Pontificia Universidad Católica de Chile
Inferencia Estadística: EYP2114
Profesor: Manuel Galea
Rafael Lorca // rafael.lorca@uc.cl
Ingeniería civil
Primer semestre 2025

Formulario Inferencia estadística

Preámbulo

Para muchos inferencia es el curso que te hace cambiar el minor (data science), para otros, es el ramo pasta del major (ingemat) pero creo que ahora existe un consenso de que tomar el ramo con Galea sirve para reforzar el caracter y tener un curso muy demandante. Cuando el profesor dijo: "No les voy a hacer un formulario, háganselo ustedes", a lo que un alumno le preguntó: ¿Qué podemos/debemos traer en el formulario? Cuando dijo todo, diapositivas, clases, el libro entero si estimábamos conveniente, supe que el curso estaría imposible. Junto con mi compañero realizamos este formulario para los valientes que decidan tomar el curso y se ahorren la lata de hacerse un formulario. Créditos a mi compañero Hudu que me ayudó con el formulario, al Rocka, Victor Marques y Martin Cooper que se mandaban unos latex muy buenos y soluciones de las ies pasadas y actuales y al ayudante del ramo David Ortega, sin él este ramo hubiese sido mucho más dificil de lo que fue.

Conceptos previos

Igualdades

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}; \quad \sum_{k=x}^{\infty} \phi^k = \frac{\phi^x}{1-\phi} \quad \text{si } |\phi| < 1;$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(\lambda); \quad \sum_{x=0}^{\infty} \binom{x+k-1}{k-1} \phi^x = \frac{1}{(1-\phi)^k} \quad \text{si } 0 < \phi < 1 \text{ y } k \in \mathbb{N}; \quad \int_{-\infty}^{\infty} e^{-z^2/2}\, dx = \sqrt{2\pi}$$

Propiedades de las funciones $\Gamma(\cdot)$ y $B(\cdot,\cdot)$

(1.) $\Gamma(k) = \int_0^\infty u^{k-1}\, e^{-u}\, du = \text{Gamma}(k)$ (2.) $\Gamma(a+1) = a\,\Gamma(a)$
(3.) $\Gamma(n+1) = n!$ si $n \in \mathbb{N}_0$ (4.) $\Gamma(1/2) = \sqrt{\pi}$
(5.) $B(q,r) = \int_0^1 x^{q-1}\,(1-x)^{r-1}\, dx$ (6.) $B(q,r) = \frac{\Gamma(q)\,\Gamma(r)}{\Gamma(q+r)} = \text{beta}(q,r)$

Distribución Gamma

$$(1)\text{Si } T \sim \text{Gamma}(k,\nu), \text{ con } k \in \mathbb{N} \Rightarrow F_T(t) = 1 - \sum_{x=0}^{k-1} \frac{(\nu t)^x\, e^{-\nu t}}{x!}$$

$$(2)\ \text{Gamma}(1,\nu) = \text{Exp}(\nu) \quad (3)\ \text{Gamma}(\eta/2, 1/2) = \chi^2(\eta)$$

Medidas descriptivas

$$\mu_X = \mathsf{E}(X),\ \sigma_X^2 = \mathsf{E}\left[(X-\mu_X)^2\right],\ \delta_X = \frac{\sigma_X}{\mu_X},\ \theta_X = \frac{\mathsf{E}\left[(X-\mu_X)^3\right]}{\sigma_X^3},\ K_X = \frac{\mathsf{E}\left[(X-\mu_X)^4\right]}{\sigma_X^4} - 3$$

$$M_X(t) = \mathsf{E}\left(e^{tX}\right), \quad \mathsf{E}[g(X)] = \begin{cases} \sum_{x\in\Theta_X} g(x)\cdot p_X(x) & \text{para variables discretas} \\ \int_{-\infty}^{\infty} g(x)\cdot f_X(x)\, dx & \text{para variables continuas} \end{cases}$$

$$\text{Rango} = \text{máx} - \text{mín}, \quad \text{RIQ} = x_{75\,\%} - x_{25\,\%}$$

$$\text{Cov}(X,Y) = \mathsf{E}[(X-\mu_X)\cdot(Y-\mu_Y)] = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y), \;\; \rho = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Teorema de Probabilidades Totales

$$p_Y(y) = \sum_{x\in\Theta_X} p_{X,Y}(x,y), \quad f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy$$

$$p_X(x) = \int_{-\infty}^{+\infty} p_{X|Y=y}(x)\cdot f_Y(y)\,dy, \quad f_Y(y) = \sum_{x\in\Theta_X} f_{Y|X=x}(y)\cdot p_X(x)$$

$$\mathsf{E}(X) = \int_{-\infty}^{+\infty} \mathsf{E}(X|Y=y)\cdot f_Y(y)\,dy\;, \quad \mathsf{E}(Y) = \sum_{x\in\Theta_X} \mathsf{E}(Y|X=x)\cdot p_X(x)$$

Teoremas de Esperanzas Iteradas

$$\mathsf{E}(Y) = \mathsf{E}[\mathsf{E}(Y|X)]\;,\text{ y } \text{Var}(Y) = \text{Var}[\mathsf{E}(Y|X)] + \mathsf{E}[\text{Var}(Y|X)]$$

Funciones de probabilidad

Suma de Normales Independientes

Consideremos X y Y variables aleatorias independientes con distribución Normal(μ_X, σ_X^2) y Normal(μ_Y, σ_Y^2) respectivamente. Si $Z = a + b\cdot X + c\cdot Y$, con a, b y c constantes, entonces:

$$Z \sim \text{Normal}(\mu, \sigma^2), \qquad \begin{cases} \mu = a + b\cdot\mu_X + c\cdot\mu_Y, \\ \sigma = \sqrt{b^2\sigma_X^2 + c^2\sigma_Y^2}. \end{cases}$$

Distribución Normal Bivariada

La función de densidad conjunta es:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}.$$

Distribuciones condicionales y marginales

- La distribución condicional de Y dado $X = x$:

$$Y|X = x \sim \text{Normal}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\; \sigma_Y^2(1-\rho^2)\right).$$

- Las distribuciones marginales son:

$$X \sim \text{Normal}(\mu_X, \sigma_X^2) \quad \text{y} \quad Y \sim \text{Normal}(\mu_Y, \sigma_Y^2).$$

Distribución	Función de Densidad/Masa	Soporte	Parámetros	Esperanza/Varianza
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, \ldots, n$	$n \in \mathbb{N},\ p \in [0,1]$	$\mu = np$ $\sigma^2 = np(1-p)$
Geométrica	$p(1-p)^{x-1}$	$x = 1, 2, \ldots$	$p \in (0,1)$	$\mu = \frac{1}{p}$ o $\dfrac{1-\theta}{\theta}$ $\sigma^2 = \frac{1-p}{p^2}$
Binomial Negativa	$\binom{r+x-1}{r-1} p^r (1-p)^x$	$x = r, r+1, \ldots$	$r \in \mathbb{N},\ p \in (0,1)$	$\mu = \frac{r(1-p)}{p}$ $\sigma^2 = \frac{r(1-p)}{p^2}$
Poisson	$\dfrac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, \ldots$	$\lambda > 0$	$\mu = \lambda$ $\sigma^2 = \lambda$
Exponencial	$\nu e^{-\nu x}$	$x \geq 0$	$\nu > 0$	$\mu = \frac{1}{\nu}$ $\sigma^2 = \frac{1}{\nu^2}$
Gamma	$\dfrac{\nu^k}{\Gamma(k)} x^{k-1} e^{-\nu x}$	$x \geq 0$	$k, \nu > 0$	$\mu = \frac{k}{\nu}$ $\sigma^2 = \frac{k}{\nu^2}$
Normal	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x \in \mathbb{R}$	$\mu \in \mathbb{R},\ \sigma > 0$	$\mu = \mu$ $\sigma^2 = \sigma^2$
Log-Normal	$\dfrac{1}{\zeta x\sqrt{2\pi}} e^{-\frac{(\ln x-\lambda)^2}{2\zeta^2}}$	$x > 0$	$\lambda \in \mathbb{R}, \zeta > 0$	$\mu = e^{\lambda+\zeta^2/2}$ $\sigma^2 = (e^{\zeta^2}-1)e^{2\lambda+\zeta^2}$
Uniforme	$\dfrac{1}{b-a}$	$x \in [a,b]$	$a < b$	$\mu = \frac{a+b}{2}$ $\sigma^2 = \frac{(b-a)^2}{12}$
Beta	$\dfrac{x^{q-1}(1-x)^{r-1}}{B(q,r)}$	$x \in [0,1]$	$q, r > 0$	$\mu = \frac{q}{q+r}$ $\sigma^2 = \frac{qr}{(q+r)^2(q+r+1)}$
Hipergeométrica	$\dfrac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\text{máx}(0, n+m-N) \leq x \leq \text{mín}(n,m)$	$N, m, n \in \mathbb{N}$	$\mu = n\frac{m}{N}$ $\sigma^2 = n\frac{m}{N}\left(1-\frac{m}{N}\right)\frac{N-n}{N-1}$

Cuadro 1: Resumen de distribuciones de probabilidad.

Otras distribuciones

Weibull Distribution

$T \sim \text{Weibull}(\eta, \beta)$ con:

$$F_T(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \quad f_T(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \quad t > 0$$

donde $\beta > 0$ es un parámetro de forma y $\eta > 0$ es uno de escala.
Si t_p es el percentil $p \times 100\,\%$, entonces:

$$\ln(t_p) = \ln(\eta) + \frac{1}{\beta} \cdot \Phi^{-1}_{\text{Weibull}}(p), \quad \text{donde} \quad \Phi^{-1}_{\text{Weibull}}(p) = \ln[-\ln(1-p)]$$

El m-esimo momento está dado por:

$$E(T^m) = \eta^m \Gamma(1 + m/\beta)$$

Media y varianza:

$$\mu_T = \eta\Gamma\left(1 + \frac{1}{\beta}\right), \quad \sigma_T^2 = \eta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]$$

Logistic Distribution

$Y \sim \text{Logistic}(\mu, \sigma)$ with:

$$F_Y(y) = \Phi_{\text{Logistic}}\left(\frac{y-\mu}{\sigma}\right), \quad f_Y(y) = \frac{1}{\sigma}\phi_{\text{Logistic}}\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty$$

where

$$\Phi_{\text{Logistic}}(z) = \frac{\exp(z)}{1+\exp(z)} \quad \text{and} \quad \phi_{\text{Logistic}}(z) = \frac{\exp(z)}{(1+\exp(z))^2}$$

are the CDF and PDF of a standard logistic distribution. Here $\mu \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is a scale parameter.
The $p \times 100\,\%$ percentile y_p is:

$$y_p = \mu + \sigma\Phi^{-1}_{\text{Logistic}}(p) \quad \text{where} \quad \Phi^{-1}_{\text{Logistic}}(p) = \ln\left(\frac{p}{1-p}\right)$$

Mean and variance:

$$\mu_Y = \mu, \quad \sigma_Y^2 = \frac{\sigma^2\pi^2}{3}$$

Log-Logistic Distribution

$T \sim \text{Log-Logistic}(\mu, \sigma)$ with:

$$F_T(t) = \Phi_{\text{Logistic}}\left(\frac{\ln(t)-\mu}{\sigma}\right), \quad f_T(t) = \frac{1}{\sigma t}\phi_{\text{Logistic}}\left(\frac{\ln(t)-\mu}{\sigma}\right), \quad t > 0$$

where $\exp(\mu)$ is a scale parameter and $\sigma > 0$ is a shape parameter.
The $p \times 100\,\%$ percentile t_p satisfies:

$$\ln(t_p) = \mu + \sigma\Phi^{-1}_{\text{Logistic}}(p)$$

For integer $m > 0$:

$$E(T^m) = \exp(m\mu)\Gamma(1+m\sigma)\Gamma(1-m\sigma)$$

The m-th moment is finite only if $m\sigma < 1$. For $\sigma < 1$:

$$\mu_T = \exp(\mu)\Gamma(1+\sigma)\Gamma(1-\sigma)$$

For $\sigma < 1/2$:

$$\sigma_T^2 = \exp(2\mu)\left[\Gamma(1+2\sigma)\Gamma(1-2\sigma) - \Gamma^2(1+\sigma)\Gamma^2(1-\sigma)\right]$$

Student's t-Distribution

A random variable T has Student's t-distribution if its PDF is:

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < t < \infty$$

Mean and variance:

$$\mu_T = 0 \text{ for } \nu > 1, \quad \sigma_T^2 = \frac{\nu}{\nu-2} \text{ for } \nu > 2$$

F-Distribution

$T \sim \text{Fisher}(\eta, \nu)$ con función de densidad:

$$f_T(t) = \frac{\Gamma\left(\frac{\eta+\nu}{2}\right)}{\Gamma(\eta/2)\Gamma(\nu/2)}\left(\frac{\eta}{\nu}\right)^{\eta/2}\frac{t^{\eta/2-1}}{\left(\frac{\eta}{\nu}t+1\right)^{(\eta+\nu)/2}}, \quad t > 0$$

Mean and variance:

$$\mu_T = \frac{\nu}{\nu-2} \text{ for } \nu > 2, \quad \sigma_T^2 = \frac{2\nu^2(\eta+\nu-2)}{\eta(\nu-2)^2(\nu-4)} \text{ for } \nu > 4$$

Mínimo y máximo

Sean $X_1, \ldots, X_n$ variables aleatorias continuas independientes con idéntica distribución (f_x y F_X), entonces para:

$$Y_1 = \text{mín}\{X_1, \ldots, X_n\} \rightarrow f_{Y_1} = n[1 - F_X(y)]^{n-1} f_X(y); \;\; Y_n = \text{máx}\{X_1, \ldots, X_n\} \rightarrow f_{Y_n} = n[F_X(y)]^{n-1} f_X(y)$$

Su distribución conjunta está dada por

$$f_{Y_1,Y_n}(u,v) = n(n-1)[F_X(v) - F_X(u)]^{n-2} f_X(v) f_X(u). \quad u \leq v$$

Aproximación de Momentos (Método Delta)

- **Aproximación de 4to orden:**

$$Y = g(X) \approx g(\mu_X) + \frac{(X-\mu_X)g'(\mu_X)}{1!} + \frac{(X-\mu_X)^2 g''(\mu_X)}{2!} + \frac{(X-\mu_X)^3 g'''(\mu_X)}{3!} + \frac{(X-\mu_X)^4 g''''(\mu_X)}{4!}$$

- **Aproximación de primer orden (multivariante):**

$$Y \approx g(\mu_{X_1}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n} (X_i - \mu_{X_i}) \frac{\partial}{\partial X_i} g(\mu_{X_1}, \ldots, \mu_{X_n})$$

$$E(Y) \approx g(\mu_{X_1}, \ldots, \mu_{X_n})$$

$$\text{Var}(Y) \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \sigma_{X_i} \sigma_{X_j} \left[\frac{\partial g}{\partial X_i} \cdot \frac{\partial g}{\partial X_j} \right] \quad \text{con } \rho_{ij} = \text{Corr}(X_i, X_j)$$

Estimador de Momento

Para $X_1, \ldots, X_n \sim$ iid p_X o f_X:

$$\mu_k = E(X^k), \quad m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k \quad \Rightarrow \quad \mu_k = m_k \quad (k = 1, 2, \ldots)$$

Estimador Máximo Verosímil

Para $\hat{\theta}$ EMV de θ:

- $E(\hat{\theta}) \rightarrow \theta$ (consistencia asintótica)
- $\text{Var}(\hat{\theta}) = \frac{1}{I_n(\theta)}$ con $I_n(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} \ln L(\theta)\right]$
- $\sqrt{I_n(\theta)}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1)$
- $\text{Var}(g(\hat{\theta})) = \frac{[g'(\theta)]^2}{I_n(\theta)}$

Error Cuadrático Medio

$$ECM(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = \text{Var}(\hat{\theta}) + \text{Sesgo}^2(\hat{\theta})$$

Distribuciones Muestrales

Para $X_1, \ldots, X_n \sim N(\mu, \sigma)$:

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0,1), \quad \frac{\overline{X}_n - \mu}{s / \sqrt{n}} \sim t(n-1), \quad s^2 \frac{(n-1)}{\sigma^2} \sim \chi^2(n-1)$$

donde $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$.

