

Clases de Probabilidad y Estadística (C), 2013

Pablo A. Ferrari

Fuentes:

Ana Bianco, Elena Martínez (2004), Probabilidades y Estadística (Computación)

Sheldon Ross (1997), A first course in Probability.

Ronald Meester (2003) A Natural introduction to Probability Theory.

Experimentos aleatorios y determinísticos

S Espacio muestral

Ejemplos:

Moneda: $\mathcal{S} = \{\text{Cara, Seca}\} = \{1, 0\}$

Dado: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$

Dos monedas

10 monedas: $\mathcal{S} = \{0, 1\} \times \dots \times \{0, 1\}$ (diez veces)

infinitas monedas: $\mathcal{S} =$ todas las sucesiones de 0 y 1.

Dos dados $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}^2$.

Tiempo de vida de una lámpara $\mathcal{S} = [0, \infty)$.

Eventos o sucesos: Subconjuntos de \mathcal{S} .

Ejemplos:

Cara sola, seca sola

Dos dados: suma par, suma igual a 7, resta menor que 2

10 monedas: por lo menos 5 caras.

lampara dura entre 3 y 5 meses

Operaciones con eventos

Unión, intersección, uniones e intersecciones numerables, complementos.

S es un subconjunto de \mathcal{S} denominado suceso cierto o seguro .

\emptyset es un subconjunto de \mathcal{S} denominado suceso imposible.

$A \cup B$ es el suceso unión. Ocurre cuando A ocurre ó B ocurre.

$A \cap B$ es el suceso intersección. Ocurre cuando ocurre A y ocurre B.

A^c o A es el opuesto o complemento de A. Ocurre cuando no ocurre A.

$A - B = A \cap B^c$ es el suceso diferencia. Ocurre cuando ocurre A y no ocurre B.

Se dice que A está contenido en B o que A implica B y se denota $A \subset B$ si la realización de A conduce a la realización de B, es decir si todo elemento de A pertenece a B.

Dos sucesos A y B se dicen 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: $A \cup B = B \cup A$, $A \cap B = B \cap A$

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:

$$\left(\cup_i A_i\right)^c = \cap_i A_i^c, \quad \left(\cap_i A_i\right)^c = \cup_i A_i^c$$

Interpretación intuitiva de la Probabilidad: 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 :

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

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

Propiedades

1) $\text{fr}(A)$ está entre 0 y 1

2) $\text{fr}(S) = 1$

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

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

cada evento A se le asocia $P(A)$ y que llamaremos probabilidad del evento A y que obedece los siguiente **axiomas**:

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

A2. $P(S) = 1$

A3a. Eventos A_1, A_2, \dots, A_n , mutuamente excluyentes, es decir si $A_i \cap A_j = \emptyset$, si $i \neq j$, entonces

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

A3b. Si A_1, A_2, \dots mutuamente excluyentes, entonces

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

$P(\{0\}) = 1 - p$, $P(\{0, 1\}) = 1$, $P(\emptyset) = 0$, con $0 \leq p \leq 1$, satisface los axiomas.

Propiedades de la Probabilidad:

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

2) $P(\emptyset) = 0$

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

Dem: Si $A \subset B \Rightarrow B = A \cup (B - A)$ y éstos dos eventos son excluyentes. Por el axioma A3a $P(B) = P(A) + P(B - A)$ Dado que, por el axioma A1, $P(B - A) \geq 0$, resulta $P(B) \geq P(A)$ y, despejando, se obtiene la segunda afirmación.

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

Dem: $A \cup B = A \cup (B - A) = A \cup (B \cap A^c)$ y estos dos eventos son excluyentes, entonces, por el axioma A3a,

$$P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \quad (1)$$

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

$$P(B) = P(B \cap A) + P(B \cap A^c) \Rightarrow P(B \cap A^c) = P(B) - P(B \cap A) \quad (2)$$

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

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

Dem: Esta propiedad se deduce inmediatamente de la propiedad anterior y del axioma A1.

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

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

b) Probar, usando inducción que, dados A_1, A_2, \dots sucesos cualesquiera,

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

Asignación de probabilidades: Si \mathcal{S} finito o infinito numerable designamos E_i a los sucesos elementales de \mathcal{S} , $\mathcal{S} = \cup_{i=1}^{\infty} E_i$.

Si conocemos $p_i = P(E_i)$, de manera que $\sum_{i=1}^{\infty} P(E_i) = 1$, entonces para cualquier suceso A ,

$$P(A) = \sum_{E_i \subset A} P(E_i) = 1$$

Ejemplos: 1) Dado equilibrado. $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ y $p_i = 1/6$ para $i = 1, \dots, 6$.

Para calcular $P(A) = P(\text{resultado par}) = P(E_2 \cup E_4 \cup E_6)$, se obtiene $P(A) = P(E_2) + P(E_4) + P(E_6) = 1/2$

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

$$P(E_1) = P(E_3) = P(E_5) = p, P(E_2) = P(E_4) = P(E_6) = 2p$$

Como $P(\mathcal{S}) = 1$, $3p + 3 \cdot 2p = 1$, entonces $p = 1/9$.

3) Arrojamus una moneda equilibrada 10 veces. Cual es la probabilidad que salgan exactamente 5 caras?

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

$$\mathcal{S} = \{(1), (0, 1), (0, 0, 1), (0, 0, 0, 1), \dots\}$$

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

El evento es $A = \{(0, 1), (0, 0, 0, 1), (0, 0, 0, 0, 1), \dots\}$

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

Espacios de equiprobabilidad: \mathcal{S} es finito y sea $n = \#\mathcal{S}$ (el símbolo $\#$ representa el cardinal del conjunto).

Diremos que el espacio es de equiprobabilidad si los n sucesos elementales tienen igual probabilidad, es decir si $P(E_i) = 1/n$, para todo i .

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

Se trata de un espacio de equiprobabilidad, $\mathcal{S} = \{1, 2, 3, 4, 5\} \times \{1, 2, 3, 4, 5\}$ entonces su cardinal es $\#\mathcal{S} = 5 \times 5 = 25$.

Supongamos que las bolillas 1 y 2 son blancas y las otras 3 rojas.

a) ¿Cuál es la probabilidad de que se extraiga al menos una bolilla roja?

b) ¿Cuál es la probabilidad de que la primera bolilla extraída sea roja y la segunda blanca?

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

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

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

2) Sucesiones de n 0 y 1. Lanzamiento de n monedas.

Si la moneda es honesta \mathcal{S} tiene 2^n elementos y todos tienen la misma proba $1/2^n$.

3) Problema de las 3 puertas. Tres puertas cerradas y un premio atrás 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?

Probabilidad condicional

100 personas

13 enfermos y no vacunados

2 enfermos y vacunados

75 sanos y vacunados

10 sanos s y no vacunados

Elijo una persona al azar de esa población y observo su estado

$\mathcal{S} = \{ev, en, sv, sn\}$, $E = \{ev, en\}$, $V = \{ev, sv\}$.

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

(cálculos hechos con casos favorables sobre posibles)

Cual es la probabilidad que una persona esté enferma?

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

Probabilidad que una persona vacunada esté enferma?

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

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

Definición de Probabilidad condicional: \mathcal{S}, P , Eventos A, B con $P(B) > 0$

$P(A|B) = P(AB)/P(B)$ es la *proba condicional* de A dado que conocemos B .

Observaciones

- $P(AB) = P(A|B)P(B)$
- $(B, P(\cdot|B))$ nuevo espacio de proba.

Ejemplos

Dados

Un dado. Calcule la probabilidad de ver un 3 dado que el resultado es a lo sumo 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

$S = \{vv, vm, mv, mm\}$, 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 conocida con dos hijos tiene por lo menos un hijo varón. Cual es la proba 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?

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

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

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

Dem: Por inducción. $P(A_1A_2) = P(A_1)P(A_2|A_1)$, por definición. $P(A_1 \dots A_n) = P(A_1 \dots A_{n-1})P(A_n|A_1 \dots A_{n-1})$ (por el caso de dos conjuntos) y la prueba sale aplicando la hipótesis inductiva a $P(A_1 \dots A_{n-1})$. □

Ejemplo Las 40 cartas de un mazo de cartas españolas son divididas en 4 pilas elegidas al azar.

Calcule la probabilidad que cada pila tenga exactamente un as.

Hay 40 lugares para poner las 40 cartas. Los primeros 10 lugares son la primera pila, etc.

Se retira la primera carta del mazo y se coloca en uno de los lugares elegido uniformemente. Después la segunda carta se coloca en uno de los lugares que quedan vacíos, elegido uniformemente, etc.

Demuestre que el orden en que se colocan las cartas no modifica la distribución final.

Empezamos colocando los 4 ases.

Defina los eventos:

A_1 = el as de espada está en cualquier pila.

A_2 = el as de bastos no está en la pila del as de espada.

A_3 = el as de copa no está en las pilas de los ases de espada y bastos.

A_4 = el as de oro no está en las pilas de los otros ases.

$A = A_1A_2A_3A_4$.

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3) = 1 \frac{30}{39} \frac{20}{38} \frac{10}{37}$$

Fórmula de la probabilidad total

Una *partición* de S es una familia de conjuntos disjuntos dos a dos B_i tal que

$$S = \dot{\cup}_i B_i$$

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

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

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

Teorema de la Probabilidad total Sea B_i una partición de S 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 $P(A) = P(\cup_i (A \cap B_i)) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$.

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:

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

Fórmula de Bayes

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

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

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

Vacunas

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

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

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

Juego de las 3 puertas B_i = premio en puerta i . $P(B_i) = 1/3$

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

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

$P(A|B_3) = 0$, $P(A|B_2) = 1$, $P(A|B_1) = 1/2$.

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

$$= \frac{1}{2} \frac{1}{3} + 1 \frac{1}{3} + 0 \frac{1}{3} = \frac{1}{2}$$

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

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

O sea que $P(\text{No cambiar de puerta y ganar}) = 1/3$ y

$P(\text{Cambiar de puerta y ganar}) = 2/3$

Simulación en R: ver [Monty Hall](#)

Independencia de eventos

Los eventos A y B son *independientes* si $P(AB) = P(A)P(B)$

porque $P(A|B) = P(A)$ etc.

Ejemplos. Dos dados. $A = \text{suma } 6$. $F = \text{primer dado } 4$. No son independientes.

$B = \text{suma } 7$. F y B son independientes.

Ejercicio: Probar que si A y B son independientes, entonces A y B^c también lo son.

Familia de eventos independientes

Tres eventos A, B, C son *independientes* si

$$P(ABC) = P(A)P(B)P(C), P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(CB) = P(C)P(B)$$

Si A, B, C son independientes entonces A es independiente de cualquier evento formado a partir de B y C .

Por ejemplo: C es independiente de $A \cup B$:

$$\begin{aligned} P(C \cap (A \cup B)) &= P(CA) + P(CB) - P(CAB) \\ &= P(C)[P(A) + P(B) - P(AB)] = P(C)P(A \cup B). \end{aligned}$$

Sea J un conjunto discreto de índices. Los eventos de una familia $(A_j, j \in J)$ son *independientes* si

$$\mathbb{P}(\bigcap_{i \in K} A_i) = \prod_{i \in K} P(A_i)$$

para cualquier subconjunto finito de índices $K \subset J$.

Ejemplo: infinitas monedas A_i = la i -ésima moneda es cara = sucesiones de 0's y 1's que tienen un 1 en la posición i .

Por ejemplo $P(A_1 A_2 \dots A_k) = \frac{1}{2^k}$.

Ejemplo dos dados son lanzados simultáneamente hasta que la suma de sus caras sea 5 o 7. Cual es la probabilidad que cuando aparece una de esas faces, sea un 5?

E_n = no aparece ni 5 ni 7 en los primeros $n - 1$ ensayos y aparece un 5 en la n -ésima tirada.

Estamos calculando

$$(*) = P(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} P(E_n)$$

porque los eventos son disjuntos. Proba de 5 es $4/36$ y proba de 7 es $6/36$. Proba de 5 o 7 en un ensayo fijo es $10/36$.

Entonces, llamando $A_j = \text{sale } 5 \text{ en la jugada } j$, B_j sale 7.

$$P(E_n) = P((A_1 \cup B_1)^c \dots (A_{n-1} \cup B_{n-1})^c A_n) = \left(1 - \frac{10}{36}\right)^{n-1} \frac{4}{36}$$

porque lo que pasa en las $n - 1$ primeras tiradas es independiente de lo que pasa en la n -ésima. Así

$$(*) = \sum_{n=1}^{\infty} \left(1 - \frac{10}{36}\right)^{n-1} \frac{4}{36} = \frac{2}{5}.$$

Solución usando proba condicional

$$P(E) = P(E|A_5)P(A_5) + P(E|A_7)P(A_7) + P(E|H)P(H)$$

donde $y H = \text{no sale ni } 5 \text{ ni } 7 \text{ en la primera jugada}$.

$$P(A_5) = \frac{4}{36}, P(A_7) = \frac{6}{36}, P(H) = \frac{26}{36}.$$

$P(E|A_5) = 1, P(E|A_7) = 0, P(E|H) = P(E)$. O sea:

$$P(E) = 1 \frac{4}{36} + 0 \frac{6}{36} + P(E) \frac{26}{36}$$

de donde $P(E) = \frac{2}{5}$.

Eventos 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.

Variable aleatoria

$$X : S \rightarrow \mathbb{R}$$

$$\text{Notación } \{X \in A\} = \{s \in S : X(s) \in A\}$$

Variable aleatoria **discreta** asume numerables valores todos con proba positiva.

Induce una partición en S : $(\{s \in S : X(s) = x\}, x \in R(X))$

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

Función de probabilidad puntual $p_X(x) = P(X = x)$ (o **distribución**)

Es una tabla.

Ejemplo Dos monedas, $S = \{00, 01, 10, 11\}$. $X = \text{número de caras}$. $X(00) = 0, X(01) = X(10) = 1, X(11) = 2$.

Induce la partición: $\{X = 0\} = \{00\}, \{X = 1\} = \{01, 10\}, \{X = 2\} = \{11\}$

Permite calcular la distribución:

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

Clase del 4 de abril

Ejemplo Suma de dos dados.

Ejemplo Geométrica.

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 $P(X = x)$.

Función de distribución acumulada

Def. $F_X(x) := P(X \leq x)$

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

i) para todo $x \in \mathbb{R}$, $F(x) \in [0, 1]$

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

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

iv) $\lim_{x \rightarrow \infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$

v) Altura del salto = probabilidad puntual: $p(x) = F(x) - F(x-)$

donde $F(x-) = \lim_{h \rightarrow 0} F(x-h)$

Uso La distribución acumulada de X caracteriza la función de probabilidad puntual. de X

$$P(a < X \leq b) = F(b) - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a-)$$

$$P(a \leq X < b) = F(b-) - F(a)$$

$$P(a < X < b) = F(b-) - F(a-)$$

Ejemplo. Distribución geométrica de parámetro p

$p \in (0, 1)$. Defino X con proba puntual

$p_X(k) = P(X = k) = (1-p)^{k-1}p$. Verifique que la suma es 1.

Éxito con proba p , fracaso con proba $1-p$.

Número de experimentos hasta el primer éxito.

$P(X > k) =$ proba de k fracasos $= (1-p)^k$.

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

Graficar la proba y la acumulada con $p = 1/2$.

Mostrar que los saltos son las probas.

Esperanza El esperanza o valor esperado de una variable aleatoria es definido como

$$EX = \sum_x xP(X = x)$$

(si la suma con el módulo existe $\sum_x |x|\mathbb{P}(X = x) < \infty$)

La suma es sobre el rango $R_X = \{x : P(X = x) > 0\}$

Ejemplos: 1) $X =$ dado.

2) número de caras en 2 monedas

3) variable Bernoulli $EX = P(X = 1)$

4) No existe: $P(X = x) = \frac{6}{\pi^2} \frac{1}{x^2}$.

Interpretaciones

Centro de gravedad.

Ley de grandes números.

Opciones ante un evento aleatorio

Billete de lotería vale \$1 con premio \$10⁶.

Probabilidad de ganar es 1/10⁷ (hay 10 millones de billetes).

$S = \{0, 1\}$, donde 1 = gana el billete, 0 = pierde el billete.

$$P(\{1\}) = \frac{1}{10^7}, \quad P(\{0\}) = 1 - \frac{1}{10^7}$$

Opción 1: comprar el billete; lucro $X(1) = 10^6 - 1$, $X(0) = -1$

$$EX = \frac{1}{10^7}(10^6 - 1) + (1 - \frac{1}{10^7})(-1) = -0,9$$

Opción 2: No comprar el billete: lucro $Y(1) = Y(0) = 0$

$$EY = 1(0) = 0,$$

“No podés perder si no jugás”.

Clase del 9 de abril (no hubo)

Clase del 11 de abril

Mintiendo con estadística

Un colegio tiene 3 aulas, con 5, 10 y 150 alumnos, respectivamente.

X = número de alumnos de un aula elegida al azar

$S = \{1, 2, 3\}$ equiprobable: $X(1) = 5$, $X(2) = 10$, $X(3) = 150$.

Cual es el tamaño promedio del aula

$$EX = \frac{1}{3}5 + \frac{1}{3}10 + \frac{1}{3}150 = \frac{165}{3} = 55$$

Número promedio de estudiantes por aula es 55.

Ahora elija *un estudiante* y vea de que tamaño es su aula.

$$S = \{1, 2, \dots, 165\}, \quad \text{equiprobable}$$

Y = tamaño del aula de un estudiante elegido al azar.

$$Y(k) = \begin{cases} 5, & \text{si } k \leq 5 \\ 10, & \text{si } 11 \leq k \leq 20 \\ 150, & \text{si } 21 \leq k \leq 165 \end{cases}$$

$$P(Y = 5) = \frac{5}{165}, P(Y = 10) = \frac{10}{165}, P(Y = 150) = \frac{150}{165}.$$

$$EY = \frac{5}{165}5 + \frac{10}{165}10 + \frac{150}{165}150 = 137$$

es el tamaño promedio del aula del estudiante elegido al azar.

Esperanza de la geométrica(p):

$$P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

$$EX = \sum_{k \geq 1} k(1 - p)^{k-1}p = -p \sum_{k \geq 1} ((1 - p)^k)' = -p \left(\sum_{k \geq 1} (1 - p)^k \right)'$$

$$= -p \left(\frac{1}{1 - (1-p)} - 1 \right)' = -p \left(\frac{1}{p} - 1 \right) = -p \left(-\frac{1}{p^2} \right) = \frac{1}{p}$$

Alternativamente: Si X asume valores naturales ≥ 0

$$EX = \sum_{x \geq 0} P(X > x)$$

Para la geométrica

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

Prueba de

$$EX = \sum_{x \geq 0} P(X > x)$$

$$\sum_{x \geq 0} \sum_{y \geq x+1} P(X = y) = \sum_{y \geq 1} \sum_{0 \leq x \leq y-1} P(X = y) = \sum_{y \geq 1} y P(X = y) = EX$$

Esperanza de una función de una v.a. $Y = g(X)$

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

Dem: Como $\{Y = y\} = \{g(X) = y\} = \dot{\cup}_{x:g(x)=y} \{X = x\}$,

$$P(Y = y) = \sum_{x:g(x)=y} P(X = x).$$

Entonces

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

Propiedades de la esperanza

1) (Linealidad) Si a y b son constantes reales, $E(aX + b) = aE(X) + b$.

