-0{,}5, 0{,}5)$.

a) Si se suman 1500 números, ¿cuál es la probabilidad de que el valor absoluto del error total exceda 15?

Si llamamos X_i al error correspondiente al i-ésimo sumando, el error total es $T_{1500} = \sum_i X_i$. Usar $EX_i = 0$ y $VX_i = 1/12$.

b) ¿Cuántos números pueden sumarse a fin de que el valor absoluto del error total sea menor o igual que 10 con probabilidad mayor o igual que 0.90? Buscamos el valor de n tal que $P(|T_n| \leq 10) \geq 0{,}9$. R: 446.

Otras Aplicaciones del TCL

1. $Y_n \sim$ Poisson(λn)

$$\frac{Y_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{D} Z$$

Dem: Defina $X_i \sim$ Poisson(λ) iid. Entonces,

$Y_n = X_1 + \cdots + X_n \sim$ Poisson(λn).

Aplique TCL y obtenga el límite.

Así la Poisson con parametro grande se aproxima por la normal.

2. $Y_n \sim$ Gama(n, λ) iid con n entero

$$\frac{Y_n - \frac{n}{\lambda}}{\sqrt{n}/\lambda} \xrightarrow{D} Z$$

$X_i \sim$ Gama$(1, \lambda)$ (exponenciales) independientes.

$X_1 + \cdots + X_n \sim$ Gama(n, λ) suma de n exponenciales independientes.

Así la suma de gamas se aproxima por la normal.

3. Un adivino acierta el color de 950 de 1500 cartas puestas al dorso. Queremos decidir si creemos que es adivino.

Sea p la probabilidad que el adivino acierte. Queremos testar $p = 1/2$ (es decir, no mejora el puro azar) contra $p > 1/2$ (tiene cierta chance de adivinar mayor que 1/2).

Supongamos que decide al azar, $p = 1/2$.

Sea $X_i = \mathbf{1}\{\text{acierta la carta } i\}$. Azar $\Rightarrow X_i \sim$ Bernoulli$(\frac{1}{2})$

Número de aciertos:

$$S_{1500} = \sum_{i=1}^{1500} X_i, \qquad \bar{X} = \frac{S_{1500}}{1500}$$

$$P(S_{1500} \geq 950) = P\left(\frac{\bar{X} - \frac{1}{2}}{\frac{0{,}5}{\sqrt{1500}}} \geq \frac{\frac{950}{1500} - \frac{1}{2}}{\frac{0{,}5}{\sqrt{1500}}}\right) \sim P(Z \geq 10{,}32) \sim 0$$

La probabilidad de acertar 950 veces con una moneda es casi 0. Aceptamos la hipótesis que el hombre es un adivino.

11. Procesos de Bernoulli y Poisson

11.1. Procesos de Bernoulli y Binomial

Para estudiar una sucesión de ensayos de Bernoulli consideramos el espacio muestral: $\Omega := \{\omega = (\omega_1, \omega_2, \dots), \omega_i \in \{0,1\}\}$. Un elemento típico $\omega \in \Omega$ tiene la forma $\omega = 11000010101100010100010110010100010100001\dots$. Sea $\mathcal{A}$ el álgebra de los eventos *cilíndricos*, eventos que dependen de un número finito de coordenadas:

$$\mathcal{A} = \big\{\{\omega \in \Omega : \omega_i = 1, i \in I; \omega_j = 0, j \in J\} : I, J \subset \mathbb{N}, \text{finitos, disjuntos}\big\}$$

P es la probabilidad "compatible con ensayos independientes" definida para $I, J \subset \mathbb{N}$ disjuntos finitos,

$$P(\omega \in \Omega : \omega_i = 1,\ i \in I, \omega_j = 0,\ j \in J) \ = \ p^{|I|}(1-p)^{|J|} \tag{43}$$

donde $|I| = \#I$ es el número de puntos en I. Por ejemplo:

$$P(\omega \in \Omega : \omega_1 = 1,\ \omega_2 = 0,\ \omega_3 = 0) \ = \ p(1-p)^2$$

Teorema 53 (de extensión de Kolmogorov) *Hay una única probabilidad P definida en $\mathcal{F}$ que coincide con* (43) *en los conjuntos cilíndricos.*

Definimos las variables aleatorias *proyección*:

$$X_i(\omega) := \omega_i.$$

La probabilidad de "éxito en i-ésimo ensayo" es

$$P(X_i = 1) = P(\omega : \omega_i = 1) = p.$$

Un **proceso estocástico** es una sucesión de variables aleatorias $X_1, X_2, \dots$ indexadas por $n \in \mathbb{N}$ o $t \in \mathbb{R}$. El proceso $X_1, X_2, \dots$ definido más arriba se llama **Proceso de Bernoulli**. Se trata de una sucesión de variables aleatorias independientes Bernoulli(p).

Un evento cilíndrico B se puede describir usando las variables proyección por

$$B = \{X_1 = b_1, \dots, X_n = b_n\}, \qquad b_i \in \{0, 1\},$$

y su probabilidad es

$$P(B) = p^{b_1}(1-p)^{1-b_1} \dots p^{b_n}(1-p)^{1-b_n} = p^{\sum b_i}(1-p)^{n-\sum b_i}$$

En particular

$$P(X_1 = 1, X_2 = 0, X_3 = 0) = p(1-p)^2.$$

El proceso de Bernoulli es **estacionario**:

$$P(X_1 = b_1, \dots, X_n = b_n) = P(X_{t+1} = b_1, \dots, X_{t+n} = b_n)$$

para todo t.

Ejemplo: En la parada del pabellón 2 sale un **colectivo 107** en cada minuto con probabilidad 1/10, en forma independiente. Cual es la probabilidad que salgan colectivos en los minutos 1,2,3? Y en los minutos 27,28,29?

$$P(X_1 = 1, X_2 = 1, X_3 = 1) = \left(\frac{1}{10}\right)^3.$$

$$P(X_{27} = 1, X_{28} = 1, X_{29} = 1) = \left(\frac{1}{10}\right)^3.$$

Proceso Binomial

Definamos las variables $S_n = X_1 + \cdots + X_n$. El proceso

$$S_1, S_2, \dots$$

es llamado *proceso Binomial.* S_n cuenta el número de éxitos hasta el n-ésimo ensayo. S_n tiene distribución Binomial(n, p) para cada $n \geq 1$.

El **incremento** del proceso en el intervalo $(m, n]$, con $n \leq m$ está dado por

$$S_n - S_m = \sum_{k=m+1}^{n} X_k.$$

El proceso binomial tiene **incrementos estacionarios**:

$$P(S_{n+m} - S_m = k) = P(S_n = k) \tag{44}$$

La distribución del número de llegadas en un intervalo depende del tamaño del intervalo y no de su localización.

El proceso binomial tiene **incrementos independientes**: Si $1 \le m \le n \le i \le j$,

$$\begin{aligned} P(S_n - S_m = k, S_j - S_i = h) &= P(S_n - S_m = k)P(S_j - S_i = h) \\ &= P(S_{m-n} = k)P(S_{i-j} = h) \end{aligned}$$

La probabilidad de incrementos k y h en los intervalos disjuntos $(m, n]$ y $(i, j]$, respectivamente, es el producto de las probabilidades. Más generalmente, vale para conjuntos finitos de intervalos:

$$\begin{aligned} &P(S_{n_1} - S_{m_1} = k_1, \dots, S_{n_\ell} - S_{m_\ell} = k_\ell) \\ &\quad = P(S_{n_1} - S_{m_1} = k_1) \dots P(S_{n_\ell} - S_{m_\ell} = k_\ell) \\ &\quad = P(S_{n_1 - m_1} = k_1) \dots P(S_{n_\ell - m_\ell} = k_\ell) \end{aligned} \tag{45}$$

si los intervalos $[m_j, n_j]$ son disjuntos dos a dos.

Instante de la primera llegada $Y_1 := \text{mín}\{k > 0 : X_k = 1\}$ tiene distribución geométrica:

$$P(Y_1 = k) = P(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1 - p)^{k-1}p$$

El evento $\{Y_1 = k\}$ depende de $X_1, \dots, X_k$, por lo tanto se puede calcular su probabilidad.

Ejemplo: Colectivo Si llego en un instante t cualquiera y defino el tiempo de espera del colectivo a partir de ese instante:

$$R_t := \text{mín}\{k > 0 : X_{t+k} = 1\}$$

$$P(R_t = k) = P(X_{t+1} = 0, \dots, X_{t+k-1} = 0, X_{t+k} = 1) = (1 - p)^{k-1}p$$

Tiene distribución geométrica igual que si empezaba en el instante 0.

Instante de la k-ésima llegada

$$Y_k := \text{mín}\{n : X_1 + \dots + X_n = k\}$$

Para $t \ge k$:

$$\begin{aligned} P(Y_k = t) &= P(k - 1 \text{ exitos en } [1, t - 1], \text{éxito en } t) \\ &= \binom{t-1}{k-1} p^{k-1}(1 - p)^{t-1-(k-1)}p \end{aligned}$$

Es decir que el instante de la k-ésima llegada tiene distribución Binomial negativa de parámetros k y p.

Dualidad entre Binomial y Binomial negativa Vimos en el Lema 10 que la k-ésima llegada ocurre antes del instante n si y sólo si el número de llegadas hasta el instante n es mayor o igual a k.

$$Y_k \le n \iff S_n \ge k,$$

donde $S_n = X_1 + \dots + X_n$ y $Y_k := \text{mín}\{n : X_1 + \dots + X_n = k\}$, tienen distribución $S_n \sim$ Binomial(n, p) y $Y_k \sim$ Binomial negativa (k, p).

Tiempo entre llegadas sucesivas Sea $T_0 := 0$ y $T_i := Y_i - Y_{i-1}$, $i \geq 1$. Ya vimos que Y_i tiene distribución binomial negativa.

Lema 54 *(T_k, $k \geq 1$) son variables independientes con distribución geométrica (p).*

Dem Escribiendo $s_j = t_1 + \cdots + t_j$ tenemos

$$\begin{aligned} P(T_1 = t_1, \ldots, T_k = t_k) &= P(\cap_{j=1}^{k} \{X_{s_{j-1}+1} = \cdots = X_{s_j - 1} = 0, X_{a_j} = 1\}) \\ &= (1-p)^{t_1 - 1} p \ldots (1-p)^{t_k - 1} p \end{aligned}$$

Sumando esa expresión sobre todos los $t_k \geq 1$ para $k \neq i$, obtenemos que la i-ésima marginal tiene distribución geométrica(p): $P(T_i = t_i) = (1-p)^{t_i - 1} p$. Concluimos que

$$P(T_1 = t_1, \ldots, T_k = t_k) = P(T_1 = t_1) \ldots P(T_k = t_k),$$

lo que prueba la independencia. □

Como corolario tenemos que

$$P(T_1 > t_1, \ldots, T_k > t_k) = P(T_1 > t_1) \ldots P(T_k > t_k)$$

Resumen de procesos

$(X_1, X_2, \ldots)$ Bernoulli(p) independientes

$(S_1, S_2, \ldots)$ donde $S_n \sim$ Binomial(n, p) con incrementos independientes y estacionarios.

$(T_1, T_2, \ldots)$, Geometricas(p) independientes.

$(Y_1, Y_2, \ldots)$, donde $Y_j \sim$ Binomial negativa (j, p).

Juego de Las Vegas Jugamos al rojo en las Vegas. Ruleta sin 0, es decir, la probabilidad de rojo en cada jugada es 1/2.

Martingala (un método infalible para ganar):
0) Fijo el lucro inicial $L = 0$
1) Fijo el monto a apostar $K = 1$.
2) Apuesto K.
3) Si sale rojo $L \leftarrow L + K$ y vuelvo a (1).
4) Si sale negro $L \leftarrow L - K$, $K \leftarrow 2K$ y vuelvo a (2).

Simulación: 1 = sale rojo, 0 = sale negro. X_i es un proceso de Bernoulli(1/2).

Apuesto	1	1	2	1	2	4	8	1	2	1	2	4	8
Sale X_i	1	0	1	0	0	0	1	0	1	0	0	0	1
Gano	1	-1	2	-1	-2	-4	8	-1	2	-1	-2	-4	8
Lucro L_i	1	0	2	1	-1	-5	3	2	4	3	1	-3	5

Se puede calcular el lucro medio si juego hasta el primer 1. El instante de aparición del primer 1 es la variable aleatoria $Y_1 = \min\{i \geq 1 : X_i = 1\} \sim$ geometrica(p). El lucro L_i en el instante i depende de $(X_1, \ldots, X_i)$. El lucro en el instante Y_1 es uno: $L_{T_1} = 1$. Dejamos como ejercicio demostrar que $L_{Y_j} = j$.

El problema es que la fortuna es finita o el casino tiene una apuesta máxima. Por ejemplo, no se permiten apuestas mayores que 2^8. En ese caso, si perdemos 9 veces seguidas perdemos $1+2^1+\cdots+2^8 = 2^9-1$ y no podemos duplicar la apuesta. En ese caso el lucro L hasta ganar 1 vez o perder 9 seguidas es

$$L = g(T_1) = 1\,\mathbf{1}\{T_1 \leq 9\} - (2^9-1)\,\mathbf{1}\{T_1 > 9\}$$

y su esperanza $EL = 1(1-2^{-9}) - (2^9-1)2^{-9} = 0$. El juego es honesto.

11.2. Proceso de Poisson

Un proceso puntual es un subconjunto discreto $S \subset \mathbb{R}^+$. Llamaremos llegadas a los puntos del proceso. Llamamos $N_t = \big|S \cap [0,t]\big|$, el número de llegadas de S en el intervalo $[0,t]$. Sea $\lambda > 0$ un parámetro.

Definición Decimos que $(N_t : t \in \mathbb{R}^+)$ es un *proceso de Poisson* de intensidad λ si su distribución satisface:

i) N_t tiene distribución Poisson(λt) para todo t.

ii) *Incrementos estacionarios.* El número de llegadas en un intervalo depende sólo del tamaño del intervalo: $N_{a+t} - N_a$ tiene la misma distribución que N_t, para todo $a > 0, t > 0$.

iii) *Incrementos independientes.* Las llegadas en intervalos disjuntos son independientes: Si $(s_i, t_i]$, $i = 1, \dots, k$, son intervalos disjuntos dos a dos, entonces las variables aleatorias $(N_{t_i} - N_{s_i})$, $i = 1, \dots, k$, son independientes.

