

[61.09] Probabilidad y Estadística

María Inés Parnisari

26 de julio de 2012

Índice

I	Probabilidad	2
1.	Introducción a la estadística y el análisis de datos	2
2.	Probabilidad	2
3.	Variables aleatorias y distribuciones de probabilidad	6
4.	Representación mediante gráficos	9
5.	Media, varianza y covarianza de variables aleatorias	10
6.	Procesos de Bernoulli, Poisson e hipergeométrico	12
7.	Distribución uniforme, normal... y otras	16
8.	Aproximaciones de distribuciones	21
9.	Funciones de variables aleatorias	22
10.	Simulación	25
II	Estadística	26
11.	Definiciones	26
12.	Estadística inferencial	26
13.	Estimadores puntuales	28
14.	Estimación por intervalo	29
15.	Test de hipótesis	33
16.	Estadística Bayesiana	36

Parte I

Probabilidad

1 Introducción a la estadística y el análisis de datos

Sean x_1, x_2, \dots, x_n las observaciones en una muestra.

Media de una muestra: es un promedio numérico. Se define como:

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + \dots + x_n}{n} \\ &= \left\{ x : F_X(x) = \frac{1}{2} \right\}\end{aligned}$$

Mediana de una muestra: si las observaciones están ordenadas en orden creciente, la mediana es:

$$\begin{aligned}\tilde{x} &= \begin{cases} x_{\frac{n+1}{2}} & \text{si } n \text{ es impar} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{si } n \text{ es par} \end{cases} \\ &= \begin{cases} P(X < \tilde{x}) \leq \frac{1}{2} \\ P(X > \tilde{x}) \geq \frac{1}{2} \end{cases}\end{aligned}$$

Moda de una muestra: es el valor que ocurre con más frecuencia en la muestra. Se define como:

$$mod_x = \{x : f_X(x) \text{ es máximo}\}$$

Rango de la muestra: $x_{\max} - x_{\min}$

Varianza de la muestra: $\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$

Desvío de la muestra: $\sigma = \sqrt{\sigma^2}$

2 Probabilidad

Probabilidad: estudio de los fenómenos aleatorios.

2.1 Espacio muestral

Espacio muestral: conjunto de todos los resultados posibles de un experimento estadístico. Se representa con la letra Ω . Hay de dos tipos:

- **Espacio muestral discreto:** contiene un número finito de posibilidades, o una serie interminable con tantos elementos como números existen.
- **Espacio muestral continuo:** contiene un número infinito de posibilidades, tantas como el número de puntos en un segmento de recta.

Punto muestral: cada resultado en un espacio muestral.

2.2 Eventos

Evento: subconjunto de un espacio muestral.

Definimos las operaciones sobre eventos:

- **Complemento:** dado un evento A de un espacio muestral Ω , es el subconjunto de todos los elementos de Ω que no están en A . Se representa con el símbolo \bar{A} .
- **Intersección:** dados dos eventos A y B , se define a la intersección como $A \cap B =$ todos los elementos que están en A y en B . Dos eventos son **mutuamente excluyentes** o **disjuntos** si $A \cap B = \emptyset$.
- **Unión:** dados dos eventos A y B , se define a la unión $A \cup B =$ todos los elementos que están en A , o en B , o en ambos.

La relación entre eventos y el correspondiente espacio muestral Ω se puede ilustrar de forma gráfica usando **diagramas de Venn**.

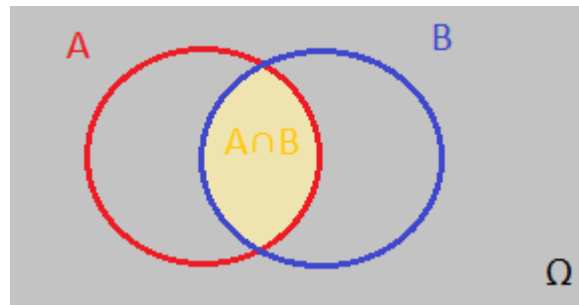


Figura 1: Diagrama de Venn

2.3 Conteo de puntos muestrales

Regla de multiplicación generalizada: si una operación se puede ejecutar en n_1 formas, y si para cada una de éstas se puede llevar a cabo una segunda operación en n_2 formas, y así sucesivamente, la serie de k operaciones se puede realizar en $\prod_{i=1}^k n_i$ formas.

Permutación simple: arreglo de todos los elementos de un conjunto de n objetos, de manera que el orden sí importa y no se repiten los elementos.

$$P_n = n!$$

Ejemplo: las permutaciones posibles de las letras a, b, c ($n = 3$) son: abc, acb, bac, bca, cab y cba ($P_3 = 3! = 6$).

Permutación con repetición: número de permutaciones distintas de n objetos de los que n_1 son de una clase, n_2 de una segunda clase, ..., n_k de una k -ésima clase.

$$P_{n[n_1, n_2, \dots, n_k]} = \frac{n!}{n_1! n_2! \dots n_k!}$$

Ejemplo: un entrenador necesita tener a 10 jugadores parados en una fila. Hay 1 jugador de primer año, 2 de segundo año, 4 de tercer año y 3 de cuarto año. El número de formas diferentes para formar la fila es $\frac{10!}{1!2!4!3!} = 12,600$.

Particiones: número de formas de partir un conjunto de n objetos en r celdas con n_1 objetos en la primera celda, n_2 objetos en la segunda, y así sucesivamente, es

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$$

donde $n_1 + n_2 + \dots + n_r = n$.

Ejemplo: la cantidad de formas de asignar 7 personas ($n = 7$) a una habitación triple y a 2 dobles ($r = 3$) es $\binom{7}{3, 2, 2} = \frac{7!}{3!2!2!} = 210$.

Combinaciones simples: número de formas de seleccionar m objetos de n sin importar el orden. Es un tipo especial de partición con 2 celdas: una con m elementos y la otra con los $n - m$ objetos restantes.

$$C_{n,m} = \binom{n}{m} = \frac{n!}{(n-m)!m!}$$

Ejemplo: el número de formas de seleccionar 3 juegos de 10 disponibles es $\binom{10}{3} = \frac{10!}{7!3!} = 120$.

Combinaciones con repetición: número de formas de seleccionar m objetos de n donde no importa el orden y sí se repiten los elementos.

$$CR_{n,m} = \binom{n+m-1}{n} = \frac{(n+m-1)!}{(m-1)!n!}$$

Ejemplo: en una bodega hay 5 tipos diferentes de botellas. ¿De cuántas formas se pueden elegir 4 botellas? La respuesta es $\binom{5+4-1}{5} = \frac{8!}{3!5!} = 56$.

Variación simple: se llama variación de m elementos tomados de n a los distintos grupos formados por m elementos de forma que sí importa el orden, y no se repiten los elementos.

$$V_{n,m} = \frac{n!}{(n-m)!}$$

Ejemplo: se tienen 7 libros y solo 3 espacios en una biblioteca. La cantidad de formas en las que se pueden colocar 3 libros elegidos es $V_{7,3} = \frac{7!}{(7-3)!} = 210$.

Variación con repetición: se llama variaciones con repetición de m elementos tomados de n a los distintos grupos formados por m elementos de manera que sí importa el orden, y sí se repiten los elementos.

$$VR_{n,m} = n^m$$

Ejemplo: ¿cuántos números de 3 cifras se puede formar con los dígitos: 1, 2, 3, 4 y 5? La respuesta es $5^3 = 125$.

2.4 Probabilidad de un evento

Probabilidad (definición axiomática): la probabilidad de un evento A es la suma de las probabilidades de todos los puntos muestrales en A . Por lo tanto:

1. $0 \leq P(A) \leq 1$
2. $P(\emptyset) = 0$
3. $P(\Omega) = 1$

Teorema de Laplace: si un experimento puede tener N diferentes resultados equiprobables, y si n de estos resultados corresponden al evento A , entonces $P(A) = \frac{n}{N}$.

Axioma de continuidad: para cada sucesión de eventos $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$ tal que $\bigcap_{i=1}^{\infty} A_i = \emptyset$, entonces $\lim_{n \rightarrow \infty} P(A_n) = 0$.

2.5 Reglas aditivas

Teorema: si A y B son dos eventos, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Corolario 1: si A_1, \dots, A_n son mutuamente excluyentes, entonces

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$$

Corolario 2: si A_1, \dots, A_n es una partición del espacio muestral Ω , entonces

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) = P(\Omega) = 1$$

Teorema: si A y \bar{A} son eventos complementarios, entonces $P(A) + P(\bar{A}) = 1$.

2.6 Probabilidad condicional

Probabilidad condicional: la probabilidad de que un evento B ocurra cuando se sabe que ya ocurrió un evento A se llama probabilidad condicional. Se denota como $P(B|A)$ y se define como

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

Eventos independientes: dos eventos A y B son independientes si y sólo si $P(B|A) = P(B)$ y $P(A|B) = P(A)$.

2.7 Reglas multiplicativas

Regla multiplicativa: si en un experimento pueden ocurrir los eventos A y B , entonces

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B \cap A) = P(B) \cdot P(A|B)$$

Eventos independientes: dos eventos A y B son independientes si y sólo si

$$P(A \cap B) = P(A) \cdot P(B)$$

2.8 Regla de Bayes

Regla de eliminación / probabilidad total: si los eventos B_1, \dots, B_k constituyen una partición del espacio muestral Ω tal que $P(B_i) \neq 0$ para $i = 1, \dots, k$, entonces para cualquier evento A de Ω :

$$P(A) = \sum_{i=1}^k P(B_i \cap A) = \sum_{i=1}^k P(B_i) \cdot P(A|B_i)$$

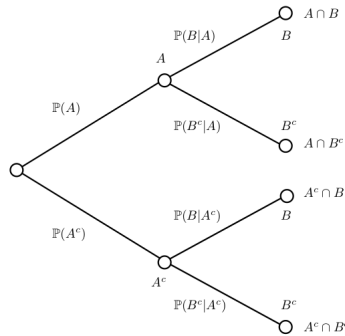


Figura 2: Diagrama de árbol

Regla de Bayes: si los eventos B_1, \dots, B_k constituyen una partición del espacio muestral Ω tal que $P(B_i) \neq 0$ para $i = 1, \dots, k$, entonces para cualquier evento A de Ω tal que $P(A) \neq 0$:

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^k P(B_i \cap A)} = \frac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^k P(B_i) \cdot P(A|B_i)}$$

2.9 Mecánica Estadística

Hay n celdas y r partículas que se distribuyen al azar en las celdas. La distribución de las partículas en las celdas:

<i>Estadística de Maxwell-Boltzmann</i>	<i>Estadística de Bose-Einstein</i>	<i>Estadística de Fermi-Dirac</i>
Las r partículas son distintas.	Las r partículas son indistinguibles.	Hay $r < n$ partículas.
Todas las ubicaciones de las partículas son igualmente posibles.	Todas las ubicaciones de las partículas son igualmente posibles.	Cada celda puede contener, a lo sumo, 1 partícula.
Evento elemental: r – <i>upla</i> (x_1, \dots, x_r) donde x_i = número de celda donde cayó la partícula i	Evento elemental: n – <i>upla</i> (r_1, \dots, r_n) donde r_i = cantidad de partículas en la celda i , $\sum_{i=1}^n r_i = r$	Evento elemental: n – <i>upla</i> (b_1, \dots, b_n) donde $b_i = \begin{cases} 1 & \text{si la celda } i \text{ está ocupada} \\ 0 & \text{si no} \end{cases}$
Cantidad de tuplas posibles: n^r	Cantidad de tuplas posibles: $\binom{r+n-1}{n-1}$	Cantidad de eventos elementales: $\binom{n}{r}$
$P[\text{evento elemental}] = \frac{1}{n^r}$	$P[\text{evento elemental}] = \frac{1}{\binom{r+n-1}{n-1}}$	$P[\text{evento elemental}] = \frac{1}{\binom{n}{r}}$

3 Variables aleatorias y distribuciones de probabilidad

3.1 Concepto de variable aleatoria

Variable aleatoria: función que asocia un número real con cada elemento del espacio muestral. Tiene la forma $f_X(x) : \Omega \rightarrow \mathbb{R}$, donde X denota la variable aleatoria, x uno de sus valores, y Ω el espacio muestral. Hay de tres tipos:

- **Discreta:** hay finitos resultados posibles, o infinitos numerables resultados posibles. Representan datos por conteo.
- **Continua:** su conjunto de valores posibles es un intervalo de números. Representan datos medidos (pesos, alturas, distancias, etc.)
- **Mixta.**

Teorema: X e Y son variables aleatorias independientes sí y sólo si $f(x, y) = f(x)f(y) \forall (x, y)$.