Potencia

Para $H_0 : \mu = \mu_0$ (σ conocido):

$$\begin{aligned}
\text{Bilateral:} &\quad 1 - \Phi\left(k_{1-\alpha/2} - \Delta\frac{\sqrt{n}}{\sigma}\right) + \Phi\left(k_{\alpha/2} - \Delta\frac{\sqrt{n}}{\sigma}\right) \\
\text{Unilateral superior:} &\quad 1 - \Phi\left(k_{1-\alpha} - \Delta\frac{\sqrt{n}}{\sigma}\right) \\
\text{Unilateral inferior:} &\quad \Phi\left(k_{\alpha} - \Delta\frac{\sqrt{n}}{\sigma}\right)
\end{aligned}$$

Comparación de Poblaciones

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias independientes con dist $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$ respectivamente.

$$\bar{X}_n = \frac{1}{n}\sum(X_i) \quad \bar{Y}_m = \frac{1}{m}\sum(Y_j)$$

$$S_X^2 = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2 \quad S_Y^2 = \frac{1}{m-1}\sum(Y_j - \bar{Y}_m)^2$$

Entonces:

- Si σ_x y σ_y son conocidos:

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

- Si σ_x y σ_y son desconocidos pero iguales:

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

  Con $S_p = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

- Si σ_x y σ_y son desconocidos:

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim t(\nu)$$

  Con $\nu = [\frac{(S_X^2/n + S_Y^2/m)^2}{\frac{(S_X^2/n)^2}{n-1} + \frac{(S_Y^2/m)^2}{m-1}}]$

- Si μ_x y μ_y son desconocidos:

$$\frac{[(n-1)S_X^2/\sigma_X^2]/(n-1)}{[(m-1)S_Y^2/\sigma_Y^2/](m-1)} = \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} \sim F(n-1, m-1)$$

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias independientes con dist $Bernoulli(p_x)$ y $Bernoulli(p_y)$ respectivamente, entonces

$$\frac{(\bar{X}_n - \bar{Y}_m) - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \sim N(0,1)$$

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias independientes con dist $Poisson(\lambda_x)$ y $Poisson(\lambda_y)$ respectivamente, entonces

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\lambda_X - \lambda_Y)}{\sqrt{\frac{\lambda_X}{n} + \frac{\lambda_Y}{m}}} \sim N(0,1)$$

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias independientes con dist $Exponencial(\nu_x)$ y $Exponencial(\nu_y)$ respectivamente, entonces

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\frac{1}{\nu_X} - \frac{1}{\nu_Y})}{\sqrt{\frac{1}{n\nu_X^2} + \frac{1}{n\nu_Y^2}}} \sim N(0,1)$$

Cuadro 2: Propiedades de suma de variables aleatorias continuas

Distribuciones	**Operación**	**Resultado**
Normales	$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ independientes	$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
Normales Correlacionadas	$X, Y \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$	$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y)$
Gamma	$X \sim \Gamma(k_1, \theta)$, $Y \sim \Gamma(k_2, \theta)$ independientes	$X + Y \sim \Gamma(k_1 + k_2, \theta)$
Gamma Ponderada	$X \sim \Gamma(k, \theta)$, $a > 0$	$aX \sim \Gamma(k, a\theta)$
Exponenciales	$X_i \sim \text{Exp}(\lambda)$ i.i.d. (n variables)	$\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$
Chi-cuadrado	$Z \sim N(0, 1)$	$Z^2 \sim \chi^2(1)$
Chi-cuadrado	$Z \sim N(0, 1)$	$Z^2 \sim \Gamma^2(\frac{1}{2}, \frac{1}{2})$
Chi-cuadrado Suma	$X_i \sim \chi^2(k_i)$ independientes	$\sum_{i=1}^n X_i \sim \chi^2\left(\sum_{i=1}^n k_i\right)$
Beta	$X \sim \Gamma(a, \theta)$, $Y \sim \Gamma(b, \theta)$ independientes	$\frac{X}{X+Y} \sim \text{Beta}(a, b)$

Explicaciones clave

- **Gamma**: Requiere mismo parámetro de escala θ para ser sumables.
- **Normales**: La correlación afecta sólo a la varianza de la suma.
- **Exponencial → Gamma**: Caso especial donde $\text{Exp}(\lambda) = \Gamma(1, \lambda)$.
- **Chi-cuadrado**: Es un caso particular de Gamma ($\chi^2(k) = \Gamma(k/2, 1/2)$).

Teorema del límite central: Sean $X_1, \ldots, X_n$ variables aleatorias *iid* tal que $\text{E}(X_i) = \mu < \infty$ y $\text{Var}(X_i) = \sigma^2 < \infty$. Sea $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Entonces

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$$

Teorema de cambio de variable:

$$f_Q(q) = f_T(g^{-1}(q))|\frac{d}{dq}g^{-1}(q)|$$

1. I1

1.1. Estimación de parámetros

Estadístico suficiente

Definición: Dada una muestra $\mathbf{X}$, un estadístico $T(\mathbf{X})$ se dice suficiente para θ, si la distribución de $\mathbf{X}$ condicionado al valor de $T(\mathbf{X})$ no depende de θ, es decir

$$P(\mathbf{X} = x|T(\mathbf{X}) = T(\mathbf{x})) \text{ no depende de } \theta$$

En la práctica, la definición anterior de puede reescribir en términos de la razón

$$\frac{f(\mathbf{x}|\theta)}{f_T(T(\mathbf{x})|\theta)} \text{ no depende de } \theta$$

donde f es la distribución de la muestra $\mathbf{x}$, y f_T es la distribución del estadístico T.

Teorema de Factorización (Fisher-Neyman): Un estadístico $T(\mathbf{X})$ es suficiente para θ si y solo si existen funciones $g(T(\mathbf{x}); \theta)$ y $h(\mathbf{x})$ tal que la función de densidad de $\mathbf{X}$ se pueda factorizar como

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta) \cdot h(\mathbf{x})$$

Definición: Un estadístico T es suficiente minimal para θ, si puede ser definido en función de cualquier otro estadístico suficiente para θ.

Teorema: Sea $T(\mathbf{X})$ un estadístico para $\mathbf{X}$ y $\mathbf{X}$ e $\mathbf{Y}$ muestras cualquiera. Si la razón

$$\frac{f(X|\theta)}{f(Y|\theta)} = \frac{L(X|\theta)}{L(Y|\theta)}$$

no depende de θ si y solo si $T(\mathbf{x}) = T(\mathbf{y})$, entonces T es un estadísitco suficiente minimal.

Información

Condiciones de regularidad. Sean $\mathbf{X} = (X_1, \ldots, X_n)$ variables aleatorias *iid* con función densidad $f(x;\theta)$ para $x \in \chi$ y $\theta \in \Theta$. Se dice que f satisface las condiciones de regularidad cuando

I. El espacio muestral χ no depende de θ.

II. El espacio paramétrico Θ es un intervalo abierto de $\mathbb{R}$.

III. La derivada parcial $\frac{\partial f}{\partial \theta}$ existe y es finita para todo $x \in \chi$ y $\theta \in \Theta$.

IV. $\frac{\partial}{\partial \theta} \int_X f(x;\theta)dx = \int_X \frac{\partial}{\partial \theta} f(x;\theta)dx, \forall \theta \in \Theta,$ para el caso continuo.

V. Para cualquier estadístico $T(X)$, $\frac{\partial^k}{\partial \theta^k} \int_X T(x) f(x;\theta)dx = \int_X T(x) \frac{\partial^k}{\partial \theta^k} f(x;\theta)dx, \forall \theta \in \Theta,$ para el caso continuo, $k = 1, 2$.

VI. $\mathbb{E}\left[\left\{\frac{\partial}{\partial \theta} \log f(X;\theta)\right\}^2\right] < \infty, \forall \theta \in \Theta.$

Definición: Función Score. Suponiendo que se cumplen las condiciones de regularidad, para cada $x \in \chi$ se define la función score como:

$$\mathcal{U}(\theta; x) = \frac{\partial log L(\theta)}{\partial \theta}$$

Bajo las condiciones de regularidad se tiene

$$\mathrm{E}[\mathcal{U}(\theta; \mathbf{X})] = 0, \quad \forall \theta \in \Theta$$

Definición: La información de Fisher o la información sobre θ contenida en $\mathbf{X}$, se define por:

$$\mathcal{I}_X(\theta) = \mathrm{E}[\{\frac{\partial}{\partial \theta} log f(X;\theta)\}^2]$$

Teorema: Bajo las condiciones de regularidad, la información de Fsiher se puede escribir como:

$$\mathcal{I}_X(\theta) = \mathrm{Var}[\mathcal{U}(\theta; X)] = -\mathrm{E}[\{\frac{\partial^2}{\partial \theta^2} log f(X;\theta\}]$$

Teorema: Sean X e Y dos variables aleatorias independientes. Si $\mathcal{I}_X$ e $\mathcal{I}_Y$ son las informaciones de Fisher acecrca de θ contenidas en X e Y, respectivamente, entonces la información de Fisher contenida en (X, Y) está dada por:

$$\mathcal{I}_{(X,Y)}(\theta) = \mathcal{I}_X(\theta) + \mathcal{I}_Y(\theta)$$

Corolario: Sean $X_1, \ldots, X_n$ variables aleatorias *iid* con función distribución $f(x;\theta)$. Se denota $\mathcal{I}_{X_1}(\theta)$ como la información contenida en la observación X_1. Entonces, la información $\mathcal{I}_X(\theta)$, contenida en la muestra aleatoria $X = (X_1, \ldots, X_n)$, está dada por:

$$\mathcal{I}_X(\theta) = n \cdot \mathcal{I}_{X_1}(\theta)$$

Teorema: Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria *iid* y $T = T(X)$ un estadístico. Entonces, se cumple que

$$\mathcal{I}_X(\theta) \geq \mathcal{I}_T(\theta), \quad \text{para todo } \theta \in \Theta.$$

Las dos medidas de información coinciden para todo θ si y solo si T es un estadístico suficiente para θ.

Ahora, se considera el caso donde θ es de dimensión k. (Las condiciones de regularidad deben ser adaptadas a la dimensionalidad de θ)

Definición: Vector Score. Para todo $x \in \chi$, al vector de derivadas parciales de la función de log-verosimilitud

$$\mathcal{U}(\theta; x) = \frac{\partial \mathcal{L}(\theta; x)}{\partial \theta} = (\frac{\partial}{\partial \theta_1} \mathcal{L}(\theta;x), \ldots, \frac{\partial}{\partial \theta_2} \mathcal{L}(\theta;x))^T$$

se denomina vector o función score y $\mathrm{E}[\mathcal{U}(\theta; X)] = 0$

Definición: Matriz de información de Fisher. A la matriz $k \times k$

$$\mathcal{I}_x(\theta) = \mathrm{Var}[\boldsymbol{\mathcal{U}}(\theta; X)]$$

se le denomina matriz de información de Fisher. Sus elementos son:

$$I_{ij} = \mathrm{E}[\frac{\partial}{\partial\theta_i}\mathcal{L}(\theta; X) \cdot \frac{\partial}{\partial\theta_j}\mathcal{L}(\theta; X)]$$

Teorema: Sea $\boldsymbol{J}(\theta; \mathbf{x})$ el negativo de la matriz de segundas derivadas de la función de log-verosimilitud, es decir:

$$J_{ij} = -\frac{\partial^2\mathcal{L}(\theta; x)}{\partial\theta_i\partial\theta_j}$$

Entonces la información de Fisher es:

$$\mathcal{I}_x(\theta) = \mathrm{E}[\boldsymbol{J}(\theta; X)]$$

Teorema: Sea X una variable aleatoria con función de distribución $f(x; \theta)$. Sea $\theta = h(\gamma)$ otra parametrización, donde h es diferenciable. Entonces la información contenida en X sobre γ está dada por:

$$\mathcal{I}^*_X(\gamma) = \mathcal{I}_X(h(\gamma))[h^{'}(\gamma)]^2$$

Ancilaridad

Un estadiśtico cuya distribucioń no depende del paraḿetro θ se llama ancilar.

1.2. Métodos de estimación

Método de los momentos

Sea $f(x; \theta)$ la densidad de una variable aleatoria X la cual depende de un vector de parámetros, θ, de dimensión k. Sea $X = (X_1, \ldots, X_n)$ una muestra *iid* desde esta distribución. Se definen los momentos poblacionales y muestrales, respectivamente, por:

I. $\mu'_j = E(X^j) = \mu'_j(\theta)$, y

II. $m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$,

para $j = 1, \ldots, k$. Formamos las k ecuaciones:

$$m_j = \mu'_j(\theta), \quad j = 1, \ldots, k$$

en las k variables $\theta_1, \ldots, \theta_k$ y sean $\hat{\theta}_1, \ldots, \hat{\theta}_k$ sus soluciones, las cuales suponemos únicas.
De esta forma, $\hat{\boldsymbol{\theta}} = (\hat{\theta_1}, \ldots, \hat{\theta}_n)$ es el estimador de momentos de $\boldsymbol{\theta}$.

Estimador de máxima verosimilitud

Recoredemos que para una muestra iid de una población con fdp $f(x; \theta)$, con $\theta = (\theta_1, ..., \theta_k)$, la función de verosimilitud está definida por:

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

Definicion: Si $\hat{\theta}(x)$ es el valor en el cual $L(\theta; x)$ alcanza su máximo como función de θ, un estimador máximo verosímil, EMV del parámetro θ es $\hat{\theta}(X)$. Esto se interpreta como el valor de θ que maximiza la probabilidad de observar los datos que ya disponemos. Es conveniente usar la función la función score (verosimilitud) para estimar el EMV.

Para cada punto muestral $(x_1, \ldots, x_n)$, si $\hat{\theta}(x)$ es el valor del parámetro en dónde $L(x|\theta)$ alcanza su máximo, entonces el EMV de la muestra X es $\hat{\theta}(x)$. Para localizar el máximo, si $L(\theta)$ es $C^2(\theta)$ entonces podemos derivar una vez e igualar a cero para encontrar el EMV y derivar por segunda vez para comprobar que es un valor máximo.

La matriz hessiana H es parecida a la J y de define similarmente:

$$H = \begin{pmatrix} \frac{\partial^2}{\partial\theta_j^2}\mathcal{L}(\theta) & \frac{\partial^2}{\partial\theta_j\theta_i}\mathcal{L}(\theta) \\ \frac{\partial^2}{\partial\theta_i\theta_j^2}\mathcal{L}(\theta) & \frac{\partial^2}{\partial\theta_j^2}\mathcal{L}(\theta) \end{pmatrix}$$

Si esta matriz evaluada en nuestros estimadores (G) para θ es semi-definifa negativa (al menos 1 derivada parcial de segundo orden negativa y determinante de G es positivo). SI lo anterior ocurre entonces dichos estimadores son EMV.

Estimadores de Bayes

En la estimación puntual bayesiana, θ en un modelo paramétrico$f(x;\theta)$ es una variable aleatoria que tiene su propia distribución $\pi(\theta,\lambda)$. $\pi(\theta,\lambda)$ se denomina distribución a priori y λ se conoce como hiperparámetro de la distribución a priori.

Las inferencias sobre θ en el enfoque bayesiano se basan en la distribución de θ dados los valores observados de una muestra aleatoria $x = (x_1,\ldots,x_n)$, que se llama distribución a posterior y se denota por $f(\theta|x)$

En particular, cuando θ es una variable aleatoria continua, el teorema de Bayes establece que la función de distribución posterior es

$$f(\theta|x) = \frac{f(x,\theta;\lambda)}{f(x)} = \frac{f(x|\theta)\pi(\theta;\lambda)}{\int_\Theta f(x|\theta)\pi(\theta;\lambda)d\theta}.$$

Cuando θ es una variable aleatoria discreta, el teorema de Bayes establece que la función de distribución posterior es

$$f(\theta|x) = \frac{f(x,\theta;\lambda)}{f(x)} = \frac{f(x|\theta)\pi(\theta;\lambda)}{\sum_{\theta\in\Theta} f(x|\theta)\pi(\theta;\lambda)}.$$

Note que en ambos casos, la distribución posterior de θ dado x es proporcional al producto de la función de verosimilitud asociada con la muestra y la distribución a priori de θ, es decir,

$$f(\theta|x) \propto L(\theta)\pi(\theta;\lambda).$$

Así, la distribución posterior combina la información sobre θ en la distribución a priori y en la función de verosimilitud, para producir una distribución actualizada que contiene toda la información disponible sobre θ. Además, en la estimación bayesiana, todas las inferencias sobre θ se basan en la distribución posterior.

Teorema: Si $X = (X_1,\ldots,X_n)$ es una muestra de variables aleatorias *iid* con distribución común $f(x|\theta)$, T un estadístico suficiente para θ y $\pi(\theta;\lambda)$ una distribución a priori para θ, entonces la distribución posterior de θ dado X depende de la muestra solo a través del estadístico suficiente T.

Definición: Para un estimador T sobre θ, una función de pérdida $\mathscr{L}(T;\theta)$ es una función real no negativa t.q $\mathscr{L}(T;\theta) = 0$. Algunos ejemplos:

- Pérdida cuadrática, $\mathscr{L}(T;\theta) = (T-\theta)^2$
- Pérdida absoluta, $\mathscr{L}(T;\theta) = |T-\theta|$
- Norma, $\mathscr{L}\mathscr{p}(T;\theta) = |T-\theta|^p$ para $p > 0$.

EL valor esperado de la función de pérdida como función de θ se conoce como **función de riesgo**.

Definición: Para un estimador $T(X)$ de θ y una función de pérdida $\mathscr{L}(T;\theta)$, la función de riesgo asociada es el valor esperado de la función de pérdida con respecto a $f(x|\theta)$;

$$T(T;\theta) = \mathrm{E}\{\mathscr{L}(T;\theta)|\theta\}.$$

Ojo: Si se tiene una muestra de v.a iid con fdp $f(x|\theta),\pi(\theta;\lambda)$ distribución a priori para θ y $\hat{\mathscr{L}}(\hat{\theta};\theta)$ es la función de pérdida **absoluta**, entonces la **mediana** de la distribución posterior es el estimador de Bayes de θ.

Definición: El riesgo de Bayes de un estimador T asociado a una función de pérdida $\mathscr{L}$ y dist a priori $\pi(\theta;\lambda)$ es $\mathsf{E}_\theta\{R(T;\theta)\}$.
Sea $\mathscr{L}$ una función de pérdida y $\pi(\theta;\lambda)$ una distribución a priori para θ. Un estimador $T^\star$ de θ con mínimo riesgo de Bayes se denomina estimador bayesiano de θ.

Así $T^\star$ es E.B de θ cuando $\mathsf{E}_\theta\{R(T^\star;\theta)\} \leq \mathsf{E}_\theta\{R(T;\theta)\}$ para cualquier otro estimador T de θ.

Distribución a priori conjugada Una clase $\mathcal{G}$ de distribuciones se denomina conjugada con respecto a la verosimilitud $L(\theta)$ si la dist posterior $f(\theta|x)$ está en $\mathcal{G}$ $\forall x$, siempre que la dist a priori $\pi(\theta;\lambda)$ esté en $\mathcal{G}$.