Dem: Sea $h(X) = aX + b$, entonces

$$\begin{aligned} E(h(X)) &= \sum_x h(x) P(X = x) = \sum_x (ax + b) P(X = x) \\ &= \sum_x ax P(X = x) + b \sum_x P(X = x) = aEX + b \end{aligned}$$

2) Si X es una v.a. tal que $P(X = c) = 1$, entonces $E(X) = c$.

Dem: $EX = cP(X = c) = c$.

Viaje 400km a velocidad aleatoria (bici o auto)

V velocidad $P(V = 20) = \frac{1}{2}$; $P(V = 100) = \frac{1}{2}$

$$\text{Velocidad promedio: } EV = \frac{1}{2}20 + \frac{1}{2}100 = 60$$

Distancia = tiempo \times velocidad \Leftrightarrow Tiempo = distancia/velocidad

$T = 400/V$. Tiempo promedio:

$$ET = \frac{1}{2} \frac{400}{20} + \frac{1}{2} \frac{400}{100} = 12$$

Distancia = tiempo por velocidad ($d = TV$) pero

$$EV ET = 60 \cdot 12 \neq 400 = E(VT)$$

$$ET = 12 \neq \frac{400}{60} = \frac{\text{distancia}}{EV}$$

Esperanza condicional

Definición de probabilidad condicional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Como r.v. definen conjuntos en S , podemos definir para $x \in R_X$ y $R \subset R_X$,

$$P(X = k|X \in R) = \frac{P(\{X = k\} \cap \{X \in R\})}{P(X \in R)} = \frac{P(X = k)}{P(X \in R)}$$

si $x \in R$.

Hay una variable aleatoria Y que tiene esas probabilidades:

$$P(Y = k) = P(X = k|X \in R)$$

La esperanza condicional de X dado $X \in R$ se define

$$E(X|X \in R) = \sum_{k \in R} k P(X = k|X \in R)$$

Por ejemplo si X asume los valores $\{2, 5, 7\}$ con probas $\frac{3}{4}, \frac{1}{8}, \frac{1}{8}$,

$$E(X|X \geq 4) = 5 \frac{1/8}{1/4} + 7 \frac{1/8}{1/4} = 6$$

Mostrar en un gráfico que lo que hacemos es tomar parte del histograma multiplicando las probabilidades remanentes por una constante para que quede una proba.

La geométrica no tiene memoria X geometrica(p). Entonces

$$P(X = k + i|X > k) = \frac{p(1-p)^{k+i-1}}{(1-p)^k} = p(1-p)^{i-1} = P(X = i)$$

Vimos que $EX = \frac{1}{p}$.

Cual es $E(X|X > k)$?

$$\begin{aligned} \sum_{j=k+1}^{\infty} j P(X = j|X > k) &= \sum_{i=1}^{\infty} (k+i) P(X = k+i|X > k) \\ &= \sum_{i=1}^{\infty} (k+i) p(1-p)^{i-1} = k \sum_{i=1}^{\infty} p(1-p)^{i-1} + \sum_{i=1}^{\infty} ip(1-p)^{i-1} = k + \frac{1}{p} \end{aligned}$$

Si esperé k minutos, en media esperaré $1/p$ minutos más

(lo mismo que tenía en media cuando llegué a la parada)

Teorema de la esperanza total A_1, \dots, A_n partición del espacio muestral. B evento.

Teorema de la proba total: $P(B) = \sum_i P(B|A_i)P(A_i)$

$$P(X = k) = \sum_i P(X = k|A_i)P(A_i)$$

Lema $EX = \sum_i E(X|A_i)P(A_i)$

Dem $EX = \sum_k kP(X = k) = \sum_k k \sum_i P(X = k|A_i)P(A_i)$

$$\sum_i \sum_k kP(X = k|A_i)P(A_i) = \sum_i E(X|A_i)P(A_i) \quad \square$$

Varianza de una v.a. discreta:

Consideremos las siguientes distribuciones:

x	-1	0	1
P(X=x)	1/3	1/3	1/3

x	-10	0	10
P(Y=y)	1/3	1/3	1/3

z	-100	0	100
P(Z=z)	1/3	1/3	1/3

Vea que $EX = EY = EZ = 0$.

Sin embargo sus histogramas están dispersos alrededor de la media de forma diferente.

Def. La **varianza** de una v.a. X es definida por

$$VX = E(X - EX)^2 = \sum_x (x - EX)^2 P(X = x) = \sigma^2$$

El **desvío standard** $\sigma := \sqrt{VX}$

Fórmula alternativa

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

Dem:

La media minimiza el desvío cuadrático medio Sea X una va discreta con distribución $p(x) = P(X = x)$.

Buscamos m tal que

$$\sum_x (x - m)^2 p(x) = \text{mín}$$

Para eso derivamos en m :

$$-2 \sum_x (x - m) p(x) = 0$$

De donde

$$m = \sum_x x p(x) = EX$$

Y la varianza es el valor del desvío cuadrático mínimo.

Ejemplos: 1) varianza de X Y Z arriba:

$$VX = \quad VY = \quad VZ =$$

2) X = número de caras pares de dos dados equilibrados

x	0	1	2
P(X=x)	1/4	1/2	1/4

3) Bernoulli.

4) Geométrica. $EX^2 - (EX)^2 = \frac{1-p}{p^2}$

Propiedades de la Varianza

$$V(aX + b) = a^2 VX$$

usar formula del estadístico inconciente

Desvio standard

$$DX = \sqrt{VX}$$

$$D(aX + b) = |a| DX$$

Si X es constante, entonces $VX = 0$.

Distribución Bernoulli y binomial Jacob Bernoulli (1654-1705), matemático suizo. Demuestra la ley débil de grandes números para variables Bernoulli.

Variable aleatoria Bernoulli: $X \in \{0, 1\}$

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

$X \sim \text{Bernoulli}(p)$.

$$EX = p, \quad VX = p(1 - p)$$

El proceso de Bernoulli

Sucesión de ensayos de Bernoulli

Espacio muestral: $S = \{(a_1, a_2, \dots, a_\ell), a_i \in \{0, 1\}\}$

Se puede pensar que $\ell = \infty$ o es muuuy grande.

Simulación de a : 11000010101100010100010110010100010100001

Modelo: El evento $B = (\text{todas las sucesiones } a_1, a_2, \dots \text{ que coinciden con } b_1, \dots, b_n \text{ en las primeras } n \text{ coordenadas})$ tiene probabilidad

$$p(b_1, \dots, b_n) = p^{\sum b_i} (1 - p)^{\sum (1 - b_i)}$$

Más fácil si definimos las variables aleatorias *proyección*:

$$X_i(a) := a_i; \quad a = (a_1, a_2, \dots)$$

Se deduce que la probabilidad de “éxito en i -ésimo ensayo” es

$$P(X_i = 1) = p,$$

Y el evento B se puede escribir

$$B = \{X_1 = b_1, \dots, X_n = b_n\}$$

Ejemplos: Los **colectivos 107** salen de la terminal cada un minuto con probabilidad $p = 1/10$. Cual es la probabilidad que pasen 3 107 seguidos en los minutos 1,2,3? Queremos calcular

$$P(X_1 = 1, X_2 = 1, X_3 = 1) = \left(\frac{1}{10}\right)^3.$$

En un casino se juega al rojo representado por 1 o negro representado por 0. Cual es la probabilidad de ganar apostando al rojo en una **ruleta** con cero? $p = 18/37$.

Propiedades del proceso de Bernoulli

Estacionario Para cada n , la sucesión futura a_{n+1}, a_{n+2}, \dots es un proceso de Bernoulli independiente del pasado.

Sin memoria El pasado es independiente del futuro. Por ejemplo si no ocurre ningún éxito hasta el instante n , el tiempo hasta el próximo éxito es geométrico.

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. Geométrica $1/2$.

Premio: $g(N) = 2^N$. $Eg(N) = \sum_{n \geq 1} 2^n 2^{-n} = \infty$.

Cuanto pagarías para participar de este juego? digamos K

$X =$ ganancia $= 2^N - K$.

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

La probabilidad de ganar lo mismo que aposté es

$$\frac{1}{2^{21}}$$

mmmmmm...

Clase del 16 de abril

Juego de Las Vegas

Jugamos al rojo en las Vegas. Ruleta sin 0.

Martingala (un método infalible para ganar):

0) Fijo $K = 0, L = 0$.

1) Apuesto 2^K .

2) Si sale rojo $L \leftarrow L + 2^K$ y vuelvo a (0).

3) Si sale negro $L \leftarrow L - 2^K, K \leftarrow 2K$ vuelvo a (1).

Cada vez que sale rojo gano \$1.

Dem: Si perdí K veces, perdí $1 + 2 + 4 + \dots + 2^{K-1} = 2^K - 1$

Apuesta actual $= 2^K$. Si sale rojo el lucro neto es

$$L = 2^K - (2^K - 1) = 1$$

Simulación: 1 = sale rojo, 0 = sale negro.

Apuesto	1	1	2	1	2	4	8	1	2	1	2	4	8
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	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: $N = \text{geométrica}(p)$.

$L = \text{lucro después del primer rojo. } L = g(N) = 1$. Como L es constante...

$$EL = \sum_{n \geq 1} L(n)P(N = n) = 1 \sum_{n \geq 1} P(N = n) = 1.$$

Problema: la fortuna es finita o no se permiten apuestas mayores que 2^8 (por ejemplo).

Si perdemos 8 veces seguidas perdemos $2^8 - 1$. Recomenzamos con 1.

En ese caso el lucro (hasta ganar 1 vez o perder 8 seguidas):

$$L = g(N) = 1 \mathbf{1}\{N \leq 8\} - (2^8 - 1) \mathbf{1}\{N > 8\}$$

$$EL = 1 - 2^8(1 - p)^8$$

Si $p = 1/2$ da $EL = 0$ (juego honesto).

Si $p < 1/2$ da $EL < 0$ (no jugar).

Si $p > 1/2$ da $EL > 0$ (conviene jugar).

Distribución binomial En los ensayos de Bernoulli. $a = (a_1, a_2, \dots)$

$S_n(a) = a_1 + \dots + a_n$ número de éxitos en n ensayos.

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

$$ES = \sum_{k=0}^n \binom{n}{k} k p^k (1 - p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1 - p)^{n-k} = np$$

$$VS = np(1 - p)$$

Hay que calcular $E(X(X - 1))$

Tiempo entre dos éxitos sucesivos $T_1(a) = \min\{k > 0 : a_k = 1\}$ tiene distribución geométrica:

$$P(T_1 = k) = P(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1 - p)^{k-1} p$$

(depende de finitas coordenadas, se puede calcular)

Note que si llego en un instante cualquiera:

$R_i = \min\{k > i : a_k = 1\} - i$ tiempo de espera del ómnibus si llego en el instante i .

$$P(R_i = k) = P(X_{i+1} = 0, \dots, X_{i+k-1} = 0, X_{i+k} = 1) = (1 - p)^{k-1} p$$

Instante de la k -ésima llegada $Y_k = \min\{n : a_1 + \dots + a_n = k\}$ instante en que ocurre el k -ésimo éxito.

Para $t \geq k$:

$$P(Y_k = t) = P(k - 1 \text{ éxitos entre } 0 \text{ y } t - 1, \text{ éxito en } t)$$

$$= \binom{t-1}{k-1} p^{k-1} (1 - p)^{t-1-(k-1)} p$$

Binomial negativa o Pascal: Dos parametros, k y p

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

$$EY_k = \frac{k}{p}, \quad VY_k = \frac{k(1 - p)}{p^2}$$

Aproximación Poisson de la binomial

S_n Binomial($n, p(n)$)

$p(n) = \lambda/n$, λ parametro.

Lemma Vale

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Dem:

$$\begin{aligned} P(S_n = k) &= \binom{n}{k} p(n)^k (1 - p(n))^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n!}{(n-k)! n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

Pero

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \frac{n!}{(n-k)! n^k} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n^k} = 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} &= 1 \end{aligned}$$

Lo que prueba el Lema. \square

Vale para $n \geq 100$ y $p < 0,01$, np “moderado”

Distribución de Poisson

Simeon-Denis Poisson (1781-1840).

$\lambda > 0$ real.

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

Recordemos que por Taylor:

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

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

Cálculo de EX, VX .

Proceso de Poisson t tiempo real.

- $N(t)$ número de llegadas en el intervalo $[0, t]$.
- $P(N(t) - N(s) = k) = \text{Poisson}(\lambda(t - s))$
- Llegadas en diferentes intervalos son independientes.
- Para δ chico

$$P(N(t + \delta) - N(t) = 1) = \lambda \delta + o(\delta)$$

$$P(N(t + \delta) - N(t) \geq 2) = o(\delta)$$

- λ es la tasa de llegada del proceso.

Proceso de Bernoulli aproxima al proceso de Poisson

n natural “grande”. Proceso de Bernoulli en \mathbb{Z}/n ,

Ponemos tn ensayos en el intervalo $[0, t]$. Cada ensayo dura $1/n$.

Probabilidad de éxito en cada ensayo $p(n) = \lambda/n$.

λ, t fijos (macroscópicos).

$p(n), n \rightarrow \infty$ (microscópicos).

El número de éxitos en el intervalo $[0, t]$ para n fijo es Binomial($nt, \lambda/n$).

En la aproximación Poisson, el número de ensayos converge a Poisson(λt) (usando lo que vimos recién)

Ejemplo de Proceso de Poisson

El número de mensajes mails que llegan a una casilla es proceso de Poisson de intensidad $\lambda = 2$ mensajes / minuto.

$N(t)$ = número de mensajes entre 0 y t .

a) ¿Cuál es la probabilidad de que no se reciba ningún mensaje entre las 12 hs y las 12:03 hs?

$N(3) \sim \text{Poisson}(2 \cdot 3) = \text{Poisson}(6)$. $P(N(3) = 0) = e^{-6} = 0,002$.

b) ¿Número esperado de mensajes en media hora? $N(30) \sim \text{Poisson}(2 \cdot 30) = \text{Poisson}(60)$.

$E(N(30)) = 60$.

c) ¿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).

Tiempo de la primera llegada

Cual es la probabilidad que la primera llegada del proceso de Poisson(λ) sea después de t ?

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

Primer ejemplo de variable aleatoria continua.

En particular, si T_1^n es el tiempo de la primera llegada en el proceso de Bernoulli discreto,

T_1^n geométrica(λ/n)

$$P(T_1^n > nt) = \left(1 - \frac{\lambda}{n}\right)^{nt} \rightarrow e^{-\lambda t}$$

Clase del 18 de abril

Variables aleatorias continuas

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

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

Cuando n es grande X_n aproxima una variable aleatoria X “esencialmente continua” (“tiempo”), $X \in [0, 1]$.

Histogramas con área total igual a 1.

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

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

$$P(X_n \in [a, b]) = [(b - a)n] \frac{1}{n} \rightarrow_{n \rightarrow \infty} b - a$$

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

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

A Boreliano medible, etc.

Para $A = [a, b]$ (intervalo)

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

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

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

Función de distribución acumulada

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx$$

Calcular la F de la variable X

Propiedades de la función de distribución acumulada:

X v.a. continua,

i) para todo $x \in \mathbb{R}$, $F(x) \in [0, 1]$.

ii) $F(x)$ es monótona no decreciente, es decir ...

iii) $F(x)$ es continua en todo punto.

iv) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$

Lema. Si X es continua y $a \leq b$ reales, vale

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

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

Lema. Si X continua con $f(x)$ y $F(x)$, entonces en todo punto 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

Distribución Uniforme: X tiene distribución uniforme en el intervalo $[A, B]$, si su función de densidad es

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

Notación: $X \sim U(A, B)$.

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 $x \notin \{A, B\}$.

Percentiles de una distribución continua: Sea X una v.a. continua con $f(x)$ y $F(x)$ y sea $0 < p < 1$. El percentil (100 p)-ésimo de la distribución de X es el valor x_p tal que

$$P(X < x_p) = p$$

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

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

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

Percentil $p = 0,25$. $x_p \in [1, 3]$:

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

2) X Uniforme(A, B). Acumulada:

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

Buscamos el percentil $p = 0,5$:

$$0,5 = F(x_{0,5}) \Rightarrow 0,5 = \frac{x_{0,5} - A}{B - A} \Rightarrow x_{0,5} = \frac{A + B}{2}$$

Mediana: Es el percentil $p = 0,5$.

Esperanza o valor esperado de una v.a. continua:

Definición: Sea X con densidad $f(x)$, la esperanza o valor esperado de X se define como

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

si $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$. Si no, decimos que no existe.

Ejemplo: Sea $X \sim$ Uniforme(A, B),

$$EX = \frac{A + B}{2}$$

Lema. Si X tiene densidad $f(x)$ y $h : \mathbb{R} \rightarrow \mathbb{R}$, entonces

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

si la integral del modulo es finita.

Porqué esa definición de esperanza? Sea $X \in [0, K]$ una variable aleatoria continua acotada por K entero y X_n una aproximación discreta de X definida por

$$X_n = h_n(X) = \frac{k}{n} \mathbf{1}\left\{\frac{k}{n} \leq X < \frac{k+1}{n}\right\}, \quad k \in \{0, \dots, nK - 1\}$$

X_n asume nK valores. Note que $|X_n - X| \leq \frac{1}{n}$.

$$\begin{aligned} EX_n &= \sum_{k=0}^{nK-1} \frac{k}{n} P\left(X_n = \frac{k}{n}\right) = \sum_{k=0}^{nK-1} \frac{k}{n} P\left(\frac{k}{n} \leq X < \frac{k+1}{n}\right) \\ &= \sum_{k=0}^{nK-1} \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x) dx = \sum_{k=0}^{nK-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} h_n(x) f(x) dx \\ &= \int_0^K h_n(x) f(x) dx \end{aligned}$$

Ahora calculemos

$$|EX_n - \int_0^K x f(x) dx| \leq \int_0^K |h_n(x) - x| f(x) dx \leq \frac{1}{n} \int_0^K f(x) dx = \frac{1}{n}$$

O sea, si X_n converge a X y es acotada, entonces EX_n converge a EX como fue definida con la integral.