Condición necesaria (pero no suficiente) para que X, Y sean independientes: el soporte de $f(x, y)$ debe ser un rectángulo o un conjunto ordenado de ellos.

3.2 Distribuciones discretas de probabilidad

Función de probabilidad de una variable aleatoria discreta X : conjunto de pares ordenados $(x, P(X = x))$ tales que para cada resultado posible x :

1. $P(X = x) \geq 0$,

$$2. \sum_{\forall x} P(X = x) = 1.$$

Función de distribución acumulada de una variable aleatoria discreta X : dada la distribución de probabilidad $P(X = x)$,

$$F_X(x) = P(X \leq x) = \sum_{t \leq x} P(X = t) \text{ para } -\infty < x < \infty$$

Propiedades de la función de distribución:

1. $P(a < X \leq b) = F_X(b) - F_X(a)$
2. $P(a \leq X \leq b) = F_X(b) - F_X(a) + P(a)$
3. $P(a \leq X < b) = F_X(b) - F_X(a) + P(a) - P(b)$
4. $P(a < X < b) = F_X(b) - F_X(a) - P(b)$
5. $F_X(x)$ es no decreciente y continua a derecha
6. $\lim_{x \rightarrow -\infty} F_X(x) = 0$; $\lim_{x \rightarrow +\infty} F_X(x) = 1$

Átomo de X : se dice que $a \in \mathfrak{R}$ es un átomo de la variable aleatoria discreta X si $P(X = a) > 0$.

3.3 Distribuciones continuas de probabilidad

Función de densidad de probabilidad de una variable aleatoria continua X : función $f(x)$ definida en el conjunto \mathfrak{R} , tal que:

1. $f(x) \geq 0$ para toda $x \in \mathfrak{R}$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $P(X = x) = 0$ para toda $x \in \mathfrak{R}$
4. $P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x) dx$

Función de distribución acumulada de una variable aleatoria continua X : dada la función de densidad $f(x)$, se define a la función de distribución como

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \text{ para } -\infty < x < \infty$$

3.4 Distribuciones mixtas de probabilidad

X es una variable aleatoria **mixta** cuando:

- Existen a tales que $P(X = a) > 0$. Dichos puntos a son los **puntos pesados** de X , es decir, que acumulan probabilidad.
- $P(a \leq X \leq b) = \left[\int_a^b f_X(x) dx \right] + P(X = a) + P(X = b)$

3.5 Distribuciones de probabilidad conjunta

3.5.1 Caso discreto

Función de probabilidad conjunta: si X e Y son dos variables aleatorias discretas, la distribución de probabilidad para sus ocurrencias simultáneas se representa mediante una función $P(x, y)$ que verifica:

1. $P(x, y) = P[(X = x) \cap (Y = y)]$
2. $P(x, y) \geq 0$ para todo (x, y)
3. $\sum_{\forall x} \sum_{\forall y} P(x, y) = 1$

Función de probabilidad marginal:
$$\begin{cases} P(X = x) = \sum_{\forall y} P(x, y) \\ P(Y = y) = \sum_{\forall x} P(x, y) \end{cases}$$

Función de probabilidad condicional:
$$\begin{cases} P(Y = y|X = x) = \frac{P(x, y)}{P(X = x)} \\ P(X = x|Y = y) = \frac{P(x, y)}{P(Y = y)} \end{cases}$$

Función de probabilidad conjunta: $P(x, y) = P(x) \cdot P(y|x) = P(y) \cdot P(x|y)$

3.5.2 Caso continuo

Función de densidad conjunta: la función $f(x, y)$ es una función de densidad conjunta de las variables aleatorias continuas X y Y si:

1. $f(x, y) \geq 0$ para todo (x, y)
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$
3. $P[(X, Y) \in A] = \iint_A f(x, y) dx dy$

Función de probabilidad marginal:
$$\begin{cases} f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \\ f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx \end{cases}$$

Función de probabilidad condicional:
$$\begin{cases} f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \\ f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \end{cases}$$

Función de probabilidad conjunta: $f(x, y) = f(x) \cdot f(y|x) = f(y) \cdot f(x|y)$

3.6 Variables truncadas

Truncado: truncar la variable aleatoria X a un conjunto $B \subset R$ significa condicionarla a tomar valores en el conjunto B .

$$F_{X|X \in B}(x) = P(X \leq x|X \in B) = \frac{P(X \leq x, X \in B)}{P(X \in B)}$$

$$f_{X|X \in B}(x) = \frac{f_X(x) \mathbf{1}\{x \in B\}}{P(X \in B)}$$

3.7 Mezcla de variables aleatorias

Sean n variables aleatorias X_i , donde cada una se asocia a un resultado R_i de un experimento aleatorio. Si $P(R_i)$ es la probabilidad de ocurrencia de cada resultado, y $\sum_{i=1}^n P(R_i) = 1$, entonces:

$$f_{X_M}(x) = f_{X_1}(x) \cdot P(R_1) + f_{X_2}(x) \cdot P(R_2) + \cdots + f_{X_n}(x) \cdot P(R_n)$$

Propiedades:

- $E[X_M] = \sum_{i=1}^n E[X_i] \cdot P(R_i)$
- $\sigma_{X_M}^2 = \sum_{i=1}^n P(R_i) \cdot \left[\sigma_{X_i}^2 + (E[X_i] - E[X_M])^2 \right]$

3.8 Suma aleatoria de variables aleatorias

Dada una cantidad aleatoria Y de términos X_i idénticamente distribuidos, y sea la variable aleatoria

$$W|Y = \sum_{i=1}^Y X_i$$

Entonces:

- $E(W) = E(Y) \cdot E(X)$
- $\sigma^2(W) = E(Y)\sigma^2(X) + E^2(X)\sigma^2(Y)$

4 Representación mediante gráficos

Distribución empírica: la función de distribución empírica $F_n(x)$ de n puntos sobre la recta x_1, \dots, x_n es la función escalera con saltos de altura $\frac{1}{n}$ en los puntos x_1, \dots, x_n .

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq x\}$$

La distribución empírica es una aproximación a la función de distribución acumulada, $F_X(x)$.

Histograma: un histograma de una muestra x_1, \dots, x_n se obtiene eligiendo una partición en m intervalos de extremos $a_0 < \dots < a_m$ con longitudes $L_j = a_j - a_{j-1}$; calculando las *frecuencias relativas*

$$p_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{a_{j-1} < x_i < a_j\}$$

y graficando la función igual a $\frac{p_j}{L_j}$ en el intervalo $(a_{j-1}, a_j]$ y a 0 fuera de los intervalos:

$$f_{x_1, \dots, x_n; a_0, \dots, a_m}(x) = \sum_{j=1}^m \frac{p_j}{L_j} \mathbf{1}\{x \in (a_{j-1}, a_j]\}$$

O sea, un conjunto de rectángulos con área p_j .

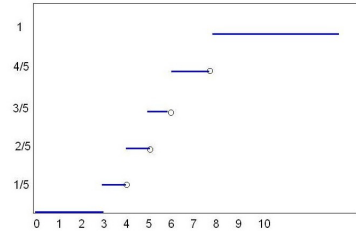


Figura 3: Distribución empírica.

Un histograma es una aproximación a la función de densidad de probabilidad, $f_X(x)$.

Cuantil- α de X : es cualquier número $x_\alpha \in \mathbb{R}$, con $\alpha \in (0, 1)$ tal que

$$P(X < x_\alpha) \leq \alpha \text{ y } P(X \geq x_\alpha) \geq \alpha$$

El cuantil- α de una variable aleatoria absolutamente continua es la única solución de la ecuación

$$\int_{-\infty}^{x_\alpha} f_X(x) dx = \alpha$$

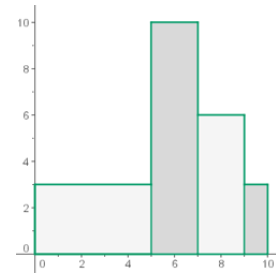


Figura 4: Histograma.

5 Media, varianza y covarianza de variables aleatorias

5.1 Media de una variable aleatoria

5.1.1 Caso univariante

Media / valor esperado / esperanza: sea X una variable aleatoria. La media o valor esperado de X es un tipo de promedio que describe el lugar donde se centra la distribución de probabilidad de X . Es un número real, que a veces no existe.

$$\mu_x = E[X] = \begin{cases} \sum_{\forall x} x \cdot P(X = x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx & \text{si } X \text{ es continua} \\ \sum_{\forall x: P(x) \neq 0} x \cdot P(X = x) + \int_{\forall x: f(x) \neq 0} x \cdot f(x) dx & \text{si } X \text{ es mixta} \end{cases}$$

Teorema: sea X una variable aleatoria. El valor esperado de la variable aleatoria $g(X)$ es:

$$\mu_{g(X)} = E[g(X)] = \begin{cases} \sum_{\forall x} g(x) \cdot P(X = x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx & \text{si } X \text{ es continua} \\ \sum_{\forall x: P(x) \neq 0} g(x) \cdot P(X = x) + \int_{\forall x: f(x) \neq 0} g(x) \cdot f(x) dx & \text{si } X \text{ es mixta} \end{cases}$$

Esperanza condicional: dada la función de densidad $f(x)$ de una variable aleatoria X , y dado un conjunto A , definimos:

$$E[X|X \in A] = \begin{cases} \sum_{\forall x} x \cdot \frac{P(X=x)}{P(X \in A)} & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} x \cdot \frac{f(x)}{P(X \in A)} dx & \text{si } X \text{ es continua} \end{cases}$$

Teorema: dada una variable aleatoria X y un conjunto A de X , tenemos que

$$E[X|A] \cdot P(A) + E[X|\bar{A}] \cdot P(\bar{A}) = E[X]$$

5.1.2 Caso bivalente

Media / valor esperado / esperanza: sean X e Y variables aleatorias con función de densidad de probabilidad conjunta $f(x, y)$. La media o valor esperado de la variable aleatoria $g(X, Y)$ es:

$$\mu_{g(X,Y)} = E[g(X, Y)] = \begin{cases} \sum_{\forall x} \sum_{\forall y} g(x, y) P(x, y) & \text{si } X \text{ e } Y \text{ son discretas} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy & \text{si } X \text{ e } Y \text{ son continuas} \end{cases}$$

Esperanza condicional: sean las variables aleatorias X, Y con función de densidad condicional $f_{Y|X}(y|x)$. La esperanza de X dado $Y = y$ es una variable aleatoria tal que, para cada $y \in \mathbb{R}$

$$E[X|Y=y] = \begin{cases} \sum_{\forall x} x \cdot P(x|y) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} x \cdot f_{X|Y}(x|y) dx & \text{si } X \text{ es continua} \end{cases}$$

Línea de regresión: considerando a x como una variable,

$$\varphi(x) = E[Y|X=x] = \int_{-\infty}^{+\infty} y \cdot f_{Y|X}(y|x) dy$$

Recta de regresión: la recta de regresión de Y basada en X es la función lineal $\hat{Y} = aX + b$ que minimiza la distancia $d(\hat{Y}, Y) = \sqrt{E[(Y - \hat{Y})^2]}$, y se calcula como:

$$\hat{Y} = \frac{\text{cov}(X, Y)}{\text{var}(X)} (X - E[X]) + E[Y]$$

5.2 Varianza y covarianza de variables aleatorias

Varianza: sea X una variable aleatoria con densidad de probabilidad $f(x)$ y media μ . La varianza de X es el número $\sigma_X^2 \geq 0$ que mide las fluctuaciones de X en torno a μ :

$$\sigma_X^2 = E[(X - \mu)^2] = \begin{cases} \sum_{\forall x} (x - \mu)^2 \cdot P(X=x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Teorema: sea X una variable aleatoria con densidad de probabilidad $f(x)$. La varianza de la variable aleatoria $g(X)$ es:

$$\sigma_{g(X)}^2 = E\{[g(X) - \mu_{g(X)}]^2\} = \begin{cases} \sum_{\forall x} [g(x) - \mu_{g(X)}]^2 P(X=x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} [g(x) - \mu_{g(X)}]^2 f(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Teorema: $\sigma_X^2 = E(X^2) - E(X)^2$

Teorema de Pitágoras: $\sigma_X^2 = \sigma^2[E(Y|X)] + E[\sigma^2(Y|X)]$

Varianza condicional: $\sigma^2(Y|X) = E[Y^2|X] - (E[Y|X])^2$

Desviación estándar: $\sigma = \sqrt{\sigma^2}$

Covarianza: sean X e Y variables aleatorias con densidad de probabilidad conjunta $f(x, y)$. La covarianza de X e Y es:

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \begin{cases} \sum_{\forall x} \sum_{\forall y} (x - \mu_X)(y - \mu_Y) \cdot P(x, y) & \text{si } X \text{ e } Y \text{ son discretas} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy & \text{si } X \text{ e } Y \text{ son continuas} \end{cases}$$

Lo que nos importa de la covarianza es su *signo*.