Como los incrementos son estacionarios, el número de llegadas en cualquier intervalo $(s,t]$ tiene distribución Poisson$(\lambda(t-s))$.

Ejemplo Los mails llegan a una computadora de acuerdo a un proceso de Poisson de intensidad $\lambda = 2$ mensajes/minuto. Sea $N_t =$ número de mensajes entre 0 y t.

a) ¿Cuál es la probabilidad de que no se reciba ningún mensaje entre las 12 y las 12:03? $N_3 \sim$ Poisson$(2{\cdot}3)$ =Poisson(6). $P(N_3 = 0) = e^{-6} = 0{,}002$.

b) ¿Cuál es la probabilidad de que no se reciba ningún mensaje entre las 13:30 hs y las 13:33 hs? Misma respuesta que en (a).

Construcción usando exponenciales independientes

Sean $\tau_1, \tau_2, \ldots$ iid Exponencial(λ). Sea $Y_0 = 0$ y $Y_n := \tau_1 + \cdots + \tau_n$. Defina $N(s) := \max\{n : Y_n \leq s\}$.

τ_n tiempos entre llegadas a un banco.

$Y_n =$ instante de la n-ésima llegada. $Y_n \sim$ Gama(n, λ).

$N(t)$ es el número de llegadas hasta el instante t

$N(s) = n$ si y sólo si $Y_n \leq s < Y_{n+1}$.

Lema 55 *$N(s) \sim$ Poisson(λs) (como variable aleatoria).*

Demostración

$$\begin{aligned}
P(N(s) = n) &= P(Y_n \leq s < Y_{n+1}) \\
&= \int_0^s f_{Y_n}(t) P(Y_{n+1} > s \,|\, Y_n = t) dt \\
&= \int_0^s f_{Y_n}(t) P(\tau_{n+1} > s - t) dt \\
&= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda(s-t)} dt, \qquad \text{(porque } Y_n \sim \text{Gama}(n, \lambda)) \\
&= \frac{(\lambda t)^n}{(n-1)!} e^{-\lambda s} \int_0^s t^{n-1} dt = e^{-\lambda s} \frac{(\lambda s)^n}{n!}. \qquad \square
\end{aligned}$$

Lema 56 *$(N(t+s) - N(s),\, t \geq 0)$ es un proceso de Poisson de tasa λ independiente de $N(r),\ 0 \leq r \leq s$.*

Demostración Si llamamos $\tilde{\tau}_1, \tilde{\tau}_2, \ldots$ los tiempos entre llegadas después de s, condicionando a $\{N(s) = n, Y_n = u\}$ con $u \in [0, s)$, tendremos $\tilde{\tau}_1 = Y_{n+1} - s$ y su distribución se puede calcular así:

$$\begin{aligned}
P(\tilde{\tau}_1 > t | Y_n = u, N(s) = n) &= P(Y_{n+1} - s > t | Y_n = u, N(s) = n) \\
&= P(Y_{n+1} - s > t | Y_n = u, Y_{n+1} > s) \\
&= P(\tau_{n+1} > t + s - u | \tau_{n+1} > s - u, Y_n = u) \\
&= P(\tau_{n+1} > t + s - u | \tau_{n+1} > s - u) = e^{-\lambda t}
\end{aligned}$$

donde la segunda y tercera identidades son consecuencia de la identidad entre los eventos condicionantes; la tercera se deduce de la independencia entre Y_n y τ_{n+1} y la tercera es la falta de memoria de la exponencial. Como, por probabilidad total tenemos

$$\begin{aligned}
&P(\tilde{\tau}_1 > t) \\
&= \sum_{n \geq 0} \int_0^s P(Y_{n+1} - s > t | Y_n = u, N(s) = n) P(N(s) = n) f_{Y_n | N(s) = n}(u) du \\
&= e^{-\lambda t} \sum_{n \geq 0} \int_0^s P(N(s) = n) f_{Y_n | N(s) = n}(u) du = e^{-\lambda t}.
\end{aligned}$$

podemos concluir que $\tilde{\tau}_1 \sim$ Exponencial(λ).

Para los intervalos entre llegadas sucesivos se hace el mismo cálculo: condicionando a $\{Y_n = u, N(s) = n\}$ tenemos $\tilde{\tau}_1 = Y_{n+1} - s = \tau_{n+1} - (s - u)$ y $\tilde{\tau}_j = \tau_{n+j}$ para $j \geq 2$. Por ejemplo:

$$\begin{aligned} &P(\tilde{\tau}_2 > t_2, \tilde{\tau}_1 > t_1 \,|\, Y_n = u, N(s) = n) \\ &\quad = \cdots = P(\tau_{n+1} > t_1 + s - u, \tau_{n+2} > t_2 | \tau_{n+1} > s - u, Y_n = u) \\ &\quad = e^{-\lambda t_1} e^{-\lambda t_2}. \end{aligned}$$

Para ver que $\tilde{\tau}_1, \tilde{\tau}_2, \ldots$ es independiente de $(N(t), t \leq s)$, basta ver que el cálculo de la distribución de $\tilde{\tau}_j$ depende de $(N(t), t \leq s)$ solamente a traves de $Y_{N(s)}$ y vimos que en realidad es independiente de $Y_{N(s)}$. □

Lema 57 *El proceso de Poisson $N(t)$ tiene incrementos independientes: si $t_0 < t_1 < \ldots < t_n$ entonces*

$$N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1}) \textit{ son independientes.}$$

Demostración El Lema 56 dice que $N(t_n) - N(t_{n-1})$ es independiente de $N(r), r \leq t_{n-1}$ y por lo tanto de $N(t_1) - N(t_0), \ldots, N(t_{n-1}) - N(t_{n-2})$. Concluya usando inducción. □

Corolario 58 *El proceso $N(t)$ construido con las exponenciales es un proceso de Poisson.*

11.3. El Proceso Binomial aproxima al Proceso de Poisson

Fijamos un $\lambda > 0$ y consideramos una sucesión de procesos de Bernoulli, indexados por $\ell > 0$, $(X_n^\ell,\ n \in \mathbb{N})$, cada uno consiste en variables iid Bernoulli(λ/ℓ). Es decir, $P(X_n^\ell = 1) = p(\ell) = \lambda/\ell$.

Vamos a introducir una familia de procesos Binomiales indexados por ℓ donde los ensayos ocurren a cada $1/\ell$ instantes y la probabilidad de éxito en cada ensayo es λ/ℓ. Sea t real positivo y defina el proceso $(S_t^\ell, t \in \mathbb{R}^+)$ por

$$S_t^\ell := \sum_{n:(n/\ell) \leq t} X_n^\ell,$$

el número de éxitos hasta el instante t. Son $[\ell t]$ ensayos de Bernoulli independientes, cada uno con probabilidad λ/ℓ de éxito. S_t^ℓ es un proceso Binomial definido en la grilla $\mathbb{N}/\ell$. El número esperado de éxitos en el intervalo $[0, t]$ es

$$ES_t^\ell = \frac{\lambda}{\ell}[t\ell] = \lambda t + O(1/\ell).$$

Vimos en (44)-(45) que para cada ℓ, el proceso binomial S_t^ℓ tiene incrementos estacionarios e independientes.

Teorema 59 *Cuando $\ell \to \infty$, las distribuciones finito-dimensionales de los procesos binomiales $(S_t^\ell, t \in \mathbb{R}^+)$ convergen a las distribuciones finito-dimensionales del proceso de Poisson $(N_t, t \in \mathbb{R}^+)$.*

Dem Veremos que el límite satisface i-iii de la definición del Proceso de Poisson.

i) Las variables S_t^ℓ convergen en distribución a variables Poisson(λ),

$$\lim_\ell P(S_t^\ell = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} = P(N_t = k), \qquad k \in \{0, 1, \dots\} \tag{46}$$

por la aproximación de la Binomial a la Poisson.

ii-iii) Por los incrementos independientes del proceso binomial (45), tenemos que si $(s_i, t_i]$, $i = 1, \dots, k$, son intervalos temporales disjuntos dos a dos, entonces las variables aleatorias $(S_{t_i}^\ell - S_{s_i}^\ell)$, $i = 1, \dots, k$, son independientes:

$$P(S_{t_i}^\ell - S_{s_i}^\ell = h_i,\ i = 1, \dots, k) = \prod_{i=1}^k P(S_{t_i}^\ell - S_{s_i}^\ell = h_i) \tag{47}$$

Sacando límite en ℓ, usando los incrementos estacionarios (44) y la convergencia a Poisson (46) obtenemos

$$\lim_{\ell\to\infty} P(S_{t_i}^\ell - S_{s_i}^\ell = h_i,\ i = 1, \dots, k) = \prod_{i=1}^k \frac{e^{-\lambda(t_i - s_i)}(\lambda(t_i - s_i))^{h_i}}{h_i!} \tag{48}$$

$$= P(N_{t_i} - N_{s_i} = h_i,\ i = 1, \dots, k) \tag{49}$$

Es decir que el vector $(S_{t_i}^\ell - S_{s_i}^\ell,\ i = 1, \dots, k)$ converge en distribución al vector $(N_{t_i} - N_{s_i},\ i = 1, \dots, k)$. $\square$

Convergencia de los tiempos entre llegadas

Sea $Y_n^\ell := \text{mín}\{t > 0 : S_t^\ell = n\}$ el tiempo de la n-ésima llegada en S_t^ℓ. Como $\ell Y_1^\ell \sim$ Geometrica(λ/ℓ),

$$P(Y_1 > t) = \lim_{n\to\infty} P(\ell Y_1^\ell > \ell t) = \lim_{\ell\to\infty} \Big(1 - \frac{\lambda}{\ell}\Big)^{[\ell t]} = e^{-\lambda t}$$

Sea $T_i^\ell := Y_i^\ell - Y_{i-1}^\ell$ el tiempo entre la $(i-1)$-ésima llegada y la i-ésima llegada del proceso Binomial S_t^ℓ, con la convención $Y_0 = 0$.

Lema 60 *Los tiempos entre llegadas* $(T_i^\ell : i \geq 0)$ *del proceso binomial* S_t^ℓ *convergen en distribución a iid Exponencial*(λ).

Dem Si ℓt_i es entero (si no es entero cometemos un error de orden $1/\ell$ que se va a cero) vimos en el Lema 54 que

$$P(T_i^\ell > t_i, i = 1, \dots, k) = \prod_{i=1}^k (1 - \lambda/\ell)^{\ell t_i - 1}\lambda/\ell,$$

es decir que T_i^ℓ tienen la misma distribución que k variables independientes geométricas de parámetro λ/ℓ, divididas por ℓ. Usando que las probabilidades puntuales caracterizan las probabilidades acumuladas, obtenemos en particular la convergencia en distribución de los primeros k tiempos entre llegadas:

$$\begin{aligned} P(T_i > t_i, 1 \le i \le k) &= \lim_{\ell\to\infty} P(T_i^\ell > t_i, 1 \le i \le k) = \lim_{\ell\to\infty} \prod_{i=1}^k P(T_i^\ell > t_i) \\ &= \lim_{\ell\to\infty} \prod_{i=1}^k (1-\lambda/\ell)^{\ell t_i} = \prod_{i=1}^k e^{-\lambda t_i}. \quad \square \end{aligned}$$

Distribución de un número fijo de puntos Supongamos que hay una única llegada en $[0,t]$. Cual es la distribución del instante de esa llegada? Sea $s \in [0,t]$ y calcule

$$P\Big(\frac{T_1^\ell}{\ell} \le s \Big| S_t^\ell = 1\Big) = \frac{P(S_s^\ell = 1)P(S_t^\ell - S_s^\ell = 0)}{P(S_t^\ell = 1)} \to s$$

Sabiendo que hay k llegadas en el intervalo $[0,t]$, cual es la distribución de los instantes de llegada?

Teorema *En un proceso de Poisson*

$$P(\{Y_1,\dots,Y_k\} \in B | N_t = k) = P(\{U_1,\dots,U_k\} \in B)$$

donde $U_1,\dots,U_k$ son variables aleatorias independientes Uniforme$[0,t]$.

Demostración Hagamos el cálculo para $n = 3$. Estamos condicionando a $\{N(t) = 3\}$.

Basta demostrar que la densidad de (T_1, T_2, T_3) condicionada a $\{N(t) = 3\}$ es la misma que la de los estadísticos de orden de 3 Uniformes en $[0,t]$.

$$\begin{aligned} &\text{“}P(T_1 = t_1, T_2 = t_2, T_3 = t_3, \tau_4 > t - t_3 \,|\, N(t) = 3)\text{”} \\ &= \frac{\lambda e^{-\lambda t_1}\lambda e^{-\lambda(t_2-t_1)}\lambda e^{-\lambda(t_3-t_2)}e^{-\lambda(t-t_3)}}{e^{-\lambda t}(\lambda t)^3/3!}\mathbf{1}\{0 < t_1 < t_2 < t_3 < t\} \\ &= \frac{3!}{t^3}\,\mathbf{1}\{0 < t_1 < t_2 < t_3 < t\} \end{aligned}$$

O sea que la distribución de (T_1, T_2, T_3) condicionada a $N(t) = 3$ es uniforme en el conjunto $\{(t_1,t_2,t_3) : 0 < t_1 < t_2 < t_3 < t\}$. Ese conjunto tiene volumen $t^3/3!$ porque $\{(t_1,t_2,t_3) : 0 < t_1, t_2, t_3 < t\}$ tiene volumen t^3 y ese es uno de los 3! posibles órdenes entre las coordenadas.

El mismo razonamiento sirve para demostrar que la densidad de $(T_1,\dots,T_n)$ dado $N(t) = n$ está dada por $n!/t^n$. $\square$

Es decir que si $N(t) = n$, la posición de los puntos es la misma que la de n uniformes en $[0,t]$.

11.4. Construcciones del proceso de Poisson

1. Exponenciales independientes

Sean $T_1, T_2, \dots$ variables aleatorias exponenciales independientes de parametro λ.

