Inferencia Estadistica

Alberto Centelles

1 Conceptos basicos de la Estadistica Matematica

Distribucion teorica o distribucion de poblacion: Distribucion desconocida F de la variable aleatoria involucrada en un problema de Inferencia Estadistica.

Espacio parametrico Θ: Subconjunto Θ de $\mathbb{R}^k$

Muestra aleatoria: Observaciones a partir de las cuales se intenta disminuir el desconocimiento de la distribucion teorica F de la variable aleatoria X en estudio.

Muestra aleatoria simple: Muestra obtenida a partir de repeticiones independientes.

Una muestra aleatoria simple, de tamano n, de una variable aleatoria X con distribucion teorica F, son n variables aleatorias $(X_1, X_2, ..., X_n)$, independientes e igualmente distribuidas, con distribucion comun F.

Consecuentemente la funcion de distribucion conjunta de una muestra aleatoria simple $(X_1, X_2, ..., X_n)$ correspondiente a una distribucion de la poblacion F, es

$$F(x_1, x_2, ..., x_n) = F(x_1)F(x_2)...F(x_n)$$

Distribucion muestral $F_n^*(x)$: Frecuencia de elementos de la muestra que son menores o iguales que x. Puesto que $F(x) = PX \leq x$, para cada $x \in \mathbb{R}$, la distribucion asociada a la muestra $(x_1, ..., x_n)$ de X se define como

$$F_n^*(x) = \frac{\text{numero de elementos muestrales} \leq x}{n}$$

Es siempre una funcion discreta cuya funcion de probabilidad es $p_n^*(x) = \frac{j}{n}$.

Espacio muestral $(\chi, \mathfrak{B})$: Conjunto de muestras posibles que pueden obtenerse al seleccionar una muestra aleatoria, de un tamano determinado, de una cierta poblacion.

Se trata siempre de un subconjunto de un espacio euclideo $\mathbb{R}^{mn}$, de manera que podemos considerar en χ la σ-algebra restringida de la σ-algebra de Borel $\mathbb{B}^{mn}$ que representaremos por $\mathfrak{B}$.

Estadistico T: Cualquier funcion $T : (\chi, \mathfrak{B}) \to (\mathbb{R}^k, \mathbb{B}^k)$ del espacio muestral $(\chi, \mathfrak{B})$ en un espacio euclideo $(\mathbb{R}^k, \mathbb{B}^k)$ que sea medible. La dimension k del espacio euclideo imagen se denomina dimension del estadistico.

Distribucion en el muestreo de un estadistico T: Distribucion de la variable aleatoria $T(X_1, X_2, ..., X_n)$. Es decir, es la medida de probabilidad que induce la distribucion de la muestra, P, mediante la funcion $T : (\chi, \mathfrak{B}) \to (\mathbb{R}^k, \mathbb{B}^k)$.

2 Propiedades de la distribucion muestral

Expresiones equivalentes de la **distribucion muestral**:

$$F_n^*(x) = \frac{\text{numero de elementos muestrales} \leq x}{n}$$

$$F_n^*(x) = \begin{cases} 0 \text{ si } x < x_{(1)} \\ \dots \\ \dfrac{j}{n} \text{ si } x_{(j)} < x < x_{(j+1)} \\ \dots \\ 1 \text{ si } x \geq x_{(n)} \end{cases}$$

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i)$$

$nF_n^*(x) = \sum_{i=1}^{n} I_{(-\infty,x])}(X_i)$ tiene distribucion binomial $B(n, F(x))$. Es decir,

$$P\{F_n^*(x) = \frac{k}{n}\} = \binom{n}{k} F(x)^k [1 - F(x)]^{n-k}$$

$$E[F_n^*(x)] = F(x) \qquad\qquad V(F_n^*(x)) = \frac{F(x)(1-F(x))}{n}$$

Puesto que $nF_n^*(x)$ es suma de variables aleatorias independientes e igualmente distribuidas, el teorema central del limite permite afirmar que cuando n es grande, la distribucion de $F_n^*(x)$ es aproximadamente $N(F(x), \sqrt{\frac{F(x)(1-F(x))}{n}})$

Momentos muestrales: Como F_n^* es una distribucion discreta, los momentos muestrales existen y valen:

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad\qquad b_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^k$$

Momento de orden k respecto al origen de la distribucion teorica: $E[a_k] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^k] = \alpha_k$

$V(a_k) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i^k) = \frac{1}{n}\sum_{i=1}^{n} V(X_1^k) = \frac{1}{n}[E[X_1^{2k}] - E[X_i^k]^2] = \frac{\alpha_{2k} - \alpha_k^2}{n} = \frac{\sigma^2}{n}$

Media muestral $\overline{x}$: $a_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$

Varianza muestral s^2: $b_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$

$E[s^2] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[X^2] = \alpha_2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2$

Cuantiles muestrales c_p: Aquellos valores que verifican simultaneamente $F_n^*(c_p) \geq p$ y $F_n^*(c_p^-) \leq p$ para cada $p \in (0,1)$.