Cuadro 3: Distribuciones conjugadas en inferencia bayesiana

Verosimilitud	**Priori conjugada**	**Posteriori**
$X \sim \text{Binomial}(n,\pi)$	$\pi \sim \text{Beta}(\alpha,\beta)$	$\pi \vert x \sim \text{Beta}(\alpha+x,\beta+n-x)$
$X \sim \text{Geométrica}(\pi)$	$\pi \sim \text{Beta}(\alpha,\beta)$	$\pi \vert x \sim \text{Beta}(\alpha+1,\beta+x-1)$
$X \sim \text{Poisson}(\lambda)$	$\lambda \sim \text{Gamma}(\alpha,\beta)$	$\lambda \vert x \sim \text{Gamma}(\alpha+x,\beta+1)$
$X \sim \text{Exponencial}(\lambda)$	$\lambda \sim \text{Gamma}(\alpha,\beta)$	$\lambda \vert x \sim \text{Gamma}(\alpha+1,\beta+x)$
$X \sim N(\mu,\sigma^2 \text{ conocida})$	$\mu \sim N(\nu,\tau^2)$	$\mu \vert x \sim N((\frac{1}{\sigma^2}+\frac{1}{\tau^2})^{-1}(\frac{x}{\sigma^2}+\frac{v}{\tau^2}),(\frac{1}{\sigma^2}+\frac{1}{\tau^2})^{-1})$
$X \sim N(\mu \text{ conocida},\sigma^2)$	$\sigma^2 \sim \text{IG}(\alpha,\beta)$	$\sigma^2 \vert x \sim \text{IG}\left(\alpha+\frac{1}{2},\beta+\frac{(x-\mu)^2}{2}\right)$

1.3. Métodos para comparar estimadores

Insesgamiento
Un estadístico T se denomina insesgado para θ si $\mathrm{E}(T) = \theta$, $\forall\theta \in \Theta$. Si el estimador es sesgado, se define el sesgo como $S(T;\theta) = \mathrm{E}(T) - \theta$

E.C.M Definimos el error cuadrático medio como:

$$\mathbf{E}([\hat{\theta} - \theta]^2) = Var(\hat{\theta}) + \mathbf{B}^2(\hat{\theta})$$

$$\mathbf{B} = \mathbf{E}(\hat{\theta}) - \theta : \text{ Sesgo}$$

Definición. Un estimador insesgado T_1 se dice más eficiente que otro estimador insesgado T_2 para estimar un parámetro θ si

$$Var(T_1) \leq Var(T_2), \forall\theta \in \Theta$$

y

$$Var(T_1) < Var(T_2)$$

para al menos un $\theta \in \Theta$

La eficiencia relativa entre ambos estimadores es la razón de sus varianzas:

$$\text{Efr} = \frac{Var(T_2)}{Var(T_1)}$$

Mejores estimadores insesgados

Cota de Cramér-Rao. Suponinedo que se satisfacen las condicones de regularidad y que la información de Fisher satisface $0 < \mathcal{I}_x(\theta) < \infty$. Sea $\tau(\theta)$, donde τ es una función de valor real continuamente diferenciable con derivada distinta de 0. Si T es un estimador insesgado para $\tau(\theta)$, entonces

$$\mathrm{Var}(T) \geq \frac{[\tau'(\theta)]^2}{\mathcal{I}_x(\theta)} \quad \forall\theta \in \Theta$$

La igualdad es válida si y solo si para $x \in \chi$ y $\forall\theta \in \Theta$

$$T(x) - \tau(\theta) = \frac{\tau'(\theta) * \mathcal{U}(\theta; x)}{\mathcal{I}_x(\theta)}$$

Donde $\mathcal{U}$ es la función score.

Definición. La eficiencia de un estimador insesgado de T se define como la razón de su varianza y la cota de Carmér-Rao, es decir

$$\mathrm{ef}(T, \theta) = \frac{[\tau'(\theta)]^2}{\mathcal{I}_x(\theta)\mathrm{Var}(T)}$$

Teorema de Rao-Blackwell Sean $X_1, \ldots, X_n$ una muestra iid con fdp $f(x;\theta)$ para $\theta \in \Theta$, S un estadístico suficiente para θ, y T un estimador insesgado de $\tau(\theta)$. Si $T^* = \mathbb{E}(T|S)$, entonces:

i) T^* es una función solo del estadístico suficiente S.

ii) $\mathbb{E}(T^*) = \tau(\theta)$.

iii) $\text{Var}(T^*) \leq \text{Var}(T)$, $\forall\theta \in \Theta$, la igualdad se cumple si $T = t(S)$.

2. I2

Propiedades asintóticas

1. Se dice que un estadístico $T_n = T(X_1, \ldots, X_n)$ es un estimador **asintóticamente insesgado** de un parámetro θ cuando: $\lim_{n\to\infty} Ses(T_n;\theta) = 0$.
2. Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria y sea $T_n = T(X)$ un estimador de θ. El estimador T_n se dice consistente en ECM (o en media cuadrática) de θ si: $\lim_{n\to\infty} ECM(T_n;\theta) = 0$
3. Si $X_1, \ldots, X_n$ es una muestra aleatoria, el estadístico T_n es consistente en ECM para θ ssi: $\lim_{n\to\infty} Var(T_n) = 0$ y $\lim_{n\to\infty} Ses(T_n;\theta) = 0$
4. Sean T_1 y T_2 dos estimadores de θ, al eficiencia relativa asintótica de T_1 a T_2 es: $\lim_{n\to\infty} \frac{ECM(T_1;\theta)}{ECM(T_2;\theta)}$

Método delta

Supongamos que disponemos de un estimador T con error estándar conocido para estimar θ; sin embargo, el objetivo es estimar una función $\tau(\theta)$. $\tau(T)$ se usa a menudo para estimar $\tau(\theta)$, pero cuando τ no es una función lineal, el error estándar del estimador $\tau(T)$ es difícil de obtener. Es por esto que se utiliza el método delta:

i) La expansión de Taylor de primer orden para $\tau(T)$ es:

$$\tau(T) \approx \tau(\mu) + (T-\mu)\tau'(\mu)$$

ii) Luego,

$$E[\tau(T)] \approx \tau(\mu) \text{ y } Var[\tau(T)] \ \approx Var(T)\cdot[\tau'(\mu)]^2$$

Teorema: Suponga que T_n es un estimador de la forma $T_n = g(S_n)$ donde la secuencia $\{S_n\}$ es asintóticamente normal, es decir

$$\sqrt{n}(S_n - \mu) \sim N(0, \sigma^2)$$

para algunas constantes μ y $\sigma^2 > 0$. Si g es función C^1 distinta de cero en μ, entonces

$$\sqrt{n}(T_n - g(\mu)) \sim N(0, [g'(\mu)]^2\sigma^2)$$

Ejemplo: Sean $X_1, \ldots, X_n$ va iid con $E(X_i) = \mu \neq 0$ y $Var(X_i) = \sigma^2 < \infty$. El parámetro $\tau(\mu) = log(\mu)$ es estimado por $\tau(\bar{X_n}) = log(\bar{X_n}$. Este estimador es consistente y asintóticamente normal. Con $g(s) = log(s) = 1/s$ obtenemos desde $\sqrt{n}(S_n - \mu) \sim N(0, \sigma^2)$, que

$$\sqrt{n}(log\bar{X_n} - log(\mu)) \sim N(0, \frac{\sigma}{\mu^2})$$

Consistencia: Diremos que una secuencia $\{T_n\}$ de estimadores para un parámetro $\tau(\theta)$ es consistente, si T_n converge en probabilidad a $\tau(\theta)$, es decir:

$$\lim_{n\to\infty} P(|T_n - \tau(\theta)| > \epsilon) = 0$$

Para verificar consistencia podemos utilizar la **desigualdad de Chebyshev**: Para una v-a Y con media finita, se tiene para todo $\epsilon > 0$

$$P(|t - E(Y)| > \epsilon) \leq \frac{Var(Y)}{\epsilon^2}$$

Sean $X_1, \dots, X_n$ variables aleatorias *iid* con función de distribución F. Entonces, se tiene que :

i) La media muestral de $\bar{X}$ converge a $\mathrm{E}(X) = \mu$: $\bar{X}_n \xrightarrow{P} \mu$

ii) La varianza muestral y la desciavión estándar muestral convergen a $\mathrm{Var}(X) = \sigma^2$ y σ, respectivamente:

$$S_n^2 \xrightarrow{P} \sigma^2 \quad S_n \xrightarrow{P} \sigma$$

iii) La frecuencia relativa de un evento A converge a su probabilidad. Sea $fr_n(A) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_i(A)$, donde $\mathbb{I}_i(A) = 1$ si A ocurre en el i-ésimo ensayo, y cero en otro caso. Entonces $fr_n(A) \xrightarrow{P} P(A)$

iv) Si el j-ésimo momento de $F\mu'_j = \mathrm{E}(X^j)$ existe, entonce el momento muestral $m_j = \frac{1}{n}\sum_{i=1}^n X_i^j$ es un estimador consistente para u'_j. Por lo tanto, el EMM para $g(\mu'_1, \dots, \mu'_k)$ es un estimador consistente si la función g es continua,

Teorema: Sean $X_1, \dots, X_n \overset{iid}{\sim} F_\theta, \theta \in \Theta \subseteq \mathbb{R}$ y sea m_k el k-ésimo momento muestral y

$$\sqrt{n}(m_k - \mu'_k) \xrightarrow{D} N(0, \sigma_M^2)$$

donde $\sigma_m^2 = \mathrm{Var}(X^k) = \mu'_{2k} - (\mu'_k)^2$ y $\mu'_k = \mathrm{E}(X^k)$. Si $\hat{\theta}_n = g(m_k)$ es el EMM de θ y g es una función continua, entonces:

i) El EMM es consistente: $\hat{\theta}_n \xrightarrow{P} \theta$

ii) $\sqrt{n}(g(m_k) - g(\mu'_k)) \xrightarrow{D} N(0, [g'(\mu'_k)]^2\sigma_M^2)$

Teorema: Sean $X_1, \dots, X_n$ variables aleatorias *iid* con función distribución $f(x;\theta)$ que satisface las condiciones de regularidad. Sea $\hat{\theta}$ el EMV de θ. Entonces,

i) $\hat{\theta}$ es un estimador asintóticamente insesgado de θ

ii) $\hat{\theta}$ es consistente en ECM para θ

iii) $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \mathcal{I}_{X_1}^{-1}(\theta))$

iv) $\hat{\theta}$ es un estimador asintóticamente eficiente de θ

Sean $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Exp}(\lambda)$ con función distribución $f = \lambda \exp(-\lambda x) I_{(0,\infty)}(x)$. Sea $\hat{\lambda} = 1/\bar{X}$ el EMV. Entonces, $\mathcal{I}_X(\lambda) = 1/\lambda^2$. Por lo tanto

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{D} N(0, \lambda^2)$$

Además, $\hat{\lambda}$ es asintóticamente insesgado, consistente en ECM y asinstóticamente eficiente para λ.

Teorema: Sea $X = (X_1, \dots, X_n)$ una muestras de variables aleatorias *iid* con función distribución $f(x;\theta)$. Sea $\mathcal{I}_x(\theta) = n\mathcal{I}_X(\theta)$ la matriz de información de Fisher y $\hat{\theta}_n$ el EMV de θ. Bajo las condiciones de regularidad, se cumple que

i) $\hat{\theta}_n$ es consistente: $\hat{\theta}_n \xrightarrow{P} \theta$

ii) $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N_k(0, \mathcal{I}_X^{-1}(\theta))$

Distribution	Moment Estimator	EMV	Sufficient Statistic
Bernoulli(θ)	$\hat{\theta}_M = \bar{X}$	$\hat{\theta}_{\text{MLE}} = \bar{X}$	$\sum X_i$
Binomial(n, θ)	$\hat{\theta}_M = \frac{\bar{X}}{n}$	$\hat{\theta}_{\text{MLE}} = \frac{\sum X_i}{nN}$	$\sum X_i$
Geometric(θ)	$\hat{\theta}_M = \frac{1}{1+\bar{X}}$	$\hat{\theta}_{\text{MLE}} = \frac{1}{1+\bar{X}}$	$\sum X_i$
Poisson(λ)	$\hat{\lambda}_M = \bar{X}$	$\hat{\lambda}_{\text{MLE}} = \bar{X}$	$\sum X_i$
Exponential(λ)	$\hat{\lambda}_M = \frac{1}{\bar{X}}$	$\hat{\lambda}_{\text{MLE}} = \frac{1}{\bar{X}}$	$\sum X_i$
Normal(μ, σ^2)	$\hat{\mu}_M = \bar{X}, \quad \hat{\sigma}^2_M = \frac{1}{n}\sum(X_i - \bar{X})^2$	$\hat{\mu}_{\text{MLE}} = \bar{X}, \quad \hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum(X_i - \bar{X})^2$	$(\sum X_i, \ \sum X_i^2)$
Uniform($0, \theta$)	—	$\hat{\theta}_{\text{MLE}} = \text{máx}\, X_i$	$\text{máx}\, X_i$
Gamma(α known, β)	$\hat{\beta}_M = \frac{\bar{X}}{\alpha}$	$\hat{\beta}_{\text{MLE}} = \frac{\bar{X}}{\alpha}$	$\sum X_i$
Laplace(θ)	$\hat{\theta}_M = \frac{1}{n}\sum \lvert X_i \rvert$	$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum \lvert X_i \rvert$	$\sum \lvert X_i \rvert$
Pareto(θ)	$\hat{\theta}_M =$	$\hat{\theta}_{\text{MLE}} = \frac{n}{\sum \log(X_i/y_0)}$	$\prod X_i$ (or $\sum \log X_i$)

Cuadro 4: Moment Estimators, MLEs, and Sufficient Statistics for Common Distributions

Cómo encontrar la distribución asintótica:

Es común que al tener un estimador se nos pida su distribución asintótica, podemos obtenerla siguiendo los siguientes pasos:

1. Definir bien el estimador $\hat{\theta}_n$. Podemos tener por ejemplo:
   - $\hat{\theta}_n = \bar{X}$ el promedio muestral (Estimador de momentos).
   - $\hat{\theta}_n = g(\bar{X})$ una función de un estimador (casi siempre EMV).
2. ¿Se puede aplicar TCL? Casi siempre aplica para estimadores de momentos, entonces:

   $$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$$

   Con $\mu = \mathbb{E}(X), \sigma^2 = Var(X)$.
3. ¿Se puede aplicar método delta? Casi siempre para EMV, entonces:

   $$\sqrt{n}(\bar{X} - \mu) \sim N(0, [g'(\mu)]^2 \cdot \sigma^2)$$

   O también:

   $$(\hat{\theta}_{EMV} - \theta) \sim N(0, I_{X_n}^{-1})$$

2.1. Métodos Bootstrap y Jacknife

Método de Bootstrap: Sea $X = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid* con función de distribución $f(x; \theta)$. Sea $\hat{\theta}$ un estimador de θ. Para evaluar la precisión de $\hat{\theta}$, el algoritmo bootstrap puede usarse para estimar el error estándar de $\hat{\theta}$ siguiendo los siguientes pasos:

1. Se escoge un número entero positivo B. Para $b = 1, \ldots, B$, se generan, de forma independiente, una muestr bootstrap $X^{(b)} = (X_1^{(b)}, \ldots, X_n^{(b)})$ cuyos elementos se extraen con remplazo de la de n objetos $X = (X_1, \ldots, X_n)$
2. Para cada muestra bootstrap $X^{(b)}$, se calcula la correspondiente estima $\hat{\theta^{(b)}} = \hat{\theta}(x^{(b)})$ de θ para $b = 1, \ldots, B$
3. Se caclula la desviación estándar muestral de las B réplicas bootsrap:

   $$\hat{es}_B = [\frac{1}{B-1}\sum_{b=1}^{B}\{\hat{\theta}^{(b)} - \hat{\theta}_B\}^2]^{1/2}$$

   la cual es una estima del error estándar de $\hat{\theta}$, donde $\hat{\theta}_B = B^{-1}\sum_{b=1}^{B}\hat{\theta^{(b)}}$. Algunas veces es $\hat{es}_B$ se le denomina estima bootstrap no-paramétrico.

Método de Bootstrap paramétrico: Si $f(x;\theta)$ depende de un parámetro θ y $\hat{\theta}$ es su estimador, entonces se toman muestras desde $f(x;\hat{\theta})$, en vez de generar muestras con reemplazo desde las observaciones $X = (X_1, \ldots, X_n)$. En este caso, el paso 1 consiste en generar muestras bootstrap $X^{(b)} = (X_1^{(b)}, \ldots, X_n^{(b)})$, $b = 1, \ldots, B$, desde $f(x;\hat{\theta})$. Estas muestras se denominan muestras bootstrap paramétricas. La distribución muestral de $\hat{\theta}$ puede se aproximada por la distribución empírica de los $\hat{\theta}^{(b)}$, $b = 1, \ldots, B$.

Método de Jackknife: Este es una técnica para estimar el sesgo y el error estándar de un estimador que tiene algunas similitudes con el aloritmo bootstrap. En lugar de elegir B muestras bootstrap, se forma una muestra jackknife dejando fuera una observación a la vez. Para $i = 1, \ldots, n$, sea

$$X_{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

la i-ésima ,muestra jackknife, que es un subconjunto de X con la i-ésima observación eliminada. Se aplica el método de estimación a la i-ésima muestra jackkinfe y se obtiene la i-ésima réplica jackkinfe de $\hat{\theta} = \hat{\theta}(X)$:

$$\hat{\theta}_{(i)} = \hat{\theta}(X_{(i)})$$

para $i = 1, \ldots, n$.

Sea $\hat{\theta}_J = n^{-1} \sum_{i=1}^{n} \hat{\theta}_{(i)}$. La estima jackknife del sesgo se define como:

$$\hat{\text{Ses}}_J = (n-1)(\hat{\theta}_J - \hat{\theta})$$

El estimador jackknife corregido por seso está dado por $\hat{\theta}_{JS} = \hat{\theta} - \hat{\text{Ses}}_J$ y puede ser escrito como:

$$\hat{\theta}_{JS} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\theta}_i$$

donde $\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$ se denomina pseudo valores.

La estima jackknife de la varianza deel estimador $\hat{\theta}$ se define como:

$$\hat{\text{Var}}_J(\hat{\theta}) = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(i)} - \hat{\theta}_J)^2 = \frac{S_J^2}{n}$$

donde

$$S_J^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\tilde{\theta}_i - \hat{\theta}_{JS})^2$$

es la varianza muestral de los pseudo valores. Bajo las condiciones de regularidad, se puede probar que $\hat{\text{Var}}_J(\hat{\theta})$ es un estimador consistente de $\text{Var}(\hat{\theta})$.

Bootsrap multiparamétrcio:
Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid* con función densidad $f(x;\theta)$ donde $\theta \in \Theta \subseteq \mathbb{R}^k$. Para cada muestra bootstrap $\mathbf{X}^{(b)}$, calcule la correspondiente estima (ahora un vector $k \times 1$) $\hat{\boldsymbol{\theta}}^{(b)} = \hat{\boldsymbol{\theta}}(\mathbf{x}^{(b)})$ de $\boldsymbol{\theta}$ para $b = 1, \ldots, B$. Entonces,

i) El estimador bootstrap de θ es $\hat{\theta}_B = B^{-1} \sum_{b=1}^{B} \hat{\theta}^{(b)}$

ii) Una estima bootstrap de la matriz de covarianza de $\hat{\theta}$ es

$$\widehat{\text{Var}}_B(\hat{\theta}) = \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}^{(b)} - \hat{\theta}_B\right)\left(\hat{\theta}^{(b)} - \hat{\theta}_B\right)^T$$

la cual corresponde a la matriz de covarianza de las B réplicas bootstrap.