Lema 61 $N_t := \text{máx}\{n : T_1 + \cdots + T_n \leq t\}$ *es un proceso de Poisson de parametro* λ.

2. Número Poisson, distribución uniforme Fije $T > 0$. Elija un número aleatorio de puntos con distribución $N_T \sim \text{Poisson}(\lambda T)$. Distribuya ese número de puntos como iid uniformemente distribuídos en el intervalo $(0, T)$. Llámelos $U_1, \ldots, U_{N_T}$. Defina:

$$N_t := \sum_{i=1}^{N_T} \mathbf{1}\{i : U_i \leq t\}$$

el número de puntos que cayeron a la izquierda de t.

Lema 62 *El proceso* $(N_t : 0 \leq t \leq T)$ *así construído es un proceso de Poisson en* $[0, T]$.

3. Extensión a dimensiones mayores

Sea $l(B)$ la medida de Lebesgue del Boreliano B.

Un proceso de Poisson de intensidad λ en $\mathbb{R}^d$ es un subconjunto aleatorio S de $\mathbb{R}^d$ cuya distribución satisface

1. $|S \cap B| \sim \text{Poisson}(\lambda l(B))$, si B es un Boreliano.

2. $|S \cap B_1|, \ldots, |S \cap B_n|$ son independientes si los B_i son Borelianos disjuntos dos a dos.

La construcción 2 se puede extender a $\mathbb{R}^d$. Sea $\lambda > 0$ y considere

• Una partición $\mathcal{J}$ de $\mathbb{R}^d$ cuyos elementos son Borelianos acotados.

• Una familia $(Y_A : A \in \mathcal{J})$, donde Y_A son iid con $Y_A \sim \text{Poisson}(\lambda l(A))$.

• Una familia de sucesiones independientes $((U_{A,j}, j \geq 1), A \in \mathcal{J})$, donde $(U_{A,j} : j \geq 1)$ son iid con distribución uniforme en A.

• Defina el proceso de Poisson como el conjunto aleatorio dado por

$$S \stackrel{def}{=} \bigcup_{A \in \mathcal{J}} \bigcup_{j \leq Y_A} \{U_{A,j}\} = \bigcup_{A \in \mathcal{J}} \{U_{A,j} : j \leq Y_A\} \tag{50}$$

El objeto aleatorio así definido se llama *Proceso de Poisson* de *intensidad* λ.

12. Cadenas de Markov

12.1. Definición

Vamos a considerar procesos estocásticos a tiempo discreto $X_1, X_2, \ldots$, asumiendo valores en un conjunto S finito o numerable llamado *espacio de estados*. El sub-índice se interpreta como tiempo. Si $X_n = x$, diremos que el proceso se encuentra en el *estado* x en el *instante* n.

En una **cadena de Markov** cuando el proceso está en el estado x en el instante n, tiene probabilidad $Q(x, y)$ de ir al estado y en el instante $n + 1$, independientemente de la trayectoria que hizo para llegar a x:

$$P(X_{n+1} = y | X_n = x,\, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = Q(x, y).$$

Los valores $Q(x,y)$ son llamados *probabilidades de transición* y conforman una **matriz de transición** $Q = (Q(x,y) : x, y \in S)$. Esta matriz tiene entradas no negativas y la suma de cada fila vale 1:

$$\sum_{y \in S} Q(x,y) = 1$$

La matriz de transición es un conjunto de parámetros que caracterizan la cadena de Markov.

Cadena de Markov con dos estados. Si hoy llueve, la probabilidad que llueva mañana es 0,8 y si hoy no llueve, esta probabilidad es 0,1. El espacio de estados es $S = \{0, 1\}$. Si interpretamos 1 cuando llueve y 0 cuando no llueve, la matriz de transición es

$$Q = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}2 & 0{,}8 \end{pmatrix} \tag{51}$$

$Q(0,0) = 0{,}1$, $Q(0,1) = 0{,}9$, $Q(1,0) = 0{,}2$, $Q(1,1) = 0{,}8$.

12.2. Construcción

Sea $U_1, U_2, \dots$ una sucesión de variables Uniformes$[0,1]$ independientes. Defina $X_0 = x \in \{0,1\}$ e, iterativamente,

$$X_{n+1} = F(X_n, U_{n+1}) \tag{52}$$

donde $F(0,u) = \mathbf{1}\{u > 0{,}9\}$ y $F(1,u) = \mathbf{1}\{u > 0{,}2\}$.

Verifique que el proceso así obtenido es una cadena de Markov con matriz de transición (51).

Definición constructiva de cadenas de Markov

Sea Q una matriz de transición en un espacio de estados S. Para cada $x \in S$ definimos una partición $\mathcal{J}_x = (J(x,y), y \in S)$ del intervalo $[0,1]$, de tal manera que

$$|J(x,y)| = Q(x,y)$$

Defina $F : S \times [0,1] \to S$ por

$$F(x,u) = \sum_{y \in S} y\, \mathbf{1}\{u \in J(x,y)\}$$

Fije un estado x y defina un proceso estocástico X_n, $n \geq 0$ por $X_0 = x$ e iterativamente,

$$X_{n+1} = F(X_n, U_{n+1}) \tag{53}$$

El proceso así definido es Markov con matriz de transición Q. En efecto,

$$\begin{aligned} &P(X_{n+1} = y | X_n = x,\, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &\quad = P(F(x, U_{n+1}) = y | F(x_{n-1}, U_n) = x, \dots, F(x_0, U_1) = x_1, X_0 = x_0)) \end{aligned}$$

Como los U_k son independientes, esa expresión es igual a

$$= P(F(x, U_{n+1}) = y) = P(U_{n+1} \in J(x,y)) = |J(x,y)| = Q(x,y).$$

12.3. Matriz de transición y distribución en el instante n

La matriz de transición sirve para calcular las probabilidades de transición a más de un paso:

$$P(X_n = y|X_0 = x) = Q^n(x, y) \tag{54}$$

Probemos esto para $n = 2$:

$$\begin{aligned} P(X_2 = y|X_0 = x) &= \sum_z P(X_2 = y, X_1 = z|X_0 = x) \\ &= \sum_z P(X_2 = y|X_1 = z, X_0 = x)P(X_1 = z|X_0 = x) \end{aligned}$$

(por las propiedades de probabilidad condicional)

$$= \sum_z P(X_2 = y|X_1 = z)P(X_1 = z|X_0 = x)$$

(por la propiedad de Markov)

$$= \sum_z Q(x, z)Q(z, y) = Q^2(x, y)$$

Para n general procedemos por inducción. Asuma (54) y calcule

$$P(X_{n+1} = y|X_0 = x) = \sum_z P(X_{n+1} = y, X_n = z|X_0 = x)$$

que por el mismo cálculo que antes es igual a

$$= \sum_z Q^n(x, z)Q(z, y) = Q^{n+1}(x, y)$$

Ecuaciones de Chapman-Kolmogorov

$$Q^{n+m}(x, y) = \sum_z Q^n(x, z)Q^m(z, y), \qquad 0 \le k \le n.$$

Demostración Por la fórmula (54),

$$Q^{n+m}(x, y) = P(X_{n+m} = y|X_0 = x) \tag{55}$$

Usando la partición $\{\{X_n = z\}, z\}$, por probabilidad total,

$$\begin{aligned} &= \sum_z P(X_{n+m} = y, X_n = z|X_0 = x) \\ &= \sum_z P(X_{n+m} = y|X_n = z, X_0 = x)P(X_n = z|X_0 = x) \end{aligned}$$

donde de nuevo usamos propiedades de probabilidad condicional,

$$= \sum_z P(X_{n+m} = y|X_n = z)P(X_n = z|X_0 = x)$$

donde usamos la propiedad de Markov. Pero usando la fórmula (54), eso es igual a

$$= \sum_z Q^n(x,z)Q^m(z,y). \qquad \square$$

Urna de Ehrenfest [4] Considere N bolillas distribuídas en dos urnas. Una bolilla es elegida al azar y es cambiada de urna. Este modelo representa el comportamiento de un gas que tiene N moléculas ocupando dos containers. Cual es la cadena de Markov que describe esta evolución temporal?

El espacio de estados es $S = \{0, 1, \dots, N\}$ que describe el número de bolillas en la primera urna. Si en un momento hay k bolillas en la primera urna, las transiciones posibles son para $k-1$ (si $k>0$) o para $k+1$ (si $k<N$) y las probabilidades de transición son

$$Q(k,k-1) = \frac{k}{N}, \qquad Q(k,k+1) = \frac{N-k}{N}$$

$$Q(x,y) = 0, \qquad \text{si } |x-y| > 1.$$

Si la primera urna tiene 4 bolillas y la segunda tiene 6, cual es la probabilidad que después de dos pasos haya 4 bolillas en la primera y 6 en la segunda?

$$Q^2(4,4) = Q(4,5)Q(5,4) + Q(4,3)Q(3,4) = \frac{6\times 5 + 4\times 7}{100}$$

Y cual es la probabilidad que despues de tres pasos haya 5 bolillas en cada urna? Hay que calcular $Q^3(4,5)$ que es igual a

$$Q(4,5)Q(5,6)Q(6,5) + Q(4,5)Q(5,4)Q(4,5) + Q(4,3)Q(3,4)Q(4,5)$$

12.4. Medidas invariantes

Considere una cadena de Markov en un espacio finito con matriz de transición Q para la cual existe un $k>0$ tal que $Q^k(x,y)>0$ para todo par de estados x,y. Suponga que existen los límites siguientes

$$\lim_n Q^n(x,y) =: \pi(y), \qquad \text{para todo par de estados } x,y$$

Si esto ocurre, decimos que la cadena olvida el valor inicial y que la distribución de X_n converge a π (convergencia en distribución) para cualquier estado inicial. Si escribimos

$$Q^{n+1}(x,y) = \sum_z Q^n(x,z)Q(z,y),$$

sacando límite en ambos miembros,

$$\pi(y) = \sum_z \pi(z)Q(z,y) \qquad \text{para todo } y$$

Estas son las **ecuaciones de balance**. Además, como $\sum_y Q^n(z,y) = 1$, para todo n, z y la suma es finita, tendremos que π es una probabilidad:

$$\sum_y \pi(y) = 1.$$

Es decir que la probabilidad π es un autovector a la izquierda de Q con autovalor 1: $\pi Q = \pi$. Una probabilidad π que satisface las ecuaciones de balance es llamada **medida invariante**. En particular,

$$\sum_x \pi(x) P(X_1 = y | X_0 = x) = \pi(y).$$

Es decir que si el estado inicial es aleatorio y con distribución π, entonces la distribución de la cadena en el instante 1 es también π. En general, para todo n, $\pi Q^n = \pi$:

$$\sum_x \pi(x) P(X_n = y | X_0 = x) = \pi(y)$$

O sea: si la distribución de X_0 es π, entonces la distribución de X_n es π para todo $n \geq 0$.

Ejemplo: lluvia. Las ecuaciones de balance son

$$\begin{aligned} \pi(0) &= 0{,}9\,\pi(0) + 0{,}2\,\pi(1), \\ \pi(1) &= 0{,}1\,\pi(0) + 0{,}8\,\pi(1) \\ \pi(0) + \pi(1) &= 1. \end{aligned}$$

Substituyendo las identidades $\pi(0) = \pi(0)(0{,}1 + 0{,}9)$ y $\pi(1) = \pi(1)(0{,}2 + 0{,}8)$ en los primeros términos, obtenemos que las ecuaciones de balance son equivalentes a

$$0{,}1\,\pi(0) = 0{,}2\,\pi(1); \qquad \pi(0) + \pi(1) = 1$$

cuya solución es

$$\pi(0) = \frac{0{,}2}{0{,}2 + 0{,}1} = \frac{2}{3}, \quad \pi(1) = \frac{0{,}1}{0{,}2 + 0{,}2} = \frac{1}{3}$$

Ejemplo: urna de Ehrenfest. Las ecuaciones de balance para $0 < k < N$ son:

$$\pi(k) = \pi(k+1)\frac{k+1}{N} + \pi(k-1)\frac{N-k+1}{N}, \quad 0 < k < N;$$

cuya solución es:

$$\pi(k) = \binom{N}{k} \big(\frac{1}{2}\big)^N$$

Teorema de existencia y unicidad de la medida invariante *Si el espacio de estados es finito y alguna potencia k de la matriz de transición tiene todas las entradas positivas, entonces la medida invariante existe y es única.*

Este teorema es un caso particular del teorema de Perron-Frobenius del Álgebra. Daremos una demostración probabilística basada en la ley de grandes números.

12.5. Ley de grandes números para cadenas de Markov

Medidas empíricas

Fijamos el estado inicial $X_0 = x$ y definimos

$$N_n(x, y) := \sum_{j=1}^{n} \mathbf{1}\{X_j = y\},$$

el número de visitas a y de la cadena empezando con $X_0 = x$ y la *distribución empirica*

$$\hat{Q}_n(x, y) := \frac{N_n(x, y)}{n}$$

Son variables aleatorias que indican la proporción de visitas al estado y hasta el instante n para la cadena que empieza en x. Para cada x y n fijos $\hat{Q}_n(x, \cdot)$ es una probabilidad aleatoria: $\sum_y \hat{Q}_n(x, y) = 1$. Calculemos la esperanza de $\hat{Q}$:

$$E\hat{Q}_n(x, y) = \frac{1}{n} \sum_{j=1}^{n} E(\mathbf{1}\{X_j = y\}|X_0 = x) = \frac{1}{n} \sum_{j=1}^{n} Q^j(x, y) \tag{56}$$

Defina el instante de la primera vuelta a t por

$$\tau(y) := \text{mín}\{n \geq 1 : X_n^y = y\}$$

donde X_n^y es la cadena con estado inicial $X_0^y = y$.

Teorema 63 *Sea X_n una cadena de Markov en un espacio de estados finito con matriz de transición Q. Asuma que existe $k > 0$ tal que $Q^k(x, y) > 0$ para todo par de estados x, y. Entonces para cada y,*

$$\text{lím}_n \hat{Q}_n(x, y) = \frac{1}{E\tau(y)}, \qquad c.s. \tag{57}$$

que no depende de x. Definiendo $\pi(y) := 1/E\tau(y)$, tenemos que π es la única medida invariante para la cadena.