Teorema: la covarianza de dos variables aleatorias X e Y es $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

Coefficiente de correlación: sean X y Y variables aleatorias con covarianza $\text{cov}(X, Y)$ y desviación estándar σ_X y σ_Y respectivamente. El coeficiente de correlación de X y Y es:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

El coeficiente de correlación mide el grado de *linealidad* entre dos variables aleatorias.

Teorema: $|\rho_{XY}| \leq 1$

5.3 Medias, varianzas y covarianzas de combinaciones lineales de variables aleatorias

Desigualdad de Cauchy-Schwartz: $E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}$

Teorema: si X, Y son variables aleatorias, $E[X] = E[E[X|Y]]$

Teorema: si X, Y son variables aleatorias, $E[X \cdot g(Y)|Y] = g(Y) \cdot E[X|Y]$

Teorema: si X e Y son independientes, $E[Y|X] = E[Y]$

Teorema: $E[X] = \begin{cases} \int_0^\infty (1 - F_X(x)) dx & \text{si } F_X(x) = 0 \text{ para } x < 0 \\ \int_0^\infty (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx & \text{si } F_X(x) \neq 0 \text{ para } x < 0 \end{cases}$

Teorema: si $a \in \mathbb{R}$, entonces $E[a] = a$.

Teorema: si $a \in \mathbb{R}$, entonces $E[a|Y] = a$.

Teorema: si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces $E[aX + b] = aE[X] + b$.

Teorema: si $a, b \in \mathbb{R}$ y X, Y, Z son variables aleatorias, entonces $E[aX + bZ|Y] = aE[X|Y] + bE[Z|Y]$.

Teorema: $E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$.

Teorema: $E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)]$.

Teorema: si X e Y son dos variables aleatorias independientes, entonces $E(XY) = E(X)E(Y)$.

Teorema: si X e Y son dos variables aleatorias independientes, entonces $cov(X, Y) = 0$. La recíproca no es cierta.

Teorema: si a y b son constantes, entonces $\sigma_{aX+b}^2 = a^2\sigma_X^2$.

Teorema: si X e Y son variables aleatorias con densidad de probabilidad conjunta $f(x, y)$ entonces $\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

Teorema: $cov(X, X) = \sigma_X^2$.

Teorema: $cov(aX, bY) = ab \cdot cov(X, Y)$.

Teorema: $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$.

6 Procesos de Bernoulli, Poisson e hipergeométrico

Notación: $Y \sim X_{BE}(p)$ significa “la variable aleatoria Y distribuye como una variable aleatoria Bernoulli de parámetro p ”.

6.1 Proceso de Bernoulli y variables asociadas

Proceso de Bernoulli: debe tener las siguientes propiedades.

1. Consistir en n ensayos.
2. Cada ensayo debe tener 2 posibles resultados, “éxito” o “fracaso”.
3. La probabilidad de éxito ($p \in (0, 1)$) debe mantenerse constante en cada ensayo.
4. Los ensayos son independientes (i.e. el muestreo se realiza *con reemplazo*).

Variable de Bernoulli: variable aleatoria dicotómica (“0” es fracaso y “1” es éxito) que verifica

$$P(X_{BE} = x) = p^x(1-p)^{1-x} = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = 0 \end{cases}$$

- Media: p

- Varianza: $p(1-p)$

Propiedades:

- Sean X_1, X_2, \dots variables aleatorias independientes tales que $X_i \sim \text{Bernoulli}(p)$. Sea $S_N = \sum_{i=1}^N X_i$. Si $N \sim \text{Poisson}(\lambda)$, entonces $S_N \sim \text{Poisson}(\lambda p)$

Variable binomial: variable aleatoria que representa la cantidad de éxitos obtenidos en n experimentos Bernoulli independientes, donde la probabilidad de éxito en un experimento es p . Su distribución de probabilidad es:

$$P(X_{BI} = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, \dots, n \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: np
- Varianza: $np(1-p)$

Propiedades:

1. La suma de n variables aleatorias Bernoulli de parámetro p es una variable aleatoria binomial de parámetros n y p .
2. Si $X \sim \text{Binomial}(n_X, p)$ y $Y \sim \text{Binomial}(n_Y, p)$ y X, Y son independientes, entonces la variable aleatoria $W = X + Y$ es $W \sim \text{Binomial}(n_X + n_Y, p)$.

Ejemplo: la probabilidad de que una persona se cure de una enfermedad es 0,4. Si se sabe que 15 contraen esa enfermedad, la probabilidad de que se curen exactamente 5 personas es $P(X_{BI} = 5) = \binom{15}{5} 0,4^5 0,6^{10} = 0,186$.

Variable geométrica: variable aleatoria que representa la cantidad de experimentos Bernoulli independientes necesarios *hasta* obtener el primer éxito (incluyéndolo). Su distribución de probabilidad es:

$$P(X_G = x) = \begin{cases} (1-p)^{x-1} p & \text{si } x = 1, 2, \dots, \infty \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: $\frac{1}{p}$
- Varianza: $\frac{1-p}{p^2}$

Propiedades:

1. Las variables geométricas tienen la propiedad de *pérdida de memoria*. Es decir, si $X \sim \text{Geom}(p)$, entonces $P(X > n + m | X > n) = P(X > m)$, donde $n, m \in \mathbb{N}$.
2. Si X es una variable aleatoria en \mathbb{N} con la propiedad de pérdida de memoria, entonces $X \sim \text{Geom}(p)$, donde $p = P(X = 1)$.

Ejemplo: en un proceso de fabricación, en promedio, 1 de cada 100 artículos está defectuoso. La probabilidad de que el 5to artículo que se inspecciona sea el primer defectuoso encontrado es $P(X_G = 5) = \frac{1}{100} \cdot \left(\frac{99}{100}\right)^4 = 0,0096$.

Variable Pascal / binomial negativa: variable aleatoria que representa la cantidad de experimentos Bernoulli independientes hasta obtener el r -ésimo éxito. Su distribución de probabilidad es:

$$P(X_{PA} = n) = \begin{cases} \binom{n-1}{r-1} p^r (1-p)^{n-r} & \text{si } n = r, r+1, \dots, \infty \\ 0 & \forall \text{ otro } n \end{cases}$$

- Media: $\frac{r}{p}$
- Varianza: $\frac{r(1-p)}{p^2}$

Propiedades:

1. La suma de n variables aleatorias geométricas de parámetro p es una variable aleatoria Pascal de parámetros n y p .
2. Si $X_n \sim \text{Binomial}(n, p)$ y $T_k \sim \text{Pascal}(k, p)$ vale que $P(S_n \geq k) = P(T_k \leq n)$.

Proceso multinomial: debe tener las siguientes propiedades.

1. Consistir en n ensayos.
2. Cada ensayo debe tener k posibles resultados.
3. La probabilidad de éxito (p_k) debe mantenerse constante para cada k .
4. Los ensayos son independientes.

Distribución multinomial: dado un experimento multinomial que consiste en n ensayos independientes, con una probabilidad p_k de obtener el resultado E_k para cada k , entonces la distribución de probabilidad de las variables aleatorias X_1, \dots, X_k (que representan el número de ocurrencias de E_1, \dots, E_k) es:

$$f(x_1, \dots, x_k; p_1, \dots, p_k, n) = \binom{n}{x_1, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

Ejemplo: un aeropuerto tiene 3 pistas. Las probabilidades de que cada una de las pistas sean utilizadas por un avión son $P_1 = \frac{2}{9}$, $P_2 = \frac{1}{16}$, $P_3 = \frac{11}{18}$. ¿Cuál es la probabilidad de que 6 aviones que llegan al azar se distribuyan de la siguiente manera: 2, 1, 3? Resolvemos: $f(2, 1, 3; \frac{2}{9}, \frac{1}{16}, \frac{11}{18}, 6) = \binom{6}{2, 1, 3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{16}\right)^1 \left(\frac{11}{18}\right)^3 = 0,1127$.

6.2 Proceso Poisson y variables asociadas

Proceso Poisson: debe tener las siguientes propiedades.

1. El número de resultados que ocurren en un intervalo o región específica es independiente del número que ocurre en cualquier otro intervalo o región del espacio disjunto. Es decir, el proceso Poisson *no tiene memoria*.
2. La probabilidad de que ocurra un solo resultado en un intervalo muy corto es proporcional a la longitud del intervalo.
3. La probabilidad de que ocurra más de un resultado en un intervalo corto es insignificante.
4. La tasa de éxito λ por unidad de continuo t debe ser constante, $\lambda > 0$.

Teorema (distribución condicional de los tiempos de llegada): dado un proceso Poisson de intensidad λ sobre \mathbb{R}^+ . Si se sabe que en un tiempo “ t ” hubo “ n ” éxitos, entonces cada éxito tiene distribución $\sim U(0, t)$ independiente de los demás.

Variable Poisson: variable aleatoria que representa la cantidad de éxitos obtenidos en un proceso Poisson. Su distribución es:

$$P(X_{PO} = x) = \begin{cases} \frac{(\lambda t)^x}{x!} e^{-\lambda t} & \text{con } x \in \mathbb{N} \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: λt
- Varianza: λt

Propiedades:

1. *Teorema de superposición:* La suma de n variables Poisson de parámetros λ y t genera otra variable Poisson de parámetros $n\lambda$ y t .
2. *Teorema de aditividad:* si $X \sim \text{Poisson}(\lambda_1)$ y $Y \sim \text{Poisson}(\lambda_2)$, y X, Y son independientes, entonces $S = X + Y$ distribuye como $S \sim \text{Poisson}(\lambda_1 + \lambda_2)$.
3. *Teorema del adelgazamiento:* si $X \sim \text{Poisson}(\lambda_x, t)$ y $Y|X \sim \text{Binomial}(p)$ entonces $Y \sim \text{Poisson}(\lambda_y = \lambda_x \cdot p, t)$ y además $X - Y \sim \text{Poisson}(\lambda_{x-y} = (1 - p) \cdot \lambda_x)$.

Variable exponencial: variable aleatoria que representa la cantidad de continuo t hasta obtener el primer éxito en un proceso Poisson, o la cantidad de continuo t entre dos éxitos. Su distribución es:

$$P(X_{EXP} = t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

- Media: $\frac{1}{\lambda}$
- Varianza: $\frac{1}{\lambda^2}$
- Función de distribución acumulada: $F_{X_E}(t) = 1 - e^{-\lambda t}$ si $t > 0$

Propiedades:

1. Las variables aleatorias exponenciales *no tienen memoria*. Es decir: $P\left(\frac{X_{EXP} > t+h}{X_{EXP} > t}\right) = P(X_{EXP} > h)$.
2. *Teorema:* sea T una variable aleatoria continua a valores en \mathbb{R}^+ . Si T pierde memoria, entonces $T \sim \text{Exp}(\lambda = -\log[P(T > 1)])$.
3. *Teorema de la competencia:* si $T_1 \sim \text{Exp}(\lambda_1)$ y $T_2 \sim \text{Exp}(\lambda_2)$, y ambas son independientes, entonces $P(\min(T_1, T_2) = T_1) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, y $P(\min(T_1, T_2) = T_2) = \frac{\lambda_2}{\lambda_1 + \lambda_2}$.
4. *Teorema de la competencia generalizado:* sean las variables aleatorias independientes $T_1 \sim \text{Exp}(\lambda_1)$, $T_2 \sim \text{Exp}(\lambda_2)$, ..., $T_n \sim \text{Exp}(\lambda_n)$. Sea $M = \min(T_1, T_2, \dots, T_n)$. Entonces $M \sim \text{Exp}(\lambda_1 + \lambda_2 + \dots + \lambda_n)$ y $P(M = T_j) = \frac{\lambda_j}{\lambda_1 + \dots + \lambda_n}$.
5. *Teorema (suma geométrica de exponenciales independientes):* sean X_1, X_2, \dots variables i.i.d tales que $X_i \sim \text{Exp}(\lambda)$. Si $S = \sum_{i=1}^N X_i$, donde $N \sim \text{Geom}(p)$, entonces $S \sim \text{Exp}(\lambda p)$

Variable Gamma: variable aleatoria que representa la cantidad de continuo t hasta obtener el k -ésimo éxito en un proceso Poisson. Su distribución es:

$$P(X_\Gamma = t) = \begin{cases} \frac{t^{k-1}}{(k-1)!} \lambda^k e^{-\lambda t} & \text{si } t > 0 \\ 0 & \forall \text{ otra } t \end{cases}$$

- Media: $\frac{k}{\lambda}$
- Varianza: $\frac{k}{\lambda^2}$

Propiedades:

1. La suma de n variables aleatorias exponenciales de parámetro λ es una variable aleatoria Gamma de parámetros (n, λ) .
2. Si $X \sim \Gamma(k_1, \lambda)$ y $Y \sim \Gamma(k_2, \lambda)$ entonces la variable $Z = X + Y$ distribuye como $Z \sim \Gamma(k_1 + k_2, \lambda)$.
3. Si $X \sim \Gamma(k, \lambda)$ entonces $P(X > x) = P(Y < k)$ donde $Y \sim \text{PO}(\lambda, x)$.