Linealidad:

Si a y b son constantes reales,

$$E(aX + b) = aE(X) + b.$$

Dem: Sea $h(X) = aX + b$,

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) f(x) dx = \int_{-\infty}^{\infty} (ax + b) f(x) dx \\ &= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx = aE(X) + b. \end{aligned}$$

Ejemplo: Dos especies compiten para controlar recurso dividido en dos partes con la distribución uniforme. Sea X : proporción del recurso controlada por la especie 1. $X \text{ Uniforme}(0,1)$:

$$f(x) = \mathbf{1}\{x \in [0, 1]\}$$

“vara rota” análogo a quebrar una vara en un punto aleatorio.

Cual es la proporción promedio que controla la especie que controla la mayoría del recurso.

La mayor proporción es la variable

$$h(X) = \max(X, 1 - X) = X\mathbf{1}\{X > 1/2\} + (1 - X)\mathbf{1}\{X \leq 1/2\}$$

y su esperanza es

$$\begin{aligned} Eh(X) &= E(X\mathbf{1}\{X > 1/2\}) + E((1 - X)\mathbf{1}\{X \leq 1/2\}) \\ &= \int_{1/2}^1 x dx + \int_0^{1/2} (1 - x) dx = 3/4 \end{aligned}$$

Fórmula para la esperanza de variables positivas

Lema. Si $X \geq 0$ es continua con densidad f y acumulada F y $\int_0^\infty x f(x) dx < \infty$, entonces

$$EX = \int_0^\infty (1 - F(x)) dx$$

Dem. Partes: $u = x$, $du = dx$, $v = -(1 - F(x))$, $dv = f(x)dx$.

$$EX = \int_0^{\infty} xf(x)dx = -[x(1 - F(x))]_0^{\infty} + \int_0^{\infty} (1 - F(x))dx$$

Veamos que $\lim_{x \rightarrow \infty} [x(1 - F(x))] = 0$:

$$\int_x^{\infty} yf(y)dy \geq x \int_x^{\infty} f(y)dy = x(1 - F(x))$$

como $\int_0^{\infty} xf(x)dx < \infty$, el lado izquierdo va a 0 cuando $x \rightarrow \infty$. \square

Varianza de una v.a. continua:

Definición: Sea X una v.a. continua con esperanza μ y densidad f , la varianza de X , que se denotará $V(X)$, σ^2

$$VX = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Desvío standard: $\sigma = +\sqrt{VX}$

Lema. Vale: $V(X) = E(X^2) - (E(X))^2$.

Ejemplos: Sea X Uniforme(A,B), $EX = (A + B)/2$

$$VX = E(X^2) - (E(X))^2 = \frac{(B - A)^2}{12}$$

Linealidad:

$$V(aX + b) = a^2VX, \quad \sigma_{aX+b} = |a|\sigma_X$$

Clase del 23 de abril

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

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 simétrica en relación a μ : $f(\mu + x) = f(\mu - x)$

Alcanza el máximo en $x = \mu$

Distribución normal standard

Def: $Z \sim N(0, 1)$ si $\mu = 0$ y $\sigma^2 = 1$.

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Tabulada: $Z \sim N(0, 1)$, el percentil 99 de la distribución es 2.33

Propiedades:

• Si $X \sim N(\mu, \sigma^2)$ entonces $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

Prueba:

$$F_Z(z) = \dots = F_X(\sigma z + \mu)$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X(\sigma z + \mu) = f_X(\sigma z + \mu) \sigma$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}\right) \sigma = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

- Si Z normal standard y $X = \sigma Z + \mu$ entonces $X \sim N(\mu, \sigma)$.

Esperanza y varianza de la normal Se calcula primero para la distribución de la normal standard Z

$$EZ = \frac{1}{\sqrt{2\pi}} \int z e^{z^2/2} dz = 0$$

Integrando impar. Integrando por partes se obtiene también:

$$VZ = EZ^2 = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) = 1$$

Se exporta para la normal $X \sim N(\mu, \sigma)$ por la formula $X = \sigma Z + \mu$:

$$EX = \mu, \quad VX = \sigma^2$$

Cálculo de probabilidades para la Normal

Para la Normal standard, por simetría:

$$P(Z < x) = P(Z > -x)$$

Defina $\Phi(z) = P(Z \leq z)$ la acumulada de la Normal standard. Está tabulada.

$X \sim N(\mu, \sigma^2)$, $(X - \mu)/\sigma \sim N(0, 1)$.

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right)$$

$$= P\left(Z \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

- Si Z normal standard y $X = \sigma Z + \mu$. Entonces los percentiles satisfacen

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

Ejemplos

1. $X \sim N(3, 9)$. Calcular $P(2 < X < 5)$, $P(X > 0)$ y $P(|X - 3| > 6)$

$$P(2 < X < 5) = \dots = \Phi\left(\frac{2}{3}\right) - \left(1 - \Phi\left(\frac{1}{3}\right)\right) \sim 0,3779$$

2. Las notas de su examen siguen una normal de media μ y varianza σ^2 . Se estima μ y σ^2 y después se dan las notas. Nota A para quien tiene nota mayor que $\mu + \sigma$, nota B entre μ y $\mu + \sigma$, nota C entre $\mu - \sigma$ y μ y nota D para aquellas menores que $\mu - \sigma$. Por ejemplo $\mu = 72$, $\sigma^2 = 100$. (A rigor, no puede haber números menores que 0 ni mayores que 100, y las notas asumen valores discretos, pero la normal aquí es usada como modelo para calcular las probabilidades de los valores discretos.)

Calcule el porcentaje de alumnos que sacará cada una de las notas.

3. (Antes de la popularización de los tests de ADN) Un experto obstetra declara en un juicio de paternidad que la gestación de un bebé tiene distribución normal con parámetros $\mu = 270$ días y $\sigma^2 = 100$. El acusado puede probar que estuvo fuera del país durante un período que comenzó 290 días antes del nacimiento y

terminó 240 días antes del nacimiento. En base a esta declaración, el juez declara que el acusado no es el padre. Cual es la probabilidad que el juez se haya equivocado? Es decir, cual es la probabilidad que si el acusado fue el verdadero padre, la madre haya tenido un ciclo de gestación compatible con la ausencia del padre?

X = número de días de gestación. $X \sim N(270, 100)$. $-X$ = fecha de comienzo del embarazo contado desde el día del nacimiento. Queremos calcular la probabilidad que $-X$ sea menor que -290 o mayor que -240 .

$$P(-X < -290) + P(-X > -240)$$

por simetría esto es igual a

$$= P(X > 290) + P(X < 240) = \dots = 0,03,$$

las cuentas se hacen standarizando las variables y usando la tabla.

Variable exponencial Decimos que X tiene distribución exponencial de parámetro λ si su densidad es

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

$$F(x) = (1 - e^{-\lambda x}) \mathbf{1}\{x \geq 0\}$$

Calculemos EX y VX

$$EX^n = \int_0^{\infty} x^n \lambda e^{-\lambda x} dx = \dots = \frac{n}{\lambda} EX^{n-1}$$

Con $n = 1$ obtenemos

$$EX = \frac{1}{\lambda}, \quad EX^2 = \frac{1}{\lambda} EX = \frac{2}{\lambda^2}$$

de donde

$$VX = \frac{1}{\lambda^2}$$

La exponencial no tiene memoria:

$$P(X > t + s | X > t) = P(X > s).$$

Ejemplo: Supongamos que el tiempo de respuesta de una terminal conectada en línea es una v.a. X con distribución exponencial con esperanza igual a 5 segundos.

- Cuál es la probabilidad de que el tiempo de respuesta sea mayor de 10 segundos?
- Cuál es la probabilidad de que el tiempo de respuesta esté entre 5 y 10 segundos?
- Cual es la probabilidad que sabiendo que ya esperé 10 segundos, tenga que esperar todavía 5 segundos más?

Clase del 25 de abril Distribución Gama Una variable aleatoria 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.

Relación de Gama con Poisson

Lema Si T_n es el instante del n -ésimo evento de un proceso de Poisson(λ), entonces T_n tiene distribución Gama(n, λ).

Dem

$$F(t) = P(T_n \leq t) = P(N(t) \geq n)$$

donde $N(t)$ es un proceso de Poisson de parámetro λ .

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

Diferenciando en t ,

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

que es la densidad de la Gama(n, λ).

Ejercicio: Calcule EX y VX .

$$EX = \int_0^{\infty} x \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} = \frac{\alpha}{\lambda}$$

VX queda como ejercicio.

Generación de números aleatorios

Cual es la probabilidad de ganar al solitario?

52 cartas. Hay 52! juegos posibles de solitario. Supongamos que tenemos una estrategia fija. Es decir, dada una de las permutaciones, hay una función $X \in \{0, 1\}$ donde X es 0 si la estrategia pierde y 1 si gana con esa permutación.

Cual es la proba de ganar? $p = P(X = 1)$.

Como hay que jugar cada permutación para saber si ganamos o perdemos, es imposible calcular la proporción de juegos en los que se gana.

Pero lo que se puede hacer es generar n juegos elegidos aleatoriamente entre las 52! permutaciones, determinar X para cada uno de los juegos y definir

$$\hat{p}_n = \frac{\#\text{juegos ganados}}{n}$$

Despues veremos que \hat{p}_n converge a p en algún sentido.

Esto motiva el interés de *simular* variables aleatorias.

Generación de números pseudo-aleatorios

Método de la congruencia Dados m, a, c y X_0 ,

$$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0$$

X_{n+1} resto entero de dividir $X_n + c$ 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: $m = 2^{64}$, $a = 6364136223846793005$, $c = 1442695040888963407$

Ver wikipedia: Linear congruential generator

Generadores de verdaderos números aleatorios

Recomiendo fuertemente visitar la página <http://www.random.org>

de donde saqué estas observaciones: PRNG son los generadores de números pseudo 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 die connected to a computer, but typically people use a physical phenomenon that is easier to connect to a computer than a die 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.”

Generación de una permutación aleatoria $n \geq 2$ números.

0. Inicialización: $k = n$, $X(i) = i$, $i = 1, \dots, n$

1. Genere una uniforme V_k en $\{1, \dots, k\}$

2. Intercambie los valores de $X(V_k)$ y $X(k)$.

3. Ponga $k \leftarrow k - 1$.

4. Si $k = 1$ imprima $X(1), \dots, X(n)$. Si no, vuelva a 1.

Ejemplo: suponga que $n = 5$ y que $V(5) = 4$, $V(4) = 2$, $V(3) = 1$, $V(2) = 1$. Entonces tendremos

12345, 12354, 15324, 35124, 53124

Lema. Los números $X(1), \dots, X(n)$ son una permutación uniforme de $1, \dots, n$.

Dem. Cada número tiene probabilidad $\frac{1}{n}$ de ser el último y por inducción ... \square

Generación de variables uniformes discretas

Sea U Uniforme en $[0, 1]$.

Sea $V_n = [Un] + 1$ (parte entera)

Veamos que V_n es uniforme en $\{1, \dots, n\}$:

$$\begin{aligned} P(V_n = k) &= P([Un] + 1 = k) = P([Un] = k - 1) \\ &= P(k - 1 \leq Un < k) = P\left(\frac{k-1}{n} \leq U < \frac{k}{n}\right) = \frac{1}{n} \end{aligned}$$

En general, para generar una variable uniforme en $\{m, \dots, m + n - 1\}$,

$$V_n = [Un] + m$$

Generación de variables aleatorias discretas Sea X una variable aleatoria discreta con probabilidad puntual

$$P(X = x) = p(x),$$

Sea U uniforme en $[0, 1]$. Sea $(J(x) : x \in R_X)$ una partición del intervalo $[0, 1]$. Defina

$$X = x \quad \text{si } U \in J(x)$$

Equivalentemente:

$$X = \sum_x x \mathbf{1}\{U \in J(x)\}$$

Defina la función inversa generalizada por

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}$$

Defina

$$X = F^{-1}(U)$$

Si definimos

$$J(x) = [F(x-), F(x))$$

$$X = x \Leftrightarrow U \in J(x)$$

Lo que implica

$$P(X = x) = P(U \in J(x)) = |J(x)| = F(x) - F(x-) = p(x)$$

Ejemplo. Simule la variable con distribución

z	1	3	9
P(Z=z)	1/2	1/4	1/4

Acoplamiento

En este contexto un acoplamiento de dos variables aleatorias X e Y es la simulación de ambas en función de un mismo número aleatorio.

Ejemplo: Queremos generar variables Y_ℓ Bernoulli con parámetro p_ℓ . Una manera es hacer lo siguiente:

$$Y_\ell = F_\ell^{-1}(U) = \mathbf{1}\{U > 1 - p_\ell\}$$

Las variables generadas tienen la distribución correcta:

$$P(Y_\ell = 1) = P(U > 1 - p_\ell) = p_\ell.$$

y satisfacen la siguiente propiedad de monotonía:

Si $p_1 \leq p_2$ entonces $Y_1 \leq Y_2$.

En general, si $1 - F_1(y) \leq 1 - F_2(y)$ para todo y y $Y_\ell := F_\ell^{-1}(U)$ entonces

$$Y_1 \leq Y_2.$$

Lo que nos da una noción de orden entre variables aleatorias.

Clase del 30 de abril Ejemplo. Sucesiones de Bernoulli

Construya un programa para generar una sucesión de variables Bernoulli de tamaño arbitrario n de 0's y 1's con parametro $p \in [0, 1]$.

Generación de variables aleatorias continuas

Método de inversión. X una va continua con densidad f y acumulada F .

Supongamos F estrictamente creciente.

U uniforme en $[0, 1]$.

Lema. La variable $Y = F^{-1}(U)$ tiene la misma distribución que X .

Obs: la F es monótona. Como no es estrictamente creciente, necesitamos la definicion de inversa generalizada.

Dem.

$$P(Y < a) = P(F^{-1}(U) < a) = P(U < F(a)) = F(a)$$

□

Generación de una exponencial λ

$$F(x) = 1 - e^{-\lambda x}, x \geq 0$$

$$F^{-1}(u) = \frac{-\log(1-u)}{\lambda}$$

Entonces la variable definida por

$$X = \frac{-\log(1-U)}{\lambda}$$

con U uniforme en $[0, 1]$ 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

Queremos generar una variable con densidad f .

Sabemos como generar una variable con densidad g

Sabemos que existe $c > 0$ tq

$$f(x) \leq cg(x) \quad \text{para todo } x$$

Algoritmo del rechazo

1. Simule X_1 con densidad g y U uniforme en $[0, 1]$
2. Si $U \leq f(X_1)/cg(X_1)$, ponga $X = X_1$ y termine.

Si no, vaya a 1.

La variable X así generada tiene densidad f .

Generación de una variable normal standard Z

No se puede usar el método de inversión.

Empezamos a generar $X = |Z|$, que tiene densidad

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \geq 0$$

Considere $g(x) = e^{-x}$, $x \geq 0$. Cuenta:

$$\frac{f(x)}{g(x)} \leq \sqrt{\frac{2e}{\pi}}$$

de donde $c = \sqrt{\frac{2e}{\pi}}$ y

$$\frac{f(x)}{cg(x)} = \exp\left(\frac{-(x-1)^2}{2}\right)$$

El algoritmo queda:

1. Genere Y exponencial de parametro 1, U uniforme en $[0, 1]$
2. Si

$$U \leq \exp\left(\frac{-(Y-1)^2}{2}\right)$$

ponga $X = Y$. Si no, vaya a (1).

Ahora defina $Z = VX - (1-V)X$, con V Bernoulli(1/2).

Z es Normal(0, 1).

Simplificación En el paso (2) Y es aceptada si

$$U \leq \exp\left(\frac{-(Y-1)^2}{2}\right)$$

que es equivalente a

$$-\log U \geq \frac{-(Y-1)^2}{2}$$

como $Y_2 = -\log U$ es exponencial (1),

1. Genere Y_1, Y_2 exponenciales (1)
2. Si $Y_2 \geq \frac{-(Y_1-1)^2}{2}$ ponga $X = Y_1$. Si no, vaya a (1).

Función generadora de momentos

Definición: momento de orden k de X , EX^k siempre que la esperanza exista.

$E(X) = \mu$ 1er momento: posición

$E(X^2) = \sigma^2 + \mu^2$ 2do momento: medida de dispersión

$E(X^3)$ 3er momento: medida de asimetría

$E(X^4)$ 4to momento: kurtosis (puntiaguda o chata)

Def: función generadora de momentos de X es una función $M_X : \mathbb{R} \rightarrow \mathbb{R}$, definida como

$$M_X(t) = E(e^{tX})$$

si existe para $t \in (-h, h)$ para algún h . Condicion tecnica para que $M(t)$ sea diferenciable en 0.

Los momentos determinan la FGM

Desarrollando en serie $e^\lambda = \sum_{k=0}^{\infty} \lambda^k/k!$, obtenemos

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} EX^k \frac{t^k}{k!}$$

El intercambio de suma con esperanza: ????

Porque generadora de momentos?

Teorema. Sea X con FGM $M_X(t)$. Entonces

$$EX^n = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$$

Dem. Prueba corta:

$$\frac{d^n}{dt^n} M_X(t) = E \left(\frac{d^n}{dt^n} e^{tX} \right) = E(X^n e^{tX})$$

(Pero hay que justificar el pase de la derivada dentro de la esperanza.) Calculando en $t = 0$ obtenemos el teorema.

Para entender mejor tratamos los casos discreto y continuo

$$\begin{aligned} \frac{d^n}{dt^n} M_X(t) &= \frac{d^n}{dt^n} E e^{tX} = \frac{d^n}{dt^n} \sum_x e^{tx} p(x) = \sum_x \frac{d^n}{dt^n} e^{tx} p(x) \\ &= \sum_x x^n e^{tx} p(x) \end{aligned}$$

que da EX^n al calcular la suma para $t = 0$.

La misma cuenta vale para el continuo:

$$\begin{aligned} \frac{d^n}{dt^n} M_X(t) &= \frac{d^n}{dt^n} \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} \frac{d^n}{dt^n} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} x^n e^{tx} f(x) dx. \end{aligned}$$

que da EX^n al calcular la integral para $t = 0$. \square

Ejemplos

Exponencial

$$M(t) = \frac{\lambda}{\lambda - t}$$

Momentos

$$EX = \frac{1}{\lambda} \quad VX = \frac{1}{\lambda^2}$$

binomial

$$M(t) = (e^{tp} + (1-p))^n$$

media varianza

Propiedad $Y = aX + b$ entonces

$$M_Y(t) = e^{bt} M_X(at)$$

Teorema de Unicidad: Si existe la función generadora de momentos de una variable aleatoria, es única. Además la función generadora de momentos determina a la función de densidad o probabilidad de la v.a. salvo a lo sumo en un conjunto de probabilidad 0.

Vamos a probar el teorema solo cuando X asume un número finito de enteros no negativos.