Dem

Defina el instante de la j-ésima visita a y por $T_j(x, y) = \text{mín}\{k \geq 1 : N_k(x, y) = j\}$. Así,

$$N_n(x, y) = j \quad \text{si y sólo si} \quad T_j(x, y) \leq n < T_{j+1}(x, y). \tag{58}$$

$$T_j(x, y) = T_1(x, y) + \tau_2(y) + \cdots + \tau_j(y) \tag{59}$$

donde $\tau_i(y)$ son iid con la misma distribución que $\tau(y) = T_1(y, y)$.

Como por hipótesis, a cada k pasos la cadena tiene por lo menos probabilidad $p := \text{mín}_{x,y} Q^k(x, y) > 0$ de visitar y,

$$P(T_1(x, y) > kn) \leq (1 - p)^n. \tag{60}$$

Por lo tanto, $\tau_i(y)$ tienen media y varianza finitas. Por la ley fuerte de grandes números tenemos

$$\text{lím}_{j\to\infty} \frac{T_j(x, y)}{j} = E\tau(y), \qquad \text{c.s.} \tag{61}$$

Usando (58) cuando $N_n = j$, tenemos

$$\frac{T_j(x,y)}{j} \leq \frac{n}{N_n(x,y)} \leq \frac{T_{j+1}(x,y)}{j}. \tag{62}$$

Por otro lado $N_n(x,y) \to \infty$, caso contrario habría un $\tau_j(y) = \infty$, pero esto no puede ser por (60). Por lo tanto, sacando límites en (62), y recordando que $\pi(y) = 1/E\tau(y)$,

$$\pi(y) \leq \liminf_n \frac{N_n(x,y)}{n} \leq \limsup_n \frac{N_n(x,y)}{n} \leq \pi(y).$$

Esto demuestra (57).

Por el teorema de convergencia acotada, si el límite c.s. existe, entonces también existe el límite de las esperanzas:

$$\lim_n E\hat{Q}_n(x,y) = \pi(y) \tag{63}$$

y usando

$$E\hat{Q}_{n+1}(x,y) = \sum_z E\hat{Q}_n(x,z)Q(z,y) + O(1/n), \tag{64}$$

que demostraremos más abajo, tenemos que el límite π satisface las ecuaciones de balance.

Unicidad. Supongamos que exista otra medida invariante π' que satisface las ecuaciones de balance $\pi' Q = \pi'$. Entonces,

$$\pi'(y) = \sum_x \pi'(x) \frac{1}{n} \sum_{k=1}^n Q^k(x,y) \to_n \sum_x \pi'(x)\pi(y) = \pi(y),$$

usando (56) y (63). □

Demostracion de (64)

$$\begin{aligned} E\hat{Q}_{n+1}(x,y) &= \frac{1}{n+1} \sum_{j=1}^n Q^{j+1}(x,y) + \frac{1}{n+1} Q(x,y) \\ &= \frac{n}{n+1} \frac{1}{n} \sum_{j=1}^n \sum_z Q^j(x,z)Q(z,y) + \frac{1}{n+1} Q(x,y) \end{aligned}$$

De donde

$$E\hat{Q}_{n+1}(x,y) = \sum_z E\hat{Q}_n(x,z)Q(z,y) + O(1/n)$$

Uso de simulaciones para estimar π Una forma de estimar π es simular la cadena de Markov por un intervalo de tiempo de tamaño n "grande" y usar las distribuciones empíricas

$$\hat{Q}_n(x,y) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}\{X_k = y\}$$

como aproximación de $\pi(y)$. Hay una teoría que permite calcular los errores al hacer esta aproximación.

12.6. Reversibilidad y balance detallado

Decimos que π satisface la condición de *balance detallado* si

$$\pi(x)Q(x,y) = \pi(y)Q(y,x)$$

Si π satisface balance detallado, decimos que es *reversible*.

Si sumamos en x obtenemos que π satisface las ecuaciones de balance:

$$\sum_x \pi(x)Q(x,y) = \pi(y) \sum_x Q(y,x) = \pi(y).$$

O sea que una distribución reversible es invariante para Q

Paseos aleatorios en grafos. Un grafo $G = (V, E)$, $V =$ vértices $E =$ aristas, subconjunto de pares de vertices. $A(u, v)$ indica si hay una arista conectando u y v, si la hay decimos que u y v son *vecinos*. Grafo no orientado. *Grado* $d(u)$ del vertice u es el número de vecinos:

$$d(v) = \sum_v A(u,v).$$

Definimos

$$p(u,v) = \frac{A(u,v)}{d(u)}$$

es una matriz de transición. Para cada constante c, $\pi(u) := cd(u)$ satisface balance detallado:

$$\pi(u)p(u,v) = cA(u,v) = cA(v,u) = \pi(v)p(v,u).$$

Para grafos con un número finito de vértices, si tomamos $c = 1/(\text{suma de los grados})$, obtenemos una distribución estacionaria.

Paseo del caballo de ajedrez. Las posibles movidas de un caballo de ajedrez inducen un grafo cuyos vértices son las 64 casillas del tablero de ajedrez y las aristas conectan casillas que se pueden alcanzar entre sí por saltos del caballo. Los vértices tienen grado entre 2 (en los rincones) hasta 8 (en el centro):

$$\begin{matrix} 2 & 3 & 4 & 4 & 4 & 4 & 3 & 2 \\ 3 & 4 & 6 & 6 & 6 & 6 & 4 & 3 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 3 & 4 & 6 & 6 & 6 & 6 & 4 & 3 \\ 2 & 3 & 4 & 4 & 4 & 4 & 3 & 2 \end{matrix}$$

La suma de los grados da 336. La distribución estacionaria es $\pi(x) = d(x)/336$.

El reverso temporal de una trayectoria estacionaria es Markov Sea $p(x, y)$ una matriz de transión con una distribución estacionaria $\pi(x)$. Sea $(X_0, X_1, \dots)$ una trayectoria de la cadena de Markov con distribución inicial π, es decir $X_0 \sim \pi$.

Teorema 64 *Fije n. Considere el proceso $Y_m = X_{n-m}$, para $0 \leq m \leq n$. Entonces Y_m es Markov con matriz de transición*

$$p^*(x,y) = P(Y_{m+1} = y | Y_m = x) = \frac{\pi(y)p(y,x)}{\pi(x)}$$

Además π es una distribución estacionaria para la matriz p^*.

Demostración

$$\begin{aligned}
&P(Y_{m+1} = y_{m+1} | Y_m = y_m, Y_{m-1} = y_{m-1}, \ldots, Y_0 = y_0) \\
&= \frac{P(Y_{m+1} = y_{m+1}, Y_m = y_m, Y_{m-1} = y_{m-1}, \ldots, Y_0 = y_0)}{P(Y_m = y_m, Y_{m-1} = y_{m-1}, \ldots, Y_0 = y_0)} \\
&= \frac{P(X_{n-(m+1)} = y_{m+1}, X_{n-m} = y_m, X_{n-(m-1)} = y_{m-1}, \ldots, X_n = y_0)}{P(X_{n-m} = y_m, X_{n-(m-1)} = y_{m-1}, \ldots, X_n = y_0)} \\
&= \frac{\pi(y_{m+1})p(y_{m+1}, y_m)P(X_{n-(m-1)} = y_{m-1}, \ldots, X_n = y_0 | X_{n-m} = y_m)}{\pi(y_m)P(X_{n-(m-1)} = y_{m-1}, \ldots, X_n = y_0 | X_{n-m} = y_m)} \\
&= \frac{\pi(y_{m+1})p(y_{m+1}, y_m)}{\pi(y_m)} = p^*(y_m, y_{m+1}).
\end{aligned}$$

Note que p^* es una matriz de transición:

$$\sum_y p^*(x,y) = \sum_y \frac{\pi(y)}{\pi(x)} p(y,x) = \frac{\pi(x)}{\pi(x)} = 1.$$

porque π es estacionaria para p. Finalmente,

$$\sum_x \pi(x)p^*(x,y) = \sum_x \pi(x)\frac{\pi(y)}{\pi(x)} p(y,x) = \pi(y),$$

lo que demuestra que π es estacionaria para p^*. $\square$

Cuando π satisface las ecuaciones de balance detallado:

$$p^*(x,y) = \frac{\pi(y)}{\pi(x)} p(y,x) = p(x,y).$$

La película yendo para atrás o para adelante tiene la misma distribución.

Algoritmo Metrópolis-Hastings El objetivo es generar muestras de una distribución π. Para eso vamos a construir una cadena de Markov cuya distribución estacionaria es π. Empezamos con una matriz de transición arbitraria $q(x,y)$ que será usada para *proponer* un salto. El salto será aceptado con probabilidad

$$r(x,y) = \min\left\{ \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1 \right\}$$

Definimos así la matriz de transición

$$p(x,y) = q(x,y)r(x,y).$$

Lema 65 *La distribución π es reversible para p.*

Demostración Supongamos que $\pi(y)q(y,x) > \pi(x)q(x,y)$. En este caso,

$$\begin{aligned}\pi(x)p(x,y) &= \pi(x)q(x,y)1\\ \pi(y)p(y,x) &= \pi(y)q(y,x)\frac{\pi(x)q(x,y)}{\pi(y)q(y,x)} = \pi(x)q(x,y).\end{aligned}$$

Es decir que se satisfacen las ecuaciones de balance detallado. □

Usando los teoremas de convergencia se pueden obtener muestras aproximadas de π, o las esperanzas en relación a π de funciones objetivo. Por ejemplo:

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to_n \sum_x f(x)\pi(x).$$

Distribución geométrica Supongamos que queremos generar muestras de la distribución geométrica $\pi(x) = \theta^x(1-\theta)$, $x = 0, 1, 2, \dots$, con $\theta < 1$. Elegimos q como el paseo aleatorio simétrico $q(0,0) = q(x, x+1) = q(x+1, x) = 1/2$, $x \geq 0$. Como q es simétrica, $r(x,y) = \text{mín}\{\pi(y)/\pi(x), 1\}$. Como $\pi(x) > \pi(x+1)$, tenemos para $x \geq 1$:

$$p(x, x-1) = \frac{1}{2} \quad p(x, x+1) = \frac{\theta}{2} \quad p(x,x) = \frac{1-\theta}{2}.$$

Cuando $x = 0$, $\pi(-1) = 0$, así que

$$p(0,-1) = 0 \quad p(0,1) = \frac{\theta}{2} \quad p(0,0) = 1 - \frac{\theta}{2}.$$

Balance detallado:

$$\pi(x)p(x, x+1) = \theta^x(1-\theta)\frac{\theta}{2} = \pi(x+1)p(x+1, x).$$

Binomial Consideremos ahora Binomial(N, θ). Es decir $\pi(x) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$, $x \in \{0, \dots, N\}$.

$$q(x,y) = \frac{1}{N+1}, \quad y \in \{0, \dots, N\}, \quad \text{uniforme}$$

Como q es simétrica r usa el cociente de las probabilidades:

$$r(x,y) = \text{mín}\Big\{1, \frac{\pi(y)}{\pi(x)}\Big\}$$

12.7. Aplicación. Algoritmo PageRank

Recomiendo leer el artículo de Wikipedia *PageRank*.

Consideramos un grafo orientado $\mathcal{G} = (V, E)$, donde $V =$ es un conjunto finito de vertices y $E \subset \{(x,y) : x, y \in V)$ es un conjunto de aristas orientadas. Vértices representan páginas web. Una arista orientada (x,y) indica que en la página x hay un link a la página y.

Queremos ranquear los vértices usando solamente la estructura del grafo. Una primera idea es usar el número de aristas que llegan a un vértice $y \in V$ y proponer el ranqueador

$$R_1(y) = \sum_{x \in V} a(x, y)$$

donde $a(x, y) = \mathbf{1}\{(x, y) \in E\}$ vale 1 cuando hay una arista que va de x a y. Pero esto le da mucho peso a los vértices que emanan muchas aristas. Para compensar, definimos el número de aristas que salen del vértice x por

$$a(x) = \sum_{y} a(x, y)$$

y dividiendo por este número obtenemos el segundo ranqueador:

$$R_2(y) = \sum_{x \in V} \frac{a(x, y)}{a(x)}$$

pero así todos los vértices que tienen el mismo número de aristas salientes envían el mismo peso, independientemente de las aristas entrantes. Mejor sería que cada vértice enviara un peso proporcional a su importancia, medida por las aristas que entran. Esto sugiere un tercer ranqueador:

$$R_3(y) = \sum_{x \in V} R_3(x) \frac{a(x, y)}{a(x)}$$

O sea que R_3 es la solución de un sistema de $|V|$ ecuaciones, una para cada vértice del grafo. Usando la notación

$$\pi = R_3, \qquad Q(x, y) = \frac{a(x, y)}{a(x)},$$

el tercer ranqueador satisface $\pi(y) = \sum_{x \in V} \pi(x)\, Q(x, y)$, las ecuaciones de balance para una cadena de Markov que se describe así:

"Cuando la cadena se encuentra en el vértice x, elige al azar, uniformemente, una de las flechas que salen de x y salta al extremo y de esa flecha"

Esta cadena no satisface la condición $Q^k > 0$ para algún k porque hay muchas páginas que no tienen links salientes. Por eso se propone una nueva matriz

$$P := dQ + (1 - d)\frac{1}{|V|}\mathbb{I}$$

donde $d \in [0, 1]$ y $\mathbb{I}$ es una matriz de la misma dimensión de Q cuyas entradas son todas 1. Esta cadena opera así: "cuando está en el estado x, con probabilidad d elige una flecha de las salientes con la matriz Q, con probabilidad $(1 - d)$ elige un vértice al azar en toda la web". La idea es que cada navegante al cabo de un tiempo deja de clicar links y se va o elige una página al azar en la red. El número de clicks que hace antes de saltar a un link aleatorio en la red es una variable aleatoria geométrica con parámetro $(1-d)$. Si $d = 0{,}85$, el número medio de clicks antes de aburrirse será $1/0{,}15 = 6{,}66$. El parámetro d se llama damping factor, o factor de amortiguamiento. El ranqueador será la solución de

$$R_4(y) = \sum_{x \in V} R_4(x)\Big(d\frac{a(x, y)}{a(x)} + (1 - d)\frac{1}{|V|}\Big) \tag{65}$$

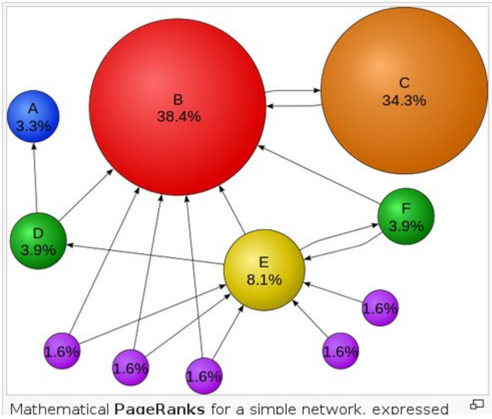


Figura 7: Medida invariante correspondiente a la matriz P con $d = 0{,}85$. Fuente PageRank de Wikipedia. Mathematical PageRanks for a simple network, expressed as percentages. (Google uses a logarithmic scale.) Page C has a higher PageRank than Page E, even though there are fewer links to C; the one link to C comes from an important page and hence is of high value. If web surfers who start on a random page have an 85 % likelihood of choosing a random link from the page they are currently visiting, and a 15 % likelihood of jumping to a page chosen at random from the entire web, they will reach Page E 8.1 % of the time. (The 15 % likelihood of jumping to an arbitrary page corresponds to a damping factor of 85 %.) Without damping, all web surfers would eventually end up on Pages A, B, or C, and all other pages would have PageRank zero. In the presence of damping, Page A effectively links to all pages in the web, even though it has no outgoing links of its own.