Jackknife multiparamétrcio:
Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid* con función densidad $f(x;\theta)$ donde $\theta \in \Theta \subseteq \mathbb{R}^k$.

En este caso los pseudo valores (vectores $k \times 1$) están dados por,

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)},$$

para $i = 1, \ldots, n$, de donde sigue que

i) El estimador jackknife de θ es $\hat{\theta}_J = n^{-1} \sum_{i=1}^{n} \tilde{\theta}_i$ y,

ii) Una estima jackknife de la matriz de covarianza de $\hat{\theta}$ es

$$\widehat{\mathrm{Var}}_J(\hat{\theta}) = \frac{V}{n}$$

donde

$$V = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\theta}_i - \tilde{\theta}_J)(\tilde{\theta}_i - \tilde{\theta}_J)^T,$$

es la matriz de covarianza empírica de los pseudo valores.

2.2. Estimación por intrervalo

Definición: Sea $X = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid*. Sean T_1 y T_2 dos estadísticos tales que $T_1 \leq T_2$. Un intervalo $[T_1, T_2]$ utilizado para estimar θ se denomina estimador de intervalo.

Definición: La probabilidad de cobertura asociada con el estimador de intervalo $[T_1, T_2]$ es $P(\theta \in [T_1, T_2])$

Definición: Sea $X_1, \ldots, X_n$ una muestra aleatoria y sean T_1 y T_2 dos estadísticos con $T_1 \leq T_2$. El intervalo $[T_1, T_2]$ se llama un intervalo de confianza del 100(1-α) % para el parámetro θ si:

$$P(\theta \in [T_1, T_2]) = P(T_1 \leq \theta \leq T_2) = 1 - \alpha$$

Teorema: Si $[T_1, T_2]$ es un intervalo de confianza del 100(1-α) % para θ y $\tau(\theta)$ es una función monótona, entonces

i) $[\tau(T_1), \tau(T_2)]$ es un intervalo de confianza del 100(1-α) % para $\tau(\theta)$ cuando τ es una función creciente

ii) $[\tau(T_2), \tau(T_1)]$ es un intervalo de confianza del 100(1-α) % para $\tau(\theta)$ cuando τ es una función decreciente

2.2.1. Método de la cantidad pivotal

Definición: Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria y sea $Q(X;\theta)$ una función de la muestra y del parámetro θ. Se dice que la variable aleatoreia $Q(X;\theta)$ es una cantidad pivotal si la distribución de $Q(X;\theta)$ no depende de θ.

Sea $Q(X;\theta)$ una cantidad pivotal. Entonces siempre es posible elegir dos constantes $a \leq b$ tales que

$$P(a \leq Q(X;\theta) \leq b) = 1 - \alpha$$

Si es posible convertir la desigualdad en la forma $T_1 \leq \theta \leq T_2$, entonces

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha$$

por lo que $[T_1, T_2]$ es un intervalo de confianza para θ.

Familia	Forma general	Pivote
Localización	$f(x-\mu)$	$\bar{X} - \mu$
Escala	$\frac{1}{\sigma}f(\frac{x}{\sigma})$	$\frac{\bar{X}}{\sigma}$
Localización y escala	$\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$	$\frac{\bar{X}-\mu}{\sigma}$

Sea $X = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid* $N(\mu, \sigma^2)$ y sea $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Entonces son cantidades pivotales:

Distribution	Parameter	Pivotal Quantity	Distribution
Bernoulli(θ)	θ	$\frac{\hat{\theta}-\theta}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$	Approximately $\mathcal{N}(0,1)$ for large n
Binomial(n,θ)	θ	$\frac{\hat{\theta}-\theta}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$	Approximately $\mathcal{N}(0,1)$
Geometric(θ)	θ	$\frac{\hat{\theta}-\theta}{\hat{\theta}(1-\hat{\theta})/\sqrt{n}}$	Approximately $\mathcal{N}(0,1)$
Poisson(λ)	λ	$\frac{\bar{X}-\lambda}{\sqrt{\bar{X}/n}}$	Approximately $\mathcal{N}(0,1)$
Exponential(λ)	λ	$2n\lambda\bar{X}$	χ^2_{2n}
Normal(μ,σ^2)	μ	$\frac{\bar{X}-\mu}{S/\sqrt{n}}$	t_{n-1}
Normal(μ,σ^2)	σ^2	$\frac{(n-1)S^2}{\sigma^2}$	χ^2_{n-1}
Uniform($0,\theta$)	θ	$\left(\frac{X_{(n)}}{\theta}\right)^n$	Uniform($0,1$)
Gamma(α known)	β	$\frac{2\sum X_i}{\beta}$	$\chi^2_{2n\alpha}$
Laplace(θ)	θ	$\frac{2\sum \lvert X_i\rvert}{\theta}$	χ^2_{2n}
Pareto(θ)	θ	$\hat{\theta}=\frac{n}{\sum\log(X_i/y_0)}$	Approximately $\mathcal{N}(\theta,\frac{\theta^2}{n})$

Cuadro 5: Pivotal Quantities for Confidence Interval Construction

i) $Q_1(X;\theta)=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\to Q_1(X;\theta)\sim N(0,1)$

ii) $Q_2(X;\theta)=\frac{\sqrt{n}(\bar{X}-\mu)}{S}\to Q_2(X;\theta)\sim t_{n-1}$

iii) $Q_3(X;\theta)=\frac{(n-1)S^2}{\sigma^2}\to Q_3(X;\theta)\sim\chi^2_{n-1}$

Utilizando las cantidades pivotales anteriores se obtienen los siguientes intervalos de confianza:

i) $[T1,T2]=[\bar{X}-b\frac{\sigma}{\sqrt{n}},\bar{X}+b\frac{\sigma}{\sqrt{n}}]$ es un IC del $100(1-\alpha)\,\%$ para μ, $b=z_{1-\alpha/2}$, $-b=z_{\alpha/2}$.

ii) $[T1,T2]=[\bar{X}-b\frac{S}{\sqrt{n}},\bar{X}+b\frac{S}{\sqrt{n}}]$ es un IC del $100(1-\alpha)\,\%$ para μ, $b=t_{n-1,1-\alpha/2}$, $-b=t_{n-1,\alpha/2}$.

iii) $[T1,T2]=[\frac{(n-1)S^2}{b},\frac{(n-1)S^2}{a}]$ es un IC del $100(1-\alpha)\,\%$ para σ^2, $b=\mathcal{X}_{n-1,1-\alpha/2}$, $a=\mathcal{X}_{n-1,\alpha/2}$.

Tamaño de la muestra: Supongamos que ahora se nos pide el tamaño de la muestra para que con una probabilidad de $1-\alpha$, la media muestral $\bar{X}$ se encuentre en un intervalo igual a ϵ unidades de μ, podemos escribir:

$$P(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\leq\bar{X}-\mu\leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})=P(|\bar{X}-\mu\leq\epsilon)=1-\alpha,$$

donde $\epsilon=z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$, y luego

$$n=(z_{1-\alpha/2}\frac{\sigma}{\epsilon})^2$$

¿Cómo construyo un IDEC? Se tiene $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ con densidad: $f(x) = \frac{1}{\lambda}e^{-x/\lambda}$. Se quiere un IDEC para λ.

1. Obtener estadístico suficiente: Usar verosimilitud. En este caso $\sum_{i=1}^{n} X_i$ es suficiente para λ, nos fijamos que la suma de exponenciales es Gamma; $T \sim Gamma(n, \lambda)$.
2. Ver la pertenencia a una familia: En este caso es de escala.
3. Construir pivote: Se sabe que $\frac{2T}{\lambda} \sim \mathcal{X}^2_{2n}$
4. Construir IDEC:

$$P\left(\chi^2_{2n,\alpha/2} \leq \frac{2T}{\lambda} \leq \chi^2_{2n,1-\alpha/2}\right) = 1 - \alpha$$

$$\left(\frac{2T}{\chi^2_{2n,1-\alpha/2}}, \frac{2T}{\chi^2_{2n,\alpha/2}}\right)$$

es un intervalo de confianza del $100(1-\alpha)\,\%$ para λ.

¿Qué hago si no tengo un pivote?: Sea T un estadístico suficiente. Lo que hay que hacer es

1. Buscar la distribución de T. Si $T = \text{mín}\{X_1, \ldots, X_m\}$ entonces calculamos $f_T(t) = n[1 - F_x(t)]^{n-1} f_x(t)$
2. Calculamos los límites de nuestras distribuciones para α; a y b: Buscamos a tal que

$$\int_{\theta}^{a} f_T(t) = \frac{\alpha}{2}$$

   De forma similar buscamos b tal que:

$$\int_{\theta}^{b} f_T(t) = 1 - \frac{\alpha}{2}$$

2.2.2. Método estadístico

Cuando no se dispone de una cantidad pivotal, aún es posible determinar un intervalo de confianza para θ cuando se conoce la distribución muestral de un estadístico suficiente para θ o la distribución muestral de un estimador de θ. En particular, si $T(x) = T$ es un estadístico suficiente o un estimador de θ con función densidad conocida, $f_T(t;\theta)$, y para cada $\theta \in \Theta$ existen funciones $h_1(\theta)$ y $h_2(\theta)$ tales que

$$P(h_1(\theta) \leq T \leq h_2(\theta)) = 1 - \alpha$$

entonces $\{\theta \in \Theta : h_1(\theta) \leq T \leq h_2(\theta)\}$ es una región de confianza del $100(1 - \alpha)\,\%$ para θ. Es posible transformar la región de confianza anterior en un intervalo de confianza para θ.

Intervalos de confianza en muestras grandes

En el caso que no se tenga una cantidad pivotal ni una distribución muestral conocida que pueda usarse para construir un intervalo de confianza para θ, si n es suficientemente grande y la función distribución subyacente de una muestra de vairables aleatoria *iid* satisface las condiciones de regularidad, es posible obtener un intervalo de confianza aproximado. En particular, un intervalo de confianza aproximada de $100(1\text{-}\alpha)\,\%$ para θ puede basarse en el EMV de θ. Sea $X = (X_1, \ldots, X_n) \sim f(x;\theta)$ con $\theta \in \Theta \subseteq \mathbb{R}$. Suponiendo que $f(x;\theta)$ satisface las condiciones de regularidad siendo T_n un estimador de θ tal que

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2(\theta))$$

Es decir, para n suficientemente grande $T_n \overset{as}{\sim} N(\theta, \sigma^2(\theta)/n)$ Luego, se tiene que

$$P(-z_{1-\alpha/2} \leq \frac{\sqrt{n}(T_n - \theta)}{\sigma(\theta)} \leq z_{1-\alpha/2}) =$$

$$P(T_n - z_{1-\alpha/2}\frac{\sigma(\theta)}{\sqrt{n}} \leq \theta \leq T_n + z_{1-\alpha/2}\frac{\sigma(\theta)}{\sqrt{n}}) \approx 1-\alpha$$

A partir de lo anterior, es posible construir un intervalo de confianza para θ con un coeficinete de confianza de $(1-\alpha)$ para θ

$$[T_1, T_2] = [T_n - z_{1-\alpha/2}\frac{\sigma(\hat{\theta})}{\sqrt{n}}, T_n + z_{1-\alpha/2}\frac{\sigma(\hat{\theta})}{\sqrt{n}}]$$

el cual se denomina intervalo de confianza de Wald. En este caso, $\sigma^2(\hat{\theta})$ es un estimador consistenete de $\sigma^2(\theta)$.

Región de verosimilitud

Definición: Una región de verosimilitud de $p\,\%$ es el conjunto $\{\theta \in \Theta : \frac{L(\theta)}{L(\hat{\theta})} \geq \frac{p}{100}\}$, donde $\hat{\theta}$ es el EMV de θ. Notar que

$$\{\theta \in \Theta : \frac{L(\theta)}{L(\hat{\theta})} \geq \frac{p}{100}\} \iff \{\theta \in \Theta : 2\{\mathcal{L}(\hat{\theta}) - \mathcal{L}(\theta)\} \leq -2\ln(p/100)\}$$

Si $X_1, \ldots, X_n$ es una muestra de variables aleatorias *iid* cond función densidad $f(x;\theta)$ que satisface las condiciones de regulardidad, entonces para n suficientemente grande, se tiene que $Q_a(X;\theta) = 2\{\mathcal{L}(\hat{\theta}) - \mathcal{L}(\theta)\} \overset{as}{\sim} \chi^2_1$, es una cantidad pivotal. Por lo tanto

$$P(\theta \in \Theta : Q_a(X;\theta) \leq -2\ln(p/100) \approx P(\chi^2_1 \leq -2\ln(p/100)$$

En general, un intervalo de confianza de la razón de verosimilitud de puede obtener invirtiendo la razón de verosimilitud. Específicamente, para un EMV $\hat{\theta}$ dado, todos los valore de θ para los cuales $Q_a(X;\theta) \leq \chi^2_{1,1-\alpha}$ forman un intervalo de confianza del 100(1-α) % aproximada para θ, puesto que

$$P(Q_a(X;\theta \leq \chi^2_{1,1-\alpha}) = P(2\{\mathcal{L}(\hat{\theta}) - \mathcal{L}(\theta)\} \leq \chi^{\in}_{\infty,\infty-\alpha}) \approx 1-\alpha$$

Notas

i) El cálculo de los límites de un intervalo de confianza de la razón de verosimilitud requiere, en general, métodos numéricos para encontrar las raíces de la ecuación

$$\{\mathcal{L}(\hat{\theta}) - \mathcal{L}(\theta)\} = \frac{1}{2}\chi^2_{1,1-\alpha}$$

ii) Los intervalos de confianza de la razón de verosimilitud son invariantes con respecto a transformaciones uno a uno del parámetro θ

iii) Los intervalos de confianza de Wald se pueden ver como intervalos de confianza aproximados de la razón de verosimilitud.

2.2.3. Intervalos de confianza de longitud mínima

Dado que los intervalos de confianza no son únicos, es posible crear criterios de optimalidad para elegir uno sobre los otros.
Sea $X = (X_1, \ldots, X_n)$ una muestra de variables aleatorias *iid* con función de distribución $f(x;\theta)$. Un método estándar para obtener un intervalo de cofianza para θ es usa una cantidad pivotal $Q(X;\theta)$. A partir de

$$P(a \leq Q(X;\theta) \leq b) = 1-\alpha$$

se construye un intervalo de confianza $[T_1, T_2]$ del 100(1-α) % para θ. Para cada $Q(X;\theta)$, a y b se pueden elegir de diferentes formas. En este caso, se escogen para obtener un intervalo de confianza más corto.

Para obtener el IC de longitud mínima, con coeficiente de confianza $1-\alpha$, minimizamos:

$$\phi(a,b,\lambda) = L_0 + \lambda(\int_a^b f_Q(q)dq - (1-\alpha))$$

con respecto a a, b, λ, donde λ es un multiplicador de Lagrange y L_0 es la longitud del IC. Las condiciones sobre a y b, para obtener un IC de longitud mínima, pueden escribirse como:

i) $\frac{dL_0}{da} = 0$

ii) $f_Q(b)\frac{db}{da} - f_Q(a) = 0$ o equivalentemente $\frac{db}{da} = \frac{f_Q(a)}{f_Q(b)}$

Para armar un IC de longitud mínima:

1. Partimos desde:
$$P(a \leq Q(X;\theta) \leq b) = 1-\alpha$$
2. Despejamos el parámetro de la cantidad pivotal y luego definimos el largo del intervalo L_0 $\propto$(algo en términos de a y b)
3. Utilizamos las condiciones y despejamos el sistema de ecuaciones.

2.2.4. Intervalos de credibilidad

Una región de credibilidad bayesiana es un estimador de intervalo basado en la distribución posterior.