Teorema de Glivenko-Cantelli: Sea $X_{ii=1}^{\infty}$ una sucesion de variables aleatorias independientes y con distribucion comun F. Si F_n^* es la funcion de distribucion muestral asociada a la muestra aleatoria simple $(X_1, X_2, ..., X_n)$ y $\Delta_n = \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)|$, entonces $\lim_{n\to\infty} \Delta_n = 0$, P-casi seguro.

Comportamiento asintotico de los cuantiles muestrales

Corolario del Teorema de Glivenko-Cantelli: Sea $\{X_i\}_i^{\infty}$ una sucesion de variables aleatorisa independientes y con distribucion comun F, que tenga un unico cuantil de orden $p \in (0,1)$. Si $c_p(n)$ es el cuantil muestral de orden p asociado a la muestra aleatoria simple $(X_1, ..., X_n)$

Lema. Si $\{T_n\}_{n=1}^{\infty}$ es una sucesion de variables aleatorias tales que

$$\sqrt{n}(T_n - \Theta) \xrightarrow{d} N(0, \sigma)$$

Y si g es una funcion derivable de $\mathbb{R}$ en $\mathbb{R}$ tal que $g'(\Theta) \neq 0$, entonces

$$\sqrt{n}(g(T_n) - g(\Theta)) \xrightarrow{d} N(0, |g'(\Theta)|\sigma)$$

Teorema. Si $c_p(n)$ es el cuantil muestral de orden p de una muestra aleatoria simple $(X_1, ..., X_n)$ de una distribucion teorica F, que tiene una funcion de densidad continua f y cuyo cuantil de orden p, x_p, es unico, se verifica que

$$\sqrt{n}(c_p(n) - x_p) \xrightarrow{d} N(0, \frac{\sqrt{p(1-p)}}{f(x_p)})$$

(ver ejercicio 2.9)

Comportamiento asintotico de los momentos muestrales

Cualquier momento muestral tiene, cuando n es grande, una distribucion muy concentrada alrededor del correspondiente momento poblacional.

Teorema. Dada una muestra aleatoria simple $(X_1, X_2, ..., X_n)$ de una poblacion con momentos finitos de orden $2k$, se verifica

$$\sqrt{n}(a_k(n) - \alpha_k) \xrightarrow{d} N(0, \sqrt{\alpha_{2k} - \alpha_k^2})$$

(Si $k = 1$, entonces $\overline{X}$ es aproximadamente $N(\mu, \frac{\sigma}{\sqrt{n}})$ cuando n es grande. Ver ejercicio 2.10)

Ademas,

$$[\sqrt{n}(a_1(n) - \alpha_1), ..., \sqrt{n}(a_k(n) - \alpha_k] \xrightarrow{d} N_k(0, \Sigma)$$

Siendo $\Sigma_{jj} = \alpha_{2j} - \alpha_j^2$, $\Sigma_j l = \alpha_{j+l} - \alpha_j \alpha_l$ los terminos de la matriz de covarianzas de la distribucion normal k-dimensional limite.

Teorema. (Momentos centrales) Si $b_k(n)$ es el momento central de orden k de una muestra aleatoria simple $(X_1, ..., X_n)$ de una poblacion con momentos finitos de orden 2_k, se verifica

$$\sqrt{n}(b_k(n) - \mu_k) \xrightarrow{d} N(0, \sqrt{\mu_{2k} - \mu_k^2 - 2k\mu_{k+1}\mu_{k-1} + k^2\mu_{k-1}^2\mu_2})$$

En particular, para la varianza muestral s^2, como $\mu_1 = E[X - \mu] = 0$, se obtiene

$$\sqrt{n}(s^2 - \sigma^2) \xrightarrow{d} N(0, \sqrt{\mu_4 - \sigma^4})$$

3 Distribuciones en el muestreo de poblaciones normales

Media y varianza muestrales

Como combinacion lineal de variables aleatorias independientes con distribucion normal, $\overline{X}$ tiene distribucion normal. Sus parametros son:

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X] = \mu$$

$$V(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = V(X) = \frac{\sigma^2}{n}$$

Luego la distribucion en el muestreo de X es $N(\mu, \frac{\sigma}{\sqrt{n}})$

Distribucion χ^2 de Pearson

Si X tiene distribucion normal $N(0,1)$, entonces X^2 tiene distribucion gamma $\gamma(\frac{1}{2}, \frac{1}{2})$. Y si $X_1, ..., X_n$ son variables aleatorias independientes con distribucion $N(0,1)$, entonces $\sum_{i=1}^{n} X_i^2$ tiene distribucion gamma $\gamma(\frac{n}{2}, \frac{1}{2})$. Esta distribucion se denomina distribucion chi^2 de Pearson con n grados de libertad.

Funcion de densidad de χ^2:

$$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{\frac{-y}{2}}$$

Grados de libertad: Las funciones de densidad y de distribucion dependen de un unico parametro n cuya denominacion de "grados de libertad" hace referencia al numero de sumandos que aportan su variabilidad a la suma. En funcion de n se expresan la media y varianza:

$$E[\sum_{i=1}^{n} X_i^2] = \sum_{i=1}^{n} E[X_i^2] = n$$

$$V(\sum_{i=1}^{n} X_i^2) = \sum_{i=1}^{n} V(X_i^2) = 2n$$

Funcion de distribucion de χ^2:

$$F(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \int_0^x y^{\frac{n}{2}-1} e^{\frac{-y}{2}}$$

No admite una expresion explicita. Sus valores estan tabulados en la tabla 4.