Vemos en http://news.netcraft.com/archives/category/web-server-survey/, que en octubre de 2015 hay más de 800 millones de páginas en la web. Con ese número de estados, el cálculo de la medida invariante en forma exacta es fisicamente imposible por el momento. La matriz está sea casi toda constituída de ceros porque cada página tiene links a unas pocas decenas o centenas de otras páginas. Cada fila de la matriz tiene tamaño 800 millones pero sólo unas pocas entradas son positivas.

Para estimar la medida invariante π para la matriz P se usa la ley de grandes números para cadenas de Markov, Teorema 63, que podemos hacer porque $P^1 \geq d\,\mathbb{I}$. Se envía un robot que circula por los vértices de acuerdo a una cadena de Markov X_k con matriz de transición P por n pasos y se estima $\pi(y)$ con la medida empírica temporal

$$\hat{P}_n(x, y) := \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{X_k = y\}$$

o simplemente se usa $\hat{P}_n$ como ranqueador.

Otra posibilidad, que no discutiremos aquí, es estimar la medida invariante π para P por una fila cualquiera de P^n para n relativamente grande. Parece que $n = 50$ funciona bastante bien.

13. Estadística

Noticias de los diarios Una semana antes del balotaje del 22 de noviembre de 2015.

1. Un estudio de Poliarquía Consultores para La Nación pronostica que, si la segunda vuelta entre Mauricio Macri y Daniel Scioli fuera hoy, el primero se impondría con más del 48 % de los votos.

Mientras que Macri tiene una intención de voto de 48,7 %, el candidato del Frente para la Victoria alcanzaría el 40,2 %. El margen de error de la encuesta es de 3,5 puntos con un nivel de confianza de 95 %.

Hay que destacar que este sondeo se hizo por teléfono a 800 personas en centros urbanos de más de diez mil habitantes y, aun sobre esa población, el número de indecisos es importante. El 6,4 % de las personas encuestadas dijo no saber todavía a quién votará el domingo 22. Asimismo el voto en blanco se elevaría al 4,7 %.

2. Según una encuesta de Hugo Haime y asociados realizada entre el 4 y el 7 de noviembre Mauricio Macri aventaja a Scioli por sólo 3,8 puntos. A once días del balotaje, el candidato de Cambiemos cuenta con un 44,2 por ciento, mientras que el del Frente para la Victoria alcanza el 40,4 por ciento.

3. El candidato de Cambiemos tiene una intención de voto de 51,8 % y el del Frente para la Victoria de 43,6 %, según una encuesta de la consultora Management & Fit exclusiva para Clarín, realizada en todo el país entre el 1 y el 5 de noviembre. A su vez, un 4,5 % asegura que impugnará el sufragio o votará en blanco. En el estudio, que muestra una distancia de 8,2 puntos entre los dos adversarios, se realizó una proyección del voto del 10,9 % de los encuestados que aún se muestra como indeciso respecto a lo que hará el 22 de noviembre.

Comentario. Los resultados del Balotaje fueron Macri 51,4 %, Scioli 48,6 %, con 2,8 puntos porcentuales de diferencia. Estos porcentajes suman 100, es decir que se sacaron respecto de los llamados votos "afirmativos", excluyendo los votos en blanco y nulos. Para compararlos con los anteriores habría que calcular los porcentajes de los votos afirmativos de las encuestadoras.

13.1. Estimación puntual

Esta Sección está basada en las notas de Ana Bianco y Elena Martinez [2]. Para obtener una estimación de la proporción de p de votantes por un candidato antes de una elección se realiza una encuesta. La encuesta consiste en tomar una muestra de electores (aleatoria en el sentido que cada posible elector tiene la misma probabilidad de entrar en la muestra) y estimar p por la proporción muestral $\hat{p}$.

Ese procedimiento se basa en un **modelo**: se considera una variable aleatoria X Bernoulli con **parámetro** p y una *muestra aleatoria* $\underline{X} = (X_1, \dots, X_n)$, conformada por n variables aleatorias independientes con la misma distribución de X. En este caso $X_i = 1$ si el i-ésimo elector de la muestra vota por el candidato, caso contrario es 0.

La **proporción muestral** es la variable aleatoria

$$\hat{p}_n(\underline{X}) := \frac{X_1 + \cdots + X_n}{n}$$

que será llamado *estimador* de p. Tiene sentido porque cuando n crece $\hat{p}_n(\underline{X}) \to p$ c.s. por la ley de grandes números.

Después de realizada la muestra tenemos valores $\underline{x} = (x_1, \dots, x_n)$ y una *estimativa* $\hat{p}(\underline{x})$.

El **error** cometido al estimar p por $\hat{p}_n(\underline{X})$ es

$$|\hat{p}_n(\underline{X}) - p|$$

que también es aleatorio. Conociendo la distribución del error, podremos hacer afirmaciones sobre la bondad de nuestra estimación.

Parámetros Así como la Bernoulli depende del *parámetro* p, otras distribuciones de probabilidad dependen de cierto número de parámetros. Por ejemplo: Poisson depende de λ, Normal depende de μ y σ^2, Binomial depende de n y p, etc. Llamaremos Θ el espacio de parámetros y $\theta \in \Theta$ un parámetro, que puede ser un vector, como en el caso de la Normal $\theta = (\mu, \sigma^2)$.

Muestras Denotamos $\underline{X} := (X_1, \ldots, X_n)$ una muestra aleatoria de una variable aleatoria X. Cualquier función de la muestra es una variable aleatoria. Por ejemplo: $\bar{X}_n$, $\max(X_1, \ldots, X_n)$, etc.

Los valores observados $(x_1, \ldots, x_n)$ serán denotados con minúsculas.

Estimación puntual paramétrica

Sea $X = X_\theta, \theta \in \Theta$ una familia de variables aleatorias con distribución $X_\theta \sim F_\theta$.

Usaremos la notación

$$E_\theta g(\underline{X}) \tag{66}$$

para denotar la esperanza de $g(X_1, \ldots, X_n)$ cuando $\underline{X}$ es una muestra de X_θ la variable con distribución F_θ.

Un **estimador puntual** de θ es una función de la muestra de X que se denota

$$\hat{\theta} = \hat{\theta}(\underline{X})$$

Cuando el experimento es realizado, la muestra observada se denota con minúsculas $\underline{x}$ y $\hat{\theta}(\underline{x})$ se llama *estimativa*.

Ejemplo: Para estudiar si un dado está bien equilibrado, se arroja el dado 100 veces obteniéndose 21 ases. ¿Qué valor podría utilizarse, en base a esa información, como estimación de la probabilidad de as?

En este caso, si llamamos p a la probabilidad que queremos estimar, usamos la proporción muestral $\hat{p} = 0{,}21$ como estimativa.

Métodos de estimación puntual

Método de momentos: Se buscan los valores de los parámetros que permiten igualar los momentos muestrales a los momentos poblacionales.

Sea $X = X_\theta$ una variable aleatoria con distribución F_θ, $\theta \in \Theta$.

Sea EX_θ^k el **momento** de orden k de X. Es una función de θ que llamamos g_k:

$$EX_\theta^k = g_k(\theta)$$

Sea $\underline{X} = (X_1, \ldots, X_n)$ una muestra de X.

Definimos el **momento muestral** de orden k por:

$$\frac{\sum_{i=1}^n X_i^k}{n}$$

Cuando la muestra observada es $(x_1, \dots, x_n)$, los momentos observados de orden k son

$$\frac{\sum_{i=1}^n x_i^k}{n}$$

Suponga que $\theta = (\theta_1, \dots, \theta_m)$. Es decir $\Theta = \mathbb{R}^m$.

Defina $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$ los parametros que se obtienen al igualar los m primeros momentos muestrales a los momentos poblacionales correspondientes. Más precisamente, $\hat{\theta}_1, \dots, \hat{\theta}_m$ es la solución de las ecuaciones

$$g_k(\theta_1, \dots, \theta_m) = \frac{\sum_{i=1}^n x_i^k}{n}, \quad k = 1, \dots, m.$$

$(\theta_1, \dots, \theta_m)$ son incógnitas y $(x_1, \dots, x_n)$ son datos. Es decir que $\hat{\theta}_i = \hat{\theta}_i(x_1, \dots, x_n) = \hat{\theta}_i(\underline{x})$ es una función de la muestra observada.

Substituyendo $\underline{x}$ por $\underline{X} = (X_1, \dots, X_n)$, tenemos las variables aleatórias $\hat{\theta}_i(\underline{X})$ que se llaman **estimadores de momentos** de $(\theta_1, \dots, \theta_m)$.

Ejemplo 1. $X \sim$ Exponencial(λ). Un parámetro, una ecuación:

$$EX = \bar{X}_n$$

Como $EX = 1/\lambda$, la ecuación queda

$$\frac{1}{\lambda} = \bar{X}_n.$$

De donde $\hat{\lambda} = 1/\bar{X}_n$.

Ejemplo 2. $X \sim$ Gama(α, λ). Dos parametros, dos ecuaciones:

$$EX = \bar{X}_n, \qquad EX^2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

Como $EX = \frac{\alpha}{\lambda}$ y $EX^2 = \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2}$, las ecuaciones quedan

$$\frac{\alpha}{\lambda} = \bar{X}_n, \qquad \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2} = \frac{\sum_{i=1}^n X_i^2}{n}$$

De aqui se despejan λ y α:

$$\hat{\lambda} = \frac{\bar{X}}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2}, \qquad \hat{\alpha} = \frac{\bar{X}^2}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2}.$$

Ejemplo 3. $U \sim$ Uniforme$[0, \theta]$. Un parametro, una ecuación:

$$EX = \bar{X}_n$$

como $EX = \frac{\theta}{2}$, la ecuación queda

$$\frac{\theta}{2} = \bar{X}_n.$$

Despejando θ obtenemos $\hat{\theta} = 2\bar{X}_n$.

Ejemplo 4. No siempre se puede usar el primer momento. Si X es Uniforme$[-\theta,\theta]$, $EX=0$ no depende de θ, por eso hay que usar el segundo momento:

$$EX^2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

como $EX^2 = \frac{4\theta^2}{12} = \frac{\theta^2}{3}$, la ecuación queda

$$\frac{\theta^2}{3} = \frac{\sum_{i=1}^n X_i^2}{n}.$$

Despejando θ, el estimador queda

$$\hat{\theta} = \sqrt{3\frac{\sum_{i=1}^n X_i^2}{n}}.$$

Método de máxima verosimilitud: Fisher en 1920.

Hallar los valores de los parámetros que maximizan la probabilidad de obtener la muestra observada.

Ejemplo: Encuesta de intención de voto de 20 personas. Queremos estimar la probabilidad p de votar por un determinado candidato. $X \sim$ Bernoulli(p). $(x_1,\dots,x_n)$ son los valores observados. La probabilidad de haber observado $(x_1,\dots,x_n)$ es

$$P\big((X_1,\dots,X_n)=(x_1,\dots,x_n)\big) = \prod_i p^{x_i}(1-p)^{1-x_i}$$

Cual es el valor de p que maximiza esa probabilidad?

$$\arg\max_p \big[\prod_i p^{x_i}(1-p)^{1-x_i}\big] = \arg\max_p \Big[\log p \sum_i x_i + \log(1-p)\sum_i (1-x_i)\Big]$$

Buscamos el punto crítico derivando en p:

$$\frac{dg(p)}{dp} = \frac{1}{p}\sum_i x_i - \frac{1}{1-p}\sum_i (1-x_i) = 0$$

de donde

$$\hat{p} = \frac{\sum_i x_i}{n}$$

Calculando la derivada segunda vemos que maximiza.

Definición de estimador de máxima verosimilitud Sea $X = X_\theta$ una familia de variables aleatorias con rango R con probabilidad puntual $p_\theta(.)$ o densidad conjunta f_θ que depende de parámetros $\theta \in \Theta$, el espacio de parámetros.