6.3 Proceso hipergeométrico y variables asociadas

Proceso hipergeométrico: posee las siguientes propiedades.

1. Dados N artículos, se seleccionan aleatoriamente n de ellos.
2. k de los N artículos se clasifican como “éxito” y $N - k$ como “fracaso”.
3. El muestreo se realiza *sin reemplazo*.

Variable hipergeométrica: variable aleatoria que representa la cantidad de éxitos obtenidos en un proceso hipergeométrico. Su distribución es:

$$P(X_{HG} = x) = \begin{cases} \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} & \text{si } x \in \{\max(0, n + m - N), \dots, \min(m, n)\} \\ 0 & \text{en otro caso} \end{cases}$$

- Media: $\frac{nk}{N}$
- Varianza: $\frac{N-n}{N-1} n \frac{k}{N} \left(1 - \frac{k}{N}\right)$

Ejemplo: tenemos un lote de 100 artículos, de los cuales 12 están defectuosos. La probabilidad de que haya 3 defectuosos en una muestra de 10 artículos es $P(X = 3) = \frac{\binom{12}{3} \binom{88}{7}}{\binom{100}{10}} = 0,08$.

Variable hiperpascal: variable aleatoria que representa la cantidad de extracciones sin reposición en un proceso hipergeométrico *hasta* obtener el r -ésimo éxito. Su distribución es:

$$P(X_{HP} = x) = \frac{\binom{k}{r-1} \binom{N-k}{n-r}}{\binom{N}{x-1}} \cdot \frac{k-r+1}{N-x+1} \text{ con } r \leq x \leq (N-k) + r$$

Distribución hipergeométrica multivariada: dado un experimento hipergeométrico multivariado que consiste en seleccionar n elementos de un universo de tamaño N , donde el universo se puede dividir en k celdas A_1, \dots, A_k con a_1, \dots, a_k elementos cada una, entonces la distribución de probabilidad de las variables aleatorias X_1, \dots, X_k (que representan el número de elementos que se seleccionan de A_k) es:

$$f(x_1, \dots, x_k; a_1, \dots, a_k; N; n) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}}$$

Ejemplo: 10 personas se usan para un estudio biológico. El grupo contiene 3 personas con sangre tipo O, 4 con sangre tipo A y 3 personas con tipo B. La probabilidad de que una muestra aleatoria de 5 personas contenga 1 persona con sangre tipo O, 2 personas con tipo A y 2 personas con tipo B es $f(1, 2, 2; 3, 4, 3; 10, 5) = \frac{\binom{3}{1} \binom{4}{2} \binom{3}{2}}{\binom{10}{5}} = 0,214$.

7 Distribución uniforme, normal... y otras

7.1 Distribución uniforme

Distribución uniforme discreta: si la variable aleatoria X toma los valores x_1, x_2, \dots, x_k con idénticas probabilidades, entonces la distribución uniforme discreta está dada por

$$P(X = x) = \begin{cases} \frac{1}{k} & \text{si } x \in \{x_1, x_2, \dots, x_k\} \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: $\frac{1}{k} \sum_{i=1}^k x_i$

- Varianza: $\frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2$

Distribución uniforme continua: la función de densidad de la variable aleatoria uniforme continua X en el intervalo (a, b) es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: $\frac{a+b}{2}$
- Varianza: $\frac{(b-a)^2}{12}$

7.2 Distribución normal

Distribución normal: la función de densidad de la variable aleatoria normal X , con media μ y desviación estándar σ es:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ para } -\infty < x < \infty$$

La probabilidad $P(a < X < b) = \int_a^b f_X(x) dx$ no se puede calcular analíticamente, porque no existe una primitiva de la función $f_X(x)$. Para efectuar dicho cálculo la variable aleatoria X deberá ser llevada a su forma **normal estándar** Z , que tiene media 0 y desvío 1:

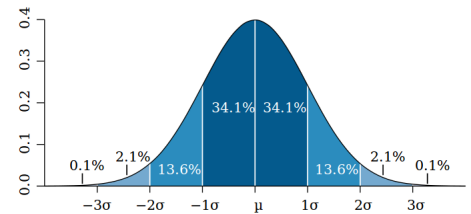


Figura 5: Distribución normal.

$$Z = \frac{X - \mu_X}{\sigma_X} \text{ siendo } f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Entonces, volviendo al cálculo anterior:

$$P(a < X < b) \underset{\text{tabulado}}{=} P\left(\frac{a - \mu_X}{\sigma_X} < Z < \frac{b - \mu_X}{\sigma_X}\right)$$

Propiedades:

1. $F_Z(z) = \Phi(z)$
2. $\Phi(-z) = 1 - \Phi(z)$
3. Si $W = \sum_{i=1}^k (c_i X_i)$ con $c_i \in R$ y X_i variables aleatorias con distribución normal de media μ_{X_i} y varianza $\sigma_{X_i}^2$, entonces W es una variable aleatoria normal con $\mu_W = \sum_{i=1}^k (c_i \cdot \mu_{X_i})$ y si las X_i son independientes entonces $\sigma_W^2 = \sum_{i=1}^k (c_i^2 \cdot \sigma_{X_i}^2)$.
4. La mezcla de variables aleatorias normales no es una variable aleatoria normal.

7.2.1 Teorema Central del Límite (TCL)

Teorema: dados n valores X_i independientes e idénticamente distribuidos, con media μ_X y desvío σ_X conocidos, si $W = \sum_{i=1}^n X_i$ y si n es suficientemente grande (generalmente $n > 25$) entonces W tiene distribución normal con media $n\mu_X$ y desvío $\sqrt{n}\sigma_X$.

7.3 Distribución chi cuadrada

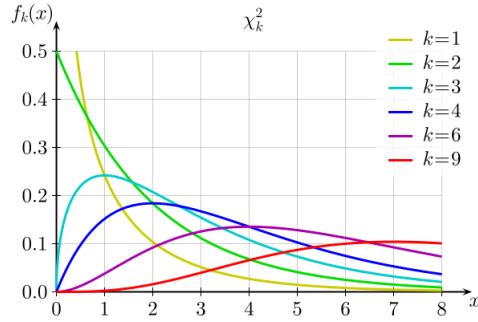


Figura 6: Distribución chi cuadrada

Distribución chi cuadrada: la variable aleatoria continua X tiene una distribución chi cuadrada, con ν grados de libertad, si su función de densidad está dada por

$$f(x; \nu) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \cdot x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}} & \text{si } x > 0 \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: ν
- Varianza: 2ν

La distribución χ_ν^2 NO es simétrica.

Teorema: si $Z \sim \mathcal{N}(0, 1)$, entonces $U = Z^2$ distribuye como $U \sim \chi_{\nu=1}^2$

Teorema: sea $T = \sum_{i=1}^n (Z_i)^2$, donde las $Z_i \sim \mathcal{N}(0, 1)$ son independientes, entonces $T \sim \Gamma(\alpha = \frac{n}{2}, \beta = \frac{1}{2})$, o lo que es lo mismo, $T \sim \chi_{\nu=n}^2$.

7.4 Distribución de Weibull

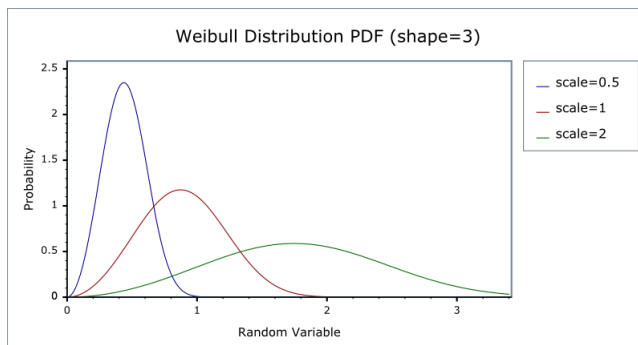


Figura 7: Distribución de Weibull

Distribución de Weibull: la variable aleatoria continua X tiene una distribución de Weibull, con parámetros $\alpha > 0$ y $\beta > 0$, si su función de densidad está dada por

$$f(x; \alpha; \beta) = \begin{cases} \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} & \text{si } x > 0 \\ 0 & \forall \text{ otro } x \end{cases}$$

- Media: $\alpha^{-\frac{1}{\beta}} \cdot \Gamma(1 + \frac{1}{\beta})$
- Varianza: $\alpha^{-\frac{2}{\beta}} \left\{ \Gamma(1 + \frac{2}{\beta}) - \left[\Gamma(1 + \frac{1}{\beta}) \right]^2 \right\}$
- Función de distribución acumulada: $F_X(x) = 1 - e^{-\alpha x^\beta}$ para $x \geq 0$

7.5 Distribución Gamma generalizada

Distribución Gamma generalizada: la variable aleatoria continua X tiene una distribución Gamma generalizada, con parámetros $\alpha > 0$ y $\beta > 0$, si su función de densidad está dada por

$$f(x; \alpha; \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \forall \text{ otro } x \end{cases}$$

donde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\Gamma(n) = (n-1)!$ y $\Gamma(0,5) = \sqrt{\pi}$.

- Media: $\alpha\beta$
- Varianza: $\alpha\beta^2$

7.6 Distribución Beta

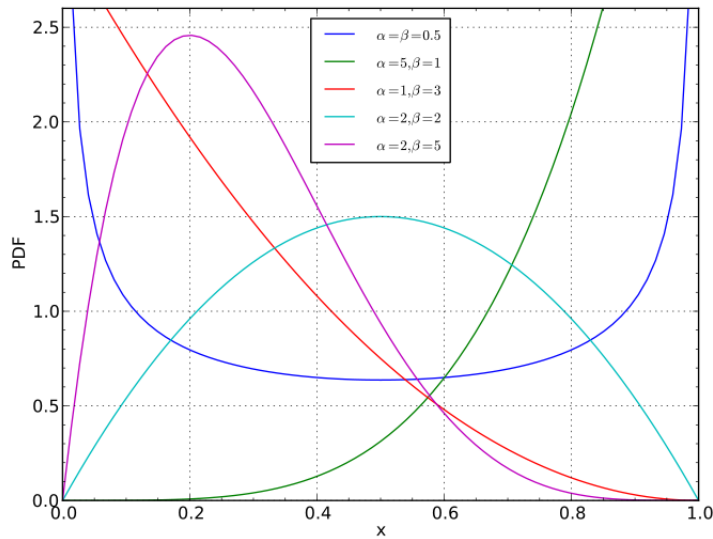


Figura 8: Distribución Beta

Distribución Beta: la variable aleatoria continua X tiene una distribución Beta, con parámetros $\alpha > 0$ y $\beta > 0$, si su función de densidad está dada por

$$f(x; \alpha; \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{si } 0 < x < 1 \\ 0 & \forall \text{ otro } x \end{cases}$$

donde $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\Gamma(n) = (n-1)!$ y $\Gamma(0,5) = \sqrt{\pi}$.

- Media: $\frac{\alpha}{\alpha+\beta}$
- Varianza: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

7.7 Distribución t de Student

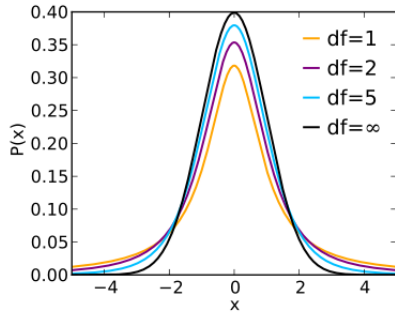


Figura 9: Distribución t de Student

Teorema: sean $Z \sim \mathcal{N}(0, 1)$ y $V \sim \chi^2_\nu$. Si Z y V son independientes, entonces la distribución de

$$T = \frac{Z}{\sqrt{V/\nu}}$$

está dada por

$$h(x) = \frac{\Gamma\left[\frac{\nu+1}{2}\right]}{\Gamma\left[\frac{\nu}{2}\right] \sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \quad \text{para } t \in \mathbb{R}$$

que es la **distribución t de Student** con ν grados de libertad.