Definición: Sea $X_1, \ldots, X_n$ una muestra aleatoria y $f(\theta|x)$ la disttribución posterior de θ para una priori $\pi(\theta;\lambda)$. Un estimador del intervalo $[T1, T2]$ es un intervalo de credibilidad bayesiano del $100(1-\alpha)$ % para θ si

$$P(T_1 \leq \theta \leq T_2|x) = \int_{t_1}^{t_2} f(\theta|x)d\theta = 1-\alpha$$

Como utilizar intervalo de credibilidad: Sea $X_1, ..., X_n|\theta \sim N(\theta, 1)$, si nos piden un intervalo de credibilidad para θ tenemos que:

1. Identificar distribución a priori de θ, por ejemplo: $\pi(\theta) = N(0,1)$
2. Buscar la posteriori:
$$\pi(\theta|x) \propto f(x|\theta)g(\theta)$$
En nuestro ejemplo:
$$\pi(\theta|x) \propto e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2} e^{-\frac{1}{2}\sum_{i=1}^n \theta^2} = e^{-\frac{n+1}{2}((n+1)\theta^n - 2\theta\sum_{i=1}^n x_i)}$$
$$\pi(\theta|x) \propto e^{-\frac{1}{2}\frac{(\theta-\frac{\sum_i^n x_i}{n+1})^2}{1/(n+1)}} \sim N(\frac{\sum_i^n x_i}{n+1}, \frac{1}{n+1}) = N(\frac{n}{n+1}\bar{X}, \frac{1}{n+1})$$
3. Armar el intervalo: $P(-z_{1-\alpha/2} \leq \sqrt{n+1}(\theta - \frac{n}{n+1}\bar{X}) \leq z_{1-\alpha/2}) = 1-\alpha$
$$ICR = [\frac{n}{n+1}\bar{X} - \frac{z_{1-\alpha/2}}{\sqrt{n+1}}, \frac{n}{n+1}\bar{X} + \frac{z_{1-\alpha/2}}{\sqrt{n+1}}]$$

2.2.5. Intervalos Bootstrap

Sea $X = (X_1, ..., X_n)$ una muestra de va $\sim$ iid $f(x;\theta)$ con $\theta \in \Theta \subseteq \mathbb{R}$.

a) Intervalo de confianza normal. Para n grande $T_n \sim N(\theta, \sigma_T^2)$, y se estima σ_T^2 por bootstrap. Estimamos:

$$\hat{\sigma_B^2} = \frac{1}{B-1}\sum_{b=1}^{B}\{\hat{\theta}^{(b)} - \hat{\theta}_B\}^2$$

Luego el IC del 100(1-α) % aprox está dado por:

$$[T_1, T_2] = [T_n - z_{1-\alpha/2}\hat{\sigma}_B, T_n + z_{1-\alpha/2}\hat{\sigma}_B]$$

a) Intervalo de confianza pivotal. Definamos el pivote como $R_n = \hat{\theta}_n - \theta$ y sea $H(r)$ la fda del pivote:

$$H(r) = P_F(R_n \leq r)$$

Sea $C_n^* = [T_1, T_2]$ con $T_1 = \hat{\theta}_n - H^{-1}(1-\alpha/2)$ y $T_2 = \hat{\theta}_n - H^{-1}(\alpha/2)$ Luego C_n^* es un IC exacto de nivel $1-\alpha$ para θ. Sin embargo T_1, T_2 dependen de la distribución H, pero la podemos estimar con bootsrap:

$$\hat{H}(r) = \frac{1}{B}\sum_{b=1}^{B} I(R_{n,b} \leq r) \quad \text{donde} \quad R_{n,b} = \hat{\theta}^{(b)} - \hat{\theta}_n$$

Sea r_p^* el cuantil p de $R_{n,1}^*, \ldots, R_{n,B}$ y θ_p^* el percentil p de $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$, las réplicas bootstrap de $\hat{\theta}_n$. Note que$r_p^* = \theta_p^* - \hat{\theta}_n$. Un IC aproximado para θ con coef de confianza $1-\alpha$ es $C_n = [\hat{a}, \hat{b}]$ donde,

$$\hat{a} = \hat{\theta}_n - \hat{H}^{-1}(\alpha/2) = 2\hat{\theta}_n - \theta_{1-\alpha/2}^*$$

$$\hat{b} = \hat{\theta}_n - \hat{H}^{-1}(\alpha/2) = 2\hat{\theta}_n - \theta_{\alpha/2}^*$$

o sea,

$$C_n = [2\hat{\theta}_n - \theta_{1-\alpha/2}^*, 2\hat{\theta}_n - \theta_{\alpha/2}^*]$$

c) Intervalo de confianza percentil. El intervalo percentil bootstrap se define por

$$C_n = [\theta_{(\alpha/2)}^*, \theta_{(1-\alpha/2)}^*]$$

d) Intervalo de confianza bootstrap-t. Se basa en la distribución $T_B = (\hat{\theta}_n - \hat{\theta})/\hat{\sigma}_B$ como una estimación de la distribución de $T = (\hat{\theta} - \theta)/\hat{\sigma}_T$ donde $\hat{\sigma}_T$ es una estimación de la desviación estándar de $\hat{\theta}$. Luego,

$$\{\theta \in \Theta : T_{\alpha/2}^B \leq \frac{\hat{\theta} - \theta}{\hat{\sigma}_T} \leq T_{(1-\alpha/2)}^B$$

$$IC: \;\; [\hat{\theta} - \hat{\sigma}_T T_{(1-\alpha/2)}^B, \hat{\theta} - \hat{\sigma}_T T_{(\alpha/2)}^B]$$

2.2.6. Regiones de confianza

En el caso multiparamétrico, una región de confianza del $100(1-\alpha_1)(1-\alpha_2)$ % para (μ, σ^2) es la región $\mathcal{R}(X)$, donde

$$\mathcal{R}(X) = \{(\mu, \sigma^2) : \bar{X} - a\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + a\frac{\sigma}{\sqrt{n}},$$

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{c} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{b}\}$$

Para n grande es posible contruir una región de confianza aproximada para (μ, σ^2).

$$\mathcal{R}_a = \{(\mu, \sigma^2) : Q \leq \mathcal{X}_{2,1-\alpha}^2\}$$

es una región del $100(1-\alpha)$ %.

2.2.7. Intervalos en poblaciones normales

Sean $X_1, \ldots, X_m$ e $T_1, \ldots, Y_n$ dos muestras aleatorias provenientes de normales independientes, con medias μ_1, μ_2 y varianzas σ_1^2, σ_2^2 respectivamente.

i) IC para diferencia de medias con varianzas iguales conocidas. Un estimador insesgado para $\sigma_1^2, \sigma_2^2 = \sigma^2$ es $S_p^2 = \frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$, donde S_1^2, S_2^2 son las varianzas muestrales de las muestras respectivas.

Tenemos que $\bar{X} \sim N(\mu_1, \sigma^2/m), \bar{Y} \sim N(\mu_2, \sigma^2/n)$ y $\frac{(m-1)S_1^2+(n-1)S_2^2}{\sigma^2} \sim \mathcal{X}_{m+n-2}^2$. También

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, 1)$$

.

De lo anterior se sigue entonces:

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

y luego

$$P(-t_{g,1-\alpha/2} \leq Q \leq t_{g,1-\alpha/2}) = 1 - \alpha$$

donde $t_{g,1-\alpha/2}$ corresponde al percentil $1 - \alpha/2$ de la t-student con $g = m + n - 2$ grados de libertad. Por lo tanto el IC para $\mu_1 - \mu_2$ está dado por

$$[T_1, T_2] = [\bar{X} - \bar{Y} - t_{g,1-\alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}, \bar{X} - \bar{Y} + t_{g,1-\alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}]$$

(Es de longitud minima, joder claro que si).

Otro IC que no está en el ppt pero sirve para varianzas conocidas:

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim N(0, 1)$$

$$IC = [T_1, T_2] = [(\bar{X} - \bar{Y}) - \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} z_{1-\alpha/2}, (\bar{X} - \bar{Y}) + \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} z_{1-\alpha/2}]$$

ii) IC para la diferencia de medias suponiendo varianzas distintas. Se tiene:

$$Q_a = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \overset{\text{ap}}{\sim} t_\nu,$$

donde

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}},$$

denominada aproximación de Welch-Satterthwaite. Luego un IC del $100(1 - \alpha)$

$$[T_1, T_2] = \left[\overline{X} - \overline{Y} - t_{\nu,1-\alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}, \overline{X} - \overline{Y} + t_{\nu,1-\alpha/2} \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}\right].$$

iii) IC para la razón de varianzas.

$$Q_F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1, n-1).$$

Un IC para $\frac{\sigma_1^2}{\sigma_2^2}$ está dado por:

$$[T_1, T_2] = [\frac{S_1^2}{S_2^2}/b, \frac{S_1^2}{S_2^2}/a]$$

donde $a = F_{\alpha/2}(m-1, n-1)$ y $b = F_{1-\alpha/2}(m-1, n-1)$

¿Cómo aplicar intervalos para comparación de poblaciones?

Normalmente preguntan por la diferencia entre la media de alguna característica de dos poblaciones distintas i.e. salario medio, altura promedio, etc. Para luego encontrar un IC del $(1-\alpha)\,\%$. Para afrontar este tipo de problemas:

1. Identificar la información disponible: Te pueden los datos muestrales para calcular $\bar{X}$ y $\bar{Y}$ o tal vez las desviaciones estándar/ calcular la varianza poblacional.

2. Identificar cual de los casos anteriores corresponde utilizar

3. I3

3.1. Prueba de Hipótesis

Una hipótesis estadística, H, es una conjetura acerca de la distribución de una o más variables aleatorias. Una prueba de una hipótesis H es una regla de decisión o procedimiento para decidir si se rechaza o no H. Finalmente, un test D de una hipótesis H se dice un test aleatorio si D está definido por la función $\Psi_D(x) = P(\text{H sea rechazada} \mid \text{x})$. Ψ_D se denomina función crítica de D.

Ejemplo: Sea $X_1, ..., X_{10}$ una ma(n) con $f(x;\theta) = \theta^x(1-\theta)^{1-x}$ para $x = 0, 1$. Si deseamos probar $H : \theta < 1/2$ un posible test es D: rechazar H si $\sum\limits_{i=1}^{10} x_i > 5$ y decidir entre rechazar y aceptar lanzando una moneda si $\sum\limits_{i=1}^{10} x_i = 5$.

D particiona a $\mathcal{X}$ en tres regiones:

$$A = \{(x_1, ..., x_{10}) : \sum_{i=1}^{10} x_i < 5\}$$

$$B = \{(x_1, ..., x_{10}) : \sum_{i=1}^{10} x_i = 5\}$$

$$C = \{(x_1, ..., x_{10}) : \sum_{i=1}^{10} x_i > 5\}$$

La función crítica de D está dada por $\Psi_D(x_1, ..., x_{10}) = \begin{cases} 1;\ \text{si}\ (x_1, ..., x_{10}) \in C \\ 1/2;\ \text{si}\ (x_1, ..., x_{10}) \in B \\ 0;\ \text{si}\ (x_1, ..., x_{10}) \in A \end{cases}$

En muchos problemas se consideran dos hipótesis: la hipótesis a probar, comunmente conocida como hipótesis nula H_0 y la segunda, hipótesis alternativa H_1. Si se rechaza H_0, pero esta es cierta, se denomina error tipo I.
El tamaño error tipo I: $P(\text{rechazar } H_0 | H_0 \text{ es cierta})$
El tamaño error tipo II: $P(\text{no rechazar } H_0 | H_0 \text{ es falsa})$

Definición: Sea D un test de la hipótesis nula H_o con región de rechazo C_D. La función de potencia del test D es una función de θ definida por $\beta_D(\theta) = P(\mathbf{X} \in C_D)$. Es decir, es la probabilidad de que H_0 sea rechazada.

Nivel de significancia

$$\alpha = \sup_{\theta \in \Theta_0} P(X \in C_D)$$

Representa el tamaño del test D con región crítica C_D.

Sea $X_1, ..., X_n \sim N(\theta, 25)$. Considere $H_0 : \theta \leq 17$ vs $H_1 : \theta > 17$ y D: rechazar H_0 si $\bar{x} > 17{+}5/\sqrt{n}$. $\Theta_0 = \{\theta : \theta \leq 17\}$. El tamaño de D es:

$$\begin{aligned}
\alpha == \sup_{\theta \in \Theta_0} P(X \in C_D) &= \sup_{\theta \leq 17} P(X \in C_D) \\
&= \sup_{\theta \leq 17} \{1 - \Phi(\frac{17 - \theta + 5/\sqrt{n}}{5/\sqrt{n}})\} \\
&= 1 - \Phi(\frac{17 - 17 + 5/\sqrt{n}}{5/\sqrt{n}}) \\
&= 1 - \Phi(1) = 0{,}159
\end{aligned}$$

Nota:

i) Note que el tamaño del error tipo II se puede escribir como

$$P(X \notin C_D | H_0 \text{ es falsa}) = 1 - P(X \in C_D | H_0 \text{ es falsa}).$$

ii) La potencia del test se puede escribir como

$$P_\theta(X \in C_D) = \begin{cases} \text{probabilidad de error tipo I} & \text{si } \theta \in \Theta_0 \\ \text{uno menos probabilidad de error tipo II} & \text{si } \theta \in \Theta_0^c \end{cases}$$

Definición: Una región crítica C de tamaño α para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ se dice región crítica insesgada si $\beta(\theta_1) \geq \beta(\theta_0)$ para cada $\theta_1 \in \Theta_1$ y $\theta_0 \in \Theta_0$.

Lema de Neyman-Pearson

Una región crítica $C^\star$ para probar una hipótesis nula simple $H_0 : \theta = \theta_0$ contra una alternativa $H_1 : \theta = \theta_1$ se dice una región crítica más potente de tamaño α si

i) $\beta_{C^\star}(\theta_0) = \alpha$

ii) $\beta_{C^\star}(\theta_1) \geq \beta_C(\theta_1)$ para cualquier otra región crítica C de tamaño α.

Lema: Sean $X_1, \ldots, X_n \sim f(x; \theta)$ con $\Theta = \{\theta_0, \theta_1\}$, $L(\theta)$ la función verosimilitud y sea $\Lambda = \Lambda(\theta_0, \theta_1) = \frac{L(\theta_0)}{L(\theta_1)}$, la razón de verosimilitudes. Entonces, $C^\star = \{x \in \mathcal{X}; \Lambda(\theta_0, \theta_1) = \frac{L(\theta_0)}{L(\theta_1)} \leq k_\alpha\}$, donde k_α es una constante tal que $\beta_{C^\star}(\theta_0) = \alpha$, es una región crítica MP de tamaño α para probar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Corolario: Sean $X_1, \ldots, X_n \sim f(x; \theta)$ con $\Theta = \{\theta_0, \theta_1\}$ y $C^\star$ la región crítica MP de Neyman-Pearson de tamaño α para probar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Entonces

i) $C^\star$ es una región crítica insesgada

ii) $C^\star$ depende de $X_1, \ldots, X_n$ sólo a través de un estadístico suficiente S para θ

Teorema Si $X_1, \ldots, X_n$ son variables aleatorias *iid* $f(x; \theta) = a(\theta)h(x)e^{b(\theta)t(x)}$ en la familia exponencial, entonces la región crítica MP para probar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ es:

i) $C^* = \left\{x \in \mathcal{X} : \sum_{i=1}^n t(x_i) \geq k_\alpha\right\}$ si $b(\theta)$ es creciente y $\theta_0 < \theta_1$,

ii) $C^* = \left\{x \in \mathcal{X} : \sum_{i=1}^n t(x_i) \leq k_\alpha\right\}$ si $b(\theta)$ es decreciente y $\theta_0 < \theta_1$,

iii) $C^* = \left\{x \in \mathcal{X} : \sum_{i=1}^n t(x_i) \leq k_\alpha\right\}$ si $b(\theta)$ es creciente y $\theta_0 > \theta_1$,

iv) $C^* = \left\{x \in \mathcal{X} : \sum_{i=1}^n t(x_i) \geq k_\alpha\right\}$ si $b(\theta)$ es decreciente y $\theta_0 > \theta_1$.

Note que cuando la fdp de $X_1, \ldots, X_n$ pertenece a la familia exponencial uniparamétrica, las regiones críticas de Neyman-Pearson se basan sólo en el estadístico suficiente $\sum_{i=1}^n t(X_i)$.

Ejemplo: Sean $X_1, \ldots, X_n \sim N(0, \theta)$ con $\Theta = \{5, 10\}$. Entonces, para probar $H_0 : \theta = 10$ versus $H_1 : \theta = 5$. Como la dist normal pertenece a la familia exponencial, podemos usar el teorema anterior para encontrar la forma de la región crítica MP con $b(\theta) = -\frac{1}{2\theta}$ y $t(x) = x^2$.

Ya que $B(\theta)$ es creciente y $\theta_0 > \theta_1$ la región crítica MP para probar $H_0 : \theta = 10$ vs $H_1 : \theta = 5$ es $C^\star = \{x \in \mathcal{X} : \sum_{i=1}^n x_i^2 \leq k_\alpha\}$. Ahora, $\frac{1}{\theta} \sum_{i=1}^n X_i^2 \sim \chi_n^2$ cuando $X_1, X_2, \ldots, X_n$ son $\sim N(0, \theta)$. Así,

$$P(C^\star|\theta) = P(\sum_{i=1}^n X_i^2 \leq k_\alpha|\theta) = P(\frac{1}{\theta} \sum_{i=1}^n X_i^2 \leq \frac{k_\alpha}{\theta}|\theta)$$

$$= P(\chi_n^2 \leq \frac{k_\alpha}{\theta})$$

Luego, para $\theta = 10$,

$$P(C^\star|\theta = 10) = P(\frac{1}{10} \sum_{i=1}^n X_i^2 \leq \frac{k_\alpha}{10}|\theta = 10)$$

$$= P(\chi_n^2 \leq \frac{k_\alpha}{10})$$

Para $n = 50$ y $\alpha = 0{,}01$, $\frac{1}{10} \sum_{i=1}^{50} X_i^2 \sim \chi_{50}^2$, y por lo tanto, el valor de $k_{0,01} = 10 \times qchisq(0{,}01, 50) = 297{,}07$. Así, la región crítica MP de tamaño $\alpha = 0{,}01$ para probar H_0 es

$$C^\star = \{x \in \mathcal{X} : \sum_{i=1}^{50} x_i^2 \leq 297{,}07\}$$

La potencia de $C^\star$ cuando $n = 50$ y $\alpha = 0{,}01$ es

$$P(C^\star|\theta = 5) = P(\sum_{i=1}^n X_i^2 \leq 297{,}07|\theta = 5) = P(\frac{1}{5} \sum_{i=1}^n X_i^2 \leq \frac{297{,}07}{5}|\theta = 5)$$

$$= P(\chi_n^2 \leq 59{,}414) = 0{,}830$$

3.2. Tests UMP

Una región crítica $C^\star$ para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ se dice uniformemente más potente, UMP, de tamaño α si

- $\sup_{\theta \in \Theta_0} \beta_{C^\star}(\theta) = \alpha$
- $\beta_{C^\star}(\theta) \geq \beta_C$ para todo $\theta \in \Theta_1$ para cualquier otra región crítica C de tamaño α.

Una región crítica UMP puede existir o no para un modelo estadístico uni-paramétrico; sin embargo, bajo ciertas condiciones, existe una región crítica UMP para una prueba unilateral.

Definición: Para una muestra aleatoria $X_1, \ldots, X_n$, una función de verosimilitud $L(\theta)$ se dice que tiene una razón de verosimilitud monótona, RVM, en un estadístico T si para cada $\theta_1 < \theta_2$ la razón de verosimilitud $\Lambda(\theta_1, \theta_2) = \frac{L(\theta_1)}{L(\theta_2)}$ es una función no decreciente o no creciente de T.

Sea $X_1, \ldots, X_n \sim U(0, \theta)$ con $\Theta = \mathbb{R}^+$. Entonces,

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} I_{0,\theta}(x_i) = \begin{cases} \theta^{-n} & \text{cuando } 0 < x_i < \theta, \quad \forall i \\ 0 & eoc \end{cases}$$

$$= \theta^{-n} I_{(0,\theta)}(t) \text{ con t = maximo}\{x_1, \ldots, x_n\}.$$

Así cuando $\theta_2 > \theta_1 > 0$:

$$\Lambda(\theta_1, \theta_2) = \frac{L(\theta_1)}{L(\theta_2)} = \frac{\theta_1^{-n} I_{(0,\theta_1)}(t)}{\theta_2^{-n} I_{(0,\theta_2)}(t)} = \begin{cases} (\theta_2/\theta_1)^n & \text{cuando } 0 < t < \theta_1 \\ 0 & \text{cuando } \theta_1 \leq t < \theta_2 \end{cases}$$

Por lo tanto, $\Lambda(\theta_1, \theta_2)$ es una función no decreciente de t, y luego $L(\theta)$ tiene una RVM en $T = \text{máximo}\{X_1, \ldots, X_n\}$.