La **función de verosimilitud** $L : \Theta \times R^n \to [0,1]$ está definida para $\theta \in \Theta$ y $\underline{x} = (x_1,\dots,x_n) \in R^n$ por

$$L(\theta,\underline{x}) = \begin{cases} p_\theta(x_1)\dots p_\theta(x_n) & \text{caso discreto} \\ f_\theta(x_1)\dots f_\theta(x_n) & \text{caso continuo} \end{cases}$$

$L(\theta,(x_1,\dots,x_n))$ es la probabilidad de observar $(x_1,\dots,x_n)$ cuando el parámetro es θ. Para cada elemento $\underline{x} \in R^n$ definimos $\hat{\theta}(\underline{x})$ como el argumento que maximiza $L(\theta,\underline{x})$:

$$\hat{\theta}(\underline{x}) := \arg\max_{\theta} L(\theta,\underline{x});$$

usualmente hay apenas un argumento que maximiza L. Substituyendo $\underline{x}$ por las variables $\underline{X} = (X_1,\dots,X_n)$ obtenemos el estimador

$$\hat{\theta}(X_1,\dots,X_n)$$

que es llamado *estimador de máxima verosimilitud*. Usualmente se escribe $L(\theta)$ en lugar de $L(\theta,\underline{x})$, subentendiendo la dependencia de $\underline{x}$.

Ejemplos

1. $(X_1,\dots,X_n) \sim$ Exponencial(λ)

$$\begin{aligned} L(\lambda) &= \lambda^n e^{-\lambda(x_1+\cdots+x_n)} \\ \log L(\lambda) &= n\log\lambda - \lambda(x_1+\cdots+x_n) \end{aligned}$$

Derivando en λ e igualando a cero

$$\frac{dL}{d\lambda} = \frac{n}{\lambda} + (x_1+\cdots+x_n) = 0$$

De donde

$$\hat{\lambda} = \frac{1}{\bar{x}_n}.$$

Verifique que es un máximo con la segunda derivada.

2. $(X_1,\dots,X_n) \sim$ Normal(μ,σ^2)

$$L(\mu,\sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\Big(-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\Big)$$

Maximizarla equivale a maximizar los logaritmos. El resultado es:

$$\hat{\mu} = \bar{x}, \qquad \hat{\sigma} = \sqrt{\frac{\sum_i (x_i-\bar{x})^2}{n}}.$$

3. $(X_1,\dots,X_n) \sim$ Uniforme($0,\theta$)

$$L(\theta) = \frac{1}{\theta^n}\prod_i I_{x_i\in[0,\theta]} = 0 I_{\theta<\text{máx}_i x_i} + \frac{1}{\theta^n} I_{\theta\geq\text{máx}_i x_i}$$

De donde $\hat{\theta} = \text{máx}_i\, x_i$

Propiedades de los estimadores

Dada una muestra $(X_1,\dots,X_n)$ de $X \sim F_\theta$ y un estimador puntual $\hat{\theta}(\underline{X})$, la diferencia

$$\hat{\theta}(\underline{X}) - \theta$$

es el **error** de estimación. Este error es una variable aleatoria dado que es función de la muestra. El **sesgo** de un estimador es su error medio:

$$b(\hat{\theta}) := E_\theta \hat{\theta}(\underline{X}) - \theta$$

Definición: Un estimador $\hat{\theta}(\underline{X})$ de θ es **insesgado** si tiene sesgo nulo. Un estimador $\hat{\theta}$ de θ es **asintóticamente insesgado** si

$$\lim_n E_\theta \hat{\theta} = \theta$$

Ejemplos.

1. $X \sim \mathit{Bernoulli}(p)$. La proporción muestral $\hat{p}$ es un estimador insesgado de p:

$$E_p \hat{p} = p.$$

2. $X \sim \mathit{Normal}(\mu, \sigma)$. Es claro que $\hat{\mu} = \bar{X}$ es insesgado pero

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$$

no es insesgado para σ^2 porque haciendo cuentas se obtiene

$$E_{\mu,\sigma^2} \hat{\sigma}^2 = \frac{n-1}{n} \sigma^2,$$

o sea que $\hat{\sigma}^2$ es estimador asintóticamente insesgado de σ^2. Las mismas cuentas dicen que S^2 es estimador insesgado de σ^2.

3. $X \sim \mathit{Uniforme}[0, \theta]$. El estimador de momentos de θ es $2\bar{X}$. Es insesgado: $E_\theta \bar{X} = \theta$. El EMV de θ es $M = \max_i X_i$. No es insesgado:

$$\begin{aligned} E_\theta M = \int_0^\theta P_\theta(M > x) dx \ &= \int_0^\theta (1 - P_\theta(M \le x)) dx \\ = \int_0^\theta \Big(1 - \Big(\frac{x}{\theta}\Big)^n\Big) dx \ &= \ \theta - \frac{\theta^{n+1}}{(n+1)\theta^n} \ = \ \frac{n}{n+1}\theta \end{aligned}$$

pero es asintoticamente insesgado.

Consistencia

Diremos que $\hat{\theta}_n(\underline{X})$ es un estimador **consistente** de θ si

$$\hat{\theta}_n(\underline{X}) \longrightarrow \theta, \quad \text{en probabilidad.}$$

Ejemplo Si X tiene media μ y varianza σ^2, entonces $\bar{X}_n$ es un estimador consistente de μ, por la ley de grandes números. Verifique que el estimador $(X_1 + X_n)/2$ no es consistente.

Lema 66 *Si un estimador es asintóticamente insesgado y su varianza va a cero, entonces es consistente.*

Dem: Inmediata si es insesgado, por Chevichev. En el caso general no lo haremos. $\square$

Ejemplo $X \sim$ Uniforme$[0,\theta]$. $\hat\theta = \max X_i$ es asintoticamente insesgado. $E_\theta(\hat\theta) = \frac{n}{n+1}\theta$. La varianza del máximo de n uniformes es

$$V(\max_i X_i) = \frac{n}{(n+1)(n+2)^2}\theta^2 \to_n 0$$

Por lo tanto $\hat\theta = \max X_i$ es consistente.

Lema 67 *S^2 es un estimador consistente de la varianza poblacional.*

Dem

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum (X_i - \bar X)^2 = \frac{1}{n-1}\sum (X_i^2 - 2X_i\bar X + \bar X^2) \\ &= \frac{1}{n-1}\Big(\sum X_i^2 - 2n\bar X^2 + n\bar X^2\Big) = \frac{n}{n-1}\Big(\sum \frac{X_i^2}{n} - \bar X^2\Big). \end{aligned}$$

Como $\bar X_n \to \mu$, $(\bar X_n)^2 \to \mu^2$. Además, por la ley de grandes números,

$$\sum_i \frac{X_i^2}{n} \to E_{\mu,\sigma^2}X^2 = \mu^2 + \sigma^2.$$

Como $n/(n-1) \to 1$,

$$S_n^2 \to \mu^2 + \sigma^2 - \mu^2 = \sigma^2. \qquad \square$$

13.2. Intervalos de confianza

Vimos estimación puntual de un parámetro, y controlamos en algunos casos el error entre el estimador y el parámetro. Otro modo es reemplazar la estimación puntual por un intervalo de valores posibles para el parámetro.

Ejemplo Si $X \sim$ Normal(μ,σ^2) con μ desconocida y σ^2 conocida. Sabemos que $\bar X_n \sim$ Normal$(\mu,\sigma^2/n)$ y que

$$Z = \frac{\bar X_n - \mu}{\sigma/\sqrt{n}} \sim \text{ Normal}(0,1),$$

de donde, usando la tabla Normal, obtenemos

$$P\Big(-1,96 \le \frac{\bar X_n - \mu}{\sigma/\sqrt{n}} \le 1,96\Big) = 0,95$$

que equivale a

$$P(\bar X - 1,96\,\sigma/\sqrt{n} \le \mu \le \bar X - 1,96\,\sigma/\sqrt{n}) = 0,95 \qquad (67)$$

Es decir que la probabilidad que el intervalo aleatorio

$$I := [\bar X - 1,96\,\sigma/\sqrt{n}, \bar X - 1,96\,\sigma/\sqrt{n}]$$

contenga μ es 0,95. Se trata de un intervalo de radio $1,96\,\sigma/\sqrt{n}$ y centro aleatorio $\bar{X}$. El intervalo I es el **intervalo de confianza para μ de confianza 0,95**.

La expresión (67) se puede interpretar así: "en 95 % de las muestras de tamaño n, el intervalo de radio $1{,}96\,\sigma/\sqrt{n}$ centrado en la media muestral $\bar{x}$ contiene a la media poblacional μ".

En general, la relación entre el radio del intervalo ε, la confianza γ y el tamaño de la muestra n está dada por la siguiente fórmula:

$$\varepsilon = z_\gamma\,\sigma/\sqrt{n} \tag{68}$$

donde z_γ está determinada por la fórmula $\Phi(z_\gamma) - \Phi(-z_\gamma) = \gamma$; ϕ es la acumulada de la normal standard.

Definición Sea X una variable aleatoria cuya distribución depende de un parámetro $\theta \in \mathbb{R}$ y $\underline{X} = (X_1, \ldots, X_n)$ una muestra de X. Si a y b son dos funciones de la muestra tales que

$$P(a(\underline{X}) \leq \theta \leq b(\underline{X})) = 1 - \alpha, \tag{69}$$

llamamos al intervalo $[a(\underline{X}), b(\underline{X})]$ el *intervalo de confianza a nivel* $1-\alpha$ para θ.

Observaciones: 1) El intervalo $[a, b]$ es aleatorio ya que sus extremos son funciones de la muestra. La expresión (69) debe leerse "La probabilidad de que el intervalo (a,b) contenga al parámetro θ es $1-\alpha$".

2) Una vez observada la muestra, el intervalo es también "observado" y ya no tiene sentido hablar de probabilidad, sino de "confianza" de que el intervalo contenga a θ. Como $(1-\alpha)100\,\%$ de las muestras producirán intervalos que contienen a θ, esa es nuestra confianza de que el intervalo observado sea uno de esos.

Intervalos de confianza asintótico para p de la Bernoulli. Sea X Bernoulli(p) y $\hat{p}_n$ el estimador puntual de p. Queremos establecer la relación entre el tamaño de la muestra n, el radio del intervalo dado por el error ε y la confianza $1-\alpha$, basados en la expresión

$$P(\hat{p}_n - \varepsilon < p < \hat{p}_n + \varepsilon) = 1 - \alpha$$

que equivale a

$$P(|\hat{p}_n - p| < \varepsilon) = 1 - \alpha$$

Standarizando obtenemos la expresión equivalente

$$P\Big(\frac{|\hat{p}_n - p|}{\sqrt{p(1-p)}/\sqrt{n}} < \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}}\Big) = 1 - \alpha$$

Por el teorema del límite central, aproximadamente

$$P\Big(|Z| < \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}}\Big) \approx 1 - \alpha$$

para $Z \sim \text{Normal}(0, 1)$. Aceptando la aproximación como identidad, obtenemos la siguiente relación:

$$z = \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}} \tag{70}$$

donde $z = z_{(1-\alpha)/2}$ satisface $P(|Z| < z) = 1 - \alpha$. Para usar la tabla, observe que $P(|Z| < z) = 1 - \alpha$ es equivalente a $\phi(z) = 1 - \alpha/2$, con ϕ la acumulada de la Normal$(0, 1)$.

Preguntas:

1) Dado el error ε y el tamaño n de la muestra, cual es la confianza $1 - \alpha$ del intervalo obtenido?

2) Dado el error ε y la confianza $1 - \alpha$ que deseamos, cual es el tamaño n que debe tener la muestra?

3) Dada la confianza $1 - \alpha$ que deseamos que tenga el intervalo obtenido y el tamaño n de la muestra, cual es el error obtenido?

Respuestas: Use (70) y la desigualdad $p(1 - p) \leq \frac{1}{4}$ para obtener:

1) Primero se obtiene z con la fórmula

$$z = \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}} \geq 2\varepsilon\sqrt{n}$$

y de ahí $1 - \alpha$ usando la tabla: $P(Z < z) = (1 - \alpha/2)$. El intervalo obtenido con este z va a tener confianza $(1 - \alpha)$, por lo menos.

2) Tenemos $1 - \alpha$ y ε y despejamos n:

$$n = \frac{z^2 p(1-p)}{\varepsilon^2} \leq \frac{z^2}{2\varepsilon^2}$$

Obtenga z usando la tabla y use el menor n mayor o igual a $\frac{z^2}{2\varepsilon^2}$ para garantizar un intervalo de radio ε con confianza $1 - \alpha$.

3) Ahora conocemos $1 - \alpha$ y n y buscamos ε. Despeje en (70):

$$\varepsilon = \frac{z\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{z}{2\sqrt{n}}$$

tomando el peor caso. Obtenemos z a partir de $1 - \alpha$ como antes y el intervalo de radio $\varepsilon = \frac{z}{2\sqrt{n}}$ va a tener confianza por lo menos $1 - \alpha$.

Ejemplos. Encuesta balotage presidencial 2015

Poliarquía Consultores entrevistó $n = 800$ personas en centros urbanos de más de diez mil habitantes y obtuvo Macri $\hat{p} = 0{,}487$, Scioli $\hat{p} = 0{,}402$. Dice que el margen de error de la encuesta es de 3,5 puntos con un nivel de confianza de 95 %. El intervalo de confianza 0.95 para la intención de voto p para cada candidato es

$$[\hat{p} - \varepsilon, \hat{p} + \varepsilon],$$

donde

$$\varepsilon = 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq 1{,}96\frac{1}{2\sqrt{800}}. = 0{,}03464$$

Así, pudimos verificar que el intervalo que propone Poliarquía es compatible con la teoría. Esto asume que la muestra es realmente aleatoria.

Método del pivote para obtener intervalos de confianza:

Sea $\underline{X} = (X_1, \dots, X_n)$ una muestra de X cuya distribución depende de un parámetro θ. Supongamos que existe una función $T(\underline{X}, \theta)$ (es decir, una función de la muestra y del parámetro) cuya distribución no depende de θ ni de ningún otro parámetro desconocido. Entonces, para cada $\alpha > 0$ existen dos valores a y b tales que

$$P(a < T(\underline{X}, \theta) < b) = 1 - \alpha.$$

A partir de esta expresión es posible obtener un intervalo de confianza para θ. T es llamado **pivote**.