- Media: 0 (para $\nu > 1$)
- Varianza: $\frac{\nu}{\nu-2}$ para $\nu > 2$
- $h(x)$ es simétrica.

Teorema: conforme $\nu \rightarrow \infty$, $t_\nu \rightarrow Z$, siendo $Z \sim \mathcal{N}(0, 1)$. En general basta con que $\nu \geq 20$.

7.8 Distribución F de Fisher

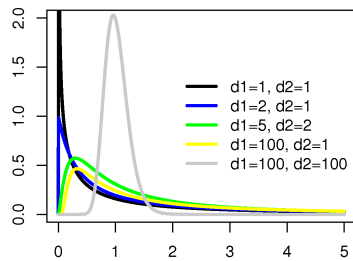


Figura 10: Distribución F de Fisher

Teorema: sean U y V dos variables aleatorias independientes tales que $U \sim \chi^2_u$ y $V \sim \chi^2_v$. La distribución de la variable aleatoria

$$F = \frac{U/u}{V/v}$$

está dada por

$$h(f) = \begin{cases} \frac{\Gamma\left[\frac{u+v}{2}\right]\left(\frac{u}{v}\right)^{u/2}}{\Gamma\left[\frac{u}{2}\right]\Gamma\left[\frac{v}{2}\right]} \frac{f^{(u/2)-1}}{(1+uf/v)^{(u+v)/2}} & \text{si } f > 0 \\ 0 & \text{si } f \leq 0 \end{cases}$$

y se conoce como la **distribución F de Fisher** con u y v grados de libertad. Se la denota como $F_{u,v}$.

- Media: $\frac{v}{v-2}$ (para $v > 2$)
- Varianza: $\frac{2v^2(u+v-2)}{u(v-2)^2(v-4)}$ (para $v > 4$)

8 Aproximaciones de distribuciones

8.1 Aproximación de la binomial mediante la normal

Si $X \sim BI(n, p)$, entonces cuando $n \rightarrow \infty$ y $p \rightarrow \frac{1}{2}$, X tiene aproximadamente una distribución normal $Y \sim \mathcal{N}(\mu = np, \sigma^2 = np(1-p))$, y además

$$\begin{aligned} P(X \leq x) &\approx P(Y \leq x + 0,5) \\ P(X = x) &\approx P(x - 0,5 < Y < x + 0,5) \end{aligned}$$

La aproximación será buena si $np \geq 5$ cuando $p \leq \frac{1}{2}$, o si $np(1-p) \geq 5$ cuando $p > \frac{1}{2}$.

8.2 Aproximación de la binomial mediante la Poisson

Si $X \sim BI(n, p)$, entonces cuando $n \rightarrow \infty$, $p \rightarrow 0$ y $np \rightarrow \mu$ permanece constante, X tiene aproximadamente una distribución Poisson $Y \sim PO(\mu = np)$.

$$P(X = x) \approx P(Y = x)$$

8.3 Aproximación de la Poisson mediante la normal

Si $X \sim PO(\lambda t)$, y $\lambda > 10$ entonces X tiene aproximadamente una distribución normal $Y \sim \mathcal{N}(\mu = \lambda t, \sigma^2 = \lambda t)$.

$$P(X = x) \approx P(x - 0,5 < Y < x + 0,579)$$

Cuando $\lambda > 1000$, $P(X \leq x) = P(Y \leq y)$

8.4 Aproximación de la hipergeométrica mediante la binomial

Si $X \sim HIP(k, N, n)$, y $\frac{n}{N} \leq 0,05$, entonces X tiene aproximadamente una distribución binomial $Y \sim BI(\mu = \frac{nk}{N}, \sigma^2 = n \cdot \frac{k}{N} (1 - \frac{k}{N}))$.

$$P(X = x) \approx P(Y = x)$$

9 Funciones de variables aleatorias

9.1 Transformación de variables aleatorias

9.1.1 Caso discreto

Teorema: sea X una variable aleatoria discreta con distribución de probabilidad $f(x)$. Si tenemos la transformación $Y = u(X)$, que es una función inyectiva $y = u(x)$ de la cual podemos resolver para x en términos de y mediante $x = w(y)$, entonces la distribución de probabilidad de Y es

$$g(y) = f[w(y)]$$

Teorema: sean X_1 y X_2 variables aleatorias discretas con distribución de probabilidad conjunta $f(x_1, x_2)$. Si tenemos las transformaciones $\begin{cases} Y_1 = u_1(x_1, x_2) \\ Y_2 = u_2(x_1, x_2) \end{cases}$, que son funciones inyectivas

$\begin{cases} y_1 = u_1(x_1, x_2) \\ y_2 = u_2(x_1, x_2) \end{cases}$, de las cuales podemos resolver x_1, x_2 en términos de y_1, y_2 mediante $\begin{cases} x_1 = w_1(y_1, y_2) \\ x_2 = w_2(y_1, y_2) \end{cases}$, entonces la distribución de probabilidad conjunta de Y_1 y Y_2 es

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]$$

9.1.2 Caso continuo

Teorema: sea X una variable aleatoria continua con distribución de probabilidad $f(x)$. Definamos con $Y = u(X)$ una transformación inyectiva entre los valores de X y Y , de manera que la ecuación $y = u(x)$ se resuelva para x en términos de y mediante $x = w(y)$. Entonces, la distribución de probabilidad de Y es

$$g(y) = f[w(y)] \cdot |J|$$

donde $J = w'(y)$ es el jacobiano de la transformación.

Teorema: suponga que X es una variable aleatoria continua con distribución de probabilidad $f(x)$. Definamos la transformación *no inyectiva* $Y = u(X)$. Si el intervalo sobre el que se define X se puede dividir en k conjuntos disjuntos, de manera que cada una de las funciones inversas $x_k = w_k(y)$ de $y = u(x)$ defina una correspondencia inyectiva, entonces la distribución de probabilidad de Y es

$$g(y) = \sum_{i=1}^k f[w_i(y)] \cdot |J_i|$$

donde $J_i = w'_i(y)$, $i = 1, 2, \dots, k$.

Teorema: suponga que X_1 y X_2 son variables aleatorias continuas con distribución de probabilidad conjunta $f(x_1, x_2)$. Definamos la transformación inyectiva $\begin{cases} Y_1 = u_1(X_1, X_2) \\ Y_2 = u_2(X_1, X_2) \end{cases}$, de manera que

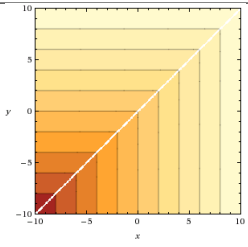
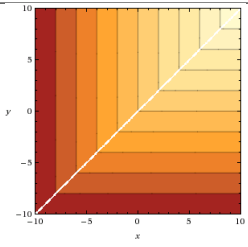
las ecuaciones $\begin{cases} y_1 = u_1(x_1, x_2) \\ y_2 = u_2(x_1, x_2) \end{cases}$ se pueden resolver para x_1, x_2 en términos de y_1, y_2 mediante $\begin{cases} x_1 = w_1(y_1, y_2) \\ x_2 = w_2(y_1, y_2) \end{cases}$. Entonces, la distribución de probabilidad conjunta de Y_1, Y_2 es

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)] \cdot |J|$$

donde $J = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix}$

9.2 Extremos de n variables aleatorias

Sean las variables aleatorias independientes e idénticamente distribuidas X_1, \dots, X_n con función de densidad $f_X(x)$ y función de distribución $F_X(x)$.

<i>Máximo</i>	<i>Mínimo</i>
$X_M = \max\{X_1, \dots, X_n\}$	$X_m = \min\{X_1, \dots, X_n\}$
	
$f_{X_M}(x) = n f_X(x) [F_X(x)]^{n-1}$	$f_{X_m}(x) = n f_X(x) [1 - F_X(x)]^{n-1}$
$F_{X_M}(x) = [F_X(x)]^n$	$F_{X_m}(x) = 1 - [1 - F_X(x)]^n$

9.3 Suma de dos variables aleatorias

Sean X, Y dos variables aleatorias con densidad conjunta $f_{X,Y}(x, y)$. Sea $Z = X + Y$. Para cada $z \in \mathbb{R}$, $B_z = \{(x, y) \in \mathbb{R}^2 : y \leq z - x\}$ vale que

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx$$

9.4 Mínimo entre dos variables aleatorias

Sean X, Y dos variables aleatorias con densidad conjunta $f_{X,Y}(x, y)$. Sea $U = \min\{X, Y\}$. Para cada $u \in \mathbb{R}$ vale que

$$F_U(u) = 1 - \int_u^{\infty} \int_u^{\infty} f_{X,Y}(x, y) dx dy$$

9.5 Cambio de variable lineal

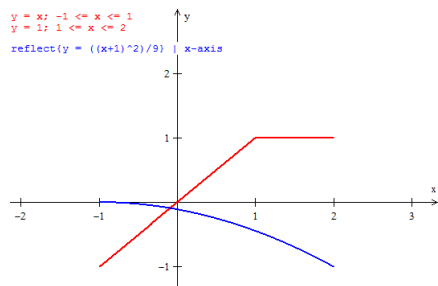
Sea la variable aleatoria X con función de densidad $f_X(x)$. La función de densidad del cambio de variable $Y = aX + b$, donde $a > 0$ y $b \in \mathbb{N}$, es:

$$f_Y(y) = \frac{1}{a} \cdot f_X\left(\frac{y-b}{a}\right)$$

9.6 Ejemplo de cambio de variable

Sea la variable aleatoria X tal que $f_X(x) = \frac{1}{9}(x+1)^2$ para $x \in [-1, 2]$.

Sea la variable aleatoria $Y = \begin{cases} X & \text{si } -1 \leq X < 1 \\ 1 & \text{si } X \geq 1 \end{cases}$



La función de distribución de Y es:

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= \begin{cases} P[X \leq y] & \text{si } -1 \leq y < 1 \\ P[1 \leq y] & \text{si } y \geq 1 \end{cases} \\
 &= \begin{cases} \int_{-1}^y \frac{1}{9} (x+1)^2 dx & \text{si } -1 \leq y < 1 \\ \int_{-1}^1 \frac{1}{9} (x+1)^2 dx & \text{si } y \geq 1 \end{cases} \\
 &= \begin{cases} 0 & \text{si } y < -1 \\ \frac{1}{27} (y+1)^3 & \text{si } -1 \leq y < 1 \\ 1 & \text{si } y \geq 1 \end{cases}
 \end{aligned}$$

La función de densidad de Y es:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{1}{9} (y+1)^2 & \text{si } -1 \leq y \leq 1 \\ 0 & \forall \text{ otro } y \end{cases}$$

9.7 Ejemplo de cambio de variable bidimensional

Sean las variables aleatorias independientes U_1, U_2 tales que $f_{U_i}(u_i) = 1$ si $u_i \in (-\frac{1}{2}, \frac{1}{2})$ para cada $i \in \{1, 2\}$.

Como son independientes, la función de densidad conjunta es $f_{U_1, U_2}(u_1, u_2) = 1$ para $u_1, u_2 \in (-\frac{1}{2}, \frac{1}{2})$.

Sea la variable aleatoria $Z = U_1 + U_2$. Es claro que el rango “posible” de Z es $z \in (-1, 1)$.

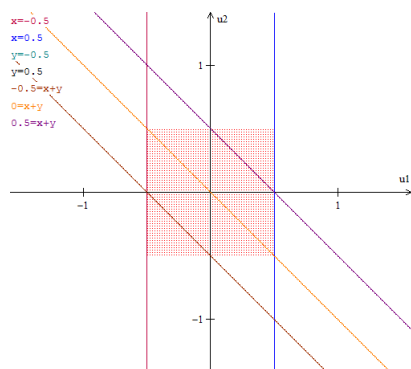


Figura 11: Algunas curvas de nivel: $U_1 + U_2 = k$ para $k \in \mathbb{R}$

La probabilidad $P(Z \leq z)$, cuando $z \leq 0$, se interpreta gráficamente como el área que está a la izquierda y abajo de la recta $U_1 + U_2 = z$ (el triángulo). Cuando $z \geq 0$, la probabilidad $P(Z \leq z)$ puede verse como el área del triángulo de vértices $(-\frac{1}{2}, -\frac{1}{2})$, $(\frac{1}{2}, -\frac{1}{2})$, $(-\frac{1}{2}, \frac{1}{2})$ más el área del

trapecio delimitado por las rectas $U_1 + U_2 = 0$ y $U_1 + U_2 = z$. O también puede pensarse como “1 menos el área del triángulo que está a la derecha y arriba de la recta $U_1 + U_2 = z$ ”.