Prueba cuando $R_X = \{0, \dots, n\}$ Fije $p(j) = P(X = j)$ y escriba

$$M(t) = \sum_{j=0}^n e^{tj} p(j)$$

$M(t)$ es un polinomio en $z = e^t$. Si definimos

$$H(z) = \sum_{j=0}^n z^j p(j)$$

H es la *función generatriz*.

H es un polinomio en z que da la misma info que M . Conocemos H si y solo si conocemos M .

Como H es un polinomio de grado n , usando Taylor:

$$p(j) = [\text{coeficiente de } z^j \text{ en } H(z)] = \frac{H^{(j)}(0)}{j!} \quad \square$$

Ejemplo Sea X con momentos $\mu_k = EX^k$ dados por

$$\mu_0 = 1, \quad \mu_k = \frac{1}{2} + \frac{2^k}{4}, \quad \text{para } k \geq 1$$

Calcule la distribución de X .

$$M(t) = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} = 1 + \frac{1}{2} \sum_{k=1}^{\infty} \frac{t^k}{k!} + \frac{1}{4} \sum_{k=1}^{\infty} \frac{(2t)^k}{k!} = \frac{1}{4} + \frac{1}{2} e^t + \frac{1}{4} e^{2t}$$

$$H(z) = \frac{1}{4} + \frac{1}{2} z + \frac{1}{4} z^2$$

De donde $p(0) = \frac{1}{4}$, $p(1) = \frac{1}{2}$, $p(2) = \frac{1}{4}$

Clase del 2 de mayo Vectores aleatorios

Ejemplo Lanzamiento de una moneda dos veces. El resultado es un vector (X, Y)

Dos tipos de estudiante: el que la tira dos veces: resultados posibles $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ con proba $1/4$ cada uno.

El fiaca tira una vez y repite el resultado: $(0, 0)$, $(1, 1)$,

Cada coordenada tiene la misma proba: $P(X = 0) = P(Y = 0) = 1/2$

Mirando sólo X o Y no podemos diferenciar entre los dos.

Hay que mirar el resultado de todo el vector (X, Y)

Def. Un *vector aleatorio* es una función $(X_1, \dots, X_n) : S \rightarrow \mathbb{R}^n$.

Función de probabilidad conjunta

$$p(x, y) = P(X = x, Y = y)$$

El *rango* del vector $R_{X,Y} = R_X \times R_Y$

$$P((X, Y) \in A) = \sum_{(x,y) \in A} p(x, y)$$

La proba conjunta satisface

$$1) p(x, y) \geq 0$$

$$2) \sum_x \sum_y p(x, y) = 1$$

Distribuciones marginales Dado vector (X, Y) ,

$$P(X = x) = \sum_y P(X = x, Y = y), \quad \text{marginal de } X$$

$$P(Y = y) = \sum_x P(X = x, Y = y), \quad \text{marginal de } Y$$

Ejemplo Sea (X, Y) vector con distribución

$$p(0, 0) = 0,4, p(0, 1) = 0,2, p(1, 0) = 0,1 \text{ y } p(1, 1) = 0,3.$$

Las marginales son

$$P(X = 0) = p(0, 0) + p(0, 1) = 0,6$$

$$P(X = 1) = p(1, 0) + p(1, 1) = 0,4$$

Toda la info en una tabla:

	0	1	X
0	0.4	0.2	0.6
1	0.1	0.3	0.4
Y	0.5	0.5	1

Independencia Dado un vector (X, Y) decimos que las variables X e Y son independientes si

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

para todo x, y . Esto implica que

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \quad \text{para todo } A, B \subset \mathbb{R}.$$

Ejemplo Tiramos una moneda 2 veces $X = 1$ si el número de caras es par. $Y = 1$ si la primera moneda es cara.

$$P(X = 0) = P(X = 1) = 1/2, \quad P(Y = 0) = P(Y = 1) = 1/2$$

$$P\{X = 0, Y = 1\} = P[\text{primera cara y número par de caras}]$$

$$= P\{(1, 1)\} = 1/4.$$

Esto es suficiente para probar que X e Y son independientes, usando que A, B indep implica A, B^c indep.

Lema. Si existen f y g tales que

$$P(X = x, Y = y) = C f(x)g(y), \quad \text{para todo } x, y$$

entonces X e Y son independientes.

Dem: Note que

$$C = \left(\sum_x f(x) \sum_y g(y) \right)^{-1}$$

Sumando sobre y tenemos

$$P(X = x) = C f(x) \sum_y g(y)$$

$$P(X = y) = C g(y) \sum_x f(x),$$

sumando sobre x . Así:

$$P(X = x)P(Y = y) = C f(x) \sum_y g(y) C g(y) \sum_x f(x) = C f(x)g(y) \quad \square$$

Ejemplo La distribución conjunta de un vector (X, Y) está dada por

$$p(k, \ell) = \frac{\lambda^k \mu^\ell e^{-\lambda-\mu}}{k! \ell!}$$

$k, \ell = 0, 1, 2, \dots; \lambda, \mu > 0$.

Claramente $p(k, \ell) = g(k)f(\ell)$, por lo tanto son independientes. La marginal de X es

$$P(X = k) = \sum_{\ell \geq 0} \frac{\lambda^k \mu^\ell e^{-\lambda-\mu}}{k! \ell!} = \frac{\lambda^k e^{-\lambda}}{k!} \sum_{\ell \geq 0} \frac{\mu^\ell e^{-\mu}}{\ell!} = \frac{\lambda^k e^{-\lambda}}{k!}$$

Es decir, $X \sim \text{Poisson}(\lambda)$. Similarmente $Y \sim \text{Poisson}(\mu)$.

Ejemplo (X, Y) tiene distribución conjunta

$$p(k, n) = C \frac{2^{-k}}{n}, \quad k = 1, 2, \dots; n = 1, \dots, k$$

C constante apropiada.

Como $p(k, n) = C 2^{-k} \frac{1}{n}$, parecería que $p(k, n)$ puede factorizarse; esto implicaría que X, Y serían independientes.

Pero no. Hay dependencia entre X e Y porque

$$p(k, n) = C \frac{2^{-k}}{n} \mathbf{1}\{n \leq k\}$$

no se puede factorizar. Así que X e Y **no** son independientes. Esta conclusión sigue también de

$$P(X = 1) > 0, P(Y = 2) > 0, \quad P(X = 1, Y = 2) = 0.$$

Distribución de la suma de dos variables Sea (X, Y) un vector aleatorio discreto con distribución conjunta p y sea $Z = X + Y$. La distribución de Z es

$$P_Z(z) = \sum_x p_{X,Y}(x, z-x) = \sum_y p_{X,Y}(z-y, y)$$

Cuando X e Y son independientes,

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

Aplicación: suma de Poisson independientes es Poisson $X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu)$. $X + Z \sim \text{Poisson}(\lambda + \mu)$.

$$\begin{aligned} P(Z = n) &= \sum_{k=0}^n p_X(k)p_Y(n-k) = \sum_{k=0}^n \frac{e^{-\lambda} \lambda^k}{k!} \frac{e^{-\mu} \mu^{n-k}}{(n-k)!} \\ &= \frac{e^{-(\lambda+\mu)} (\lambda + \mu)^n}{n!} \sum_{k=0}^n \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(\frac{\mu}{\lambda + \mu}\right)^{n-k} \end{aligned}$$

Distribución condicional Dado vector (X, Y) , La distribución condicional de X dado Y está dada por

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Esperanza condicional

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

Ejemplo X, Y Poisson independientes con λ y μ . $Z = X + Y$ Poisson con suma.

$$P(X = k|Z = k + m) = \text{binomial}(k + m, \lambda/(\lambda + \mu))$$

Teorema. Vale

$$E(X) = \sum_y E(X|Y = y)P(Y = y)$$

Ejemplo Gallina produce N huevos Poisson λ . Cada huevo produce un pollo con proba p independiente de los otros. Sea K el número de pollos.

Calcule $E(K|N = n)$ y $E(K)$.

Note que

$$P(K = k|N = n) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Así

$$E(K|N = n) = np$$

$$EK = \sum_n E(K|N = n)P(N = n) = \sum_n npP(N = n) = pEN = \lambda p$$

Se puede calcular también $P(K = k)$ directamente.

Se puede calcular $P(N = n|K = k)$ y $E(N|K = k)$.

Juego de los sobres Dos sobres. Uno contiene a pesos y el otro b pesos; $a < b$. Desconocemos los valores a y b .

Usted elige uno de los sobres, lo abre y observa el valor que contiene.

Le ofrezco la oportunidad de elegir el otro sobre.

Tiene sentido cambiarse de sobre?

Más precisamente: hay una estrategia que le permita elegir el sobre con b pesos con proba estrictamente mayor que $1/2$?

Estrategia: Sea X_1 : valor en el sobre elegido.

$$P(X_1 = a) = P(X_1 = b) = 1/2$$

Sea $Y \sim \text{exponencial}(1)$, una variable independiente de X_1

Observe X_1 y simule Y .

Si $X_1 < Y$ cambie de sobre; si $X_1 > Y$ no cambie.

X_2 : valor en el sobre final (después de un eventual cambio).

Sabemos calcular las probabilidades condicionales siguientes:

$$P(X_2 = b|X_1 = b) = P(Y < b) = 1 - e^{-b},$$

$$P(X_2 = b|X_1 = a) = P(Y > a) = e^{-a}.$$

Usando el teorema de la probabilidad total:

$$P(X_2 = b) = P(X_2 = b|X_1 = b)P(X_1 = b) + P(X_2 = b|X_1 = a)P(X_1 = a)$$

$$= \frac{1}{2}(1 - e^{-b}) + \frac{1}{2}e^{-a} = \frac{1}{2} + \frac{1}{2}(e^{-a} - e^{-b}) > \frac{1}{2}$$

Clase del 7 de mayo Vectores aleatorios continuos

Def. Un vector aleatorio $X = (X_1, \dots, X_d)$ es continuo con densidad conjunta g si

$$P(a_i \leq X_i \leq b_i, i = 1, \dots, d) = \int_{a_1}^{b_1} \dots \int_{a_d}^{b_d} g(x_1, \dots, x_d) dx_1 \dots dx_n$$

Así, para $A \subset \mathbb{R}^n$:

$$P((X_1, \dots, X_d) \in A) = \int_A g(x_1, \dots, x_d) dx_1 \dots dx_n$$

Esto vale para A donde se pueda calcular la integral. En ese caso, en teoría de la medida se dice que A es *medible*.

Distribución acumulada

La distribución acumulada de un vector continuo se define para $x = (x_1, \dots, x_d)$ como

$$\begin{aligned} F(x) &= F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f(x_1, \dots, x_d) dx_1 \dots dx_d \end{aligned}$$

Lema La distribución acumulada de un vector caracteriza la distribución del vector.

Dem. Basta mostrar que la acumulada conjunta determina la densidad conjunta. Lo hacemos para el caso de dos dimensiones. De la definición sigue que

$$f(x, y) = \frac{\partial F(x, y)}{\partial x \partial y}. \quad \square$$

y “a lo físico”:

$$\begin{aligned} P(x \leq X \leq x + dx, y \leq Y \leq y + dy) &= \int_x^{x+dx} \int_y^{y+dy} f(z, w) dz dw \\ &\sim f(x, y) dx dy \end{aligned}$$

Distribuciones marginales Sea $X = (X_1, \dots, X_d)$ un vector continuo con densidad f_X . Entonces cada X_i es una variable continua con densidad

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{d-1}} f_X(x_1, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

f_{X_i} es la densidad *marginal* de X_i que (por la fórmula de arriba) se obtiene integrando la densidad conjunta en todas las otras variables.

Ejemplo Sea (X, Y) vector con densidad conjunta

$$f(x, y) = \frac{1}{y} e^{-y - \frac{x}{y}} \quad x, y > 0$$

La marginal de Y está dada por

$$f_Y(y) = \int f(x, y) dx = e^{-y}$$

para todo $y > 0$. O sea que $Y \sim \exp(1)$.

Calcule $P(X < Y)$ y $P(X < a)$

$$P(X < Y) = P((X, Y) \in A) = \int_0^\infty \int_0^y f(z, w) dz dw = \dots = \frac{1}{3}$$

$$P(X < a) = \int_0^\infty \int_0^a f(z, w) dz dw = \dots = 1 - e^{-a}.$$

Ejemplo (X, Y) con densidad

$$f(x, y) = \frac{1}{x} \mathbf{1}\{0 < y \leq x \leq 1\}$$

La marginal de X :

$$f_X(x) = \int_0^x f(x, y) dy = \mathbf{1}\{0 < x \leq 1\}$$

Así X tiene distribución uniforme en $(0, 1]$.

La densidad de Y :

$$f_Y(y) = \int_y^1 f(x, y) dx = -\log y \mathbf{1}\{0 < y \leq 1\}$$

Independencia de variables aleatorias continuas

Def X e Y son *independientes* si y solo si para todo x, y ,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

Lema las variables continuas X e Y con densidad f_X, f_Y , respectivamente son *independientes* si y sólo si

$$f_X(x)f_Y(y) = f(x, y), \text{ para todo } x, y$$

Dem: Ejercicio.

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.

Def Una familia $(X_i : i \in J)$ de *vectores* aleatorios es *independiente* (mutuamente independientes) si para todo subconjunto finito de índices $K \subset J$,

$$P(X_i \leq a_i, i \in K) = \prod_{i \in K} P(X_i \leq a_i), \quad \forall a_i \in \mathbb{R}$$

Ejemplos

1. Encuentros casuales. Dos personas deciden encontrarse un día entre las 5 y las 6. Cada uno llega en instantes independientes distribuidos uniformemente en ese intervalo y espera 15 minutos. Cual es la probabilidad que se encuentren?

Definiendo

$$A := \{(x, y) \in [0, 60]^2 : |x - y| \leq 15\}$$

queremos calcular $P((X, Y) \in A)$, con (X, Y) uniforme en $[0, 60]^2$:

$$f(x, y) = \frac{1}{60^2} \mathbf{1}\{(x, y) \in [0, 60]^2\}$$

$$P((X, Y) \in A) = \frac{\text{area}(A)}{60^2} = 1 - \frac{\text{area}(A^c)}{60^2} = 1 - \frac{45^2}{60^2} = \frac{7}{9}$$

2. Permutaciones. Sean X_1, \dots, X_n una familia de n variables continuas independientes con densidad común f y acumulada F . Muestre que la familia $(F(X_1), \dots, F(X_n))$ es una familia de variables uniformes en $[0, 1]$ independientes.

Sean S_1, \dots, S_n las estadísticas de orden definidas por

$$S_1 < \dots < S_n; \quad \{X_1, \dots, X_n\} = \{S_1, \dots, S_n\} \text{ (como conjuntos)}$$

es decir, $S_1 = \min_i S_i$, $S_n = \max_i S_i$, etc. Sea $K(i)$ el lugar de X_i cuando las variables son ordenadas: $X_i = S_{K(i)}$.

Muestre que $(K(1), \dots, K(n))$ es una permutación aleatoria de $(1, \dots, n)$.

3. Records. Sean X_1, X_2, \dots una familia de variables continuas independientes. Sea $Y_n = \mathbf{1}\{X_n > X_i, \text{ para todo } 1 \leq i < n\}$. Y_n es uno si hay un record en el instante n . Pregunta: Y_1, Y_2, \dots son variables independientes?

4. Aguja de Buffon En un piso de tabla corrida, las líneas determinadas por las tablas son paralelas y están a distancia D . Una aguja de longitud $L < D$ es lanzada al azar sobre ese piso y se considera el evento $A =$ “la aguja interseca una de las líneas”. El evento complementario es $A^c =$ “la aguja está totalmente dentro de una de las tablas”.

Veremos que la probabilidad de A depende del número π . Las variables relevantes son:

$X =$ distancia del centro de la aguja a la paralela más cercana

$\theta =$ ángulo entre la recta que contiene la aguja y la recta perpendicular a las tablas que contiene el centro de la aguja.

$$X \sim \text{Uniforme}[0, D/2]. \quad f_X(x) = \frac{2}{D} \mathbf{1}\{x \in [0, D/2]\}.$$

$$\theta \sim \text{Uniforme}[0, \pi/2]. \quad f_\theta(y) = \frac{2}{\pi} \mathbf{1}\{y \in [0, \pi/2]\}.$$

X y θ son independientes.

La aguja interseca una de las paralelas si

$$X < \frac{L}{2} \cos \theta,$$

que equivale a

$$\begin{aligned} (X, \theta) &\in \left\{ (x, y) \in \left[0, \frac{D}{2}\right] \times \left[0, \frac{\pi}{2}\right] : x < \frac{L}{2} \cos y \right\} \\ &= \left\{ (x, y) : 0 < y < \frac{\pi}{2}, 0 < x < \frac{L}{2} \cos y \right\} \end{aligned}$$

Entonces

$$\begin{aligned} P(A) &= P\left(X < \frac{L}{2} \cos \theta\right) = \int_0^{\pi/2} \int_0^{\frac{L}{2} \cos y} f_X(x) f_\theta(y) dx dy \\ &= \frac{4}{\pi D} \int_0^{\pi/2} \int_0^{\frac{L}{2} \cos y} dx dy = \frac{4}{\pi D} \int_0^{\pi/2} \frac{L}{2} \cos y dy = \frac{2L}{\pi D} \end{aligned}$$

Esto se usa para “estimar” π usando

$$\pi = \frac{2L}{P(A)D}$$

Llamemos $p = P(A)$. Repitiendo el experimento muchas veces y tomando la proporción muestral \hat{p} de éxitos, se estima π por $\hat{\pi} = \frac{2L}{\hat{p}D}$.

Suma de variables continuas X, Y va continuas con f . $Z = X + Y$. Entonces

$$P(Z \leq z) = \int \int_{\{(x,y): x+y \leq z\}} f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dx dy$$

substituya $u = x, v = y + x$:

$$= \int_{-\infty}^{\infty} \int_{-\infty}^z f(u, v - u) du dv$$

de donde

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) dx$$

Caso independiente:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

La densidad de la suma de dos variables independientes es la convolución de las densidades de las variables.

Clase del 9 de mayo Gama X_1, \dots, X_n exponenciales indep. $Z_n = X_1 + \dots + X_n$. Entonces

$$f_Z(z) = \frac{\lambda^n}{(n-1)!} z^{n-1} e^{-\lambda z} \quad \text{Gama}(n, \lambda)$$

Inducción. Suponga que $T = X_1 + \dots + X_{n-1}$ es $\text{Gama}(n-1, \lambda)$. Como T y X_n son independientes:

$$\begin{aligned} f_Z(z) &= \int_0^z \frac{\lambda^{n-1}}{(n-2)!} x^{n-2} e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx \\ &= \frac{\lambda^n}{(n-2)!} e^{-\lambda z} \int_0^z x^{n-2} dx = OK \end{aligned}$$

Distribución condicional de variables continuas

(X, Y) vector aleat con densidad f .