Si $n > 30$, la distribucion χ^2 se puede aproximar mediante una distribucion normal.

$$\frac{\sum_{i=1}^{n} X_i^2 - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$$

de forma que para n grande, χ^2 se aproxima a $N(n, \sqrt{2n})$.

Existe una aproximacion mejor: Sea Y una variable con distribucion χ^2,

$$\sqrt{2Y} - \sqrt{2n-1} \xrightarrow{d} N(0,1)$$

Teorema de Fisher: Si $(X_1, ..., X_n)$ es una muestra aleatoria simple de una poblacion $N(\mu, \sigma)$, entonces s^2 y $\overline{X}$ son variables aleatorias independientes y la distribucion en el muestreo de $\frac{ns^2}{\sigma^2}$ es χ^2_{n-1} (mientras que $\overline{X}$ tiene distribucion $N(\mu, \frac{\sigma}{\sqrt{n}})$)

Distribucion t de Student

Saber que $\overline{X}$ tiene distribucion en el muestreo $N(\mu, \sigma/\sqrt{n})$ o equivalentemente $\sqrt{n}\dfrac{\overline{X}-\mu}{\sigma}$ tiene distribucion $N(0,1)$ resulta de poca utilidad si la varianza poblacional σ^2 es desconocida. La idea de Student viene al considerar que s^2 y σ^2 tendran valores similares para muestras grandes.

Cuasivarianza muestral: $S^2 = \dfrac{ns^2}{n-1}$

Estadistico t de Student:

$$t = \sqrt{n-1}\frac{\overline{X}-\mu}{s} = \sqrt{n}\frac{\overline{X}-\mu}{S}$$

Distribucion t de Student con n grados de libertad: Si $(X_1, ..., X_n)$ son variables aleatorias independientes y con distribucion $N(0, \sigma)$, la distribucion de

$$\frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}$$

de densidad

$$\frac{1}{\sqrt{n\pi}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{-(n+1)/2}$$

para $t \in \mathbb{R}$ se denomina distribucion t de Student con n grados de libertad.

Teorema de Student: Si $(X_1, ..., X_n)$ es una muestra aleatoria simple de una poblacion $N(\mu, \sigma)$, el estadistico de Student:

$$t = \sqrt{n-1}\frac{\overline{X}-\mu}{s} = \sqrt{n}\frac{\overline{X}-\mu}{S}$$

tiene distribucion t de Student con $n-1$ grados de libertad.

Distribucion de la diferencia de medias muestrales

Sea (X, Y) una variable aleatoria bidimensional. Puesto que trata de comparar ambas poblaciones X e Y, y en primer lugar sus medias desconocidas μ_1 y μ_2, parece natural que el analisis este basado en sus dos medias muestrales $\overline{X}$ e $\overline{Y}$ y mas concretamente en su diferencia $\overline{X}-\overline{Y}$.

1. $\overline{X}-\overline{Y}$ **siendo** X, Y **independientes.** μ_1, μ_2 **desconocidas.** σ_1^2, σ_2^2 **conocidas**:

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}$$

   tiene distribucion $N(0,1)$

2. $\overline{X}-\overline{Y}$ **siendo** X, Y **independientes.** μ_1, μ_2 **desconocidas.** σ_1^2, σ_2^2 **desconocidas (pero iguales)**:

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

   tiene distribucion t_{m+n-2}.

   Si los tamanos muestrales no son muy pequenos $(n, m \geq 15)$:

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{S_1^2/n+S_2^2/m}}$$

   sera aproximadamente $N(0,1)$

(Welch) Si los tamanos muestrales son muy pequenos ($n, m \leq 15$):

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}}$$

tiene aproximadamente distribucion t_f, siendo f el numero de grados de libertad, esto es, el entero mas proximo a

$$\frac{(S_1^2/n + S_2^2/m)^2}{\frac{1}{n+1}(S_1^2/n)^2 + \frac{1}{m+1}(S_2^2/m)^2} - 2$$

3. $\overline{X} - \overline{Y}$ **siendo** X, Y **no necesariamente independientes.** μ_1, μ_2 **desconocidas.** $\sigma_1^2, \sigma_2^2, \sigma_{11}$ **conocidas**:

$$\begin{pmatrix} \sigma_1^2 & \sigma_{11} \\ \sigma_{11} & \sigma_2^2 \end{pmatrix}$$

entonces

$$\sqrt{n}\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}}}$$

tiene distribucion $N(0, 1)$

4. $\overline{X} - \overline{Y}$ **siendo** X, Y **no necesariamente independientes.** μ_1, μ_2 **desconocidas.** $\sigma_1^2, \sigma_2^2, \sigma_{11}$ **desconocidas**: Por el teorema de Student

$$\sqrt{n-1}\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{s^*}$$

tiene distribucion t_{n-1}, donde s^* es ahora la varianza muestral de la muestra $(X_i - Y_i)$, es decir $s^{*2} = \frac{1}{n}\sum_{i=1}^n (X_i - Y_i)^2 - (\overline{X} - \overline{Y})^2$.