Teorema: Si $X_1, \ldots, X_n \sim f(x;\theta) = a(\theta)h(x)e^{b(\theta)t(x)}$, perteneciente a la familia exponencial, entonces $L(\theta)$ tiene una RVM en el estadístico suficiente $\sum_{i=1}^{n} t(X_i)$.

Teorema de Karlin-Rubin

Sea $\mathbf{X} = (X_1, \ldots, X_n) \sim f(x;\theta)$ una fdp uni-paramétrica y espacio paramétrico Θ, y función de verosimilitud $L(\theta)$ con RVM en un estadístico $T(X)$.

1. Si $L(\theta)$ tiene una RVM no decreciente en T, entonces una región crítica UMP de tamaño α para probar $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ es $C^\star = \{x \in \mathcal{X} : T(x) \leq k_\alpha\}$

2. Si $L(\theta)$ tiene RVM no creciente en T, entonces una región crítica de tamaño α para probar $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ es $C^\star = \{x \in \mathcal{X} : T(x) \geq k_\alpha\}$.

El teorema también se puede utilizar para hipótesis de cola inferior, $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$, con las desigualdades invertidas en las regiones críticas UMP. Por ejemplo, para la hipótesis anterior cuando la función de verosimilitud tiene RVM no creciente en T, la región crítica UMP es $C^\star = \{x \in \mathcal{X} : T(x) \leq k_\alpha\}$.

Cuadro 6: Estadísticos suficientes en distintas familias

Distribución	**Parámetro(s)**	**Estadístico suficiente mínimo**
Normal $\mathcal{N}(\mu, \sigma^2)$ (σ^2 conocida)	μ	$\bar{x}$
Normal $\mathcal{N}(\mu, \sigma^2)$ (σ^2 desconocida)	μ, σ^2	$(\bar{x}, \sum(X_i - \bar{x})^2)$
Bernoulli / Binomial	p	$\sum X_i$
Poisson	λ	$\sum X_i$
Exponencial	θ	$\sum X_i$
Uniforme $U(0,\theta)$	θ	$\text{máx}(X_1, \ldots, X_n)$
Gamma $\Gamma(k,\theta)$ (k fijo)	θ	$\sum X_i$

Sean $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$. Dado que la distribución pertenece a la familia exponencial y $\sum X_i$ es un estadísitco suficiente, $L(\theta)$ tiene RVM en $\sum X_i$. Además, como $b(\theta) = \theta$ es una función creciente, $L(\theta)$ tiene RVM no creciente en $\sum X_i$. Luego, por el teorema de Karlin-Rubin, la región crítica UMP de tamaño α para probar $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$ es $C^\star = \{x \in \chi : \sum x_i \geq k_\alpha\}$

Sea $X_1, \ldots, X_n \overset{iid}{\sim} N(0, \theta)$. Encuentre un test UMP de tamaño α para probar $H_0 : \theta \leq 1$ vs $H_1 : \theta > 1$. El estadísitco suficiente es $T = \sum X_i^2$. Dado que la función de verosimlitud es no creciente en T, se tiene que $C^\star = \{x \in \chi : T \geq k_\alpha\}$ con $\alpha = P(T \geq k_\alpha | H_0)$ tal que $k_\alpha = \chi^2_{n,1-\alpha}$

Sea $X_1, \ldots, X_{20}$ una muestra aleatoria desde una distribución Poisson. Muestre que el test D : rechace H_0 si $\sum x_i \geq 5$ es UMP para probar $H_0 : \lambda = 1/10$ vs $H_1 : \lambda > 1/10$. ¿Cuál es la probabilidad de error tipo I?. El test es equivalente a $H_0 : \lambda = 1/10$ vs $H_1 : \lambda > 1/10$. Se tiene que Λ es no creciente en $T = \sum x_i$, por lo que la región UMP es de la forma $T \geq c$. El error tipo I corresponde a $P(T \geq c | H_0)$ y se sabe que $T \sim Poisson(n\lambda) = Poisson(2)$.

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(\theta, 16)$. Para encontrar el tamaño de la muestra n y un test D UMP, para probar $H_0 : \theta = 25$ vs $H_1 : \theta < 25$, tal que $\beta_D(25) = 0{,}1$ y $\beta_D(23) = 0{,}9$, se tiene que la RVM es no creciente en $T = \bar{x}$ por lo que la región es de la forma $C^\star = \{\bar{X} \geq c\}$. Para encontrar c y n, se tiene que $\bar{X} \sim N(\theta, \frac{\sigma^2}{n})$ tal que $P(\bar{x} \geq c|\theta) = P(\frac{\sqrt{n}(\bar{x}-\theta)}{\sigma} \geq \frac{\sqrt{n}(c-\theta)}{\sigma}|\theta) = P(Z \geq \frac{\sqrt{n}(c-\theta)}{\sigma}|\theta)$. De lo anterior se tiene

Sea $X_1 \ldots, X_n$ una muestra aleatoria dedsde una distribución con densidad dada por

$$f(x) = \begin{cases} (m/\theta)x^{m-1}\exp(-x^m/\theta) & x > 0 \\ 0 & e.o.c \end{cases}$$

donde m es una constante. Para esta distribución se sabe que $2\sum X_i^m/\theta \sim \chi^2(2n)$. Para encontrar un test UMP, D, para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, se tiene que la RVM es no creciente en $T = \sum X_i^m$. Es decir, el test

UMP tiene región crítica dada por $C^\star = \{T \geq c\}$. Si para el test anterior se tiene que $\theta_0 = 100$, la probabilidad de error tipo I es 0.05 y $\beta_D(400) = 0{,}95$, entonces, para obtener el tamaño de la muestra y la región crítica se tiene que

$$P(T \geq c|\theta = 100) = 0{,}05 \implies P(\frac{2T}{100} \geq \frac{2c}{100}) = 0{,}05 \implies \frac{2c}{100} = \chi^2_{2n,0{,}95}$$

Además, se tiene que

$$P(T \geq c|\theta = 400) = 0{,}95 \implies P(\frac{2T}{400} \geq \frac{2c}{400}) = 0{,}05 \implies \frac{2c}{400} = \chi^2_{2n,0{,}05}$$

Definición: Una región crítica $C^\star$ es UMP insesgada de tamaño α para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ si

1. $\sup_{\theta\in\Theta_0} \beta_{C^\star}(\theta) = \alpha$
2. $C^\star$ es una región crítica insesgada.
3. $\beta_{C^\star}(\theta) \geq \beta_C(\theta)$ para cada $\theta \in \Theta_1$ y para cualquier otra región crítica insesgada de tamaño α.

Teorema: Si $X_1, \ldots, X_n \sim f(x; \theta) = a(\theta)h(x)e^{\theta t(x)}$ para $\theta \in \Theta$, entonces para probar $H_0; \theta = \theta_0$ vs $H_1; \theta \neq \theta_0$ (prueba bilateral) existe una región crítica UMP insesgada de tamaño α de la forma,

$$C^\star = \{x \in \mathcal{X} : \sum_{i=1}^{n} t(x_i) \leq k_{\alpha_1} \ \vee \sum_{i=1}^{n} t(x_i) \geq k_{\alpha_2}\},$$

donde $P(\sum_{i=1}^{n} t(x_i) \leq k_{\alpha_1}|\theta_0) = \alpha_1$, $P(\sum_{i=1}^{n} t(x_i) \geq k_{\alpha_2}|\theta_0) = \alpha_2$ y $\alpha_1 + \alpha_2 = \alpha$.

Test de la razón de verosimilitud

Existen parámetros que tenemos que considerar pero que no son de nuestro interés estimar, se le dicen parámetros molestos. Un test que los utiliza es el test de la razón de verosimilitud generalizada, RVG.

Sea $X_1, \ldots, X_n$, una ma con función de verosimilitud $L(\theta)$. Para probar $H_0 : \theta \in \Theta$ vs $H_1 : \theta \in \Theta_1$, la RVG se define como $\Lambda(x) = \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta_1} L(\theta)}$.

La RVG se puede escribir como $\Lambda(x) = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$ donde $\hat{\theta}_0$ corresponde al EMV de θ bajo Θ_0 y $\hat{\theta}$ es el EMV sobre Θ (EMV no restringido). Ya que $\hat{\theta}$ maximiza la función de verosimilitud, se tiene que $L(\hat{\theta}_0) \leq L(\hat{\theta})$, y por lo tanto, $0 \leq \Lambda(x) \leq 1$. Con el test RVD, se **rechaza** H_0 cuando el valor observado de Λ es pequeño.

Definición: Sea $X_1, \ldots, X_n$ una muestra de va iid con función de verosimilitud $L(\theta)$, la región crítica del test de la RVG, de tamaño α, para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ está dada por

$$C_\Lambda = \{x \in \mathcal{X} : \Lambda(x) \leq k_\alpha\}$$

Sean $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\theta)$ con $\Theta = \mathbb{R}^+$. Para probar $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$ el test de ka RVG tiene región crítica dada por $C_\Lambda = \{x : \Lambda(x) \leq k_\alpha\}$. El EMV de θ sin restricciones es $\hat{\theta} = \hat{x}$ y el EMV bajo H_0 es $\hat{\theta}_0 = 1$. Aquí $\Theta_o = \{1\}$. Así

$$\Lambda(x) = \frac{L(\hat{\theta}_0)}{L(\bar{x})} = \frac{\prod e^{-x_i}}{\prod \frac{1}{\bar{x}} e^{\frac{-x_i}{\bar{x}}}}$$

Nota: Si $X_1, \ldots, X_n$ es una muestr iid uniparamétrica el RVG para porbar hipótesis simples i.e: $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ es equivalente al test MP de Newyman-Pearson.

Teorema: Sea $X_1, \ldots, X_n$ una muestra de v.a iid cuya fdp satisface las condiciones de regularidad. Para probar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, si $\theta \in \Theta_0$ entonces $T_{RV}(X) = -2\ln\Lambda_n(X) \xrightarrow{d} \chi^2_r$, donde $r = dim(\Theta) - dim(\Theta_0)$, y $\Lambda_n(X)$ es el test de la RVG para una muestra de tamaño n.

Nota:

- La región crítica del test de la RGV, para muestras grande, de tamaño α, está dada por

$$C_\Lambda = \{x \in \mathcal{X} : -2\ln\Lambda_n(x) \geq k_\alpha\}$$

- Notemos que $T_{RV}(x) - 2\ln\Lambda_n(x) = 2\{\mathcal{L}(\hat{\theta}) - \mathcal{L}(\hat{\theta}_0)\}$

Cuadro 7: Test de la RVG para probar $H_0 : \mu = \mu_0$ con nivel de significancia α. Con $t = \sqrt{n}(\bar{x} - \mu_0)/s$

Hipótesis alternativa	**Criterio de rechazo**
$H_1 : \mu \neq \mu_0$	Rechazar H_0 cuando $\lvert t\rvert \geq t_{n-1,1-\alpha/2}$
$H_1 : \mu > \mu_0$	Rechazar H_0 cuando $t \geq t_{n-1,1-\alpha}$
$H_1 : \mu < \mu_0$	Rechazar H_0 cuando $t \leq t_{n-1,\alpha}$

Cuadro 8: Prueba de hipótesis para $H_0 : \sigma^2 = \sigma_0^2$ con nivel de significancia α, con $u = \frac{(n-1)s^2}{\sigma_o^2}$

Hipótesis alternativa	**Criterio de rechazo**
$H_1 : \sigma^2 \neq \sigma_0^2$	Rechazar H_0 cuando $u \leq \chi^2_{n-1,\alpha/2}$ o $u \geq \chi^2_{n-1,1-\alpha/2}$
$H_1 : \sigma^2 > \sigma_0^2$	Rechazar H_0 cuando $u \geq \chi^2_{n-1,1-\alpha}$
$H_1 : \sigma^2 < \sigma_0^2$	Rechazar H_0 cuando $u \leq \chi^2_{n-1,\alpha}$

Test de Wald

Sea $X = (X_1, \ldots, X_n)$ una muestra de va iid con fdp común $f(x;\theta)$ donde $\theta \in \mathbb{R}^k$. Sea $\mathcal{I}_x(\theta) = n\mathcal{I}_x(\theta)$ la matriz de información de Fisher y $\hat{\theta}_n$ el EMV de θ. Bajo las condiciones de regularidad, se tiene,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N_K(0, \mathcal{I}_X^{-1}(\theta))$$

O sea, para muestras grandes $\hat{\theta}_n - \theta \overset{\text{aprox}}{\sim} N_k(0, \Sigma_\theta)$, donde $\Sigma_\theta = \mathcal{I}_X^{-1}(\theta)$, se sigue que $(\hat{\theta}_n - \theta)^T \hat{\Sigma_\theta}^{-1}(\hat{\theta}_n - \theta) \overset{\text{aprox}}{\sim} \chi^2_k$.
Consideremos la hipótesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. El estadístico para el test de Wald se define por

$$T_W(X) = (\hat{\theta}_n - \theta_0)^T \hat{\Sigma}_\theta^{-1}(\hat{\theta}_n - \theta_0)$$

También es común utilizar:

$$T_W = \frac{(\hat{\theta} - \theta_0)^2}{Var(\hat{\theta})}$$

y la región crítica está dada por

$$C_W = \{x \in \mathcal{X} : T_W(x) > \chi^2_{k,1-\alpha}\}$$

Nota:

i) Note que la distribución asintótica de los tests de la RVG y de Wald coinciden, y en algunos casos estos estadísticos tienen distribuciones exactas.

ii) Para probar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, en el caso escalar ($k = 1$), la región crítica para el test de Wald es:

$$C_W = \left\{ x \in \mathcal{X} : \left| \frac{\hat{\theta} - \theta_0}{\sqrt{\hat{V}}} \right| > z_{1-\frac{\alpha}{2}} \right\}$$

donde $\hat{V}$ es un estimador consistente de la varianza de $\hat{\theta}$ y z_p es el percentil $100p$ % de una $N(0, 1)$.

Test Score

Otro test para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, que tiene la misma distribución asintótica que los test de Wald y RVG, es el test Score de Rao. Sea $X = (X_1, \ldots, X_n)$ una muestra de va iid con fdp $f(x;\theta)$ donde $\theta \in \mathbb{R}^k$. Sean $\mathcal{L}(\theta)$ y $\mathcal{U}(\theta)$ la log-verosimilitud y la función score. El test se basa en el estadístico

$$T_S(X) = \mathcal{U}^T(\hat{\theta}_0)\mathcal{I}_x^{-1}\mathcal{U}(\hat{\theta}_0)$$

Se suele utilizar:

$$T_S = \frac{\mathcal{U}(\theta_0)^2}{\mathcal{I}_x(\theta_0)}$$

Bajo las condiciones de regularidad se tiene que $T_S(X) \overset{\text{aprox}}{\sim} \chi^2_k$, para muestras grandes, y la región crítica del test está dada por

$$C_S = \{x \in \mathcal{X} : T_S(x) > \chi^2_{k,1-\alpha}\}$$

Test Gradiente

Otro test para probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, que tiene la misma distribución asintótica que los test de Wald, Score y RVG, es el test Gradiente. Sea $X = (X_1, \ldots, X_n)$ una muestra de va iid con fdp $f(x; \theta)$ donde $\theta \in \Theta \subseteq \mathbb{R}^k$. Sea $\mathcal{U}(\theta)$ la función score. El test se basa en el estadístico

$$T_{Gr}(X) = \mathcal{U}^T(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0)$$

Bajo las condiciones de regularidad se tiene que $T_{Gr}(X) \overset{\text{aprox}}{\sim} \chi^2_k$, para muestras grandes, y la región crítica del test está dada por

$$C_{Gr} = \{x \in \mathcal{X} : T_{Gr}(x) > \chi^2_{k,1-\alpha}\}$$

Valor - p

Sea $X = (X_1, \ldots, X_n) \overset{iid}{\sim} f(x; \theta), \theta \in \Theta$ y $H_0 : \theta \in \Theta$. El valor-p asociado con la muestra X y una región crítica de tamaño α, C_α, es el valor más pequeño de α para el cual se rechaza H_0, es decir

$$p(x) = inf\{\alpha \in (0, 1) : x \in C_\alpha\}$$

Consideremos una región crítica de la forma $C_\alpha = \{x \in \mathcal{X} : T(x) \geq k_{1-\alpha}\}$. Sea G_0 la fda de $T(X)$ bajo H_0. Entonces

$$T(x) \geq k_{1-\alpha} \implies G_0(T(x)) \geq G_0(k_{1-\alpha}) = 1 - \alpha$$

o sea $\alpha \geq 1 - G_0(T(x))$. Por lo tanto el valor-p asociado es

$$P(x) = 1 - G_0(T(x)) = 1 - P(T \leq T_{obs}) = P(T > T_{obs})$$

Nombre del Test	Estadístico / Fórmula	Condiciones de Aplicabilidad
Z-Test para media	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	Muestra normal o grande ($n \geq 30$), varianza poblacional σ^2 conocida
t-Test para media	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	Muestra pequeña ($n < 30$), población normal, varianza desconocida
Test para varianza (Chi-cuadrado)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	Datos normales, hipótesis sobre la varianza poblacional
Test F para comparación de varianzas	$F = \frac{S_1^2}{S_2^2}$	Dos muestras normales independientes, para comparar varianzas
Z-Test para diferencia de medias	$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	Varianzas poblacionales conocidas, muestras normales o grandes, independientes
t-Test para diferencia de medias	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$	Varianzas desconocidas pero iguales, muestras normales independientes
Test de Welch (t para varianzas distintas)	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	Muestras normales independientes, varianzas desconocidas y distintas
Test de proporciones	$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	Datos binomiales, muestra grande, prueba sobre proporción
Test para diferencia de proporciones	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$	Datos binomiales, muestra grande, prueba sobre dos proporciones
Test de bondad de ajuste χ^2	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$	Frecuencias observadas vs esperadas, cada $E_i \geq 5$ preferentemente
Test de independencia χ^2	$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$	Tablas de contingencia, prueba de independencia entre variables categóricas
Test de Kolmogorov-Smirnov	$D_n = \sup_x \lvert F_n(x) - F_0(x) \rvert$	Comparación de distribución empírica con una teórica, datos continuos
Test de Anderson-Darling	Basado en $A^2 = -n - \frac{1}{n}\sum(2i-1)[\log F(X_i) + \log(1 - F(X_{n+1-i}))]$	Mayor sensibilidad en las colas que K-S, para bondad de ajuste
Test de Wilcoxon de una muestra	Suma de rangos positivos/negativos	Datos ordinales o no normales, prueba sobre la mediana
Test de Mann-Whitney (U-test)	$U = \text{mín}(U_1, U_2)$, basado en rangos	Comparación de dos muestras independientes sin suponer normalidad
Test de signos	Cuenta de signos positivos y negativos respecto a una mediana	Distribución simétrica, alternativa no paramétrica al t-test
Test de razón de verosimilitud (LRT)	$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$	Aplicable en modelos paramétricos generales, prueba poderosa asintótica

Regresión lineal

Especificación del modelo Se quieren escribir n ecuaciones para n observaciones, tal que la variables de respuesta y sea una combinación líneal de los predictores x:

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$$

Definiendo los vectores y matriz

$$y = \begin{pmatrix} y_i \\ \vdots \\ y_n \end{pmatrix} \quad y \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ . & \dots & . \\ . & \dots & . \\ x_{n1} & \dots & x_{np} \end{pmatrix}$$

De esta forma, en notación matricial

$$y = X\beta + \epsilon$$

Supuestos modelo de regresión lineal clásico:

1. Esperanza de los errores: Los errores tienen media cecro, es decir, $\mathrm{E}(\epsilon_i) = 0$.

2. Varianzas y estructuras de correlación de los errores. Se asume una varianza de error constante tal que $\text{Var}(\epsilon_i) = \sigma^2$.
3. Los errores y las covariables son independientes. Por ejemplo, $\text{E}(\epsilon|X) = 0$ o $\text{Cov}(\epsilon|X)=\sigma^2 I_n$.
4. Errores gaussianos. Se asume una distribución normal para los errores, es decir, $\epsilon_i \sim N(0, \sigma^2)$ o en notación matricial $\epsilon \sim N_n(0, \sigma^2 I_n)$.