Queremos definir $P(Y \leq y | X = x)$

Si X es continua, $P(X = x) = 0$. Procedimiento límite:

$$\begin{aligned} \star &= P(Y \leq y | x \leq X \leq x+h) = \frac{P(Y \leq y, x \leq X \leq x+h)}{P(x \leq X \leq x+h)} \\ &= \frac{\int_{-\infty}^y \int_x^{x+h} f(u, v) du dv}{\int_x^{x+h} f_X(v) dv} \end{aligned}$$

dividiendo arriba y abajo por h y sacando límite,

$$\lim_{h \rightarrow 0} \star = \int_{-\infty}^y \frac{f(x, v)}{f_X(x)} dv$$

Así definimos $f_{Y|X=x}(y) = f(x, y) / f_X(x)$ para x tal que $f(x) \neq 0$.

$f_{Y|X=x}$ es una densidad: $\int f_{Y|X=x}(y) dy = \int \frac{f(x, y)}{f_X(x)} dy = 1$.

Es la densidad de una nueva variable con esperanza:

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

Valen las siguientes fórmulas:

$$P(Y \leq y) = \int_{-\infty}^{\infty} P(Y \leq y | X = x) f_X(x) dx$$

$$EY = \int_{-\infty}^{\infty} E(Y|X = x) f_X(x) dx$$

Ejemplos

1. (X, Y) tienen densidad conjunta $f(x, y) = e^{-y}$, $0 < x < y$

(a) Calcule la distribución marginal de Y .

(b) Pruebe que $f_{X|Y=y}(x) = 1/y$, para $0 < x < y$.

(c) Calcule $E(X|Y = y)$ y use el resultado para calcular $E(X)$.

2. $f(x, y) = 2(x + 2y)I_T(x, y)$ con $T = \{0 \leq x \leq 1, 0 \leq y \leq 1 - x\}$

Calcular las marginales de X e Y .

$$f_X(x) = 2(1 - x)I_{[0,1]}(x)$$

$$f_Y(y) = (1 + 2y - 3y^2)I_{[0,1]}(y)$$

Calcular $P(X \leq 1/2 | Y \leq 1/4) = 8/19$

$$P(X \leq 1/2 | Y = 1/4) = \int_0^{1/2} \frac{f(x, 1/4)}{f_Y(1/4)} dx$$

Densidad condicional e Independencia

X e Y son indep si $f(x, y) = f_X(x)f_Y(y)$.

En función de proba condicional:

$$f_X(x) = f_{X|Y=y}(x)$$

Dem: Por la def de la densidad condicional, $f(x, y) = f_Y(y)f_{X|Y=y}(x)$.

Por lo tanto las variables son independientes si y solo si $f_X(x) = f_{X|Y=y}(x)$

Para probar que dos variables continuas **no son independientes** basta exhibir un rectángulo $[a, b] \times [c, d]$ tal que

$$\int_a^b \int_c^d f(x, y) dx dy \neq \int_a^b f_X(x) dx \int_c^d f_Y(y) dy$$

Si $R_{X,Y} \neq R_X \times R_Y$, las variables no son independientes.

Otra forma de probar que X e Y no son independientes es encontrar un punto (u, v) en \mathbb{R}^2 tal que $f(x, y)$, $f_X(x)$ y $f_Y(y)$ sean todas continuas en ese punto y $f(x, y) \neq f_X(x)f_Y(y)$.

Por continuidad, la condición se cumplirá en un entorno rectangular del punto.

Esperanza de funciones de vectores

$$Eh(X, Y) = \int \int h(x, y) f(x, y) dx dy$$

$$E(aX + bY) = aEX + bEY$$

Si X e Y son independientes:

$$E(XY) = EX EY$$

Contraejemplo de funciones con $EXY = EX EY$ pero no son independientes:

$$f(x, y) = C \mathbf{1}\{x^2 + y^2 \leq 1\}$$

No son independientes porque el rango del vector no es el producto cartesiano de los rangos. La esperanza de cada variable es 0 y también lo es la esperanza del producto.

El juego de los sobres. II

Dos sobres con plata Y_1, Y_2 . iid Uniformes en $[0, 10]$.

Abro un sobre y veo y . Debo cambiar de sobre?

Estrategia 1: Fijo $K \in (0, 10)$. Si $y > K$, me quedo con y . Si no, cambio.

Sea X_1 valor del primer sobre.

X_2 valor obtenido despues de aplicar la estrategia.

$$X_2 = Y_1 \mathbf{1}\{Y_1 > K\} + Y_2 \mathbf{1}\{Y_1 \leq K\}$$

$$EX_2 = E(Y_1 \mathbf{1}\{Y_1 > K\}) + EY_2 P(Y_1 \leq K)$$

$$= \int_K^{10} yf(y)dy + 5 P(Y \leq K) = \left[\frac{y^2}{2} \right]_K^{10} + 5 \frac{K}{10}$$

$$= 5 - \frac{K^2}{20} + 5 \frac{K}{10} = 5 + \frac{K}{10} \left(5 - \frac{K}{2} \right)$$

EX_2 asume un máximo en $K = 5$.

Para verlo, multiplique por 2 y vea que $g(K) = K(10 - K)$ es una parabola con inclinación para abajo que pasa por 0 y 10, por lo tanto asume su máximo en 5.

En resumen, la estrategia queda:

Miro Y_1 , si es mayor que 5, me quedo. Si no, me paso a Y_2 .

La media para $K = 5$ queda

$$EX_2 = 6,25$$

Clase del 21 de mayo Covarianza y correlación Sean X e Y dos v.a. con esperanzas EX y EY respectivamente, la *covarianza* entre X e Y se define como

$$E(X - EX)(Y - EY) = \text{caso continuo y discreto}$$

Observación: $\text{Cov}(X, X) = V(X)$.

Idea intuitiva: Si X e Y tienen una fuerte relación positiva, en el sentido que valores grandes de X aparecen asociados con valores grandes de Y y valores pequeños de X aparecen asociados con valores pequeños de Y , entonces los productos serán positivos y por lo tanto la covarianza será positiva.

Por otra parte, si X e Y tienen una fuerte relación negativa, en el sentido que valores grandes de X aparecen asociados con valores pequeños de Y y valores pequeños de X aparecen asociados con valores grandes de Y , entonces la mayoría de los productos serán negativos y por lo tanto la covarianza será negativa.

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

Probarlo para discreto. Continuo igual.

Ejemplo discreto:

	0	1	2	X
0	0.4	0.1	0.1	0.6
1	0.1	0.2	0.1	0.4
Y	0.5	0.3	0.2	1

Ejemplo continuo: $f(x, y) = \frac{6}{5}(x + y^2) \mathbf{1}\{(x, y) \in [0, 1]^2\}$.

$$\text{Cov}(X, Y) = -\frac{1}{100}$$

Propo Si X e Y son independientes, $\text{Cov}(X, Y) = 0$. La reciproca no es verdadera.

Dem Como las variables son independientes las funciones de probabilidad en el caso discreto y las densidades en el caso continuo factorizan. Por ejemplo en el caso continuo.

$$EXY = \int_{\mathbb{R}^2} xy f_X(x) f_Y(y) dx dy = \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_Y(y) dy$$

Contraejemplo: X e Y tienen covarianza cero pero no son indep:

	-1	0	1	X
-1	1/8	0	1/8	1/4
0	0	1/2	0	1/2
1	1/8	0	1/8	1/4
Y	1/4	1/2	1/4	1

Ejercicio: Contraejemplo continuo Buscar una densidad que satisfaga: $f(x, y) = f(x, -y) = f(-x, y) = f(-x, -y)$ que garantiza que $E(XY) = 0$ y $EX = EY = 0$ pero que no sea el producto de dos funciones.

Verifique que por ejemplo $f(x, y)$ uniforme en una bola centrada en 0 satisface.

Coefficiente de correlación Sean X e Y dos v.a. con esperanzas EX y EY respectivamente y varianza positiva, el coeficiente de correlación entre X e Y se define como

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Propo. 1. Sean a, b, c y d números reales, $a \neq 0, c \neq 0$ y X e Y v.a. con varianza positiva, entonces

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

donde sg denota la función signo.

2. $-1 \leq \rho(x, y) \leq 1$

3. $|\rho(X, Y)| = 1$ si Y es función lineal de X .

Dem: 1. Cuentas.

2. Asumamos $EX = EY = 0$.

Defina $g(t) = E(X - tY)^2$

Claramente $g(t) \geq 0$

$$g(t) = EX^2 - 2tE(XY) + t^2EY^2$$

Polinomio de segundo grado en t . $a = EY^2, b = -2E(XY), c = EX^2$.

$$\text{Discriminante } b^2 - 4ac = 4(E(XY))^2 - 4EX^2EY^2 \leq 0$$

Por lo tanto

$$\frac{(E(XY))^2}{EX^2EY^2} \leq 1$$

es decir $\rho^2 \leq 1$, lo que implica $-1 \leq \rho \leq 1$.

Caso general: basta ver que $\rho(X, Y) = \rho(X - EX, Y - EY)$.

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

$$E(X - t_0Y)^2 = 0$$

Como X e Y tienen esperanza cero, $X - t_0Y = 0$ con probabilidad 1.

Caso general, substituyendo

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

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

Recíprocamente, si $Y = AX + B$ entonces $|\rho| = 1$ (cuenta).

Esperanzas de funciones de variables aleatorias

Se aplican las fórmulas siguientes que se pueden probar como lo hicimos para el caso de una variable:

Caso discreto:

$$Eg(X_1, \dots, X_n) = \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n)$$

(si la suma está bien definida)

Caso continuo. Vector (X_1, \dots, X_n) con densidad conjunta f .

$$Eg(X_1, \dots, X_n) = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1, \dots, dx_n$$

(si la integral está bien definida)

Las fórmulas valen también para vectores infinitos (si las sumas e integrales están bien definidas).

Esperanzas y varianzas de sumas de variables aleatorias

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i EX_i$$

$$V\left(\sum_i a_i X_i\right) = \sum_i a_i^2 VX_i + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)$$

Si son independientes, como las covarianzas son 0,

$$V\left(\sum_i a_i X_i\right) = \sum_i a_i^2 VX_i$$

Muestra. Una *muestra* de una variable aleatoria X es un vector X_1, \dots, X_n de variables aleatorias independientes idénticamente distribuidas (iid) con $X_i \sim X$.

Defina la **media muestral** de una muestra por

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$$

Si $EX = \mu$ y $VX = \sigma^2$, obtenemos

$$E\bar{X}_n = \mu, \quad V\bar{X}_n = \sigma^2/n$$

Desigualdad de Markov. Sea X una variable aleatoria no negativa con esperanza finita. Entonces,

$$P(X > \varepsilon) \leq \frac{EX}{\varepsilon}.$$

Dem:

$$X = X \mathbf{1}_{X > \varepsilon} + X \mathbf{1}_{X \leq \varepsilon} \geq \varepsilon \mathbf{1}_{X > \varepsilon}$$

porque $X \geq 0$. Sacando esperanzas,

$$EX \geq \varepsilon E(\mathbf{1}_{X > \varepsilon}) = \varepsilon P(X > \varepsilon). \quad \square$$

Desigualdad de Chebichev:

$$P(|X - EX| > \varepsilon) \leq \frac{VX}{\varepsilon^2}$$

Dem. Ejercicio.

La cota que provee la desigualdad de Chebyshev puede ser grosera o, peor aún, no informativa, por ejemplo, si $\varepsilon^2 \leq \sigma^2$

Ejemplo: Sea $X \sim U(0, 10)$, entonces $E(X) = 5$ y $V(X) = 100/12$.

Aplicando la desigualdad de Chebyshev,

$$P(|X - 5| > 4) \leq 0,52$$

Verdadero valor:

$$P(|X - 5| > 4) = 0,20$$

Clase del 23 de mayo Convergencia en probabilidad: Sea $X_n, n \geq 1$, una sucesión de variables aleatorias, diremos que X_n converge en probabilidad a la v.a. X si para todo $\varepsilon > 0$

$$\lim_n P(|X_n - X| > \varepsilon) = 0$$

Ley de grandes números:

Sea X una variable aleatoria con $EX = \mu$. Se desea estimar μ por \bar{X}_n , la media muestral de una muestra de X .

Teorema. Sean X_1, X_2, \dots iid. $EX = \mu$ $VX = \sigma^2$. Entonces \bar{X}_n converge a μ en probabilidad.

Dem: Ya vimos que $E\bar{X}_n = \mu$, $V\bar{X}_n = \sigma^2/n$.

Chevichev:

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad \square$$

Versión Bernoulli de la Ley de los Grandes Números:

Consideremos n repeticiones independientes de un experimento aleatorio y sea A un evento con probabilidad $P(A) = p$, constante en las n repeticiones. Si llamamos \hat{p}_n la proporción muestral de A (número de veces que ocurre A en las n repeticiones dividido n), entonces \hat{p}_n converge en probabilidad a p .

Dem: Note que $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$, donde $X_i = 1$ si A ocurre en el i -ésimo ensayo y $X_i = 0$ si no ocurre.

$X_i \sim X \sim \text{Bernoulli } p$.

$EX = p, VX = p(1 - p)$.

$$\bar{X}_n = \hat{p}_n$$

y se obtiene:

$$P(|\hat{p}_n - p| > \varepsilon) \leq \frac{p(1 - p)}{n\varepsilon^2} \rightarrow 0, \text{ con } n.$$

Ejemplo: Cuántas repeticiones del experimento deberían hacerse para que la frecuencia relativa difiera de p en menos de 0.01 con probabilidad mayor o igual que 0,95? En este caso, $\varepsilon = 0,01$ y queremos encontrar n tal que

$$P(|f_A - p| < 0,01) \geq 0,95$$

que equivale a

$$P(|f_A - p| \geq 0,01) \leq 0,05$$

Chevichev: $0,05 = p(1 - p)/(0,01^2 n)$ y se despeja n :

$$n \geq \frac{p(1 - p)100^2}{0,05^2}$$

Tomando el mayor valor posible de $p(1 - p) \leq \frac{1}{4}$, es suficiente tomar

$$n \geq \frac{1}{4} 10,000 \frac{10,000}{25} = \frac{10^8}{100} = 1,000,000.$$

Distribucion de Sumas de variables independientes

Teorema Si X_i son variables aleatorias independientes con FGM $M_i(t)$ entonces:

$$M_{X_1+\dots+X_n}(t) = M_1(t) \dots M_n(t)$$

Dem. Por independencia,

$$\begin{aligned} M_{X_1+\dots+X_n}(t) &= E(e^{t(X_1+\dots+X_n)}) = E(e^{tX_1} \dots e^{tX_n}) \\ &= Ee^{tX_1} \dots Ee^{tX_n} = M_1(t) \dots M_n(t). \quad \square \end{aligned}$$

Otras propiedades:

1) $M_{aX+b}(t) = e^{tb} Ee^{atX} = e^{tb} M_X(at)$

2) Si $Z \sim N(0, 1)$, entonces $M_Z(t) = e^{t^2/2}$

3) Si $X \sim N(\mu, \sigma^2)$, entonces $M_X(t) = M_{\sigma Z+\mu}(t) = e^{\mu t} e^{\sigma^2 t^2/2}$

4) Si X_1, \dots, X_n son iid media μ varianza σ^2 y $S_n = X_1 + \dots + X_n$,

$$M_{S_n} = (M_X(t))^n$$

5) Si $T_n = S_n/\sqrt{n}$,

$$M_{T_n} = (M_X(t/\sqrt{n}))^n$$

6) Suma de normales independientes es normal con media = suma de las medias y varianza igual a la suma de las varianzas.

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

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

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

Teorema de unicidad de la FGM. Si la FGM de una variable aleatoria existe, entonces es única. Además la FGM de X determina la función de distribución acumulada F_X .

Convergencia en distribución es equivalente a convergencia de las FGM:

$$Y_n \rightarrow Y \text{ en distribución si } \lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t).$$

Teorema central del límite. Sean X_i iid con media μ y varianza σ^2 y sea $S_n := X_1 + \dots + X_n$. Entonces

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow Z, \text{ en distribución,}$$

donde $Z \sim N(0, 1)$.

Observaciones:

1) Z_n tiene media 0 y varianza 1 para todo n .

2) Convergencia en distribución es Convergencia de las acumuladas.

3) Uso: para n grande trate Z_n como si fuera $N(0, 1)$.

Historia:

1733: TCL para Bernoulli(1/2) por Abraham de Moivre

1823: Pierre-Simon Laplace extiende de Moivre's para aproximar la Binomial(n, p) por la normal.

1901: Aleksandr Lyapunov demuestra rigurosamente el TCL.

Demostración del TCL:

Asumimos que la FGM $M = M_{X_i}$ de X_i existe e inicialmente tomamos $\mu = 0$ y $\sigma^2 = 1$ (despues vemos como se extiende).

Calculamos la FGM de Z_n en función de M :

$$M_{Z_n}(t) = Ee^{t(X_1+\dots+X_n)/\sqrt{n}} = (M(t/\sqrt{n}))^n$$

Sea

$$L(t) = \log M(t)$$

y note que

$$L(0) = 0, \quad L'(0) = \frac{M'(0)}{M(0)} = \mu = 0$$

$$L''(0) = \frac{M(0)M''(0) - (M'(0))^2}{(M(0))^2} = EX^2 = 1$$

Para probar el teorema, necesitamos probar que

$$\lim_{n \rightarrow \infty} (M(t/\sqrt{n}))^n = e^{t^2/2}$$

que es equivalente a probar que

$$\lim_{n \rightarrow \infty} nL(t/\sqrt{n}) = t^2/2.$$

Calculemos:

$$\lim_{n \rightarrow \infty} \frac{L(t/\sqrt{n})}{n^{-1}} = \lim_{n \rightarrow \infty} \frac{L'(t/\sqrt{n})tn^{-3/2}}{2n^{-2}}$$

(por L'Hopital)

$$= \lim_{n \rightarrow \infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} = \lim_{n \rightarrow \infty} \frac{L''(t/\sqrt{n})t^2n^{-3/2}}{2n^{-3/2}}$$

(de nuevo por L'Hopital)

$$= \lim_{n \rightarrow \infty} L''(t/\sqrt{n}) \frac{t^2}{2} = \frac{t^2}{2}.$$

Esto termina la demostración para media cero y varianza 1.