Distribucion del cociente de cuasivarianzas muestrales

El metodo para obtener informacion acerca de la relacion entre σ_1^2 y σ_2^2 esta basado en las cuasivarianzas muestrales S_1^2 y S_2^2. Pero en vez de considerar $S_1^2 - S_2^2$, conviene utilizar el estadistico S_1^2/S_2^2.

Distribucion F de Snedecor: Si $X_1, ..., X_n$, $Y_1, ..., Y_n$ son variables aleatorias independientes, con distribucion $N(0, \sigma)$ la distribucion de

$$\frac{1/n\sum_{i=1}^n X_i^2}{\sum_{i=1}^m Y_i^2}$$

se denomina distribucion F de Snedecor con n y m grados de libertad y tiene densidad

$$\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}(\frac{n}{m})^{n/2}t^{(n-2)/2}(1 + \frac{n}{m}t)^{-(n+m)/2}$$

para $t \geq 0$

Dicho de otra manera, $F_{n,m}$ es la distribucion del cociente de dos χ^2 independientes, de n y m grados de libertad respectivamente, divididas cada una de ellas por sus grados de libertad.

Momento de orden r:

$$\frac{\Gamma(\frac{n}{2} + r)\Gamma(\frac{m}{2} - r)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}(\frac{m}{n})^r$$

Teorema. Si S_1^2 y S_2^2 son las cuasivarianzas de sendas muestras aleatorias simples, de dos poblaciones normales de varianzas σ_1^2 y σ_2^2 respectivamente, el estadistico

$$U = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

tiene distribucion F de Snedecor con $n - 1$ y $m - 1$ grados de libertad (ver ejercicio 3.4).

Coeficiente de correlacion muestral

Dada una muestra aleatoria simple $((X_1, Y_1), ..., (X_n, Y_n)$, se trata de comprobar si el coeficiente de correlacion teorico ρ es cero y se verifica la independencia entre X y Y.

Cuasicovarianza muestral: $S_{11} = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_1 - \overline{Y})$

Coeficiente de correlacion muestral: $R = \dfrac{S_{11}}{S_1 S_2}$

Para abordar este tipo de problemas, partimos de la hipotesis de que ρ es cero.

Teorema. Si R es el coeficiente de correlacion muestral de una distribucion normal bidimensional, con coeficiente de correlacion $\rho = 0$,

$$R^* = \sqrt{n-2} \frac{R}{\sqrt{1 - R^2}}$$

tiene distribucion t de Student con $n - 2$ grados de libertad.

4 Intervalos de confianza

De una poblacion descrita por una variable aleatoria X, cuya distribucion teorica F pertenece a una familia parametrica $\mathfrak{F} = \{F_\theta | \theta \in \Theta \subset \mathbb{R}^2\}$ se considera una muestra aleatoria $(X_1, ..., X_n)$ con distribucion P_θ. Sea $g(\theta)$ una funcion del parametro, con valores reales, y $T_1 \leq T_2$ dos estadisticos unidimensionales tales que

$$P_\theta\{T_1(X_1, ..., X_n) \leq g(\theta) \leq T_2(X_1, ..., X_n)\} \geq 1 - \alpha$$

para cada $\theta \in \Theta$. Entonces, para cualquier muestra, el intervalo

$$[T_1(x_1, ..., x_n), T_2(x_1, ..., x_n)]$$

se denomina **intervalo de confianza** para $g(\theta)$, de **nivel de confianza** $1 - \alpha$.

Metodo de la cantidad pivotal

Supongamos que $T(X_1, ..., X_n; \theta)$ es una funcion real de la muestra y del parametro, cuya distribucion en el muestreo **no** depende de θ. En tal caso, fijado cualquier nivel de confianza $1 - \alpha$ entre 0 y 1, se pueden determinar constantes, c_1 y c_2 (que no seran unicas) tales que

$$P_\theta\{c_1 \leq T(X_1, ..., X_n; \theta) \leq c_2\} \geq 1 - \alpha$$

Si es posible despejar $g(\theta)$ en las desigualdades

$$c_1 \leq T(X_1, ..., X_n; \theta) \qquad\qquad T(X_1, ..., X_n; \theta) \leq c_2$$

obtendremos sendos valores $T_1(X_1, ..., X_n)$ y $T_2(X_1, ..., X_n)$ tales que para cualquier valor $\theta \in \Theta$

$$P_\theta\{T_1(X_1, ..., X_n; \theta) \leq g(\theta \leq T(X_1, ..., X_n; \theta))\} \geq 1 - \alpha$$

De manera que $[T_1(X_1, ..., X_n; \theta), T_2(X_1, ..., X_n; \theta)]$ sera un intervalo de confianza para $g(\theta)$, de nivel de confianza $1 - \alpha$

Determinacion de un estadistico que permita llevar a cabo el metodo de la cantidad pivotal: Si $(X_1, ..., X_n)$ es una muestra aleatoria simple de una poblacion unidimensional cuya distribucion teorica pertenece a una familia $\mathfrak{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}\}$ entonces el estadisitico

$$T(X_1, ..., X_n; \theta) = -\sum_{i=1}^{n} log F_\theta(X_i)$$

tiene distribucion en el muestreo independiente de θ. Para cualquier valor θ, $-\log F(X_i)$ tiene distribucion exponencial de parametro 1. Luego $-\sum_{i=1}^{n} \log F(X_i)$ tiene distribucion gamma $\gamma(n, 1)$.