La función de distribución de Z es:

$$\begin{aligned}
 F_Z(z) &= P(Z \leq z) \\
 &= \begin{cases} 0 & \text{si } z < -1 \\ \int_{-0,5}^{z+0,5} \int_{-0,5}^{z-x} 1 \, dy \, dx & \text{si } -1 \leq z < 0 \\ 1 - \int_{z-0,5}^{0,5} \int_{z-x}^{0,5} 1 \, dy \, dx & \text{si } 0 \leq z < 1 \\ 1 & \text{si } z > 1 \end{cases} \\
 &= \begin{cases} 0 & \text{si } z < -1 \\ \int_{-0,5}^{z+0,5} (z - x + 0,5) \, dx & \text{si } -1 \leq z < 0 \\ 1 - \int_{z-0,5}^{0,5} (0,5 - z + x) \, dx & \text{si } 0 \leq z < 1 \\ 1 & \text{si } z > 1 \end{cases} \\
 &= \begin{cases} 0 & \text{si } z < -1 \\ 0,5z^2 + z + 0,5 & \text{si } -1 \leq z < 0 \\ -0,5z^2 + z + 0,5 & \text{si } 0 \leq z < 1 \\ 1 & \text{si } z > 1 \end{cases}
 \end{aligned}$$

Por lo tanto, la función de densidad de Z es:

$$f_Z(z) = \frac{d(F_Z(z))}{dz} = \begin{cases} z + 1 & \text{si } -1 \leq z < 0 \\ 1 - z & \text{si } 0 \leq z < 1 \\ 0 & \forall \text{ otro } z \end{cases}$$

10 Simulación

Sea U una variable aleatoria tal que $U \sim \mathcal{U}(0, 1)$. Sea X una variable aleatoria tal que su función de distribución es $F_X(x)$. Si

$$X = \begin{cases} F_X^{-1}(U) & \text{si } F_X \text{ es creciente} \\ \sup \{x \in \mathbb{R} : F_X(x) < u\} & \text{en otro caso} \end{cases}$$

entonces X es una variable aleatoria tal que su función de distribución es $F_X(x)$.

Ejemplo: sea X una variable aleatoria exponencial de parámetro $\lambda = 1$.

$$\begin{aligned}
 f_X(x) &= e^{-x} \\
 F_X(x) &= 1 - e^{-x} \\
 x &= F_X^{-1}(u) \\
 x &= -\ln(1 - u)
 \end{aligned}$$

Parte II

Estadística

Objetivo: a partir de las observaciones de un fenómeno aleatorio, hacer inferencias sobre la distribución de probabilidades subyacente.

11 Definiciones

Población: totalidad de las observaciones en las que estamos interesados, de número finito o infinito.

Muestra: subconjunto de una población.

Muestra aleatoria: sean X_1, X_2, \dots, X_n variables aleatorias independientes, cada una con la misma distribución de probabilidad $f_X(x)$. Definimos X_1, X_2, \dots, X_n como una muestra aleatoria de tamaño n de la población $f_X(x)$, y su distribución de probabilidad conjunta es $f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n)$.

Estadístico: variable aleatoria que es función de variables aleatorias que forman una muestra aleatoria.

Estimador: estadístico que se usa para estimar el valor de un parámetro desconocido de una distribución de probabilidad.

12 Estadística inferencial

12.1 Propiedades de los estimadores

Sea X una variable aleatoria que depende de un parámetro desconocido θ . Sea $\hat{\theta}$ un estimador de θ basado en una muestra aleatoria de tamaño n .

1. **Sesgo:** $B[\hat{\theta}] = E[\hat{\theta}] - \theta$

- Si $B[\hat{\theta}] = 0$ para el parámetro θ , el estimador $\hat{\theta}$ es **insesgado**.
- Si $\lim_{n \rightarrow \infty} B[\hat{\theta}] = 0$, el estimador $\hat{\theta}$ es **asintóticamente insesgado**, y por lo tanto,
 $ECM[\hat{\theta}] = var[\hat{\theta}]$

2. **Varianza:** $var[\hat{\theta}] = E[(\hat{\theta} - \theta)^2]$

3. **Error cuadrático medio:**

$$\begin{aligned} ECM[\hat{\theta}] &= E[(\hat{\theta} - \theta)^2] \\ &= var[\hat{\theta}] + B^2[\hat{\theta}] \end{aligned}$$

4. **Consistencia:** a medida que aumenta el tamaño de la muestra, el estimador se aproxima al verdadero valor del parámetro.

a) **Consistencia débil:** para cada $\epsilon > 0$, debe cumplir que

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

Teorema: si $\hat{\theta}$ es asintóticamente insesgado y $var[\hat{\theta}] \xrightarrow{n \rightarrow \infty} 0$, entonces $\hat{\theta}$ es débilmente consistente.

b) **Consistencia fuerte:**

$$P\left(\lim_{n \rightarrow \infty} \hat{\theta} = \theta\right) = 1$$

5. **Eficiencia:** si $\text{var}[\hat{\theta}_1] < \text{var}[\hat{\theta}_2]$ entonces $\hat{\theta}_1$ es mejor estimador que $\hat{\theta}_2$.

12.2 Comparación de estimadores

El estimador $\hat{\theta}_1$ es mejor que el estimador $\hat{\theta}_2$ si $ECM[\hat{\theta}_1] \leq ECM[\hat{\theta}_2]$ para todo θ , con desigualdad estricta para al menos un valor de θ .

12.3 Algunos estadísticos importantes

Sean X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de una población con distribución $X \sim \mathcal{N}(\mu, \sigma^2)$.

■ Media de la muestra:

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

- Es insesgado.
- $ECM[\bar{X}] = \frac{\sigma^2}{n}$
- Distribución muestral de \bar{X} : $\bar{X} \sim \mathcal{N}\left(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\right)$.
(Si $X \not\sim \mathcal{N}(\mu, \sigma^2)$, la distribución anterior vale sólo para n suficientemente grande).

■ Varianza de la muestra:

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n(n-1)} \left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \end{aligned}$$

- Es insesgado.
- *Teorema:* si S^2 es la varianza de una muestra aleatoria de tamaño n que se toma de una población normal que tiene varianza σ^2 , entonces el estadístico $U = \frac{(n-1) \cdot S^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$ tiene una distribución chi cuadrada con $\nu = n - 1$ grados de libertad.

■ Proporción de la muestra:

$$\hat{P} = \frac{X}{n}$$

donde X es la cantidad de éxitos en n experimentos.

- Distribución muestral de \hat{P} : $\hat{P} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

■ Diferencia de dos medias:

$$\Delta = \bar{X}_1 - \bar{X}_2$$

- Distribución muestral de Δ : $\Delta \sim \mathcal{N}\left(\mu_{\Delta} = \mu_1 - \mu_2, \sigma_{\Delta}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

■ Diferencia de dos proporciones:

$$\Delta = \hat{P}_1 - \hat{P}_2$$

- Distribución muestral de Δ : $\Delta \sim \mathcal{N}\left(\mu_{\Delta} = p_1 - p_2, \sigma_{\Delta}^2 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$

13 Estimadores puntuales

Sea X una variable aleatoria cuya distribución pertenece a la familia paramétrica $F = \{F_\theta(x) : \theta \in \Theta\}$. Sean los valores $\bar{x} = (x_1, x_2, \dots, x_n)$ de una muestra aleatoria $\bar{X} = (X_1, X_2, \dots, X_n)$.

Familia regular: la familia $F = \{F_\theta(x) : \theta \in \Theta\}$ es regular si cumple que:

1. El conjunto de parámetros Θ es abierto.
2. El soporte de $f_X(x)$ no depende de θ .
3. La función $f_X(x)$ es derivable respecto a θ .

Ejemplo: algunas familias no regulares:

- Bernoulli discreta (no se cumple 1)
- Uniforme continua (no se cumple 2)

Estimador puntual del parámetro θ : variable aleatoria $\hat{\theta}(\bar{X})$ que depende de la muestra aleatoria pero no del parámetro.

13.0.1 Estimadores de máxima verosimilitud

Estimador de máxima verosimilitud: valor $\hat{\theta}_{mv} \in \Theta$ que maximiza la función de verosimilitud

$$L(\theta|\bar{x}) = \prod_{i=1}^n f(x_i|\theta)$$

*Cálculo del emv para familiares **no** regulares:* debe hacerse “a mano”.

Cálculo del emv para familias regulares: $\hat{\theta}_{mv}$ es solución del siguiente sistema de ecuaciones:

$$\begin{cases} \sum_{i=1}^n \psi_j(\theta|x_i) = 0 & \text{si } j = 1, \dots, d \\ \psi_j(\theta|x) = \frac{\partial \log f(x|\theta)}{\partial \theta_j} & \text{para cada } x, \text{ y para } j = 1, \dots, d \end{cases}$$

Esta condición es necesaria pero no suficiente. Para asegurarse que θ es un máximo, debe verificarse las condiciones de segundo orden, y verificar que sea un máximo absoluto.

Dada una muestra aleatoria $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de tamaño n :

Distribución	Estimador de máxima verosimilitud
$X \sim \mathcal{U}(0, \theta)$	$\hat{\theta}_{mv}(\mathbf{X}) = \max(X_1, X_2, \dots, X_n)$
$X \sim \text{Bernoulli}(p)$	$\hat{p}_{mv}(\mathbf{X}) = \bar{X}$
$X \sim \text{Binom}(n, p)$	$\hat{p}_{mv}(\mathbf{X}) = \frac{\bar{X}}{n}$, X = cantidad éxitos
$X \sim \text{Geom}(p)$	$\hat{p}_{mv}(\mathbf{X}) = \frac{1}{\bar{X}}$
$X \sim \text{Pascal}(r, \theta)$	$\hat{\theta}_{mv}(\mathbf{X}) = \frac{r}{\bar{X}}$
$X \sim \text{Poisson}(\lambda t)$	$\hat{\lambda} t_{mv}(\mathbf{X}) = \bar{X}$
$X \sim \text{Exp}(\lambda)$	$\hat{\lambda}_{mv}(\mathbf{X}) = \frac{1}{\bar{X}}$
$X \sim \text{Gamma}(\lambda, k)$, k conocido	$\hat{\lambda}_{mv}(\mathbf{X}) = \frac{nk}{\sum_{i=1}^n X_i}$
$X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma^2 > 0$ conocida	$\hat{\mu}_{mv}(\mathbf{X}) = \bar{X}$
$X \sim \mathcal{N}(\mu, \sigma^2)$, μ conocida	$\hat{\sigma}_{mv}^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$
$X \sim \mathcal{N}(\mu, \sigma^2)$	$\hat{\mu}_{mv}(\mathbf{X}) = \bar{X}$ $\hat{\sigma}_{mv}^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

Teorema (principio de invarianza): sea X_1, X_2, \dots, X_n una muestra aleatoria de una variable aleatoria X cuya distribución pertenece a la familia paramétrica $F = \{F_\theta : \theta \in \Theta\}$. Sea $g : \Theta \rightarrow \Lambda$

una función biunívoca de Θ sobre Λ . Si $\hat{\theta}_{mv}$ es un estimador de máxima verosimilitud para θ , entonces el estimador de máxima verosimilitud para $\lambda = g(\theta)$ es

$$\hat{\lambda}_{mv} = g(\hat{\theta}_{mv})$$

Teorema: sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de una variable aleatoria X perteneciente a una *familia exponencial uniparamétrica* con función de probabilidad de la forma

$$f_{X|\theta}(x) = c(x) \cdot e^{a(\theta) \cdot T(x) + b(\theta)}$$

El estimador de máxima verosimilitud de θ basado en la muestra es solución de la ecuación

$$-\frac{b'(\theta)}{a'(\theta)} = \frac{1}{n} \sum_{i=1}^n T(X_i)$$

siempre y cuando la solución pertenezca al espacio paramétrico Θ .

14 Estimación por intervalo

Sea una muestra aleatoria $\mathbf{X} = (X_1, X_2, \dots, X_n)$ de la variable aleatoria X cuya función de distribución pertenece a la familia paramétrica de funciones $F = \{F_\theta : \theta \in \Theta\}$, $\Theta \in \mathbb{R}$.