Si μ y σ^2 son cualesquiera,

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \left(\frac{X_1 - \mu}{\sigma} + \dots + \frac{X_n - \mu}{\sigma} \right)$$

y se aplica la demostración anterior a las variables $X_i^* = \frac{X_i - \mu}{\sigma}$ que son centradas y tienen varianza 1. \square

Formas alternativas del TCL:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow Z$$

y dividiendo numerador y denominador por n , obtenemos

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow Z$$

Una razón matemática para el TCL:

$$Z_{2n} = \frac{S_{2n}}{\sqrt{2n}} = \frac{S_n + S_{2n} - S_n}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left(\frac{S_n}{\sqrt{n}} + \frac{S_n^*}{\sqrt{n}} \right),$$

donde S_n^* tiene la misma distribución que S_n pero es independiente de S_n .

O sea que el límite, si existe tiene que satisfacer:

$$Z \sim \frac{Z + Z^*}{\sqrt{2}} \quad (*)$$

para Z y Z^* idénticamente distribuídas e independientes. En términos de la FGM esa ecuación es equivalente a

$$M_Z(t) = (M_Z(t/\sqrt{2}))^2$$

que es satisfecha por la normal:

$$M_Z(t) = e^{t^2/2} = (e^{(t/\sqrt{2})^2/2})^2 = (M_Z(t/\sqrt{2}))^2$$

Para obtener una demostración del TCL usando este argumento falta probar: (1) que el límite de Z_n existe y (2) que la normal es la *única* distribución que satisface la “ecuación” (*).

Clase del 28 de mayo Comentarios sobre el TCL. 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 independientes y con distribución $U(-0.5, 0.5)$.

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 $T_{1500} = \sum_i X_i$ y queremos calcular $P(|T_{1500}| > 15)$. Como $EX_i = 0$ y $VX_i = 1/12$, $ET_{1500} = 0$ y $VT_{1500} = \frac{1500}{12} = 125$. Entonces

$$P(|T_{1500}| > 15) = P(|Z| > 15/\sqrt{125}) = P(|Z| > 1,34) = 0,18$$

(usando la tabla de la Normal)

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 $P(|T_n| \leq 10) \geq 0,9$.

$$P(|T_n| \leq 10) \geq 0,9 \Leftrightarrow P(|Z| \leq 10/\sqrt{n/12}) \geq 0,9$$

Buscamos z tal que $P(|Z| \leq z) = 0,9$, que por tabla es $z = 1,64$. Así

$$10/\sqrt{n/12} = 1,64, \text{ de donde } n \geq 446.$$

Otras Aplicaciones del TCL

1. Si $Y_n \sim \text{Poisson}(\lambda n)$ entonces

$$\frac{Y_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{D} Z$$

Dem: considere $X_i \text{ Poisson}(\lambda)$ iid.

$Y_n = X_1 + \dots + X_n \text{ Poisson}(\lambda n)$. Aplique TCL y obtenga el límite.

Así la Poisson con parametro grande se aproxima por la normal.

2. $Y_n \sim \text{Gama}(n, \lambda)$ iid con n entero

$$\frac{Y_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{D} Z$$

$X_i \sim \text{Gama}(1, \lambda)$ (exponenciales) independientes.

$X_1 + \dots + X_n \text{ Gama}(n, \lambda)$ 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 probabilidad 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 $\Rightarrow X_i \sim \text{Bernoulli}(\frac{1}{2})$

Número de aciertos:

$$S_{1500} = \sum_{i=1}^{1500} X_i, \quad \bar{X} = \frac{S_{1500}}{1500}$$

$$P(S_{1500} \geq 950) = P\left(\frac{\bar{X} - \frac{1}{2}}{0,5/\sqrt{1500}}\right) \geq \frac{950/1500}{0,5/\sqrt{1500}}$$

$$\sim P(Z \geq 10,32) \sim 0$$

La proba de acertar 950 veces con una moneda es casi 0. Aceptamos la hipótesis que el hombre es un adivino.

Porqué convergencia en puntos de continuidad de F ?

Considere una sucesión de variables aleatorias X_n con acumuladas $F_n(x) = \mathbf{1}\{x \geq 1/n\}$.

X_n es una variable aleatoria constante: $P(X_n = 1/n) = 1$.

Cuando $n \rightarrow \infty$, la distribución de X_n aproxima la distribución de una variable aleatoria X concentrada en 0: $P(X = 0) = 1$. Sin embargo, si F es la acumulada de X , vemos que $F_n(0)$ no converge a $F(0)$.

De hecho, $F_n(0) = 0$ para todo n , pero $F(0) = 1$.

Cadenas de Markov

Un **proceso estocástico** (a tiempo discreto) es una sucesión de variables aleatorias X_1, X_2, \dots que asumen valores en un conjunto S finito o numerable.

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** cada vez que el proceso está en el estado x tiene probabilidad $p(x, y)$ de ir al estado y en el instante siguiente:

$$P(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = p(x, y).$$

Los valores $p(x, y)$ son llamados probabilidades de transición y conforman una **matriz de transición** $P = (p(x, y) : x, y \in S)$.

Cadena de Markov con dos estados Si hoy llueve, la probabilidad que llueva mañana es α y si hoy no llueve, esta probabilidad es β . El espacio de estados es $S = \{0, 1\}$; interpretamos 0 cuando llueve y 1 cuando no llueve. La matriz de transición es

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad (1)$$

O sea $p(0, 0) = \alpha$, etc.

Ejemplo constructivo de cadena de Markov Sea U_1, U_2, \dots una sucesión de variables uniformes en $[0, 1]$ independientes. Defina $X_0 = x$ e, iterativamente,

$$X_{n+1} = F(X_n, U_{n+1}) \quad (2)$$

donde $F(0, u) = \mathbf{1}\{u > \alpha\}$ y $F(1, u) = \mathbf{1}\{u > \beta\}$. Verifique que el proceso así obtenido es una cadena de Markov con matriz de transición (1).

En general, si X_n es una cadena de Markov con matriz P , entonces podemos definir para cada x una partición $\mathcal{J}_x = (J(x, y), y \in S)$ del intervalo $[0, 1]$, de tal manera que

$$|J(x, y)| = p(x, y)$$

y si definimos

$$F(x, u) = \sum_{y \in S} y \mathbf{1}\{u \in J(x, y)\}$$

se demuestra que el proceso definido por la ecuación (2) es Markov con matriz de transición P . En efecto,

$$\begin{aligned} &P(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= P(F(x, U_{n+1}) = y) = P(U_{n+1} \in J(x, y)) = |J(x, y)| = p(x, y). \end{aligned}$$

Cálculo de la distribución en el instante n . La matriz de transición sirve para calcular las probabilidades de transición a más de un paso:

$$P(X_n = y | X_0 = x) = P^n(x, y)$$

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 proba condicional)

$$= \sum_z P(X_2 = y | X_1 = z) P(X_1 = z | X_0 = x)$$

(por la propiedad de Markov)

$$= \sum_z p(x, z) p(z, y) = P^2(x, y)$$

Clase del 30 de mayo

Urna de Ehrenfest Considere N bolillas distribuidas en dos urnas. Una bolilla es elegida al azar y es cambiada de urna. 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

$$p(k, k - 1) = \frac{k}{N}, \quad p(k, k + 1) = \frac{N - k}{N}$$

y las probabilidades $p(x, y) = 0$ si $|x - y| > 1$.

Este modelo representa el comportamiento de un gas que tiene N moléculas ocupando dos containers.

Ecuaciones de Chapman-Kolmogorov Un argumento igual prueba que para $0 \leq k \leq n$,

$$P^n(x, y) = \sum_z P^k(x, z) P^{n-k}(z, y)$$

Usando esta fórmula podemos calcular la distribución de X_n si conocemos la de X_1 :

$$P(X_n = y) = \sum_x P^n(x, y) P(X_1 = x)$$

Medidas invariantes Se puede probar el siguiente resultado:

Si $p(x, y) > 0$ para todo par de estados x, y , entonces existe una probabilidad π tal que

$$\lim_n P^n(x, y) = \pi(y), \quad \text{para todo } x$$

es decir que la cadena olvida el valor inicial y la distribución de X_n converge a π (convergencia en distribución) para cualquier estado inicial.

En ese caso, obtenemos

$$P^{n+1}(x, y) = \sum_z P^n(x, z)P(z, y)$$

Sacando límite en ambos miembros,

$$\pi(y) = \sum_z \pi(z)P(z, y) \quad \text{para todo } y$$

Estas son las **ecuaciones de balance**. La probabilidad π se llama **medida invariante** y es la única solución de las ecuaciones de balance.

Propiedades de la medida invariante:

π es un autovector a la izquierda de P con autovalor 1: $\pi P = \pi$. Esto quiere decir que

$$\sum_x \pi(x)P(X_1 = y|X_0 = x) = \pi(y)$$

y en general, para todo n , $\pi P^n = \pi$:

$$\sum_x \pi(x)P(X_n = y|X_0 = x) = \pi(y)$$

O sea: si la distribución de X_0 es π , entonces la distribución de X_n es π para todo $n \geq 0$.

Ejemplo de la lluvia. Las ecuaciones de balance son

$$\pi(0) = \alpha\pi(0) + \beta\pi(1),$$

$$\pi(1) = (1 - \alpha)\pi(0) + (1 - \beta)\pi(1)$$

con $\pi(0) + \pi(1) = 1$, que tiene como solución

$$\pi(0) = \frac{\beta}{1 - \alpha + \beta}, \quad \pi(1) = \frac{1 - \alpha}{1 - \alpha + \beta}$$

Ley de grandes números para cadenas de Markov Se puede demostrar que los promedios temporales convergen a una distribución:

$$\lim_n \frac{1}{n} \sum_{k=1}^n P(X_k = y|X_0 = x) = \mu(y)$$

Aceptemos ese límite, que se puede escribir como

$$\lim_n \frac{1}{n} \sum_{k=1}^n P^k(x, y) = \mu(y)$$

$$\lim_n \frac{1}{n+1} \sum_{k=1}^{n+1} P^k(x, y) = \lim_n \frac{1}{n+1} \sum_z \sum_{k=1}^n P^k(x, z)P(z, y)$$

Es decir, como el límite es μ ,

$$\mu(y) = \sum_z \mu(z)P(z, y)$$

Es decir que el límite μ satisface las ecuaciones de balance. Como la solución de las ecuaciones de balance es única e iguales a π , tendemos que $\mu = \pi$.

Una forma entonces de encontrar π es hacer muestras de la cadena de Markov y tomar proporciones muestrales.

Ejemplo de urna de Ehrenfest Las ecuaciones de balance para este problema son: para $0 < k < N$,

$$\pi(k) = \pi(k+1)p(k+1, k) + \pi(k-1)p(k-1, k)$$

(las otras transiciones son cero) o sea,

$$\pi(k) = \pi(k+1)\frac{k+1}{N} + \pi(k-1)\frac{N-k+1}{N}, \quad 0 < k < N;$$

y en los bordes:

$$\pi(0) = \pi(1)\frac{1}{N}, \quad \pi(N) = \pi(N-1)\frac{1}{N}$$

cuya solución es:

$$\pi(k) = \binom{N}{k} \left(\frac{1}{2}\right)^N$$

Ejemplo: ranqueo de páginas de Google Grafo orientado: nodos representan páginas web. Aristas orientadas representan links (direccionados). $\mathcal{G} = (V, E)$, $V =$ conjunto de vertices. $E \subset \{(x, y) : x, y \in V\}$, conjunto de aristas orientadas.

Queremos ranquear los nodos. Para eso podemos usar el número de aristas que llegan a un nodo $y \in V$:

$$R_1(y) = \sum_{x \in V} a(x, y)$$

donde $a(x, y) = \mathbf{1}\{(x, y) \in E\}$.

Pero esto le da mucho peso a los nodos que distribuyen muchas aristas. Para compensar, definimos el número de aristas que salen del nodo 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 en este ranqueador todos los nodos que que tienen el mismo número de aristas salientes envían el mismo peso, independientemente de las aristas entrantes.

Más interesante sería que cada nodo enviara un peso proporcional a su importancia (medida por las aristas que entran). Esto nos lleva a plantear el tercer ranqueador:

$$R_3(y) = \sum_{x \in V} R_3(x) \frac{a(x, y)}{a(x)}$$

Así vemos que los rankings satisfacen las ecuaciones de balance para una cadena de Markov que es un paseo aleatorio en el grafo.

Escribiendo $p(x, y) = \frac{a(x, y)}{a(x)}$ y $\pi(x) = R_3(x)$, el tercer ranqueador coincide con la medida invariante para una cadena de Markov que se comporta así:

“Cuando el proceso se encuentra en el nodo x , elige uno de los nodos uniformemente entre los que reciben una flecha saliendo de x y salta a ese nodo”

Obtener el ranking R_3 es entonces equivalente a obtener la medida invariante π para la cadena de Markov con transiciones p .

Como estamos hablando de un espacio de estados de miles de millones, la obtención analítica es físicamente imposible.

Para estimar π (que nos da el ranking), se usa la ley de grandes números para cadenas de Markov. Se envía una o más robots que circulan por los nodos con transiciones p y se estima π con la media temporal

$$\pi(x) \sim \frac{1}{T} \sum_{t=1}^T \mathbf{1}\{X_t = x\}.$$

Clase del 4 de junio Paseos aleatorios

Contando caminos Un camino de longitud n es un vector (s_0, s_1, \dots, s_n) ,

$$s_k = x_1 + \dots + x_k$$

donde los incrementos $x_i \in \{-1, 1\}$.

Hay 2^n caminos de longitud n . Si $s_0 = 0$ y $s_n = x$, entonces los a incrementos positivos y los b incrementos negativos deben satisfacer:

$$a + b = n, \quad a - b = x.$$

Es decir:

$$a = \frac{n + x}{2}, \quad b = \frac{n - x}{2}.$$

Así, $N_{n,x}$ el número de caminos de longitud n que van de 0 a x es

$$N_{n,x} = \binom{a+b}{a} = \binom{a+b}{b}$$

Consideraremos $N_{n,x} = 0$ cuando no se puede alcanzar x en n pasos.

Ejemplo Elecciones. Supongamos que en una elección el candidato A saca a votos y el candidato B saca b votos, con $a > b$ (es decir A gana la elección).

Cual es la probabilidad que durante todo el escrutinio A esté por delante de B ?

Podemos representar la ventaja de A por un camino: cada vez que sale un voto para A sumamos 1 y cada vez que sale un voto para B restamos 1. O sea que $x_i = 1$ si el i -ésimo voto computado sale para A y $x_i = -1$ en caso que sea para B . La ventaja de A después de computar el k -ésimo voto es

$$s_k = x_1 + \dots + x_k$$

A lidera todo el escrutinio si para todo $0 < k \leq n$,

$$s_1 > 0, s_2 > 0, \dots, s_k > 0.$$

Asumimos que todos los posibles caminos de tamaño n que terminan en $a - b$ son igualmente probables. (todas las permutaciones de los votos son igualmente probables)

Principio de reflexión

Considere puntos espacio-temporales (k, x) y (n, y) .

$0 \leq k < n, x > 0, y > 0$.

El punto **reflejado** de (k, x) es $(k, -x)$

Consideraremos caminos que van de (k, x) a (n, y) .

Principio de reflexión El número de caminos que van de (k, x) a (n, y) que toca o cruza el eje de las abscisas es igual al número de caminos que van de $(k, -x)$ a (n, y) .

Dem Considere un camino $x = s_k, s_{k+1}, \dots, s_n = y$ que toque el eje de las abscisas. Sea T el primer instante en que eso sucede:

$$T = \min\{i \in [k, n] : s_i = 0\}$$

El camino

$$-x = -s_k, -s_{k+1}, \dots, -s_{T-1}, 0, s_{T+1}, \dots, s_n = y$$

va de $(k, -x)$ a (n, y) .

Como las secciones $(k, x), \dots, (t, 0)$ y $(k, -x), \dots, (t, 0)$ son reflejadas una de la otra, existe una biyección entre esos dos pedazos. Esto implica que el número de caminos es el mismo. \square

Lema (del escrutinio) Sean n y x enteros positivos. Hay exactamente $\frac{x}{n} N_{n,x}$ caminos $(s_1, \dots, s_n = x)$ desde el origen a (n, x) tal que $s_1 > 0, \dots, s_n > 0$.

Dem Claramente hay tantos caminos admisibles como caminos desde $(1, 1)$ a (n, x) que no tocan el eje de las abscisas. Por el lema de la reflexión, ese número es

$$N_{n-1, x-1} - N_{n-1, x+1} = \binom{a+b-1}{a-1} - \binom{a+b-1}{a}$$

con a y b satisfaciendo que $a+b = n$ y $a-b = x$. Una cuenta muestra que ese número es igual a $\frac{x}{n} N_{n,x}$. \square

Paseos aleatorios son cadenas de Markov Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes con distribución

$$P(X_i = 1) = \frac{1}{2}, \quad P(X_i = -1) = \frac{1}{2}.$$

Se define **paseo aleatorio** al proceso

$$S_n = X_1 + \dots + X_n, \quad n \geq 0$$

S_n es una **cadena de Markov** con transiciones

$$q(x, x+1) = \frac{1}{2}, \quad q(x, x-1) = \frac{1}{2}.$$

Así, la probabilidad que el paseo esté en x en el instante n es

$$p_{n,x} = P(S_n = x) = \binom{n}{\frac{n+x}{2}} 2^{-n}$$

(se interpreta como 0 si $\frac{n+x}{2}$ no es un entero entre 0 y n .)

Una **vuelta al origen** ocurre en el instante $2k$ si $S_{2k} = 0$. La vuelta sólo puede ocurrir en instantes pares.

Definimos $u_{2k} = P(S_{2k} = 0)$.

$$u_{2k} = \binom{n}{\frac{k}{2}} 2^{-2k}$$

Ejercicio Use la aproximación de Stirling para probar que

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}$$

Eso quiere decir que

$$\lim_{k \rightarrow \infty} u_{2k} \sqrt{\pi k} = 1$$

El TCL nos dice que

$$\lim_n P(S_n \leq r\sqrt{n}) = \phi(r)$$

donde ϕ es la función de distribución acumulada de la Normal standard.

El **primer retorno al origen** ocurre en el instante $2k$ si

$$S_1 \neq 0, \dots, S_{2k-1} \neq 0, S_{2k} = 0$$

y su probabilidad se denota f_{2k} .

Lema Las probabilidades u_{2k} y f_{2k} se relacionan por

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0$$

Dem Use el teorema de la probabilidad total. \square

Sea $T := \min\{n > 0 : S_n = 0\}$ instante del primer retorno al origen.