Metodo de Neyman

En este caso, T es cualquier estadistico unidimensional. Sea $L \in \mathbb{R}$ su recorrido y $(X_1, ..., X_n)$ una muestra aleatoria con distribucion P_θ.

Fijado un nivel de confianza $1 - \alpha$, para cada $\theta \in \Theta$ se pueden determinar dos valores $c_1(\theta) < c_2(\theta)$ en L tales que

$$P_\theta\{T < c_1(\theta)\} \leq \alpha_1 \qquad\qquad P_\theta\{T < c_2(\theta)\} \leq \alpha_2$$

siendo $\alpha_1, \alpha_2 > 0$ y $\alpha_1 + \alpha_2 = \alpha$.

Sera entonces, para cada θ,

$$P_\theta\{c_1(\theta) \leq T(X_1, ..., X_n) \leq c_2(\theta)\} \geq 1 - \alpha$$

Para cada valor t del estadistico T:

$$P_\theta\{\theta_1(T(X_1, ..., X_n)) \leq \theta \leq \theta_2(T(X_1, ..., X_n))\} \geq 1 - \alpha$$

y por tanto $[\theta_1(T(X_1, ..., X_n)), \theta_1(T(X_1, ..., X_n))]$ es un intervalo de confianza para θ de nivel de confianza $1 - \alpha$.

Intervalos de confianza para los parametros de distribuciones normales

1. **Intervalo de confianza para la media si la varianza poblacional es conocida**: Puesto que $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ tiene distribucion en el muestreo $N(0,1)$, sera

$$P_\mu\{-z_{\alpha 2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\} = 1-\alpha$$

de forma que $(\overline{x}-z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}, \overline{x}+z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}})$ es un intervalo de confianza para μ de nivel de confianza $1-\alpha$.

2. **Intervalo de confianza para la media si la varianza poblacional es desconocida**: Puesto que $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ tiene distribucion en el muestreo t_{n-1}, sera

$$P_\mu\{-t_{n-1;\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{n-1;\alpha/2}\} = 1-\alpha$$

de forma que $(\overline{x}-t_{n-1;\alpha/2}\dfrac{S}{\sqrt{n}}, \overline{x}+t_{n-1;\alpha/2}\dfrac{S}{\sqrt{n}})$ es un intervalo de confianza para μ de nivel de confianza $1-\alpha$.

3. **Intervalo de confianza para σ^2 si la varianza poblacional es conocida**: Puesto que $\sum_{i=1}^n \dfrac{(X_i-\mu)^2}{\sigma^2}$ tiene distribucion en el muestreo χ^2, sera

$$P_\sigma\{\chi^2_{n;1-\alpha/2} < \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2} < \chi^2_{n;1-\alpha/2}\} = 1-\alpha$$

de forma que $(\dfrac{1}{\chi^2_{n;\alpha/2}}\sum_{i=1}^n (x_i-\mu)^2, \dfrac{1}{\chi^2_{n;1-\alpha/2}}\sum_{i=1}^n (x_i-\mu)^2)$ es un intervalo de confianza para σ^2 de nivel de confianza $1-\alpha$.

4. **Intervalo de confianza para σ^2, si la media poblacional es desconocida**: Mediante la afirmacion del teorema de Fisher, puesto que $\sum_{i=1}^n \dfrac{(X_i-\overline{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2}$ tiene distribucion χ^2_{n-1}, sera

$$P_{\mu,\sigma}\{\chi^2_{n-1;1-\alpha/2} \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\alpha/2}\} = 1-\alpha$$

de forma que $(\dfrac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}, \dfrac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}})$ es un intervalo de confianza para σ^2 de nivel de confianza $1-\alpha$.

5. **Region de confianza para la media y varianza poblacional**: El teorema de Fisher, al proporcionar la distribucion conjunta de los estadisticos

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \qquad\qquad \frac{(n-1)S^2}{\sigma^2}$$

permite construir una region de confianza simultaneamente para μ y σ^2. Como ambos estadisticos son independientes y

$$P_{\mu,\sigma}\{-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\} = 1-\alpha$$

$$P_{\mu,\sigma}\{\chi^2_{n-1;1-\beta/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\beta/2}\} = 1-\beta$$

la region del espacio muestral

$$\{-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}, \chi^2_{n-1;1-\beta/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\beta/2}\}$$

tiene probabilidad $(1-\alpha)(1-\beta)$.

6. **Intervalo de confianza para la diferencia de medias con σ_1 y σ_2 conocidas**: Como

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}$$

tiene distribucion en el muestreo $N(0,1)$, resulta directamente que

$$(\overline{x}-\overline{y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, \overline{x}-\overline{y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}})$$

es un intervalo de confianza para $\mu_1-\mu_2$ de nivel de confianza $1-\alpha$.