Intervalo de confianza para θ de nivel $1 - \alpha$: intervalo aleatorio $I(\mathbf{X})$ que depende de una muestra aleatoria \mathbf{X} , tal que

$$P(\theta \in I(\mathbf{X})) = 1 - \alpha$$

Cota inferior de confianza para θ de nivel $1 - \alpha$: variable aleatoria $a(\mathbf{X})$ que depende de una muestra aleatoria \mathbf{X} , tal que

$$P(a(\mathbf{X}) \leq \theta) = 1 - \alpha$$

Cota superior de confianza para θ de nivel $1 - \alpha$: variable aleatoria $a(\mathbf{X})$ que depende de una muestra aleatoria \mathbf{X} , tal que

$$P(\theta \leq a(\mathbf{X})) = 1 - \alpha$$

Pivote: dada una variable aleatoria X_θ y dada una muestra $\mathbf{X} = \mathbf{x}$, un pivote es una variable aleatoria de la forma $Q(\mathbf{X}, \theta)$ cuya distribución es conocida, y no depende de θ .

Método de construcción de intervalo de confianza: sea la variable aleatoria X que depende de un parámetro θ . Sea una muestra $\mathbf{X} = \mathbf{x}$. Para construir un intervalo de confianza para θ , $P(\theta \in I(\mathbf{X})) = 1 - \alpha$, los pasos a seguir son:

1. Obtener un estimador puntual del parámetro, $\hat{\theta}$, en general por máxima verosimilitud.
2. Caracterizar la distribución de $\hat{\theta}$.
3. Hallar un pivote de θ .
4. Poner cotas al pivote y despejar el parámetro.

Error: si tenemos una estimación $\hat{\theta}$ del parámetro desconocido θ , $\epsilon = |\hat{\theta} - \theta|$ es el error cometido en la estimación.

Teorema llave: sea $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Sea la muestra aleatoria $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Valen las siguientes afirmaciones:

1. $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ tiene distribución $Z \sim \mathcal{N}(0, 1)$.
2. $U = \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$ tiene distribución $U \sim \chi_{v=n-1}^2$.

3. Z y U son variables aleatorias independientes.

Teorema: sean X_1, X_2, \dots, X_n variables aleatorias independientes tales que $X \sim \mathcal{N}(\mu, \sigma)$. Sean los estadísticos

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\end{aligned}$$

Entonces la variable aleatoria $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ es $T \sim t_{\nu=n-1}$.

Tenemos las muestras X_1, X_2, \dots, X_n . Notar que $[a \mp b] = [a - b, a + b]$.

<i>Dist de muestra</i>	<i>Parámetro</i>	<i>Pivote para el parámetro</i>	<i>Dist del pivote</i>	<i>I. de C. de nivel $1 - \alpha$</i>	<i>Error</i>
$X_i \sim \mathcal{U}(0, \theta)$	θ	$X_{(n)} = \max(X_i)$ $Y = \frac{X_{(n)}}{\theta}$	$f_Y(y) = ny^{n-1}$ $y \in [0, 1]$	$\left[X_{(n)}, \frac{X_{(n)}}{\sqrt[n]{\alpha}} \right]$	
$X_i \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 conocida	μ	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[\bar{X} \mp \frac{\sigma}{\sqrt{n}} \cdot z_{(1-\alpha/2)} \right]$	$\epsilon \leq z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}$
$X_i \sim \mathcal{N}(\mu, \sigma^2)$, μ conocida	σ^2	$U = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$	$U \sim \chi_{v=n}^2$	$\left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, (1-\alpha/2)}^2}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, (\alpha/2)}^2} \right]$	
$X_i \sim \mathcal{N}(\mu, \sigma^2)$ μ, σ^2 desconocidos	μ, σ^2	$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ $Y = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ $U = \frac{(n-1)}{\sigma^2} S^2$	$Y \sim t_{v=n-1}$ $U \sim \chi_{v=n-1}^2$	$I(X)_\mu = \left[\bar{X} \mp t_{v=n-1, (1-\alpha/2)} \frac{s}{\sqrt{n}} \right]$ $I(X)_{\sigma^2} = \left[\frac{(n-1)s^2}{\chi_{v=n-1, (1-\alpha/2)}^2}, \frac{(n-1)s^2}{\chi_{v=n-1, (\alpha/2)}^2} \right]$	
$X_i \sim \text{Bernoulli}(p)$ $n \gg 1$	p	$Z = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[\frac{z_{(1-\alpha/2)}^2 + 2n\bar{X}}{2z_{(1-\alpha/2)}^2 + 2n} \mp \frac{z_{(1-\alpha/2)} \sqrt{z_{(1-\alpha/2)}^2 + 4n\bar{X}(1-\bar{X})}}{2z_{(1-\alpha/2)}^2 + 2n} \right]$	$n \geq \frac{z_{(1-\alpha/2)}^2}{\epsilon^2}$
$X_i \sim \text{Binom}(n, p)$	p	$Z = \frac{\hat{P} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[\hat{P} \mp z_{(\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$	Con confianza $(1 - \alpha)100\%$: $\epsilon \leq z_{(1-\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Con confianza $\geq (1 - \alpha)100\%$: $\epsilon \leq z_{(1-\alpha/2)} \sqrt{\frac{1}{n}}$
$X_i \sim \text{Exp}(\lambda)$	λ	$X = \sum_{i=1}^n X_i$ $U = 2\lambda X$	$U \sim \chi_{v=2n}^2$	$\left[\frac{\chi_{v=2n, (\alpha/2)}^2}{2X}, \frac{\chi_{v=2n, (1-\alpha/2)}^2}{2X} \right]$	
$X_i \sim \text{Poisson}(\lambda)$	λ	$X = \sum_{i=1}^n X_i$ $Z = \frac{X - n\lambda}{\sqrt{n\lambda}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[\frac{X}{n} + \frac{z_{(1-\alpha/2)}^2}{2n} \mp \frac{z_{(1-\alpha/2)}}{\sqrt{n}} \sqrt{\frac{X}{n} + \frac{z_{(1-\alpha/2)}^2}{4n}} \right]$	
$A_i \sim \mathcal{N}(\mu_A, \sigma_A^2)$ $B_i \sim \mathcal{N}(\mu_B, \sigma_B^2)$ σ_A^2, σ_B^2 conocidas	$\Delta = \mu_A - \mu_B$	$Z = \frac{(\bar{A} - \bar{B}) - \Delta}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[(\bar{A} - \bar{B}) \mp z_{(1-\alpha/2)} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \right]$	

<i>Dist de muestra</i>	<i>Parámetro</i>	<i>Pivote para el parámetro</i>	<i>Dist del pivote</i>	<i>I. de C. de nivel $1 - \alpha$</i>	<i>Error</i>
$A_i \sim \mathcal{N}(\mu_A, \sigma_A^2)$ $B_i \sim \mathcal{N}(\mu_B, \sigma_B^2)$ $\sigma_A^2 = \sigma_B^2$ desconocidas	$\Delta = \mu_A - \mu_B$	S_A^2, S_B^2 son los estimadores de varianza de A y B $S_P^2 =$ $\frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A+n_B-2}$ $T = \frac{(\bar{A}-\bar{B})-\Delta}{\sqrt{S_P^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$	$T \sim t_{v=n_A+n_B-2}$	$\left[(\bar{A} - \bar{B}) \mp t_{v=15, (1-\alpha/2)} \sqrt{S_P^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)} \right]$	
$A_i \sim \mathcal{N}(\mu_A, \sigma_A^2)$ $B_i \sim \mathcal{N}(\mu_B, \sigma_B^2)$ $\sigma_A^2 \neq \sigma_B^2$ desconocidas	$\Delta = \mu_A - \mu_B$	$T = \frac{(\bar{A}-\bar{B})-\Delta}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$	$\nu = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^2}{\frac{\left(\frac{S_A^2}{n_A}\right)^2}{\frac{n_A}{n_A+1}} + \frac{\left(\frac{S_B^2}{n_B}\right)^2}{\frac{n_B}{n_B+1}}}$ $T \sim t_\nu$		
$A_i \sim \mathcal{N}(\mu_A, \sigma_A^2)$ $B_i \sim \mathcal{N}(\mu_B, \sigma_B^2)$ μ_A, μ_B desconocidas	$R = \frac{\sigma_A^2}{\sigma_B^2}$	$F = \frac{1}{R} \left(\frac{S_A^2}{S_B^2} \right)$	$F \sim F_{n_A-1, n_B-1}$		
$X_i \sim \text{Bern}(p_1)$ $Y_i \sim \text{Bern}(p_2)$ $n_1 \approx n_2$, grandes	$\Delta = p_1 - p_2$	$Z = \frac{\bar{X}-\bar{Y}-\Delta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n_1} + \frac{\bar{Y}(1-\bar{Y})}{n_2}}}$	$Z \sim \mathcal{N}(0, 1)$	$\left[\bar{X} - \bar{Y} \mp z_{(1-\alpha/2)} \sqrt{\frac{\bar{X}(1-\bar{X})}{n_1} + \frac{\bar{Y}(1-\bar{Y})}{n_2}} \right]$	

15 Test de hipótesis

Objetivo: formar un procedimiento de decisión que se base en datos muestrales, para ofrecer una conclusión acerca de una conjetura. Dicho procedimiento de decisión debe hacerse con la noción de la probabilidad de una conclusión errónea.

15.1 Definiciones

Hipótesis estadística: conjetura con respecto a una o más poblaciones. En toda conjetura hay dos hipótesis:

- H_0 : hipótesis **nula**. Se opone a H_1 y a menudo (*no siempre*) es su complemento lógico. A menudo representa el “status quo” contrario a una nueva idea. Las conclusiones no implican una “aceptación” literal de H_0 .

$$H_0 : \theta \in \Theta_0$$

- H_1 : hipótesis **alternativa**. Por lo general, representa la teoría que debe probarse.

$$H_1 : \theta \in \Theta_1$$

Tipos de hipótesis:

- Hipótesis **simple**: es del tipo $H : \theta = \theta_0$, donde $\theta_0 \in \mathbb{R}$.
- Hipótesis **compuesta**: es del tipo $H : \theta \in \Theta$, donde Θ es un subconjunto de \mathbb{R} .

Tipos de pruebas:

1. **Prueba de una sola cola:** prueba donde la alternativa es *unilateral*, es decir

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \text{ o } \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

2. **Prueba de dos colas:** prueba donde la alternativa es *bilateral*, es decir

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

Test / Regla de Decisión: función $\delta : \mathbb{R}^n \rightarrow \{0, 1\}$ que le asigna a cada posible realización de la muestra \mathbf{x} una y solo una de las hipótesis.

$$\delta(\mathbf{X}) = \begin{cases} \text{rechazamos } H_0 & \text{si } \mathbf{X} \in \mathcal{R} \\ \text{aceptamos } H_0 & \text{si } \mathbf{X} \notin \mathcal{R} \end{cases} = \begin{cases} 1 & \text{si } \mathbf{X} \in \mathcal{R} \\ 0 & \text{si } \mathbf{X} \notin \mathcal{R} \end{cases}$$

Región crítica: región del espacio \mathbb{R}^n sobre la que rechazamos H_0 .

$$\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^n : \delta(\mathbf{x}) = 1\}$$

Región de aceptación: región sobre la que no rechazamos H_0 .

Valor crítico: valor x_c para el cual si $x > x_c$, rechazamos H_0 .

Estadístico de prueba: variable aleatoria X que depende de una muestra aleatoria, sobre la cual se basa la decisión.

15.2 Errores

Situaciones posibles al probar una hipótesis estadística:

Conclusión \ Realidad	H_0 es verdadera	H_0 es falsa
No rechace H_0	Decisión correcta	Error tipo II
Rechace H_0	Error tipo I	Decisión correcta

$$\alpha = P[\text{error tipo I}] = P[\text{rechazar } H_0 | H_0 \text{ es verdadera}]$$

$$\beta = P[\text{error tipo II}] = P[\text{no rechazar } H_0 | H_0 \text{ es falsa}]$$

Propiedades:

- Los errores tipo I y II están relacionados. En general, cuando disminuye la probabilidad de uno, aumenta la probabilidad del otro.
- El tamaño de la región crítica, y por lo tanto la probabilidad de cometer un error tipo I, se puede reducir al ajustar el(los) valor(es) crítico(s).
- Si aumentamos el tamaño muestral n , entonces α y β se reducen de forma simultánea.
- Para un valor n fijo, si aumenta α disminuye β .
- Si H_0 es falsa, β es máximo cuando el valor real del parámetro se aproxima al valor hipotético.

Nivel de significancia del test: sea δ un test para decidir entre las hipótesis $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$. El nivel de significación del test es la máxima probabilidad de rechazar la hipótesis H_0 cuando ésta es verdadera.

$$\alpha = NS = \max \{P[\text{rechazar } H_0 | H_0 \text{ es V}]\} = \max_{\theta \in \Theta_0} \beta(\theta)$$

Su elección es arbitraria, pero generalmente $\alpha = 0,05; 0,01; 0,1$, etc.