Lema Sea $n > 0$, entonces

$$P(T > 2n) = P(S_{2n} = 0)$$

Clase del 06 de junio

(La parte en azul no fue dada) **Dem** Por simetría,

$$\begin{aligned} P(T > 2n) &= P(S_1 > 0, \dots, S_{2n} > 0) + P(S_1 < 0, \dots, S_{2n} < 0) \\ &= 2P(S_1 > 0, \dots, S_{2n} > 0) \end{aligned}$$

Por el teorema de la probabilidad total:

$$P(S_1 > 0, \dots, S_{2n} > 0) = \sum_{x \geq 1} P(S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2x)$$

Por el lema de reflexión,

$$\begin{aligned} &P(S_1 > 0, \dots, S_{2n-1} > 0, S_{2n} = 2x) \\ &= 2^{-2n} (N_{2n-1, 2x-1} - N_{2n-1, 2x+1}) = \frac{1}{2} (p_{2n-1, 2x-1} - p_{2n-1, 2x+1}) \end{aligned}$$

Sumando (telescopicamente),

$$\sum_{x \geq 1} \frac{1}{2} (p_{2n-1, 2x-1} - p_{2n-1, 2x+1}) = \frac{1}{2} p_{2n-1, 1} = \frac{1}{2} u_{2n} \quad \square$$

Máximo El **máximo** M_n está definido por

$$M_n(S_0, \dots, S_n) = \max\{S_0, \dots, S_n\}$$

Lema Sea y un entero tal que $n \geq y > 0$. La probabilidad de un camino de $(0, 0)$ a $(2n, 0)$ con un máximo mayor o igual a y es igual a $p_{2n, 2y} = P(S_{2n} = 2y)$.

Dem Queremos calcular $P(M_{2n} \geq y, S_{2n} = 0)$. El número de caminos de $(0, 0)$ a $(2n, 0)$ que tocan o cruzan y es igual al número de caminos de $(0, y)$ a $(2n, y)$ que tocan 0. Por el Lema de reflexión, ese número es igual a $N_{2n, 2y}$. Multiplicando por 2^{-2n} , obtenemos

$$P(M_{2n} \geq y, S_{2n} = 0) = p_{2n, 2y}. \quad \square$$

Observe que

$$p_{2n,2y} = \binom{2n}{\frac{2n+2y}{2}} = \binom{2n}{n+y}$$

Lema

$$\lim_{n \rightarrow \infty} P(M_{2n} \geq b\sqrt{2n} \mid S_{2n} = 0) = e^{-2b^2}$$

Dem Dividiendo la expresión obtenida para $p_{2n,2y}$ por $p_{2n,0} = \binom{2n}{n}2^{-2n}$, cancelan los $(2n)!$ y las potencias de 2 y obtenemos

$$\begin{aligned} P(M_{2n} \geq y \mid S_{2n} = 0) &= \frac{p_{2n,2y}}{p_{2n,0}} = \frac{n!n!}{(n-y)!(n+y)!} \\ &= \frac{n(n-1)\dots(n-y+1)}{(n+y)(n+y-1)\dots(n+1)} \end{aligned}$$

dividiendo cada uno de los términos del denominador por el el correspondiente término del numerador, obtenemos

$$= \left(\left(1 + \frac{y}{n}\right) \left(1 + \frac{y}{n-1}\right) \dots \left(1 + \frac{y}{n-y+1}\right) \right)^{-1}$$

Substituyendo $y = b\sqrt{2n}$, y

$$\begin{aligned} &= \left(\left(1 + \frac{b\sqrt{2}}{\sqrt{n}}\right) \left(1 + \frac{b\sqrt{2}}{\sqrt{n} - \frac{1}{\sqrt{n}}}\right) \dots \left(1 + \frac{b\sqrt{2}}{\sqrt{n} - \frac{b\sqrt{2}+1}{\sqrt{n}}}\right) \right)^{-1} \\ &\sim \left(1 + \frac{b\sqrt{2}}{\sqrt{n}}\right)^{-b\sqrt{2}\sqrt{n}} \rightarrow e^{-2b^2} \quad \square \end{aligned}$$

Inferencia estadística - Estimación puntual

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 con la encuesta se obtiene una **muestra aleatoria** X_1, \dots, X_n de X . $X_i = 1$ si el i -ésimo elector de la muestra vota por el candidato.

La **proporción muestral** es la variable aleatoria

$$\hat{p}_n = \frac{X_1 + \dots + X_n}{n}$$

El **error** cometido al estimar p por \hat{p}_n es

$$|\hat{p}_n - p|$$

que por supuesto también es aleatorio.

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.

Los parámetros se estiman a partir de la muestra.

Cualquier función de la muestra es una variable aleatoria. Por ejemplo: \bar{X}_n , $\max(X_1, \dots, X_n)$, etc.

Una vez obtenida la muestra los valores observados (x_1, \dots, x_n) serán denotados con minúsculas.

Estimación puntual.

Definición: Un estimador puntual de un parámetro θ de la distribución de X es una función de la muestra de X :

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$$

Ejemplo: Con el fin de estudiar si un dado es o no equilibrado, se arroja el dado 100 veces en forma independiente, 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 una variable aleatoria que depende de parámetros $\theta_1, \dots, \theta_m$. Sea EX^k el **momento** de orden k de X . Es una función g_k de los parámetros:

$$EX^k = g_k(\theta_1, \dots, \theta_m)$$

Sea X_1, \dots, X_n una muestra de X .

Momento muestral de orden k :

$$\frac{\sum_{i=1}^n X_i^k}{n}$$

Cuando la muestra observada es (x_1, \dots, x_n) , los momentos observados de orden k son

$$\frac{\sum_{i=1}^n x_i^k}{n}$$

Defina $\hat{\theta}_1, \dots, \hat{\theta}_m$ los parámetros que se obtienen al igualar los primeros momentos muestrales a los momentos poblacionales. Más precisamente, $\hat{\theta}_1, \dots, \hat{\theta}_m$ es la solución de las ecuaciones

$$g_k(\theta_1, \dots, \theta_m) = \frac{\sum_{i=1}^n x_i^k}{n}, \quad k = 1, \dots, m.$$

Es decir que $\hat{\theta}_i = \hat{\theta}_i(x_1, \dots, x_n)$ es una función de la muestra observada.

Substituyendo (x_1, \dots, x_n) por (X_1, \dots, X_n) , obtenemos las variables aleatorias $\hat{\theta}_i(X_1, \dots, X_n)$ que se llaman **estimadores de momentos** de $(\theta_1, \dots, \theta_m)$.

Ejemplo 1. $X \sim \text{exponencial}(\lambda)$. 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. $X \sim \text{Gama}(\alpha, \lambda)$. Dos parámetros, dos ecuaciones:

$$EX = \bar{X}_n, \quad EX^2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

Como $EX = \frac{\alpha}{\lambda}$ y $EX^2 = \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2}$, las ecuaciones quedan

$$\frac{\alpha}{\lambda} = \bar{X}_n, \quad \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2} = \frac{\sum_{i=1}^n X_i^2}{n}$$

De aqui se despejan λ y α :

$$\hat{\lambda} = \frac{\bar{X}}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2}$$

$$\hat{\alpha} = \frac{\bar{X}^2}{\frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2}$$

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 :

$$\hat{\theta} = 2\bar{X}_n$$

Ejemplo 4. No siempre se puede usar el primer momento. X Uniforme en $[-\theta, \theta]$. $EX = 0$ no depende de θ , así 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

$$\frac{\theta^2}{3} = \frac{\sum_{i=1}^n X_i^2}{n}$$

Y 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 opinión con muestra de 20 personas. Se le formula una única pregunta que será respondida por SÍ o por NO. Queremos estimar la probabilidad p de SÍ.

$X \sim$ Bernoulli(p). (x_1, \dots, x_n) son los valores observados.

Probabilidad de haber observado (x_1, \dots, x_n) :

$$P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \prod_i p^{x_i} (1-p)^{1-x_i}$$

Cual es el valor de p que maximiza esa proba?

$$\arg \max_p \left[\prod_i p^{x_i} (1-p)^{1-x_i} \right]$$

$$= \arg \max_p \left[(\log p) \sum_i x_i + \log(1-p) \sum_i (1-x_i) \right]$$

Buscamos el punto crítico derivando en p :

$$\frac{\partial g(p)}{\partial p} = \frac{1}{p} \sum_i x_i - \frac{1}{1-p} \sum_i (1-x_i) = 0$$

$$\hat{p} = \frac{\sum_i x_i}{n}$$

Calculando la derivada segunda vemos que maximiza.

Definición de estimador de máxima verosimilitud Sea X una variable aleatoria con probabilidad $p(\cdot)$ o densidad conjunta f que depende de parámetros $(\theta_1, \dots, \theta_m)$.

La **función de verosimilitud** está definida por

$$L(\theta_1, \dots, \theta_m) = \begin{cases} p(x_1) \dots p(x_n) & \text{caso discreto} \\ f(x_1) \dots f(x_n) & \text{caso continuo} \end{cases}$$

$L(\theta_1, \dots, \theta_m)$ es la probabilidad de observar (x_1, \dots, x_n) cuando los parámetros son $(\theta_1, \dots, \theta_m)$.

El estimador de máxima verosimilitud es el vector $(\hat{\theta}_1, \dots, \hat{\theta}_m)$ que maximiza L .

Como variables aleatorias, el EMV es el que se obtiene al reemplazar x_i por las va X_i .

Ejemplos

1. (X_1, \dots, X_n) exponencial λ

$$L(\lambda) = \lambda^n e^{-\lambda(x_1 + \dots + x_n)}$$

$$\log L(\lambda) = n \log \lambda - \lambda(x_1 + \dots + x_n)$$

Derivando e igualando a cero

$$\frac{\partial L}{\partial \lambda} = \frac{n}{\lambda} - (x_1 + \dots + x_n) = 0$$

De donde

$$\hat{\lambda} = \frac{1}{\bar{x}_n}$$

(verifique que es un máximo con la segunda derivada)

2. (X_1, \dots, X_n) Normal (μ, σ^2)

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2\right)$$

Maximizarla equivale a maximizar los logaritmos.

El resultado es:

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

2. (X_1, \dots, X_n) Uniforme $(0, \theta)$

$$L(\theta) = \frac{1}{\theta^n} \prod_i I_{x_i \in [0, \theta]}$$

$$L(\theta) = 0 I_{\theta < \max_i x_i} + \frac{1}{\theta^n} I_{\theta \geq \max_i x_i}$$

De donde $\hat{\theta} = \max_i x_i$

Clase del 11 de junio

Propiedades de los estimadores

Dada una muestra (X_1, \dots, X_n) de $X \sim F_\theta$, un estimador puntual de θ es una función de la muestra $\hat{\theta}$. La diferencia

$$\hat{\theta} - \theta$$

es el error de estimación y una estimación será más precisa cuanto menor sea este error.

Este error es también una variable aleatoria dado que es función de la muestra.

Propiedad deseable: que la esperanza del error sea 0, es decir que “en promedio” el error obtenido al estimar a partir de diferentes muestras sea cero.

Definición: Un estimador $\hat{\theta}$ de θ es **insesgado** si

$$E_{\theta}\hat{\theta} = \theta$$

Si el estimador no es insesgado, el **sesgo** se define por

$$b(\hat{\theta}) = E_{\theta}\hat{\theta} - \theta$$

Un estimador $\hat{\theta}$ de θ es **asintóticamente insesgado** si

$$\lim_n E_{\theta}\hat{\theta} = \theta$$

Ejemplos. 1. $X \sim \text{Bernoulli}(p)$. Usamos la proporción muestral \hat{p} como estimador de p . Como

$$E_p\hat{p} = p$$

\hat{p} es insesgado.

2. **Normal.** $X \sim N(\mu, \sigma)$. 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.

S^2 es estimador insesgado de σ^2 .

$\hat{\sigma}^2$ es estimador asintóticamente insesgado de σ^2 .

3. $X \sim \text{Uniforme}[0, \theta]$.

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:

$$\begin{aligned} E_{\theta}M &= \int_0^{\theta} P_{\theta}(M > x)dx = \int_0^{\theta} (1 - P_{\theta}(M \leq x))dx = \int_0^{\theta} \left(1 - \left(\frac{x}{\theta}\right)^n\right)dx \\ &= \theta - \frac{\theta^{n+1}}{(n+1)\theta^n} = \frac{n}{n+1}\theta \end{aligned}$$

El EMV no es insesgado pero es asintóticamente insesgado.

Consistencia

Sea θ_n un estimador de θ . Diremos que θ_n es un estimador **consistente** de θ si

$$\theta_n \rightarrow \theta, \quad \text{en probabilidad}$$

Es decir si para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P_{\theta}(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

Ejemplo Si X tiene media μ y varianza σ^2 , entonces \bar{X}_n es un estimador consistente de μ . Ya lo vimos, usando Chebichev.

Verifique que $(X_1 + X_n)/2$ no es consistente.

Lema Si un estimador es asintóticamente insesgado y su varianza va a cero, entonces es consistente.

Dem: Inmediata si es insesgado, por Chebichev. En el caso general no lo haremos. \square

Ejemplo $X \sim \text{Uniforme } [0, \theta]$. $\hat{\theta} = \max X_i$ es asintóticamente insesgado. $E_{\theta}(\hat{\theta}) = \frac{n}{n+1}\theta$.

Calcular la varianza del máximo de n uniformes dá

$$\frac{n}{(n+1)(n+2)^2}\theta^2 \rightarrow_n 0$$

Por lo tanto $\hat{\theta} = \max X_i$ es consistente.

Lema S^2 es un estimador consistente de la varianza poblacional.

Dem

$$S^2 = \dots = \frac{n}{n-1} \sum_i \left(\frac{X_i^2}{n} - \bar{X}^2 \right)$$

Como $\bar{X}_n \rightarrow \mu$, $(\bar{X}_n)^2 \rightarrow \mu^2$.

Por la LGN:

$$\sum_i \frac{X_i^2}{n} \rightarrow E_{\mu, \sigma^2} X^2 = \mu^2 + \sigma^2$$

Como $n/(n-1) \rightarrow 1$,

$$S_n^2 \rightarrow \mu^2 + \sigma^2 - \mu^2 = \sigma^2. \quad \square$$

Intervalos de confianza Hasta ahora 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 $X \sim N(\mu, \sigma^2)$ con μ desconocida y σ^2 conocida. Sabemos que $\bar{X}_n \sim N(\mu, \sigma^2/n)$ y que

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

De donde,

$$P(-1,96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq 1,96) = 0,95$$

que equivale a

$$P(\bar{X} - 1,96 \sigma/\sqrt{n} \leq \mu \leq \bar{X} + 1,96 \sigma/\sqrt{n}) = 0,95$$

Es decir que la proba que el intervalo

$$[\bar{X} - 1,96 \sigma/\sqrt{n}, \bar{X} + 1,96 \sigma/\sqrt{n}]$$

contenga μ (el verdadero valor) es 0,95.

Se llama **intervalo de confianza para μ de confianza 0,95**.

Definición Sea X una variable aleatoria cuya distribución depende de un parámetro θ . Sea (X_1, \dots, X_n) una muestra de X . Dadas dos funciones a y b de la muestra tales que

$$P(a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n)) = 1 - \alpha$$

$[a, b]$ se denomina intervalo de confianza a nivel $1 - \alpha$ para el parametro θ .

Observaciones: 1) El intervalo $[a, b]$ es aleatorio ya que sus extremos son funciones de la muestra. "La probabilidad de que el intervalo (a,b) contenga al parámetro 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 con parámetro p (desconocido). Sea \hat{p}_n el estimador puntual de p . Queremos establecer la relación entre el radio del intervalo dado por el error ε y la confianza $1 - \alpha$ en la expresión

$$P(\hat{p}_n - \varepsilon < p < \hat{p}_n + \varepsilon) = 1 - \alpha$$

que equivale a

$$P(|\hat{p}_n - p| < \varepsilon) = 1 - \alpha$$

Standardizando obtenemos la expresión equivalente

$$P\left(\frac{|\hat{p}_n - p|}{\sqrt{p(1-p)}/\sqrt{n}} < \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}}\right) = 1 - \alpha$$

Por el teorema del límite central, aproximadamente

$$\sim P(|Z| < \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}}) = 1 - \alpha$$

para $Z \sim N(0, 1)$. Aceptando la aproximación como identidad, obtenemos la siguiente relación:

$$z = \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}} \tag{3}$$

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 $N(0, 1)$.

El **error** es el radio del intervalo de confianza y se denota ε .

Preguntas:

- 1) Dado el error ε y el tamaño n de la muestra, cual es la confianza del intervalo obtenido?
- 2) Dado el error ε y la confianza que deseamos que tenga el intervalo obtenido, cual es el tamaño n de la muestra?
- 3) Dada la confianza que deseamos que tenga el intervalo obtenido y el tamaño n de la muestra, cual es el error obtenido?

Clase del 13 de junio

Respuestas: Use la identidad (3) para obtener lo siguiente:

1) Se obtiene z con la fórmula

$$z = \frac{\varepsilon}{\sqrt{p(1-p)}/\sqrt{n}} \geq \frac{\varepsilon\sqrt{n}}{1/2}$$

que es el peor caso para $p(1-p)$. Entonces calculamos

$$z = 2\varepsilon\sqrt{n}$$

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 buscamos n .

A partir de (3) despeje n :

$$n = \frac{z^2 p(1-p)}{\varepsilon^2} \geq \frac{z^2}{2\varepsilon^2}$$

dado que $1/2$ es el mayor valor posible para $\sqrt{p(1-p)}$.

Obtenga z usando la tabla: es el valor que satisface $\phi(z) = 1 - \alpha/2$ y substituya arriba para obtener el valor de n mínimo.

3) Ahora conocemos $1 - \alpha$ y n y buscamos ε . Despeje en (3):

$$\varepsilon = \frac{z\sqrt{p(1-p)}}{\sqrt{n}} \geq \frac{z}{2\sqrt{n}}$$

tomando el peor caso.

Obtenemos z a partir de $1 - \alpha$ como antes y listo.

Intervalo de confianza asintótico para la media de variables con varianza conocida

Sea X una variable aleatoria con media μ (desconocida) y varianza σ^2 conocida.

Usamos que la distribución asintótica de

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

es aproximadamente $N(0,1)$ para obtener el siguiente intervalo de confianza asintótica $1 - \alpha$:

$$\left[\bar{X} - z \frac{\sigma}{\sqrt{n}}, \bar{X} + z \frac{\sigma}{\sqrt{n}} \right]$$

donde $P(Z < z) = 1 - \alpha/2$

Test de Hipotesis

Una empresa fabrica motores que gastan en media 10 litros a cada 100 km.

Se fabrican prototipos de un nuevo modelo de motor y se decide que el nuevo modelo se va a fabricar si su consumo es menor o igual a 10 (litros/100 km).

El consumo del nuevo motor se modela por una variable aleatoria X (litros/100 km).

Asumimos $X \sim N(\mu, 1)$. Varianza conocida.

Necesitamos saber si

$H_0 : \mu = 10$, en ese caso el nuevo proyecto será descartado.

$H_1 : \mu < 10$, en ese caso el nuevo motor será fabricado.