Ejemplo X Exponencial(λ).

$$X_1 + \cdots + X_n \sim \text{ Gama}(n, \lambda)$$

Como $\lambda X \sim$ Exponencial(1), se puede demostrar que

$$T = 2\lambda(X_1 + \cdots + X_n) \sim \chi^2_{2n}$$

Con eso se puede obtener un intervalo de confianza para λ con la tabla de la χ^2 con $2n$ grados de libertad. De la tabla de la Chi cuadrado:

$$P(\chi^2_{2n,1-\frac{\alpha}{2}} < T < \chi^2_{2n,\frac{\alpha}{2}}) = 1 - \alpha$$

se despeja λ para obtener el intervalo

$$P\Big(\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{2\sum X_i} < \lambda < \frac{\chi^2_{2n,\frac{\alpha}{2}}}{2\sum X_i}\Big) = 1 - \alpha$$

Ejemplo. Sea X Uniforme$[0, \theta]$. El EMV de θ es $\hat{\theta} = \text{máx}\, X_i$. La distribución de $T = \hat{\theta}/\theta$ no depende de θ. De hecho, la distribucion de X_i/θ es Uniforme$[0, 1]$ y la distribución de T es la del máximo entre n Uniforme$[0, 1]$. La acumulada de T es

$$F_T(x) = x^n, \quad x \in [0, 1]$$

Derivando obtenemos la densidad

$$f_T(x) = nx^{n-1} I_{[0,1]}, \quad x \in [0, 1]$$

Usando T como pivote, tenemos

$$P(a < T < b) = 1 - \alpha \tag{71}$$

obtenemos el siguiente intervalo de confianza $1 - \alpha$:

$$\Big[\frac{\text{máx}\, X_i}{b}, \frac{\text{máx}\, X_i}{a}\Big]$$

Cómo elegir a y b? Son soluciones de (71) para α fijo:

$$\int_a^b nx^{n-1} dx = 1 - \alpha,$$

que tiene infinitas soluciones. Podemos buscar a y b que minimicen la longitud promedio del intervalo de confianza que es $(E\, \text{máx}\, X_i)[\frac{1}{a} - \frac{1}{b}]$. Se obtiene $b = 1$ y $a = \sqrt{\alpha}$.

13.3. Test de Hipotesis

En una isla hay dos tribus. Una de aborígenes amigables de altura media 170 cm y otra de caníbales de altura media 150 cm.

Al llegar a la isla un explorador encuentra 9 aborígenes y quiere que decidir si son amigables o caníbales.

La altura de los aborígenes es una variable aleatoria X, en cm. Asumimos $X \sim$ Normal$(\mu, 100)$. Varianza conocida. μ desconocida, o es 150 o 170. Necesitamos decidir entre una de las dos hipótesis:

$H_0 : \mu = 150$, los aborígenes son caníbales.

$H_1 : \mu = 170$, los aborígenes son amigables.

Obtenemos una muestra aleatoria $(x_1, \dots, x_9)$ de X y calculamos su media muestral $\bar{x}$.

Regla de decisión: Decidimos que si $\bar{x} \geq 160$, se rechaza H_0 y se desembarca en la isla. En caso contrario, no se rechaza H_0 y no se desembarca. Por el momento 160 es un valor arbitrario.

Es decir testeamos la hipótesis H_0 con el criterio "si la media muestral está arriba de 160, la rechazamos; si no, la aceptamos".

Región de rechazo (RR) para $\bar{x}$ es el intervalo

$$RR = [160, \infty).$$

Por ejemplo, si observamos $\bar{x} = 162$ que está en la región de rechazo, rechazamos H_0. Si en otra muestra observamos $\bar{x} = 157$, no rechazamos H_0.

Podemos cometer dos errores:

Error de tipo 1: Rechazar H_0 cuando H_0 es verdadera.

Error de tipo 2: Aceptar H_0 cuando H_0 es falsa.

Observe la gravedad de cada uno de los errores. Es más serio el error de tipo 1 (desembarcar cuando son caníbales) que el de tipo 2 (retirarse cuando son amigables).

Cálculo de la probabilidad del error 1

La media muestral tiene distribución normal $\bar{X} \sim$ Normal$(\mu, 100/9)$ y la región de rechazo para $\bar{X}$ es $RR = [160, \infty)$. Llamando α al error de tipo 1, tenemos

$$\begin{aligned}
\alpha &= P(\text{error tipo } 1) \\
&= P(\bar{X} > 160 \,|\, H_0 \text{ verdadera}) \\
&= P(\bar{X} \geq 160 \,|\, \mu = 150) \\
&= P\Big(\frac{\bar{X} - 150}{10/3} \geq \frac{160 - 150}{10/3} \,\Big|\, \mu = 150\Big)
\end{aligned}$$

pero, como bajo $\mu = 150$, $Z = (\bar{X} - 150)/(10/3) \sim$ Normal$(0, 1)$, tendremos

$$\alpha = P(Z > 3) = 0{,}0013 \quad \text{(por la tabla Normal)}$$

α es llamado **nivel de significancia** del test.

Qué quiere decir $\alpha = 0{,}0013$? Que de cada 10000 muestras que provienen de una población con H_0 verdadera (es decir con $\mu = 150$), rechazaremos (equivocadamente) H_0 en 13 de los tests.

De 10000 expediciones a esa isla que observan una muestra de 9 caníbales (pero sin saber que son caníbales), habrá 13 que desembarcarán (y serán comidos por los caníbales).

Por el momento no estamos considerando las expediciones que observan muestras de nativos amigables.

Definición Dadas dos hipótesis H_0 y H_1 relativas a parámetros de la distribución de una variable aleatoria X, un **test** es una **regla de decisión** basada en un estadístico de una muestra de X y en una región de rechazo para ese estadístico. Si el estadístico observado pertenece a la región de rechazo, entonces se rechaza la hipótesis nula H_0.

En el ejemplo anterior el estadístico era $\bar{X}$ y la región de rechazo el intervalo $[160, \infty)$.

La región de rechazo se puede fijar en función del error α que estamos dispuestos a cometer.

La regla de decisión es aleatoria, porque depende del valor del estadístico.

Podemos equivocarnos. Por ejemplo podemos rechazar H_0 aún siendo $\mu = 150$.

Es imposible construir tests en los cuáles estemos absolutamente seguros de tomar la decisión correcta

Tipos de error:

Tipo 1: Se rechaza H_0 cuando H_0 es cierta

Tipo 2: No se rechaza H_0 cuando H_0 no es cierta

$\alpha = P($ error tipo 1$)$ Nivel de significancia.

$\beta = P($ error tipo 2$)$

¿Cómo se elige la región de rechazo?

Usualmente se elige el valor α y se calcula la región de rechazo que tiene probabilidad de error tipo 1 igual a α. En el ejemplo, para $\alpha = 0,05$, buscamos z tal que $\phi(z) = 1 - 0{,}05$ y rechazamos H_0 si $\frac{\bar{X}-150}{10/3} > z$ que corresponde a $z = 1{,}64$ y

$$\bar{x} \geq 150 + 1{,}64\frac{10}{3} = 150 + 5{,}4 = 154,4$$

Para $\alpha = 0,05$ rechazamos si $\bar{x} \geq 154,4$.

Elección de estadísticos En el test precedente podríamos haber usado directamente el estadístico

$$T := \frac{\bar{X} - 150}{10/3}$$

y la región de rechazo para T:

$$RC = [1{,}64, \infty)$$

P-valor Otra manera de hacer el test es considerar un estadístico llamado P-valor. Si estamos considerando el estadístico T y observamos t_{obs}, el P-valor es el α correspondiente a la región de rechazo para T cuyo extremo es t_{obs}. En particular, para el ejemplo anterior con el estadístico $T = \bar{X}$, si se la muestra observada es $x_1, \ldots, x_n$ y la media muestral observada es $\bar{x} = \bar{x}_{\mathrm{obs}} = 156$, el P-valor es

$$\begin{aligned}\text{P-valor}(x_1, \ldots, x_n) &= P(\bar{X} > \bar{x} \mid H_0) \\ &= P(\bar{X} > 156 \mid \mu = 150) = P(Z > 1{,}8) = 0{,}0359.\end{aligned}$$

(por la tabla) Esto quiere decir que si hacemos un test con $\alpha < 0{,}0359$, no podremos rechazar H_0.

Se rechaza a nivel α cuando P-valor$(x_1, \ldots, x_n) < \alpha$.

Substituyendo $(x_1, \ldots, x_n)$ por $(X_1, \ldots, X_n)$, obtenemos el estadístico

$$\text{P-valor}(X_1, \ldots, X_n)$$

El P-valor es una función de la muestra, por lo tanto es un estadístico.

Para rechazar H_0, el P-valor observado tiene que ser menor que el α deseado. O sea, la región de rechazo para el P-valor es $[0, \alpha]$.

Error tipo 2

Supongamos que en nuestro ejemplo, observamos una altura media 154 en la muestra de tamaño 9 y trabajamos con el test de nivel 0,05. En este caso,

$$\bar{x} = 154 \leq 154{,}4$$

que está fuera de la región de rechazo $[154{,}4, \infty)$. Por lo tanto **no rechazamos** H_0.

Podríamos estar cometiendo un error de tipo 2. Por ejemplo, si los aborigenes observados no son caníbales y tienen altura media 160, ¿cuál es la probabilidad de cometer un error tipo II?

$$\begin{aligned}P(\text{error tipo 2}) &= P(\text{aceptar } H_0 \mid H_1 \text{ verdadera, con } \mu = 160) \\ &= P(\bar{X} < 154{,}4 \mid H_1 \text{ verdadera, con } \mu = 160) \\ &= P\Big(\frac{\bar{X} - 160}{10/3} < \frac{154{,}4 - 160}{10/3} \Big| \mu = 160\Big) \\ &= P(Z < -1{,}68) = 1 - 0{,}9535 = 0{,}0465,\end{aligned}$$

usando la tabla. El error de tipo 2 es una **función** del valor alternativo de H_1 y de la región de rechazo. En este caso $\beta(160) = 0{,}0465$.

Analogía con el sistema de justicia

Una persona es acusada de un crimen. La hipótesis nula es que la persona es inocente. La hipótesis alternativa es que el acusado es culpable.

El test de hipótesis es un juicio con pruebas presentadas por las dos partes.

Una vez consideradas las presentaciones de la acusación y la defensa, el jurado toma la decisión de "culpable" o "no culpable".

Nunca declara *inocente* al acusado, a lo sumo concluye que las pruebas presentadas no son suficientes para declararlo culpable.

El objetivo del juicio es determinar si hay pruebas suficientes para declararlo culpable.

El error de tipo 1 corresponde a declarar culpable a un inocente.

El error de tipo 2 es liberar a una persona culpable.

El error de tipo 1 es el más serio ("somos todos inocentes hasta que se demuestre lo contrario").

Se busca que la probabilidad de ese error sea muy chica.

En juicios criminales, lo usual es declarar culpable al acusado cuando hay poco espacio para la duda.

Tipos de hipotesis

Las hipótesis alternativas pueden ser unilaterales o bilaterales. Las regiones de rechazo dependen del tipo de test.

Ejemplo, el test para μ de la normal con σ^2 conocida.

Hay tres posibles tests para μ:

1) H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$; (contra menor)

2) H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$; (contra mayor)

3) H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$; (bilateral)

Usamos el estadístico

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma},$$

Como bajo H_0, $T \sim$ Normal$(0, 1)$, las regiones de rechazo a nivel α son, respectivamente:

1) $RC = (-\infty, -z_\alpha]$

2) $RC = [z_\alpha, \infty)$

3) $RC = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$

donde z_α satisface $P(Z < z_\alpha) = 1 - \alpha$.

Tests para la media cuando la varianza es desconocida: Supongamos ahora que la varianza es desconocida y consideremos las mismas hipótesis sobre μ:

1) H_0: $\mu = \mu_0$, H_1: $\mu < \mu_0$; (contra menor)

2) H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$; (contra mayor)

3) H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$; (bilateral)

Estadístico: $T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$

Bajo $\mu = \mu_0$ $T \sim t_{n-1}$ (t de Student con $n - 1$ grados de libertad). .

Regiones de rechazo son:

1) $RC = (-\infty, -t_\alpha]$

2) $RC = [t_\alpha, \infty)$

3) $RC = (-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty)$

donde t_α satisface $P(T < z_\alpha) = 1 - \alpha$, que se encuentra en la tabla de la t de Student.

Tests para la varianza cuando la media es desconocida: Las hipótesis a testear son

1) H_0: $\sigma^2 = \sigma_0^2$, H_1: $\sigma^2 < \sigma_0^2$; (contra menor)

2) H_0: $\sigma^2 = \sigma_0^2$, H_1: $\sigma^2 > \sigma_0^2$; (contra mayor)

3) H_0: $\sigma^2 = \sigma_0^2$, H_1: $\sigma^2 \neq \sigma_0^2$; (bilateral)

Estadístico: $T = \frac{(n-1)S^2}{\sigma_0^2}$

Bajo la hipótesis H_0 ($\sigma^2 = \sigma_0^2$) el estadístico $T \sim \chi^2_{n-1}$ (Qui-cuadrado con $n-1$ grados de libertad).

Regiones de rechazo son:

1) $RC = (-\infty, -x_\alpha]$

2) $RC = [\chi^2_{1-\alpha}, \infty)$

3) $RC = (-\infty, x_{\alpha/2}] \cup [x^+_{1-\alpha/2}, \infty)$

donde x_α satisface $P(\chi^2_{n-1} < x_\alpha) = \alpha$. Esos valores se encuentran tabla de la χ^2 con $n-1$ grados de libertad.

Ejemplo Se toman 25 determinaciones de la temperatura en cierto sector de un reactor, obteniéndose

$\bar{x} = 243^oC$ y $s = 2{,}8^oC$

Interesa saber, a nivel $\alpha = 0{,}05$

a) si existe evidencia para decidir que la temperatura media en ese sector del reactor es menor que 250^oC .

b) si existe evidencia para decidir que la varianza de la temperatura en ese sector del reactor es mayor que $(2^oC)^2$.

a) Las hipótesis a testear son H_o: $\mu = 250$ (ó $\mu \geq 250$) vs H_1: $\mu < 250$.

El estadístico del test será $T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$ y la región de rechazo para ese estadístico será $(-\infty, -t_{n-1,0{,}05}]$.