Modelo Lineal Clásico
El modelo

$$y = X\beta + \epsilon$$

se denomina modelo de regreasión lineal clásico, si se cumplen las siguientes suposiciones:

1. $\text{E}(\epsilon)=0$
2. $\text{Cov}(\epsilon) = \text{E}(\epsilon\epsilon^T)=\sigma^2 I_n$
3. La matriz X es de rango completo, es decir, $r(X) = p$.
4. $\epsilon \sim N_n(0, \sigma^2 I_n)$

Mínimos Cuadrados
El método de mínimos cuadrados es útil para obtener un estimador de β. Este método consiste en minimizar $\sum \epsilon_i^2$ respecto a β. Es decir, se minimiza la suma de cuadrados del error:

$$\text{MC}(\beta) = \epsilon^T\epsilon = (y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta) = ||y - \mathbf{X}\beta||$$

Notar que:

$$\text{MC}(\beta) = (y - \mathbf{X}\beta)^T(y - \mathbf{X}\beta) = y^T y - 2y^T X\beta + \beta^T X^T X\beta$$

Ahora, se puede minimizar el MC haciendo el vector de primeras derivadas con respecto a β igual a 0 y mostrando que la matriz de segundas derivadas es definida positiva.

$$\frac{\partial \text{MC}}{\partial \beta} = -2X^T y + 2X^T X\beta$$

$$\frac{\partial^2 MC}{\partial\beta\partial\beta} = 2X^T X$$

Como las columnas de la matriz X son linealmente independientes, la matriz $X^T X$ es definida positiva. Así, se obtiene que el estimador de mínimos cuadrados de $\hat{\beta}$ es

$$(X^T X)\hat{\beta} = X^T y$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Usando el estimador de MC $\hat{\beta}$, se puede estimar la media (condicional) de y por

$$\hat{\text{E}}(y) = \hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy$$

Donde

$$H = X(X^T X)^{-1} X^T$$

Los elementos del vector

$$y - \hat{y} = y - X\hat{\beta} = (I_n - H)y$$

son denominados resíduos y se denotan por $\boldsymbol{e}$. El valor mínimo de $\boldsymbol{e}^T\boldsymbol{e}$ es
Teorema: Propiedad del EMC. Sea $y = X\beta + \epsilon$, donde X es de rango completo, $\text{E}(\epsilon) = 0$ y $\text{Cov}(\epsilon) = \sigma^2 I_n$. Entonces, el EMC $\hat{\beta} = (X^T X)^{-1} X y$ tiene las siguientes propiedades

i) $\text{E}(\hat{\beta}) = \beta$, es decir $\hat{\beta}$ es un estimador insesgado de β.
ii) $\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.
iii) Entre todos los estimadores lineales insesgados $\hat{\beta}_L$, el EMC tiene varianza mínima.

4. Pautas Ies 2025-1

I1

Estimación de parámetros

La vida útil de una bombilla tiene la función de densidad de probabilidad:

$$f(x;\theta) = \theta^2 x \exp(-\theta x);$$

para $x > 0$, donde θ es una constante positiva desconocida con $E(X) = 2/\theta$ y $Var(X) = 2/\theta^2$.

i) Demuestre que el valor esperado de X es el doble de la moda.

$$\rightarrow \log f(x;\theta) = 2\log(\theta) + \log(x) - x\theta$$

$$\frac{\partial \log}{\partial x} = \frac{1}{x} - \theta = 0 \implies x = \frac{1}{\theta} = Mo = 2E(X)$$

ii) Encuentre e interprete la cota de Cramer - Rao para estimadores insegados de la moda.

La CCR es $\frac{\tau'(\theta)^2}{\mathcal{I}_x(\theta)}$; $\tau(\theta) = \frac{1}{\theta}, \tau'(\theta) = -\frac{1}{\theta^2}$. Ahora,

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = \frac{2}{\theta} - x = u(\theta) \implies V(u(\theta)) = V(X) = \frac{2}{\theta^2} \rightarrow n \cdot \frac{2}{\theta^2} = \mathcal{I}_x(\theta)$$

$$CCR = \frac{[1/\theta^4]}{2n/\theta^2} = \frac{1}{2n\theta^2}$$

iii) Encuentre un estimador eficiente de la moda y estime su varianza.

El EMV satisface $\sum_{i=1}^{n}(\frac{2}{\theta^2} - x_i) = 0$.

$$\rightarrow \hat{\theta} = \frac{2}{\bar{X}}, \hat{Mo} = \frac{\bar{X}}{2} \implies E(\hat{Mo}) = Mo$$

$$V(\hat{Mo}) = \frac{1}{4}V(\bar{X}) = \frac{1}{2n\theta^2} = CCR$$

$$V(\hat{Mo}) = \frac{1}{2n(2/\bar{X})^2} = \frac{\bar{X}^2}{8n} = 6582{,}82$$

Estimadores de Bayes

Sea $x = X_1, \ldots, X_n$ una muestra de va iid, $N(\theta, 1)$ con $\Theta = \mathbb{R}$. Suponga que $\theta \sim N(x_0, 1)$. En este caso la distribución posterior de θ es también normal con media $\sum_{i=0}^{n} x_i/(n+1)$ y varianza $1/(n+1)$. Encuentre:

i) Estimador de Bayes de θ para la pérdida cuadrática.

El estimador de Bayes para pérdida cuadrática es la media posterior:

$$\hat{\theta}_B = \frac{\sum_{i=0}^{n} x_i}{n+1}$$

ii) Estimador de Bayes θ para pérdida absoluta.

Para pérdida absoluta, el estimador es la mediana posterior (que coincide con la media por ser normal simétrica):

$$\hat{\theta}_B = \frac{\sum_{i=0}^{n} x_i}{n+1}$$

iii) Estimador de Bayes de $r(\theta) = 4\theta + \theta^2$ para pérdida cuadrática.

El estimador de Bayes para $r(\theta)$ es:

$$\hat{r}(\theta) = 4E(\theta/X) + E(\theta^2/X) = 4E(\theta/X) + V(\theta/X) + (E(\theta/X))^2$$

$$= 4\sum_{i=0}^{n} x_i/(n+1) + \frac{1}{n+1} + [\sum_{i=0}^{n} x_i/(n+1)]^2$$

Estimadores de momentos e información de Fischer

Se pone a prueba n aparatos eléctricos y se observa cuantas semanas tardan en fallar. Sea $X = X_1, \ldots, X_n$ la muestra aleatoria correspondiente la cual asumimos que proviene de una $Gamma(\alpha, \beta)$ con densidad

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp\{-x/\beta\}, \quad x > 0.$$

$E(X) = \alpha\beta$ y $Var(X) = \alpha\beta^2$.

i) Determine estimadores de momentos de los parámetros y estime la vida útil media de la población.

Igualamos momentos:

$$\bar{X} = \alpha\beta, \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \alpha\beta^2 + \alpha^2\beta^2$$

$$\alpha = \bar{X}/\beta. \text{ Luego, } \tilde{\beta} = S_n^2/\bar{X},\ S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\ \tilde{\alpha} = \bar{X}^2/S_n^2.$$

ii) Sea $\mathcal{I}_x(\alpha, \beta)$ la información de Fischer contenida en la muestra X. Se propone nueva parametrización de modo que $E(X) = \mu$ y $Var(X) = \mu\phi$; encuentre $\mathcal{I}_x(\mu, \phi)$.

Tenemos $\theta^\top = (\alpha, \beta) = (\mu/\phi, \phi) = (h_1, h_2)$.

$$A = \begin{bmatrix} \frac{\partial h_1}{\partial \mu} & \frac{\partial h_1}{\partial \phi} \\ \frac{\partial h_2}{\partial \mu} & \frac{\partial h_2}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \frac{1}{\phi} & \frac{-(1)\mu}{\phi^2} \\ 0 & 1 \end{bmatrix}$$

Luego $\mathcal{I}_x(\mu, \phi) = A^\top \mathcal{I}_x(h(\gamma))A$

EMV e insesgamiento

Supongamos dos observaciones independientes X_1, X_2 provenientes de una función discreta

$$f(x; \theta) = (\frac{\theta}{2})^{|x|}(1-\theta)^{1-|x|}, \quad x = -1, 0, 1; 0 \leq \theta \leq 1$$

- **Especificación del modelo estadístico:**

  Aqui $F = \{f(x_1, x_2; \theta\}, 0 \leq \theta \leq 1$, donde $f(x_1, x_2; \theta) = (\frac{\theta}{2})^{(|x_1|+|x_2|)}(1-\theta)^{2-(|x_1+x_2|)} = (\frac{\theta}{2})^t(1-\theta)^{2-t}$

- **Muestre que EMV de θ es $T = \hat{\theta} = (|X_1)| + |X_2|)/2$.**

$$\mathcal{L}(\theta) = \log f(x_1, x_2; \theta) = t\log(\theta/2) + (2-t)\log(1-\theta)$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \frac{2t}{\theta} \cdot \frac{1}{2} + \frac{2-t}{1-\theta}(-1) = \frac{t}{\theta} - \frac{2-t}{1-\theta} = 0 \implies \hat{\theta} = \frac{t}{2}$$

Note que $\hat{\theta}$ toma valores en $\{0, \frac{1}{2}, 1\}$, con probabilidades:

$$P(\hat{\theta} = 0) = P(0, 0) = (1-\theta)^2$$

$$P(\hat{\theta} = \frac{1}{2}) = P(\{(0,-1),(0,1),(-1,0),(1,0)\} = 4 \cdot \frac{\theta(1-\theta)}{2} = 2\theta(1-\theta)$$

$$P(\hat{\theta} = 1) = P(\{(-1,-1),(-1,1),(1,-1),(1,1)\}) = 4\frac{\theta^2}{4} = \theta^2$$

Luego $E(\hat{\theta} = \theta(1-\theta) + \theta^2 = \theta$, sesgo = 0. $E(\hat{\theta}^2) = \frac{\theta^2}{2} + \frac{\theta}{2}, V(\hat{\theta}) = \theta(1-\theta)/2$.

- **Demuestre que $S = I(X_1 = 1) + I(X_2 = 1)$, donde $I(\cdot)$ es la función indicadora es insesgado para θ. Luego demuetre que su varianza es mayor que la de $\hat{\theta}$.**

  Notemos
  $$f(s;\theta) = \begin{cases} \theta^2/4 & ; s = 2 \\ \theta^2/2 + \theta(1-\theta) & ; s = 1 \\ (1-\theta)^2 + \theta(1-\theta) + \theta^2/4 & ; s = 0 \end{cases}$$

  Luego $E(S) = \frac{\theta^2}{2} + \frac{\theta^2}{2+\theta(1-\theta)=\theta}$. $\therefore$ S es un E.Insesgado de θ.

  $E(S^2) = \frac{\theta^2}{2} + \theta \implies V(S) = \frac{\theta^2}{2} + \theta - \theta^2 = \frac{1}{2}\theta(2-\theta)$, luego $V(S) = V(\hat{\theta}) + \frac{\theta}{2}$ y $V(S) > V(\hat{\theta})$.

Estimadores insesgados

El número de fallas semanaes de un sistema informático es una variable aleatoria X con distribución Poisson con media μ. Se dispone de una muestra $X_1, \ldots, X_n$. El costo semanal de reparar las fallas es de $C = 3X + X^2$

- **Suponga $n = 2$ y el estimador insesgado para $\mu, T = (X_1 + 2X_2)/3$, ¿Es T un estadístico suficiente y eficiente para μ?**

  T no es suficiente ya que

  $$P(X_1 = 0, X_2 = 1 | X_1 + 2X_2 = 2) = \frac{P(X_1 = 0, X_2 = 1)}{P(X_1 = 0, X_2 = 1) + P(X_1 = 1, X_2 = 0)}$$

  $$= \frac{e^{-\mu}e^{-\mu}\mu}{e^{-\mu}e^{-\mu} \cdot \mu + e^{-\mu}e^{-\mu}\frac{\mu^2}{2!}} = \frac{2}{\mu + 2}$$

  , que depende de μ. Tampoco es eficiente ya que su varianza $V(T) = \frac{1}{9}(\mu + 4\mu) = \frac{5}{9}\mu$ no alcanza la CCR $= \mu/2$.

- **Sugiera estimador insesgado para μ basado en un estadístico suficiente**. Es eficiente?

  $\bar{X}$ es un E.I para μ basado en un estadístico suficiente. También es ediciente ya que $V(\bar{X}) = \mu/n = a$, la CCR $= \mathcal{I}_x^{-1}(\mu) = \mu/n$.

- **Demuestre que $E(C) = 4\mu + \mu^2$ y encuentre estimador insesgado para $E(C)$. Encuentre el menor valor que puede alcanzar la varianza de estimadores insesgados de $4\mu + \mu^2$.**

  $E(C) = 3E(X) + E(X^2) = 4\mu + \mu^2$. El EMV de $E(C)?\tau(\mu)$ es $\tau(\bar{X}) = 3\bar{X} + \bar{X}^2$. Ahora $E(\tau(\bar{X})) = E\{3\bar{X} + \bar{X}\} = 4\mu + \frac{\mu}{n} + \mu^2 \neq 4\mu + \mu^2$. SI¿in embargo el estimador $4\bar{X} + \bar{X}^2 - \bar{X}/n$ es insesgado

  $$E(\bar{X} + \bar{X}^2 - \bar{X}/n) = 4\mu + \frac{\mu}{n} + \mu^2 - \frac{\mu}{n} = 4\mu + \mu^2$$

  El menor valor que puede alcanzar la varianza de E.I de $\tau(\mu)$ es la CCR $= \mu(4 + 2\mu)^2/n$.

  Note que $P(x \geq 1) = 1 - P(x = 0 = 1 - e^{-\mu}$. Sabemos que $(\frac{n-1}{n})^s, S \sum_{1=1}^{n} x_i$ es un EI de $e^{-\mu}$ basado en un estadístico suficiente. Luego $1 - (\frac{n-1}{n})^s$ es un EI, basado en un estadístico suficiente de $P(X \geq 1)$.

I2

(a) Intervalo de confianza del 95 % para la diferencia de medias

Datos:

- Muestra X: $n_x = 20$, $\bar{x} = 1021{,}3$ hr, $\sigma_0 = 30$ hr
- Muestra Y: $n_y = 25$, $\bar{y} = 1005{,}7$ hr, $\sigma_0 = 30$ hr

Calculamos la varianza de la diferencia:

$$\sigma^2_{\bar{x}-\bar{y}} = \frac{\sigma_0^2}{n_x} + \frac{\sigma_0^2}{n_y} = 30^2\left(\frac{1}{20} + \frac{1}{25}\right) = 900 \times 0{,}09 = 81 \text{ hr}^2$$

El error estándar es $\sqrt{81} = 9$ hr. El valor crítico para 95 % de confianza es $z_{0,025} = 1{,}96$.
Intervalo de confianza:

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \text{SE} = (1021{,}3 - 1005{,}7) \pm 1{,}96 \times 9 = 15{,}6 \pm 17{,}64 \text{ hr}$$

Resultado final:

$$(-2{,}04 \text{ hr},\ 33{,}24 \text{ hr})$$

(b) Estimador jackknife de $\gamma_x = \sigma_0/\mu_x$

El estimador jackknife se construye eliminando una observación a la vez:
1. Estimador completo:

$$\hat{\gamma} = \frac{\sigma_0}{\bar{X}}$$

2. Estimadores parciales:

$$\hat{\gamma}_{(i)} = \frac{\sigma_0}{\bar{X}_{(i)}} \quad \text{con} \quad \bar{X}_{(i)} = \frac{1}{n-1}\sum_{j\neq i} X_j$$

3. Estimador jackknife:

$$\hat{\gamma}_J = n\hat{\gamma} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\gamma}_{(i)}$$

(c) Región de confianza del 95 % para (μ_x, μ_y)

La región está dada por la elipse:

$$\left(\frac{\mu_x - \bar{x}}{\sigma_0/\sqrt{n_x}}\right)^2 + \left(\frac{\mu_y - \bar{y}}{\sigma_0/\sqrt{n_y}}\right)^2 \leq \chi^2_{2,0,05} = 5{,}991$$

Sustituyendo valores:

$$\left(\frac{\mu_x - 1021{,}3}{30/\sqrt{20}}\right)^2 + \left(\frac{\mu_y - 1005{,}7}{30/\sqrt{25}}\right)^2 \leq 5{,}991$$

Forma final:

$$\left(\frac{\mu_x - 1021{,}3}{6{,}708}\right)^2 + \left(\frac{\mu_y - 1005{,}7}{6}\right)^2 \leq 5{,}991$$

(a) Estimador de momentos $\hat{\theta}_1$

Partiendo de $\mathbb{E}(X^2) = \theta$:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Consistencia:

$$\hat{\theta}_1 \xrightarrow{p} \mathbb{E}(X^2) = \theta \quad \text{(Ley de Grandes Números)}$$

Distribución asintótica:

$$\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{d} N(0, \theta^2) \quad \text{(TCL, usando Var}(X^2) = \theta^2\text{)}$$

(b) Estimador máximo verosímil $\hat{\theta}_2$

Función de verosimilitud:

$$L(\theta) = \prod_{i=1}^{n} \frac{2X_i}{\theta} e^{-X_i^2/\theta}$$

Log-verosimilitud:

$$\ell(\theta) = n \ln 2 + \sum \ln X_i - n \ln \theta - \frac{T}{\theta}, \quad T = \sum X_i^2$$

Derivando:

$$\frac{d\ell}{d\theta} = -\frac{n}{\theta} + \frac{T}{\theta^2} = 0 \implies \hat{\theta}_2 = \frac{T}{n} = \hat{\theta}_1$$

Distribución asintótica:

$$\sqrt{n}(\hat{\theta}_2 - \theta) \xrightarrow{d} N(0, \theta^2)$$

(c) Comparación de estimadores

Ambos estimadores coinciden ($\hat{\theta}_1 = \hat{\theta}_2$) y tienen la misma eficiencia asintótica.