Obtenemos una muestra aleatoria (X_1, \dots, X_9) de X y calculamos su media muestral \bar{X}_9 .

Test de hipótesis: Si $\bar{x}_9 < 9,5$, se rechaza H_0 y se emprende la fabricación del nuevo motor.

En caso contrario, se acepta H_0 y no se fabrica el nuevo motor.

Es decir testeamos la hipótesis H_0 con el criterio "si la media muestral está abajo de 9,5, la rechazamos; si no, la aceptamos".

Región crítica (o de rechazo) para \bar{x} es el intervalo $(-\infty, 9,5]$.

Por ejemplo, si observamos $\bar{x} = 9,2$. Qué hacemos?

Como el valor observado está en la región crítica (es menor que 9,5), 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.

Cual es la probabilidad de cometer el error de tipo 1?

Usaremos que bajo H_0 conocemos la distribución de \bar{X}_9 .

La media muestral tiene distribución normal: $\bar{X}_9 \sim N(\mu, 1/9)$.

Cálculo de la probabilidad del error 1

$$\begin{aligned}\alpha &= P(\text{error tipo 1}) = P(\bar{X}_9 < 9,5 | H_0 \text{ verdadera}) \\ &= P(\bar{X} \leq 9,5 | \mu = 10) \\ &= P((\bar{X} - 10)/(1/3) \leq (9,5 - 10)/(1/3) | \mu = 10)\end{aligned}$$

pero, como bajo $\mu = 10$, $Z = (\bar{X} - 10)/(1/3) \sim N(0, 1)$,

$$= P(Z < -1,5) = 0,07 \quad (\text{por la tabla})$$

α es el **nivel de significancia** del test.

Si por el contrario observamos 9,7, no rechazamos H_0 .

Qué quiere decir $\alpha = 0,07$? Que de cada 100 muestras que provienen de una población con H_0 verdadera (es decir $\mu = 10$), rechazaremos (equivocadamente) H_0 en 7 de los tests.

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 o función de una muestra de X y en una zona de rechazo, es decir un conjunto de valores para los cuáles se rechaza la hipótesis nula H_0 .

En el ejemplo anterior el estadístico era \bar{X} y la zona de rechazo el intervalo $(-\infty, 9,5]$.

La zona de rechazo es también una función de la muestra.

La regla de decisión es aleatoria, porque depende del valor del estadístico.

Podemos equivocarnos. Por ejemplo podemos rechazar H_0 aún siendo $\mu = 10$.

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

$\alpha = P(\text{error tipo 1})$ Nivel de significancia.

$\beta = P(\text{error tipo 2})$

¿Cómo se elige la zona de rechazo?

Elegiremos la zona de rechazo del test de manera que la probabilidad de error tipo 1 sea un valor α predefinido.

En el ejemplo, para $\alpha = 0,05$, buscamos z tal que $\phi(z) = 1 - 0,05$ y rechazamos H_0 si $\frac{\bar{X}-10}{1/3} < -z$ que corresponde a $-z = -1,64$ y

$$\bar{x} \leq 10 - \frac{1,64}{3} = 10 - 0,54 = 9,46$$

Para $\alpha = 0,10$ rechazamos si $\bar{x} \leq 9,46$.

P-valor Otra manera de hacer el test es considerar un estadístico llamado P -valor.

Si estamos considerando el estadístico T y observamos $t_{\text{observado}}$, el P -valor es el α correspondiente a la región crítica para T cuyo extremo es $t_{\text{observado}}$.

En particular, para el ejemplo anterior con el estadístico $T = \bar{X}$, si se la muestra observada es x_1, \dots, x_n y la media muestral observada es $\bar{x} = \bar{x}_{\text{observado}} = 9,5$, el P -valor es

$$\begin{aligned} P\text{-Valor}(x_1, \dots, x_n) &= P(\bar{X} < \bar{x} | H_0) \\ &= P(\bar{X}_9 < 9,5 | \mu = 10) = P(Z < 31,5) = 0,7. \end{aligned}$$

Esto quiere decir que si hacemos un test con $\alpha < 0,7$, no podremos rechazar H_0 .

Substituyendo (x_1, \dots, x_n) por (X_1, \dots, X_n) , obtenemos el estadístico $P(X_1, \dots, X_n)$. 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 crítica para el P -valor es $[0, \alpha]$.

Error tipo 2

Supongamos que en nuestro ejemplo, observamos un consumo promedio en la muestra de tamaño 9 igual a 9.5 litros y trabajamos con el test de nivel 0.05.

En este caso,

$$\bar{x} = 9,5 \geq 9,46$$

que está fuera de la región crítica $(-\infty, 9,46]$. Por lo tanto no rechazamos H_0 .

Podríamos estar cometiendo un error de tipo 2.

Por ejemplo, si el nuevo motor consume a 9.3 litros cada 100 km, ¿cuál es la probabilidad de cometer un error tipo II?

$$\begin{aligned} P(\text{error tipo 2}) &= P(\text{aceptar } H_0 | H_1 \text{ verdadera, con } \mu = 9,3) \\ &= P(\bar{X}_9 > 9,46 | H_1 \text{ verdadera, con } \mu = 9,3) \\ &= P\left(\frac{\bar{X} - 9,3}{1/3} > \frac{9,46 - 9,3}{1/3} \middle| \mu = 9,3\right) \\ &= P(Z > 0,69) = 1 - 0,7549 = 0,2451 \end{aligned}$$

(usando la tabla).

El error de tipo 2 es una **función** del valor alternativo de H_1 y de la región crítica.

En este caso $\beta(9,3) = 0,2451$. Depende de la región crítica y del valor alternativo

Clase del 18 de junio

Analogía con el sistema de justicia

Suponga que alguien es acusado 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”. El juicio 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”). Por ese motivo 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.

Función de potencia de un test, Fijada la región crítica, se llama **potencia** $\pi(\mu)$ a la función que da la probabilidad de rechazar la hipótesis nula cuando el valor verdadero del parámetro es μ .

Utilizando la función de potencia es posible obtener una expresión general para los dos tipos de errores, pues

$$\pi(\mu) = \alpha(\mu)I\{\mu \in H_0\} + (1 - \beta(\mu))I\{\mu \in H_1\}$$

Tipos de hipótesis

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 $H_0, T \sim N(0, 1)$, 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.

La distribución **t de Student** es la distribución de probabilidad del cociente

$$\frac{Z}{\sqrt{Y/q}}$$

donde Z tiene una distribución normal de media nula y varianza 1

Y tiene una distribución qui-cuadrado con q grados de libertad

Z e Y son independientes

La distribución χ^2 (de Pearson), llamada **qui cuadrado**, es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria

$$Y = Z_1^2 + \dots + Z_k^2$$

donde Z_i son variables aleatorias normales independientes de media cero y varianza uno.

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 $T \sim \chi_{n-1}^2$ (Qui-cuadrado con $n - 1$ grados de libertad).

Regiones de rechazo son:

1) $RC = (-\infty, -x_\alpha]$

2) $RC = [\chi_{1-\alpha}^2, \infty)$

3) $RC = (-\infty, x_{\alpha/2}] \cup [x_{1-\alpha/2}^+, \infty)$

donde x_α satisface $P(\chi_{n-1}^2 < x_\alpha) = \alpha$. 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

$\bar{x} = 243^\circ C$ y $s = 2,8^\circ C$

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 $250^\circ C$.

b) si existe evidencia para decidir que la varianza de la temperatura en ese sector del reactor es mayor que $(2^\circ C)^2$.

a) Las hipótesis a testear son $H_0: \mu = 250$ (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 $250^\circ C$.

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 $[\chi_{n-1,0,05}^2, \infty)$.

En nuestro caso, $n = 25$ y por lo tanto $\chi_{24,0,05}^2 = 36,42$. 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 $(2^\circ C)^2$.

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 $N(0, 1)$ por el TCL.

Se toma n "grande" y se trabaja como en el caso de $X \sim N(\mu, \sigma^2)$. 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 $P(Z < z_\alpha) = 1 - \alpha, Z \sim N(0, 1)$.

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

$$T = \sqrt{n} \frac{\bar{X} - p_0}{\sqrt{p(1-p)}},$$

Como bajo $H_0, T \sim N(0, 1)$ asintóticamente (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 $X \sim N(\mu, \sigma^2)$. 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

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$$

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 .

clase del 25 de junio

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
Probab	1/4	1/2	1/4

Hay 3 categorías: 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:

$$E_{AA} := 100 P(AA) = 100 \frac{1}{4} = 25$$

$$E_{Aa} := 100 P(Aa) = 100 \frac{1}{2} = 50$$

$$E_{aa} := 100 P(aa) = 100 \frac{1}{4} = 25$$

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 , $i = 1, \dots, k$.

Sea p_i la probabilidad asociada a la categoría i .

El objetivo es testear las hipótesis

$$H_0 : p_1 = p_{o1}, \dots, p_k = p_{ok}$$

H_1 : existe por lo menos una diferencia.

Aquí p_{oi} 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

$$E_i = np_{oi}, \quad i = 1, \dots, k.$$

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

$$\chi_{k-1}^2(\underline{O}) = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

donde $\underline{O} = (O_1, \dots, O_k)$ 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,

$$P(\underline{o}) = P(\chi_{k-1}^2(\underline{O}) \geq \chi_{k-1}^2(\underline{o})),$$

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 ($k = 3$):

$$\chi_{k-1}^2(\underline{o}) = 0,04 + 0,50 + 0,64 = 1,18$$

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

Hipótesis que serán testeadas

Test de independencia

H_0 : X e Y son variables independientes.

H_1 : X e Y no son independientes.

Cuántas observaciones debería haber en cada celda de la tabla si X e Y fueran independientes?

En ese caso las probabilidades conjuntas deberían ser iguales al producto de las probabilidades marginales:

$$p_{ij} = P(X = i, Y = j) = P(X = i)P(Y = j)$$

y el número esperado de observaciones 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.

$n_{(i\cdot)}$:= número de observaciones de $X = i$.

$n_{(\cdot j)}$:= número de observaciones de $Y = j$.

n_{ij} := número de observaciones de $X = i$ conjunto con $Y = j$.

El estadístico propuesto bajo la suposición de independencia está dado por:

$$\chi_q^2(Q) = \sum_{i,j} \frac{(E_{ij} - O_{ij})^2}{E_{ij}}$$

donde $O_{ij} = n_{ij}$ representa el número total de observaciones en la celda (i, j) .

Bajo la hipótesis de independencia $\chi_q^2(Q)$ 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

$$P(o) = P(\chi_q^2(Q) \geq \chi_q^2(o))$$

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-cuadrado observado es

$$\chi_q^2(o) = \dots \text{cuentas} \dots = 36,62$$

Determinación del número de grados de libertad:

Categorías de renta: $f = 3$

Categorías de número de hijos: $c = 4$

$$q = (f - 1)(c - 1) = 2 \cdot 3 = 6$$

El P -valor observado es $P(\underline{a}) = P(\chi_6^2 \geq 36,62) = 0,000$ (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.)

Modelos no paramétricos Basado en el *Curso de modelos no paramétricos* de Pedro Delicado, Universidad de Cataluña.

Modelos paramétricos versus no paramétricos X sigue un modelo paramétrico si su distribución de probabilidad F pertenece a una familia de distribuciones indexada por un parámetro θ de dimensión finita:

$$X \sim F, \quad F \in \{\mathcal{F}_\Theta = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^k\}\}$$

La familia de distribuciones \mathcal{F}_Θ recibe el nombre de modelo estadístico paramétrico.

Diremos que X sigue un modelo estadístico no paramétrico si sobre su distribución F unicamente se suponen algunas condiciones de regularidad. Por ejemplo: F es una función de distribución continua.

Métodos no paramétricos Son métodos de inferencia estadística válidos cuando no se hacen hipótesis paramétricas sobre la distribución de los datos.

Test de bondad de ajuste

Sea X v.a. con función de distribución F desconocida.

Sea F_0 una función de distribución conocida. Se desea testear

$$H_0 : F = F_0$$

$$H_1 : F \neq F_0$$

También se pueden considerar las hipótesis alternativas unilaterales:

$$H_1 : F(x) < F_0(x) \text{ para todo } x$$

$$H_1 : F(x) > F_0(x) \text{ para todo } x$$

Disponemos de una muestra (X_1, \dots, X_n) de X .

Vamos a estudiar el test de **Kolmogorov-Smirnov**.

Distribución empírica: Definimos $F_n = F_n(\underline{x}, x)$ por

$$F_n(\underline{x}, x) = \frac{1}{n} \sum_i \mathbf{1}\{x_i \leq x\}$$

cuenta la proporción de observaciones x_i que son menores o iguales a x .

Es fácil ver que para \underline{x} fijo, $F_n(\underline{x}, x)$ como función de x es una función de distribución: Está entre 0 y 1, el límite a la izquierda es 0, el límite a la derecha es 1 y es no decreciente.

Como $F_n(\underline{x}, \cdot)$ depende de \underline{x} , $F_n(\underline{X}, \cdot)$ es una función del vector aleatorio \underline{X} y por lo tanto es una función de distribución aleatoria.

Para cada x fijo cada término $\mathbf{1}\{X_i \leq x\}$ es una variable aleatoria de Bernoulli con probabilidad de éxito $p = P(\mathbf{1}\{X_i \leq x\} = 1) = P(X_i \leq x) = F(x)$

Escribimos $F_n(x)$ en lugar de $F_n(\underline{X}, x)$.

$F_n(x)$ es una variable aleatoria y $nF_n(x)$ tiene distribución binomial con parámetros n y $p = F(x)$.

Propiedades

- 1) $EF_n(x) = F(x)$ para cada $x \in \mathbb{R}$.
- 2) Por la ley de grandes números $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ en probabilidad, para cada $x \in \mathbb{R}$.
- 3) Por el Teorema Central del Límite,

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{F_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} = Z \quad \text{en distribución}$$

donde $Z \sim N(0, 1)$.

Definición

$$D_n^+ := \sup_{x \in \mathbb{R}} (F_n(x) - F(x)), \quad D_n^- := \sup_{x \in \mathbb{R}} (F(x) - F_n(x))$$

$$D_n := \max\{D_n^+, D_n^-\} = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

- 4) Teorema de Glivenko Cantelli.

$$\lim_{n \rightarrow \infty} D_n = 0$$

Esto no lo probaremos.

- 5) Para $z > 0$ tenemos las siguientes convergencias en distribución

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n^\pm > z) = e^{-2z^2}$$

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n > z) = 2 \sum_{i=0}^{\infty} (-1)^{i-1} e^{-2i^2 z^2}$$

- 6) Para n “grande”

$$4n(D_n^+)^2 \sim \chi_2^2$$

Es decir que el supremo de la diferencia converge a una distribución chi-cuadrado de 2 grados de libertad.

Vamos a establecer la región crítica y el P -valor para los tres tests de bondad de ajuste

H_0	H_1	RC (α)	P -valor
$F = F_0$	$F \neq F_0$	$D_n(\underline{x}) \geq d_{n,\alpha}$	$P(D_n \geq D_n(\underline{x}))$
$F = F_0$	$F > F_0$	$D_n^+(\underline{x}) \geq d_{n,\alpha}^+$	$P(D_n^+ \geq D_n^+(\underline{x}))$
$F = F_0$	$F < F_0$	$D_n^-(\underline{x}) \geq d_{n,\alpha}^-$	$P(D_n^- \geq D_n^-(\underline{x}))$

donde $D_n(\underline{x})$ son los valores observados, $d_{n,\alpha}$ está definido por $P(D_n > d_{n,\alpha}) = \alpha$, etc.

Ejemplo Queremos saber si los valores $\{1; 7; 2; 5; 5,3\}$ vienen de una distribución mayor que la uniforme en $[0, 10]$.

$H_0 : F(x) = F_0(x) = \frac{x}{10}$ en $[0, 10]$, etc.

$H_1 : F(x) > F_0(x)$.

Ordenamos los datos: 1;2;5;5.3;7

Calculamos la distribución empírica:

F_n	$F_n - F$	intervalo
0	0	$x < 0$
0	$-\frac{x}{10}$	$0 \leq x < 1$
$\frac{1}{5}$	$\frac{1}{5} - \frac{x}{10}$	$1 \leq x < 2$
$\frac{2}{5}$	$\frac{2}{5} - \frac{x}{10}$	$2 \leq x < 5$
$\frac{3}{5}$	$\frac{3}{5} - \frac{x}{10}$	$5 \leq x < 5,3$
$\frac{4}{5}$	$\frac{4}{5} - \frac{x}{10}$	$5,3 \leq x < 7$
1	$1 - \frac{x}{10}$	$7 \leq x < 10$
1	0	$10 \leq x$

de donde $d_n^+(x) = \sup_x F_n(x) - F(x) = \frac{3}{10}$.

$$4n(d_n^+(x))^2 = 4 \times 5 \times \frac{9}{100} = 1,8$$

P -valor = $P(\chi_2^2 > 1,8) = 0,4$. No se puede rechazar H_0 .

Dos muestras

Queremos testear si dos muestras del mismo tamaño X_1, \dots, X_n de X y Y_1, \dots, Y_n de Y vienen de la misma distribución.

$$H_0 : F_X = F_Y$$

$$H_1 : F_X(x) > F_Y(x) \text{ para todo } x.$$

Supongamos F_X continua. Todas las observaciones son distintas.

Para construir el estadístico, primero ordenamos las muestras. Definiendo

$$A = \{X_1, \dots, X_n, Y_1, \dots, Y_n\}$$

$$T_k = \min(A \setminus \{T_1, \dots, T_{k-1}\}), \quad k = 1, \dots, 2n.$$

Y construimos la trayectoria de un paseo aleatorio: $S_0 = 0$

$$S_k = S_{k-1} + \mathbf{1}\{T_k \in \underline{X}\} - \mathbf{1}\{T_k \in \underline{Y}\}$$

Vamos recorriendo las observaciones ordenadas y subiendo uno cuando la observación viene de la muestra X y bajando 1 cuando viene de la muestra Y .

Como el tamaño de las muestras es el mismo, el paseo aleatorio termina en 0 en el instante $2n$.

Bajo la hipótesis H_0 todas las combinaciones de subidas y bajadas tienen la misma probabilidad $1/2^n$ y el máximo

$$M_{2n} = \max\{S_k, k = 0, \dots, 2n\}$$

del paseo aleatorio S_n satisface el siguiente límite asintótico (como lo probamos en la sección de paseos aleatorios):

$$\lim_{n \rightarrow \infty} P\left(\frac{M_{2n}}{\sqrt{2n}} \geq b \mid S_{2n} = 0\right) = e^{-2b^2}$$

Por otra parte, asintoticamente,

$$\frac{M_{2n}^2}{2n} \sim \chi_2^2$$

Con esto en manos podemos construir nuestro test.