7. **Intervalo de confianza para la diferencia de medias con varianzas desconocidas pero iguales**: El estadistico

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{ns_1^2+ms_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

tiene distribucion en el muestreo t_{n+m-2}. De manera que el intervalo de extremos

$$\overline{x}-\overline{y} \pm t_{n+m-2;\alpha/2}\sqrt{\frac{ns_1^2+ms_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}$$

es un intervalo de confianza para $\mu_1-\mu_2$ de nivel de confianza $1-\alpha$.

Analogo resultado se obtiene con el estadistico

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{S_1^2/n+S_2^2/m}}$$

utilizando la aproximacion de Welch en el caso en que σ_1^2 y σ_2^2 no pudiesen suponerse iguales y alguna de las muestras fuese de pequeno tamano. El intervalo de confianza de nivel $1-\alpha$ tendria entonces por extremos

$$\overline{x}-\overline{y} \pm t_{f;\alpha/2}\sqrt{S_1^2/n+S_2^2/m}$$

siendo f el entero mas proximo a

$$\frac{(S_1^2/n+S_2^2/m)^2}{\frac{1}{n+1}(S_1^2/n)^2+\frac{1}{m+1}(S_2^2/m)^2}-2$$

8. **Intervalo de confianza para el cociente de varianzas poblacionales**: Puesto que $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ tiene distribucion $F_{n-1,m-1}$ sera

$$P\{F_{n-1,m-1;1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n-1,m-1;\alpha/2} = 1-\alpha\}$$

y por tanto $\dfrac{S_1^2/S_2^2}{F_{n-1,m-1;\alpha/2}, \dfrac{S_1^2/S_2^2}{F_{n-1,m-1,1-\alpha/2}}}$ es un intervalo de confianza para σ_1^2/σ_2^2 de nivel de confianza $1-\alpha$.

9. **Intervalo de confianza para la diferencia de medias de dos poblaciones normales no independientes**: Puesto que

$$\sqrt{n-1}\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{s^*}$$

tiene distribucion en el muestreo t_{n-1}, y

$$(\overline{x}-\overline{y}-t_{n-1;\alpha/2}\frac{s^*}{\sqrt{n-1}},\overline{x}-\overline{y}+t_{n-1;\alpha/2}\frac{s^*}{\sqrt{n-1}})$$

es un intervalo de confianza para $\mu_1-\mu_2$ de nivel de confianza $1-\alpha$ donde s^* representa la varianza muestral de las diferencias X_1-Y_i.

Intervalos de confianza basados en distribuciones asintoticas

Si disponemos de una sucesion T_n de estadisticos, correspondientes a los sucesivos tamanos muestrales n, tales que $\frac{T_n-\theta}{\sigma_n(\theta)} \xrightarrow{d} N(0,1)$, donde θ representa el parametro que caracteriza la distribucion teorica y $\sigma_n(\theta)$ depende en general de n y del parametro poblacional, podemos obtener intervalos de confianza aproximados para el parametro θ. Si n es suficientemente grande sera, para cada θ

$$P_\theta\{-z_{\alpha/2}<\frac{T_n-\theta}{\sigma_n(\theta)}<z_{\alpha/2}\}\approx 1-\alpha$$

de manera que despejando θ, obtenemos un intervalo de confianza de nivel de confianza aproximado $1-\alpha$

5 Estimacion puntual

Estimador: Estadistico $T(X_1, ..., X_n)$ independiente del parametro θ cuyo valor se utiliza para obtener su estimacion puntual.

Funcion de perdida: $L : \Theta \times \Theta \to \mathbb{R}$ cuyos valores indican el coste en que se incurre si se da como estimacion el valor t, cuando realmente el parametro que identifica la distribucion teorica vale θ.

Ejemplos de funciones de perdida:

- $L_1(\theta, t) = |\theta - t|$ mide directamente el error en la estimacion
- $L_2(\theta, t) = (\theta - t)^2$, error cuadratico de estimacion
- $L_3(\frac{|\theta - t|}{\theta})$, error relativo en la estimacion
- $L_4(\theta, t) = \begin{cases} c \text{ si } |\theta - t| > \epsilon \\ 0 \text{ si } |\theta - t| \leq \epsilon \end{cases}$, que penaliza con un coste c los errores mayores que ϵ.

Perdida media $R_T(\theta)$: Mide el riesgo del estimador T. Es una funcion positiva de θ

$$R_T(\theta) = E_\theta[L(\theta, T(X_1, ..., X_n))] = \int_{\mathbb{R}} L(\theta, t) H_\theta(dt) = \int_{\mathcal{X}} L(\theta, T(X_1, ..., X_n)) F_\theta(dx_1) F_\theta(dx_2) ... F_\theta(dx_n)$$

donde $H_\theta(t)$ representa la distribucion en el muestreo del estadistico T correspondiente al valor θ del parametro.