En nuestro caso, $n = 25$ y por lo tanto $-t_{24,0{,}05} = -1{,}71$. Como el valor observado de T es $-12{,}5$, se rechaza H_0, es decir hay evidencia de que la temperatura media del reactor es menor que 250^oC.

b) Las hipótesis a testear son H_0: $\sigma^2 = 4$ (ó $\sigma^2 \leq 4$) vs H_1: $\sigma^2 > 4$

El estadístico del test será $T = \frac{(n-1)S^2}{\sigma_0^2}$ y la región de rechazo $[\chi^2_{n-1,0{,}05}, \infty)$.

En nuestro caso, $n = 25$ y por lo tanto $\chi^2_{24,0{,}05} = 36{,}42$. Como el valor observado de T es $47{,}04$, se rechaza H_0. Es decir, hay evidencia de que la varianza de la temperatura del reactor es mayor que $(2^oC)^2$.

Tests de hipótesis de nivel aproximado (o asintótico) α para la media de una distribución cualquiera: Queremos testear la media μ asumiendo la varianza σ^2 finita pero desconocida.

Usaremos el estadístico $T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$ que tiene distribución asintótica Normal$(0,1)$ por el TCL.

Se toma n "grande" y se trabaja como en el caso de $X \sim$ Normal(μ, σ^2). Las regiones de rechazo son

1) $RC = (-\infty, -z_\alpha]$

2) $RC = [z_\alpha, \infty)$

3) $RC = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$

donde z_α satisface $P(Z < z_\alpha) = 1 - \alpha$, $Z \sim$ Normal$(0,1)$.

Test de hipótesis asintótico para p de la Bernoulli

Hay tres posibles tests para p:

1) H_0: $p = p_0$, H_1: $p < p_0$; (contra menor)

2) H_0: $p = p_0$, H_1: $p > p_0$; (contra mayor)

3) H_0: $p = p_0$, H_1: $p \neq p_0$; (bilateral)

Usamos el estadístico

$$T = \sqrt{n}\frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)}},$$

Como bajo H_0, $T \sim$ Normal$(0,1)$ asintoticamente (TCL), las regiones de rechazo a nivel α son, respectivamente:

1) $RC = (-\infty, -z_\alpha]$

2) $RC = [z_\alpha, \infty)$

3) $RC = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$

donde z_α satisface $P(Z < z_\alpha) = 1 - \alpha$.

Ejemplo del adivino Un adivino acierta el color de 850 de 1600 cartas puestas al dorso. Queremos decidir si creemos que es adivino.

Sea p la probabilidad que el adivino acierte. Queremos testar

$H_0:\ p = 1/2$ (es decir, no mejora el puro azar) contra $H_1:\ p > 1/2$ (tiene probabilidad de adivinar mayor que 1/2).

Usando que bajo H_0 el parámetro es $p_0 = 1/2$, el estadístico observado es

$$t_{\text{obs}} = \sqrt{n}\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}} = \sqrt{1600}\frac{\frac{850}{1600} - \frac{1}{2}}{\frac{1}{2}} = 2{,}5$$

que corresponde a un P-valor de 0,005 (por la tabla de la normal). Es decir que podemos rechazar H_0 para cualquier $\alpha > 0{,}005$.

Si el adivino hubiese adivinado 825 cartas el estadístico sería

$$t_{\text{obs}} = \sqrt{1600}\frac{\frac{820}{1600} - \frac{1}{2}}{\frac{1}{2}} = 1{,}25$$

Aquí el P-valor es 0,105 que nos deja en duda.

Relación entre intervalos de confianza y tests bilaterales

Asumamos $X \sim$ Normal(μ, σ^2). Sea $X_1, \dots, X_n$ una muestra aleatoria de X.

Sabemos que el intervalo de confianza para μ de confianza $1-\alpha$ está dado por

$$IC = \left[\bar{X} - z\frac{\sigma}{\sqrt{n}},\ \bar{X} + z\frac{\sigma}{\sqrt{n}}\right]$$

Supongamos que queremos testear las hipótesis

$$H_0:\ \mu = \mu_0, \qquad H_1:\ \mu \neq \mu_0$$

Si μ_0 no pertenece al intervalo de confianza, sospechamos que H_0 es falsa.

De hecho,

$$P_{\mu_0}(IC \not\ni \mu_0) = 1 - P(IC \ni \mu_0) = 1 - (1-\alpha) = \alpha$$

O sea que rechazar H_0 si μ_0 no pertenece al intervalo de confianza $(1-\alpha)$ nos dá un test de nivel de significancia α.

Tests no paramétricos Basado en notas del Curso de Estadística del Instituto de Matemática y Estadística de la Universidad de San Pablo.

Tests de adherencia Objetivo: Testear si un modelo probabilístico es adecuado para un conjunto de datos observados.

Exemplo 1: Genética – Equilibrio de Hardy-Weinberg

Supongamos que consideramos los hijos de una pareja que tiene genotipos Aa el padre y Aa la madre.

El **modelo teórico** dice que las probabilidades de los genotipos de los hijos son:

Tipo	AA	Aa	aa
Probabilidad	1/4	1/2	1/4

Hay 3 categorias: AA, Aa, aa

En una población se estudian 100 descendientes de una pareja con esos genotipos y se observan

Genotipo	AA	Aa	aa	Total
Frecuencia observada	26	45	29	100

Objetivo: Verificar si el modelo genético propuesto es adecuado para esa población.

Si el modelo es adecuado, las frecuencias esperadas de descendientes para cada genotipo se calculan así:

$E_{AA} := 100\, P(AA) = 100\frac{1}{4} = 25$

$E_{Aa} := 100\, P(Aa) = 100\frac{1}{2} = 50$

$E_{aa} := 100\, P(aa) = 100\frac{1}{2} = 50$

Tenemos una tabla para las frecuencias esperadas y observadas:

Genotipo	AA	Aa	aa	Total
Frecuencia observada O_i	26	45	29	100
Frecuencia esperada E_i	25	50	25	100

Podemos afirmar que los valores observados están suficientemente cerca de los esperados, de tal manera que el modelo de Hardy-Weinberg es adecuado a esta población?

Test de Adherencia – Metodología

Considere una tabla de frecuencias observadas de $k \geq 2$ categorías de resultados en n observaciones:

Categorías	1	2	...	k	Total
Frecuencia observada	O_1	O_2	...	O_k	n

donde O_i es el total de individuos observados en la categoría i, $i = 1, \ldots, k$.

Sea p_i la probabilidad asociada a la categoría i.

El objetivo es testear las hipótesis

$H_0 : p_1 = p_{o1}, \ldots, p_k = p_{ok}$

H_1 : existe por lo menos una diferencia.

Aquí p_{0i} es la probabilidad asociada al modelo que estamos testeando.

Si E_i es el número esperado de individuos en la categoría i cuando H_0 es verdadera, entonces

$$E_i = np_{oi}, \quad i = 1, \ldots, k.$$

La tabla de frecuencias observadas y esperadas es

Categorías	1	2	...	k	Total
Frecuencia observada	O_1	O_2	...	O_k	n
Frecuencia esperada	E_1	E_2	...	E_k	n

Definimos el estadístico

$$\chi^2_{k-1}(\underline{O}) = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

donde $\underline{O} = (O_1, \ldots, O_k)$ son funciones de la muestra aleatoria y por lo tanto variables aleatorias.

Suponiendo que H_0 es verdadera, ese estadístico tiene distribución asintótica Chi-cuadrado con $k - 1$ grados de libertad. Sus probabilidades están tabuladas.

Este resultado es válido grosso modo para n grande y para valores esperados $E_i \geq 5$.

Basamos la regla de decisión en el P-valor. En ese caso,

$$P(\underline{o}) = P(\chi^2_{k-1}(\underline{O}) \geq \chi^2_{k-1}(\underline{o})),$$

Si para α fijado obtenemos $P(\underline{o}) \leq \alpha$, rechazamos H_0, si no, no rechazamos.

En el ejemplo, las hipótesis son:

H_0 : el modelo de Hardy-Weinberg es adecuado a la situación.

H_1 : el modelo no es adecuado.

Equivalentemente,

H_0: $p_0(AA) = 1/4$, $p_0(Aa) = 1/2$ e $p_0(aa) = 1/4$

H_1: por lo menos una de las tres igualdades no se verifica.

La tabla presenta los valores observados y esperados calculados antes.

Genotipo	AA	Aa	aa	Total
Frecuencia observada O_i	26	45	29	100
Frecuencia esperada E_i	25	50	25	100

Cálculo del valor del estadístico del test ($k = 3$):

$$\chi^2_{k-1}(\underline{o}) = 0{,}04 + 0{,}50 + 0{,}64 = 1{,}18$$

Usando la distribución de qui-cuadrado con $k - 1 = 2$ grados de libertad, el P-valor es

$$P = P(\chi^2_2 \geq 1{,}18) = 0{,}5543$$

Conclusión: Para $\alpha = 0{,}05$, como $P = 0{,}5543 > 0{,}05$, no rechazamos H_0, es decir que no hay evidencia que la población no siga el equilibrio de Hardy-Weinberg.

Tests de Independencia

Objetivo: Verificar si hay independencia entre dos variables.

Ejemplo: Queremos verificar si hay dependencia entre renta y número de hijos en las familias de una ciudad.

Son elegidas 250 familias al azar y se obtiene la tabla siguiente:

Renta \ # de hijos	0	1	2	≥ 3	Total
menos de 2000	15	27	50	43	135
2000 a 5000	25	30	12	8	75
más de 5000	8	13	9	10	40
Total	48	70	71	61	250

Los datos se refieren a dos variables aleatorias X e Y observadas en una muestra de tamaño n en forma de tabla

Test de independencia

H_0: X e Y son variables independientes.

H_1: X e Y no son independientes.

Cuantas observaciones debería haber en cada celda de la tabla si X e Y fueran independientes? Como en ese caso las probabilidades conjuntas deberían ser iguales al producto de las probabilidades marginales,

$$p_{ij} = P(X = i, Y = j) = P(X = i)P(Y = j),$$

el número esperado de observaciones en la celda ij debería ser

$$E_{ij} = np_{ij} = np_{(i\cdot)}p_{(\cdot j)} = \frac{n_{(i\cdot)}n_{(\cdot j)}}{n}$$

bajo la hipótesis de independencia. Estamos usando la notación

$n_{(i\cdot)}$:= número de observaciones de $X = i$.

$n_{(\cdot j)}$:= número de observaciones de $Y = j$.

n_{ij} := número de observaciones de $X = i$ e $Y = j$.

El estadístico propuesto bajo la suposición de independencia está dado por:

$$\chi_q^2(\underline{O}) = \sum_{i,j} \frac{(E_{ij} - O_{ij})^2}{E_{ij}}$$

donde $O_{ij} = n_{ij}$ representa el número total de observações en la celda (i, j).

Bajo la hipótesis de independencia $\chi_q^2(\underline{O})$ tiene distribución asintótica Chi-cuadrado de q grados de libertad.

$q := (f-1)(c-1)$, f :=número de filas; c := número de columnas.

La regla de decisión se basa en el P-valor

$$P(\underline{o}) = P(\chi_q^2(\underline{O}) \geq \chi_q^2(\underline{o}))$$

Si para α fijo obtenemos $p \geq \alpha$, rechazamos H_0, en caso contrario no podemos rechazar.

Continuación del ejemplo: renta y número de hijos. $n = 250$.

H_0 : renta y número de hijos son variables independientes.

H_1 : existe dependencia entre esas variables.

Valores esperados bajo independencia:

Renta \ # de hijos	0	1	2	≥ 3	Total
menos de 2000	25.92	37.80	38.34	32.94	135
2000 a 5000	14.40	21	21.30	18.30	75
más de 5000	7.68	11.20	11.36	9.76	40
Total	48	70	71	61	250

Donde, por ejemplo:

$$11{,}20 = \frac{70 \times 40}{250}$$

El estadístico chi-quadrado observado es

$$\chi_q^2(\underline{o}) = \ldots \text{cuentas} \cdots = 36,62$$

Determinación del número de grados de libertad:

Categorías de renta: $f = 3$

Categorías de número de hijos: $c = 4$

$$q = (f-1)(c-1) = 2\,3 = 6$$

El P-valor observado es $P(\underline{o}) = P(\chi_6^2 \geq 36,62) = 0,000$ (por la tabla de la χ_6^2)

Como $P = 0,000 < \alpha = 0,05$ (por ejemplo), rechazamos la independencia entre el número de hijos y la renta familiar a nivel 0,05. (Y para muchos otros valores de α menores.)

Referencias

[1] Robert B. Ash. *Probability and measure theory.* Harcourt/Academic Press, Burlington, MA, second edition, 2000. With contributions by Catherine Doléans-Dade.

[2] Ana M. Bianco and Elena J. Martínez. Probabilidades y estadística. Available at `http://www.dm.uba.ar/materias/probabilidades_estadistica_C/2011/2/apuntes.html` (2015/11/17).

[3] Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[4] Paul Ehrenfest and Tatiana Ehrenfest. *The conceptual foundations of the statistical approach in mechanics.* Translated by M. J. Moravcsik. Cornell University Press, Ithaca, N.Y., 1959.

[5] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[6] Hans-Otto Georgii. *Stochastics.* de Gruyter Textbook. Walter de Gruyter & Co., Berlin, extended edition, 2013. Introduction to probability and statistics, Translated from the German original [MR2397455] by Marcel Ortgiese, Ellen Baake and Georgii.

[7] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and random processes.* Oxford University Press, New York, third edition, 2001.

[8] Barry R. James. *Probabilidade: um curso em nível intermediário*, volume 12 of *Projeto Euclides [Euclid Project].* Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981.

[9] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms.* Addison-Wesley, third edition, 1997. Section 3.2.1: The Linear Congruential Method, pp. 10–26.

[10] Leonardo T. Rolla. Introdução à probabilidade. Available at `http://mate.dm.uba.ar/~leorolla/teaching/intro-probab.pdf` (2015/11/17).

[11] Sheldon Ross. *A first course in probability.* Pearson Prentice Hall, eigth edition, 2010.

[12] Hermann Thorisson. *Coupling, stationarity, and regeneration.* Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[13] Víctor J. Yohai. Notas de probabilidades y estadística. Available at `http://www.dm.uba.ar/materias/probabilidades_estadistica_M/2003/2/notas_proba.pdf` (2015/11/17).