Nivel de confianza del test:

$$NC = 1 - \alpha$$

Función de potencia:

$$\beta(\theta) = P[\text{rechazar } H_0 | \theta]$$

- Un test será “bueno” si para $\theta \in \Theta_0$, $\beta(\theta) \approx 0$, y si para $\theta \in \Theta_1$, $\beta(\theta) \approx 1$.

Potencia del test: es la probabilidad de NO cometer error de tipo II. Lo ideal es que tienda a ser 1. La potencia se ve afectada por tres factores:

1. Tamaño de la muestra (n): a mayor n , mayor potencia.
2. Nivel de significancia: a mayor nivel de significancia, mayor potencia, y menor es la región de aceptación.
3. El verdadero valor del parámetro que está bajo prueba. Cuanto mayor es la diferencia entre ese valor y el valor especificado por H_0 , mayor potencia, mayor potencia.

Característica operativa: función $\mathcal{L}(\theta) = 1 - \beta(\theta)$.

15.3 Construcción de reglas de decisión

1. Si $\theta_1(\mathbf{X})$ es una cota inferior de confianza de nivel $1 - \alpha$ para θ , entonces

$$\delta(\mathbf{X}) = \begin{cases} 1 & \text{si } \theta_0 < \theta_1(\mathbf{X}) \\ 0 & \text{si } \theta_0 \geq \theta_1(\mathbf{X}) \end{cases}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \leq \theta_0 \text{ contra } H_1 : \theta > \theta_0$$

2. Si $\theta_2(\mathbf{X})$ es una cota superior de confianza de nivel $1 - \alpha$ para θ , entonces

$$\delta(\mathbf{X}) = \begin{cases} 1 & \text{si } \theta_0 > \theta_2(\mathbf{X}) \\ 0 & \text{si } \theta_0 \leq \theta_2(\mathbf{X}) \end{cases}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta \geq \theta_0 \text{ contra } H_1 : \theta < \theta_0$$

3. Si $[\theta_1(\mathbf{X}), \theta_2(\mathbf{X})]$ es un intervalo de confianza de nivel $1 - \alpha$ para θ , entonces

$$\delta(\mathbf{X}) = \begin{cases} 1 & \text{si } \theta_0 \notin [\theta_1(\mathbf{X}), \theta_2(\mathbf{X})] \\ 0 & \text{si } \theta_0 \in [\theta_1(\mathbf{X}), \theta_2(\mathbf{X})] \end{cases}$$

es un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \text{ contra } H_1 : \theta \neq \theta_0$$

4. Un test de nivel α para decidir entre las hipótesis

$$H_0 : \theta = \theta_0 \text{ contra } H_1 : \theta > \theta_0$$

es

$$\delta(\mathbf{X}) = \begin{cases} 1 & \text{si } Q\left(\hat{\theta}(\mathbf{X}), \theta_0\right) > q_{1-\alpha} \\ 0 & \text{si } Q\left(\hat{\theta}(\mathbf{X}), \theta_0\right) \leq q_{1-\alpha} \end{cases}$$

15.4 Test de la χ^2 para bondad de ajuste

Objetivo: decidir a favor o en contra de una hipótesis H que afirma que los datos observados constituyen una muestra aleatoria $\mathbf{X} = (X_1, \dots, X_n)$ de una distribución F .

Tipos de hipótesis:

- Hipótesis **simple**: F está completamente especificada.
- Hipótesis **compuesta**: F pertenece a una familia paramétrica, que depende de r parámetros desconocidos.

Procedimiento: dada una muestra aleatoria de tamaño n , $\mathbf{X} = (x_1, \dots, x_n)$, donde $X \sim F_X$ y la hipótesis $H_0 : F_X = F_r$; dado un nivel de significancia de α .

1. Si $r \geq 1$, estimar los parámetros desconocidos con EMV, y establecer que $F_X = F_{\hat{r}_{mv}}$.

- Dividir el rango de la variable aleatoria X en una cantidad k de conjuntos disjuntos (*clases* C_i) tales que

$$\begin{aligned} p_i &= \mathbb{P}[x \in C_i | H_0 \text{ es verdadera}] \\ \sum_{i=1}^k p_i &= 1 \\ p_i &> 0 \\ np_i &\geq 5 \text{ (regla de Fisher)} \end{aligned}$$

Para evitar que haya clases sub o sobre dimensionadas, utilizar el criterio $p_i \approx p_j$ para $C_i \neq C_j$.

- Agrupar los datos de la muestra según las clases establecidas, y contar las frecuencias observadas ($n_i, i = 1 \dots k$).
- Calcular el valor del estadístico:

$$D^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

- Test:

$$\delta(\mathbf{X}) = \mathbf{1} \left\{ D^2 > \chi_{\nu=k-r-1, (1-\alpha)}^2 \right\}$$

16 Estadística Bayesiana

Objetivo: sea X una variable aleatoria cuya función de densidad es conocida y depende de un parámetro θ . La esencia del enfoque Bayesiano consiste en realizar estimaciones acerca de θ suponiendo que θ es una *variable aleatoria* con cierta función de densidad de probabilidades a priori.

Regla de Bayes:

- Caso discreto: $P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|\bar{A})P(\bar{A})}$
- Caso continuo: $f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x) \cdot f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y=y}(x) \cdot f_Y(y) dy}$

Función de densidad *a priori*: función de densidad del parámetro desconocido θ *antes* de realizar el experimento. Es subjetiva.

$$\pi_\theta(t), t \in \Theta$$

Función de densidad *a posteriori*: sea la variable aleatoria X con función de densidad $f(\mathbf{x}|\theta)$, donde la función de densidad *a priori* de θ es $\pi_\theta(t)$. La densidad del parámetro desconocido θ *después* de obtener una muestra \mathbf{x} es:

$$\pi_\theta(t|\mathbf{x}) = \frac{f(\mathbf{x}|t) \cdot \pi_\theta(t)}{\int_{\Theta} f(\mathbf{x}|t) \cdot \pi_\theta(t) dt}$$

Si la muestra es $\bar{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ entonces la función de densidad *a posteriori* es:

$$\pi_\theta(t|\bar{\mathbf{x}}) = \frac{\prod_{i=1}^n f(\mathbf{x}_i|t) \cdot \pi_\theta(t)}{\int_{\Theta} \prod_{i=1}^n f(\mathbf{x}_i|t) \cdot \pi_\theta(t) dt}$$

Tener en cuenta que el espacio paramétrico Θ puede cambiar luego de realizar el experimento. Ejemplo: sea $X \sim \mathcal{U}(0, \theta)$. A priori, $\theta \sim \mathcal{U}(2, 3)$. El espacio paramétrico *a priori* es $\Theta \in (2, 3)$. Si la muestra obtenida es $\mathbf{x} = 8$, entonces el espacio paramétrico *a posteriori* es $\Theta = (8, \infty)$.

Distribuciones predictivas: sea $X = (X_1, \dots, X_n)$ una muestra aleatoria de una distribución indexada por θ . Se observa que $X = x$ y se quiere predecir el comportamiento de una nueva observación $Y \sim g(y|\theta)$, donde Y depende de θ . La función de densidad predictiva (o *incondicional*) de la nueva observación Y será:

$$g(y|\mathbf{x}) = \int g(y|t) \cdot \pi(t|\mathbf{x}) dt$$

Teorema: si la distribución *a priori* del parámetro θ es $\sim \mathcal{U}(a, b)$, la densidad *a posteriori* es proporcional a la función de verosimilitud.

$$\pi(t|\mathbf{x}) \propto L(t|\mathbf{x})$$

16.1 Estimadores bayesianos

Se trata de obtener una estimación puntual o por intervalo para θ , después de haber obtenido una muestra aleatoria \mathbf{X} , donde X tiene función de densidad $f(\mathbf{x}|\theta)$.

- Estimación bayesiana por esperanza condicional: $\hat{\theta}(\mathbf{X}) = E[\theta|\mathbf{X}] = \int t \cdot \pi(t|\mathbf{X}) dt$
- Estimación bayesiana por máximo a posteriori: $\hat{\theta}_{map}(\mathbf{X}) = \max_{t \in \Theta} \pi(t|\mathbf{X})$
- Estimación por intervalo para parámetro continuo: dada la muestra aleatoria \mathbf{X} , para encontrar un intervalo $[a, b]$ tal que $P(\theta \in [a, b] | \mathbf{X}) = 1 - \alpha$ y que la longitud de ese intervalo sea la mínima posible, hay que resolver las siguientes ecuaciones:

- $\int_a^{\hat{\theta}(x)} \pi(t|\mathbf{x}) dt = \frac{1-\alpha}{2}$
- $\int_{\hat{\theta}(x)}^b \pi(t|\mathbf{x}) dt = \frac{1-\alpha}{2}$

16.2 Estadísticos bayesianos conocidos

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria de tamaño n correspondiente a una variable aleatoria X .

1. Si $X \sim \text{Bernoulli}(\theta)$ y la distribución *a priori* de θ es $\sim \mathcal{U}(0, 1)$...
 - a) Distribución a posteriori de θ : $\sim \beta(k+1, n-k+1)$ donde k es la cantidad de éxitos que hubo en la muestra
 - b) Función de probabilidad predictiva: $\sim \text{Bernoulli}\left(\frac{k+1}{n+2}\right)$
 - c) Estimadores bayesianos:

$$\hat{\theta}(x) = \frac{1}{n+2} (1 + \sum_{i=1}^n x_i)$$

$$\hat{\theta}_{map}(x) = \frac{k}{n} = \bar{x}$$
 - d) Estimación por intervalo: $[0, 1 - \sqrt[n+1]{\alpha}]$
2. Si $X \sim \mathcal{N}(\theta, \sigma^2)$ (σ^2 conocido) y la distribución *a priori* de θ es $\sim \mathcal{N}(\mu, \rho^2)$...
 - a) Distribución a posteriori de θ : $\sim \mathcal{N}\left(\frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2}, \frac{\sigma^2\rho^2}{n\rho^2 + \sigma^2}\right)$
 - b) Función de probabilidad predictiva: $\sim \mathcal{N}(\mu_*, \sigma^2 + \rho_*^2)$ con

$$\mu_* = \frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2} \text{ y }$$

$$\rho_*^2 = \frac{\sigma^2\rho^2}{n\rho^2 + \sigma^2}$$
 - c) Estimadores bayesianos: $\hat{\theta} = \frac{n\rho^2\bar{x} + \sigma^2\mu}{n\rho^2 + \sigma^2}$

- d) Estimación por intervalo:
3. Si $X \sim Poisson(\theta)$ y la distribución *a priori* de θ es $\sim \Gamma(\nu, \lambda) \dots$
- a) Distribución a posteriori de θ : $\sim \Gamma(\sum_{i=1}^n x_i + \nu, n + \lambda)$
- b) Función de probabilidad predictiva: $f(x|\mathbf{x}) = \frac{\Gamma(\nu(\mathbf{x})+x)}{\Gamma(\nu(\mathbf{x}))x!} \left(\frac{1}{n+\lambda+1}\right)^x \left(\frac{n+\lambda}{n+\lambda+1}\right)^{\nu(\mathbf{x})}$
- c) Estimadores bayesianos:
- $$\hat{\theta} = \frac{\sum_{i=1}^n x_i + \nu}{n + \lambda}$$
- $$\hat{\theta}_{map} = \frac{\sum_{i=1}^n x_i + \nu - 1}{n + \lambda}$$
- d) Estimación por intervalo: $\left[\frac{\chi_{2\nu(\mathbf{x}), \alpha/2}^2}{2(n+\lambda)}, \frac{\chi_{2\nu(\mathbf{x}), 1-\alpha/2}^2}{2(n+\lambda)} \right]$
4. Si $X \sim Binom(n, P)$ y la distribución *a priori* de P es $\sim \mathcal{U}(0, 1) \dots$
- a) Distribución a posteriori de P : $\sim \beta(n, x)$ donde x es la cantidad de éxitos observados en la muestra y n es el tamaño de la muestra.
5. Si $X \sim Poisson(\lambda t = \theta)$ y la densidad *a priori* de θ es $\sim \Gamma(\lambda, \nu) \dots$
- a) Distribución a posteriori de θ : $\sim GammaGeneralizada(\lambda_\Gamma = n + \lambda, k_\Gamma = (\sum_{i=1}^n x_i) + \nu)$
- b) Estimadores bayesianos:
- $$\hat{\theta} = \frac{n + \nu - 1}{\lambda + \sum_{i=1}^n x_i}$$