El concepto de riesgo proporciona un criterio para la comparacion de estimadores; de hecho, un estimador T_1 sera preferible a otro estimador T_2 si

$$R_{T_1}(\theta) \leq R_{T_2}(\theta) \text{ para cualquier } \theta \in \Theta \text{ y}$$

$$R_{T_1}(\theta) < R_{T_2}(\theta) \text{ para algun } \theta \in \Theta$$

Error cuadratico medio del estimador T: Funcion de riesgo definida por: $ECM_t(\theta) = E_\theta[|\theta - T|^2]$.
Estimadores admisibles: Aquellos para los cuales no existe otro estimador preferible a el
Error cuadratico medio de T **como estimador de** $g(\theta)$: $E_\theta[\big|g(\theta) - T(X_1, ..., X_n)^2\big|]$

Propiedades deseables de los estimadores

- **Estimadores insesgados**:

  Sesgo del estimador T **como estimador de** $g(\theta)$: Diferencia $b_T(\theta) = E_\theta[T] - g(\theta)$

  Estimador insesgado o centrado en $g(\theta)$: Sesgo nulo, esto es, $b_T(\theta) = 0$, luego $E_\theta[T] = g(\theta)$ para cada $\theta \in \Theta$.

- **Estimadores consistentes**: La consistencia de los estimadores hace referencia a su comportamiento cuando el tamano muestral n crece hacia infinito.

  Una sucesion T_n de estimadores se denomina **consistente** para estimar una funcion $g(\theta)$ del parametro poblacional si cuando $n \to \infty$ se verifica $T_n \xrightarrow{P_\theta} g(\theta)$. O, lo que es lo mismo,

  $$P_\theta\{|T_n - g(\theta)| - \epsilon \to 0\}$$

  para todo $\epsilon > 0, \theta \in \Theta$.

  La consistencia en media cuadratica significa que, para estimar $g(\theta)$, $ECM_{T_n}(\theta) \to 0$ para cualquier $\theta \in \Theta$; o bien $V_\theta(T_n) \to 0$ y $b_{T_n} \to 0$.

- **Estimadores invariantes**:
  - **Estimador invariante por traslaciones**:

    $$T(x_1 + c, ..., x_n + c) = T(x_1, ..., x_n) + c$$

    para toda muestra $(x_1, ..., x_n)$ y cualquier $c \in \mathbb{R}$.

    Requisitos:
    * $\Theta = \mathbb{R}$
    * La familia de distribuciones continuas $F = \{F_\theta | \theta \in \Theta\}$ ha de ser invariante por traslaciones, i.e $F_\theta(x) = F_{\theta+c}(x + c) = F_0(x - \theta)$, lo cual indica que la variable aleatoria que describe la poblacion se puede escribir: $X = \theta + Z$ siendo Z una variable aleatoria con distribucion fija F_0.

    Ejemplos:
    * $X_{(1)} = min_{1 \leq i \leq n} X_i$
    * $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
    * $\frac{1}{2}(X_{(1)} + X_{(2)})$

    No son invariantes por traslaciones:
    * Media geometrica $(\prod_{i=1}^{n} X_i)^{1/n}$
    * Media armonica $\frac{n}{\sum_{i=1}^{n} X_i^{-1}}$
  - **Estimador invariante por cambios de escala**:

    $$T(bx_1, ..., bx_n) = bT(x_1, ..., x_n)$$

    para toda muestra $(x_1, ..., x_n)$ y cualquier $b > 0$.

    Requisitos:
    * $\Theta = (0, \infty)$
    * La familia de distribuciones continuas $F = \{F_\theta | \theta \in \Theta\}$ ha de ser invariante por cambios de escala, i.e $F_\theta(x) = F_{b\theta}(bx) = F_1(\frac{x}{\theta})$, lo cual indica que la variable aleatoria que describe la poblacion se puede escribir: $X = \theta Z$ siendo Z una variable aleatoria con distribucion fija F_1.

    Ejemplos:
    * $\overline{X}$
    * $X_{(n)} = max_{1 \leq i \leq n} X_i$

    No son invariantes por traslaciones:
    * $log(\frac{1}{n} \sum_{i=1}^{n} e^{X_i})$

  Bajo condiciones de invarianza, el critero del minimo riesgo es capaz de seleccionar un estimador invariante preferible a todos los demas, supeusto que la funcion de perdida se elige adecuadamente para no destruir la invarianza del problema.

- **Estidisticos suficientes**: Dada una muestra aleatoria $(X_1, X_2, ..X_n)$ de una poblacion con distribucion teorica perteneciente a una familia parametrica $\{F_\theta | \theta \in \Theta\}$, un estadistico $T(X_1, ..., X_n)$ se denomina suficiente si la distribucion de la muestra, condicionada por el valor del estadistico T, no depende de θ.

  No es necesario que las distribuciones F_θ sean discretas, ni que la muestra sea aleatoria simple.

  Con poblaciones discretas, siempre se puede descomponer la funcion de probabilidad de la muestra en la forma:

  $$P_\theta\{X_1 = x_1, ..., X_n = x_n\} = P_\theta\{T = t\} P_\theta\{X_1 = x_1, ..., X_n = x_n | T = t\}$$

Teorema de factorizacion: Un estadistico $T(X_1, ..., X_n)$ es suficiente si y solo si

$$f_\theta(x_1, ..., x_n) = g_\theta(T(x_1, ..., x_n))h(x_1, ..., x_n)$$

siendo $g\theta$ una funcion que solo dependen de $x_1, ..., x_n$ a traves del valor $T(x_1, ..., x_n)$ del estadistico y $h(x_1, ..., x_n)$ una funcion que no depende de θ

Estadisticos suficientes minimales: Es conveniente disponer de un estadistico suficiente de la menor dimension posible, para eliminar el maximo de informacion superflua.

Un estadistico T se denomina minimal suficiente si, para cualquier otro estadistico suficiente T', existe una funcion (medible) ρ tal que $T = \rho(T')$.

La determinacion de estadisticos suficientes por aplicacion del teorema de factorizacion suele conducir a estadisticos minimales suficientes, aunque no siempre.

T es un estadistico minimal suficiente si y solo si

$$\begin{cases} T(x_1, ..., x_n) = T(x_1^{'}, ..., x_n^{'}) \text{ si } \dfrac{f_\theta(x_1, ..., x_n)}{f_\theta(x_1^{'}, ..., x_n^{'})} \text{ no depende de } \theta \\ T(x_1, ..., x_n) \neq T(x_1^{'}, ..., x_n^{'}) \text{ si } \dfrac{f_\theta(x_1, ..., x_n)}{f_\theta(x_1^{'}, ..., x_n^{'})} \text{ depende de } \theta \end{cases}$$

Criterios de seleccion de estimadores

- **Estimadores minimax**: Puesto que el riesgo $R_T(\theta)$ expresa la perdida esperada, con el estimador T, cuando el valor del parametro es *theta*, la manera de asegurar una perdida esperada pequena, cualquiera que sea θ, consiste en controlar el $\max_{\theta\in\Theta} R_T(\theta)$ y preferir un estimador T_1 a otro T_2 si

$$max_{\theta\in\Theta} R_{T_1}(\theta) < max_{\theta\in\Theta} R_{T_2}(\theta)$$

  Sera optimo con este criterio cualquier estimador comparable que alcanzase el valor $min_T \max_{\theta\in\Theta} R_T(\theta)$. Consecuentemente, los estimadores $T*$ tales que

$$max_{\theta\in\Theta} R_{T*}(\theta) = \min_T \max_{\theta\in\Theta} R_T(\theta)$$

  se denominan estimadores minimax y por extension el criterio de comparar los estimadores por el maximo de su riesgo se denomina **criterio minimax**.

  El criterio minimax no siempre conduce a estimadores centrados y tampoco puede asegurarse que de estimadores basados en un estadistico suficiente minimal. Por el criterio minimal un estimador con poco riesgo para casi todos los valores de θ pero un alto riesgo R para algunos de ellos es desechado frente a un estimador de riesgo constante igual a $R - \epsilon$.

- **Estimadores Bayes**: La utilizacion del criterio Bayes requiere comportarse como si el parametro θ fuese una variable aleatoria de la que se conoce su distribucion a priori, $\pi(\theta)$. El **riesgo de Bayes** de un estimador T frente a la distribucion a priori π es el promedio

$$r_T(\pi) = \int_\Theta R_T(\theta)\pi(d\theta)$$

  y un estimador T_1 es preferido a T_2 si $r_{T_1}(\pi) < r_{T_2}(\pi)$. Lo ideal es entonces encontrar un estimador $T*$ que proporcione el minimo riesgo Bayes, es decir, tal que $r_{T^*}(\pi) = inf_T r_T(\pi)$. En caso de existir, se denomina **estimador Bayes frente a la distribucion a priori** π.

  Los elementos $F_\theta(x)$ de la familia parametrica de distribuciones teoricas pueden interpretarse como la distribucion de la variable poblacional X condicionada por el valor θ del parametro aleatorio. Por consiguiente, puede considerarse la distribucion conjunta de θ y X; formada a partir de la distribucion marginal π y de la distribucion condicionada $F_\theta(x)$.

$$f(\theta, x) = \pi(\theta)f_\theta(x)$$

La densidad marginal de $(X_1, ..., X_n)$ sera entonces $f(x_1, ..., x_n) = \int_\Theta f_\theta(x_1, ..., x_n)\pi(\theta)d\theta$, y la distribucion de θ condicionada por $(X_1, ..., X_n)$ tendra por densidad la llamada **distribucion a posteriori**:

$$\pi(\theta \mid x_1, ..., x_n) = \frac{\pi(\theta) f_\theta(x_1, ..., x_n)}{\int_\Theta f_\theta(x_1, ..., x_n)\pi(\theta)d\theta}$$

La distribucion a posteriori expresa la probabilidad de que se haya producido cada posible valor del parametro una vez se han realizado las observaciones de la poblacion.

La busqueda del estimador Bayes frente a la distribucion a priori π se reduce a determinar para cada muestra $(x_1, ..., x_n)$ el valor que minimice

$$\int_\Theta L(\theta, t)\pi(\theta \mid x_1, ..., x_n)d\theta$$

El metodo Bayes no produce, en general, estimadores insesgados. En cambio, siempre da lugar a estimadores admisibles.

La distribucion a posteriori es funcion del estadistico minimal suficiente.

Si T es el estimador Bayes, frente a una cierta distribucion a priori π_0 y su riesgo $R_T(\theta)$ es independiente de θ, entonces T es tambien el estimador minimax. Por tanto, el metodo Bayes sirve tambien como procedimiento para determinar el estimador minimax.