(d) IC exacto para θ

Usando $Q = 2T/\theta \sim \chi^2_{2n}$:

$$P\left(\chi^2_{1-\alpha/2,2n} \leq \frac{2T}{\theta} \leq \chi^2_{\alpha/2,2n}\right) = 1 - \alpha$$

Intervalo:

$$\left(\frac{2T}{\chi^2_{\alpha/2,2n}}, \frac{2T}{\chi^2_{1-\alpha/2,2n}}\right)$$

(e) IC aproximado basado en EMV

$$\hat{\theta}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}_2^2}{n}}$$

(a) Estimación de $P(X > 3) = e^{-3/\theta_1}$

Estimador puntual:

$$\widehat{P(X > 3)} = e^{-3/\hat{\theta}_1} = e^{-3/10,5} \approx 0{,}7515$$

Error estándar (método delta):

$$\text{EE} \approx \left| \frac{3e^{-3/\theta_1}}{\theta_1^2} \right| \cdot \sqrt{\frac{\theta_1^2}{n}} = \frac{3e^{-3/\theta_1}}{\theta_1 \sqrt{n}} \approx 0{,}068$$

(b) Cantidad pivotal

Para $X \sim \text{Exp}(\theta_1)$, $Y \sim \text{Exp}(\theta_2)$:

$$\frac{2n\bar{X}}{\theta_1} \sim \chi^2_{2n}, \quad \frac{2m\bar{Y}}{\theta_2} \sim \chi^2_{2m}$$

La razón:

$$Q = \frac{\bar{X}/\theta_1}{\bar{Y}/\theta_2} \sim F(2n, 2m)$$

es pivotal.

(c) IC para θ_2/θ_1

Con $F_{0,025}(20, 20) = 0{,}3958,\ F_{0,975}(20, 20) = 2{,}526$:

$$0{,}3958 \leq \frac{10{,}5}{14{,}0} \cdot \frac{\theta_2}{\theta_1} \leq 2{,}526$$

Despejando:

$$0{,}5277 \leq \frac{\theta_2}{\theta_1} \leq 3{,}368$$

I3

Región crítica, error tipo I, potencia, especificación de modelo

Región crítica de Newymann-Pearson

Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra aleatoria (iid) de una distribución normal con varianza conocida $\sigma^2 = 100$, es decir, $X_i \sim N(\theta, 100)$. Se desean probar las hipótesis simples: $H_0 : \theta = 2 \quad \text{vs.} \quad H_1 : \theta = 10,$, utilizando el lema de Neyman-Pearson.

i) Describa la región crítica C del test.

El lema de Neyman-Pearson requiere obtener Λ:

$$\Lambda(\theta_0, \theta_1) = \frac{L(\theta_0)}{L(\theta_1)} = \frac{\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \theta_0)^2\right)}{\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \theta_1)^2\right)} = \exp\left(\frac{\sum(x_i - \theta_1)^2 - \sum(x_i - \theta_0)^2}{2\sigma^2}\right)$$

Simplificando:

$$\Lambda(\theta_0, \theta_1) = \exp\left(\frac{-n\bar{x}(\theta_1 - \theta_0) + n\frac{\theta_1^2 - \theta_0^2}{2}}{\sigma^2}\right)$$

La región crítica del test está dada por:

$$C^* = \left\{\mathbf{x} : \exp\left(\frac{-n\bar{x}(\theta_1 - \theta_0) + n\frac{\theta_1^2 - \theta_0^2}{2}}{\sigma^2}\right) \leq k_\alpha\right\}$$

Aplicando logaritmo y simplificando:

$$C^* = \{\mathbf{x} : \bar{x} \geq k_\alpha\}$$

Nota: Como la normal es familia exponencial, el estadístico suficiente es $\sum x_i$, y la región crítica es equivalente a $\bar{x} \geq k_\alpha$.

ii) Determine el número de observaciones n necesario para que el error tipo I sea 0.01 y la potencia sea 0.99.

La potencia se calcula como:

$$\beta(\theta) = P(\bar{X} \geq k_\alpha | \theta) = 1 - \Phi\left(\frac{\sqrt{n}(k_\alpha - \theta)}{\sigma}\right)$$

Sabemos que:

$$\bar{X} \sim N(\theta, \frac{400}{n}); \quad Z = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} \sim N(0, 1)$$

Para el error tipo I $\theta = 2$:

$$0{,}01 = 1 - \Phi\left(\frac{\sqrt{n}(k_\alpha - 2)}{10}\right) \implies \frac{\sqrt{n}(k_\alpha - 2)}{10} = \Phi^{-1}(0{,}99) \approx 2{,}33$$

Para la potencia $\theta = 10$:

$$0{,}99 = 1 - \Phi\left(\frac{\sqrt{n}(k_\alpha - 10)}{10}\right) \implies \frac{\sqrt{n}(k_\alpha - 10)}{10} = \Phi^{-1}(0{,}01) \approx -2{,}33$$

Resolviendo el sistema:

$$k_\alpha - 2 = -(k_\alpha - 10) \implies k_\alpha = 6$$

Sustituyendo:

$$\sqrt{n}(6-2) = 10 \cdot 2{,}33 \implies \sqrt{n} = \frac{23{,}3}{4} = 5{,}825 \implies n \approx 34$$

Error tipo I, potencia

Sea X una muestra de tamaño uno de una distribución con función de densidad:

$$f(x;\theta) = \frac{2}{\theta^2}(\theta - x) \cdot I_{(0,\theta)}(x),$$

donde $\theta \in (0,\infty)$. Se considera la prueba de hipótesis: $H_0 : \theta = 1$ vs. $H_1 : \theta > 1$. Sea D el test con región crítica $C_D = \{X > 1\}$. Calcule:

- La probabilidad de error tipo I.
- La función de potencia $\beta(\theta)$.
- Grafique $\beta(\theta)$ contra θ.

La probabilidad de error tipo I es:

$$\beta(1) = P(X > 1|\theta = 1) = 1 - P(x \leq 1|\theta = 1) := 1 - \int_0^1 2(1-x)dx = 1 - 2 + 1 = 0$$

La función de potencia es:

$$\beta(\theta) = P(X > 1|\theta) = \int_1^\theta \frac{2}{\theta^2}(\theta - x)dx = (1 - \frac{1}{\theta})^2 = (\frac{\theta - 1}{\theta})^2$$

La gráfica de $\beta(\theta)$ muestra que la función es positiva, con un mínimo en $\theta = 1$ y tiende a 1 cuando $\theta \to \infty$.

Especificación del modelo de regresión lineal

Sean:

$$Y_1 = \beta_1 + \beta_2 + \epsilon_1, \quad Y_2 = 2\beta_1 - 2\beta_2 + \epsilon_2, \quad Y_3 = 2\beta_1 - \beta_2 + \epsilon_3,$$

donde $E(\epsilon_i) = 0$ y $\mathrm{Var}(\epsilon_i) = \sigma^2$ para $i = 1, 2, 3$.

- ¿Es este un modelo lineal? ¿Cuáles son los supuestos?
- Escriba el modelo en forma matricial.

Sí, es un modelo lineal bajo el supuesto de que $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$.
En forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

Otra forma de decir el supuesto es decir que $\epsilon_1, \epsilon_2, \epsilon_3$ con va iid $(0, \sigma^2)$; el modelo es lineal.

Tests UMP, Neyman-Pearson

Considere el siguiente modelo estadístico:

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

(a)

Encuentre el test de Neyman-Pearson de tamaño α (basado en una muestra de tamaño 1) para:

$$H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta = \theta_1, \quad \text{con } \theta_1 > 1.$$

La razón de verosimilitud es:

$$\Lambda(1,\theta_1) = \frac{1}{\theta_1 x^{\theta_1 - 1}}$$

La región crítica es:

$$C^* = \{x : f(x;\theta_1) > k_\alpha f(x;1)\} = \{x : x > \tilde{k_\alpha} | \theta = 1\}$$
$$\int_{\tilde{k_\alpha}}^{1} dx = 1 - \tilde{k_\alpha} = \alpha \implies \tilde{k_\alpha} = 1 - \alpha \to \{x : x > 1 - \alpha\}$$

(b)

¿Es sesgado el test encontrado en (a)?

El test es insesgado porque para $\theta_1 > 1$, $\beta(\theta_1) \geq \beta(1)$, $\beta(\theta) = \int\limits_{1-\alpha}^{1} \theta x^{\theta-1} dx = 1 - (1-\alpha)^\theta$ es monótona creciente.

(c)

Para las hipótesis del punto (a), dado $\theta_1 > 1$, calcule la potencia máxima que puede tener una prueba de tamaño α.
La potencia es:

$$\beta^*(\theta) = 1 - k_\alpha^\theta = 1 - (1-\alpha)^\theta$$

Ya que N-P garantiza que el test sea UMP, entonces para cualquier otro test de tamaño α, β^* es la potencia máxima.

(d)

Deduzca de (a) una prueba de nivel α uniformemente más potente (UMP) para: $H_0 : \theta = 1 \quad \text{vs.} \quad H_1 : \theta > 1$.

Usando el teorema de Karlin-Rubin, la región crítica UMP es:

$$C^* = \{x : x \geq k_\alpha\}$$

TRV y Test Score

Suponga que un ingeniero desea comparar el número de reclamos por semana que registran dos turnos diferentes en una planta manufacturera. Se obtienen cien observaciones independientes del número de quejas por semana, lo que arroja medias muestrales de $\bar{x}_1 = 20$ para el turno 1 y $\bar{x}_2 = 22$ para el turno 2. Suponga que el número de quejas por semana en el i-ésimo turno tiene una distribución de Poisson con media θ_i, para $i = 1, 2$. Utilice el test de la razón de verosimilitudes y el test score para probar la hipótesis:

$$H_0 : \theta_1 = \theta_2 \quad \text{vs.} \quad H_1 : \theta_1 \neq \theta_2,$$

con $\alpha \approx 0{,}01$. ¿Qué concluye?

Sea $\theta(\theta_1, \theta_2)^\top$. El test de razón de verosimilitudes calcula:

$$L(\theta) = (1/k)\theta_1^{n\bar{X}_1} \exp(-n\theta_1) \cdot \theta_2^{n\bar{X}_2} \exp(-n\theta_2)$$

con $k = x_{11}! \dots x_{n1}! \cdot x_{12}! \dots x_{n2}!$; $n = 100$. Luego, tomando el logaritmo:

$$\mathcal{L}(\theta) = \log(1/k) + n\bar{X}_1 \log\theta_1 - n\theta_1 + n\bar{X}_2 \log\theta_2 - n\theta_2$$

Derivando e igualando a 0 se obtiene: $\hat{\theta} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix}$ y $\hat{\theta}_0 = \frac{1}{2}\begin{pmatrix} \bar{X}_1 + \bar{X}_2 \\ \bar{X}_1 + \bar{X}_2 \end{pmatrix}$

$$\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \left(\frac{21}{20}\right)^{2000} \left(\frac{21}{22}\right)^{2200} \approx e^{-4{,}76} \approx 0{,}0085656$$

El estadístico es:

$$-2\ln\Lambda \approx 9{,}52$$

Como $9{,}52 > 6{,}63$ (valor crítico de χ_1^2 para $\alpha = 0{,}01$), se rechaza H_0.
El test score calcula:

$$\mathcal{U}(\theta) = n\begin{pmatrix} \frac{\bar{X}_1}{\theta_1} \\ \frac{\bar{X}_2}{\theta_2} \end{pmatrix} - n\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$Var(\mathcal{U}) = n\begin{pmatrix} 1/\theta_1 & 0 \\ 0 & 1/\theta_2 \end{pmatrix} = \mathcal{I}_X$$

Luego, $\mathcal{U}(\hat{\theta}_0) = 100\begin{pmatrix} 20/21 \\ 22/21 \end{pmatrix} - 100\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -4{,}8 \\ 4{,}8 \end{pmatrix}$.

$$\mathcal{I}_X^{-1}(\hat{\theta}_0) = \frac{1}{100}\begin{pmatrix} 21 & 0 \\ 0 & 21 \end{pmatrix} = \frac{21}{100} I_2 \to T_S(x) = \mathcal{U}^\top(\hat{\theta}_0)\mathcal{I}_X^{-1}(\hat{\theta}_0)\mathcal{U}(\hat{\theta}_0)$$

Bajo H_0:

$$T_{SC} = (-n_1 + \frac{\Sigma x_i}{\theta} \quad -n_2 + \frac{\Sigma x_i}{\theta})\begin{pmatrix} \theta/n_1 & 0 \\ 0 & \theta/n_2 \end{pmatrix}\begin{pmatrix} -n_1 + \frac{\Sigma x_i}{\theta} \\ -n_2 + \frac{\Sigma x_i}{\theta} \end{pmatrix} \approx 9{,}52$$

Tambien se puede usar el estadístico:

$$S = \frac{(\hat{X}_1 - \hat{X}_2)^2}{\bar{X}/n_1 + \bar{Y}/n_2} \approx \frac{(\hat{\theta}_1 - \hat{\theta}_2)^2}{Var(\hat{\theta}_1 - \hat{\theta}_2)}$$

Y como el MLE es la media global: $\hat{\theta} = \frac{n_1\hat{X} + n_2\hat{X}}{n_1 + n_2} = 21$, reemplazando en S se obtiene 9.52 .
Como $9{,}5238 > 9{,}21$ (valor crítico de χ_2^2 para $\alpha = 0{,}01$), también se rechaza H_0.
En conclusión, ambos tests rechazan H_0, indicando que las medias de quejas difieren entre los turnos.

Ayudantías

A6

Si $X_i \sim$ Poisson(λ), el EMV para $P(X = 0)$

$$P(X_i = 0) = e^{-\lambda} = \tau(\lambda)$$

Si $\hat{\lambda} = \bar{X}$ es el EMV de λ, entonces $\tau(\hat{\lambda})$ es el EMV de $\tau(\lambda)$. Así, el EMV de $e^{-\lambda}$ es $e^{\bar{X}}$. Del TLC se tiene que

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$$

Utilizando el método delta

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{D} N(0, \sigma^2) \implies \sqrt{n}(g(\bar{X}) - g(\lambda)) \xrightarrow{D} N(0, (g'(\lambda))^2\sigma^2), \quad g(\lambda) = e^{-\lambda},\ g'(\lambda) = -e^{-\lambda},\ \mathrm{Var}(X) = \lambda$$

$$\implies \sqrt{n}(e^{-\bar{X}} - e^{-\lambda}) \xrightarrow{D} N(0, \lambda e^{-2\lambda})$$

A8

Sea $f(x|\theta) = 1, \theta - \frac{1}{2} < x < \theta + \frac{1}{2}$. Para encontrar un IC $1-\alpha$ para θ, se puede notar que $X \sim \text{Uniforme}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, por lo que $Z = X - \theta \sim \text{Uniforme}(-\frac{1}{2}, \frac{1}{2})$. Ahora, se buscan $L(X) \leq U(X)$ tal que $P(L \leq \theta \leq U) = 1 - \alpha$. Dado que es de la familia de localización, se propone $L = X - b$ y $U = X - a$. Así

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

Dado que $Z = X - \theta \sim \text{Unif}$

$$P(a \leq X - \theta \leq b) = b - a \implies b - a = 1 - \alpha$$

Una opción en particular $a = -\frac{1}{2} + \frac{\alpha}{2}, b = \frac{1}{2} - \frac{\alpha}{2}$

Sea $f(x|\theta) = \frac{2x}{\theta^2}, 0 \leq x \leq \theta$. En este caso, la distribución es de una familia de escala. Cambiando a la variable $Z = \frac{X}{\theta}$ y usando $L = aX$ y $U = bX$

$$P(aX \leq \theta \leq bX) = P(a \leq \frac{\theta}{X} \leq b) = P(\frac{1}{b} \leq \frac{X}{\theta} \leq \frac{1}{a}) = \int_{\frac{1}{b}}^{\frac{1}{a}} 2zdz = (\frac{1}{a})^2 - (\frac{1}{b})^2 \implies (\frac{1}{a})^2 - (\frac{1}{b})^2 = 1 - \alpha$$

Sea $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$. Dado que la distribución es de la familia exponencial, un estadístico suficiente es $T = \sum x_i \sim \text{Gamma}(n, \lambda)$. Como pivote se propone $Q = \frac{2T}{\lambda}$. Usando el cambio de variable

$$g(T) = Q = \frac{2T}{\lambda} \to g^{-1}(q) = T = \frac{q\lambda}{2} \to T' = \frac{\lambda}{2}$$

$$f_Q(q) = \frac{1}{\Gamma(n) 2^n} q^{n-1} e^{-q/2} \sim \text{Gamma}(n, 2)$$

Que es independiente de λ. Así, usando el método de la cantidad pivotal

$$P(a \leq \frac{2T}{\lambda} \leq b) = 1 - \alpha \to a = \chi^2_{2n}(\frac{\alpha}{2}), \quad b = \chi^2_{2n}(1 - \frac{\alpha}{2})$$

Sea $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda_1)$ e $Y_1, \ldots, Y_m \overset{iid}{\sim} \text{Exp}(\lambda_2)$. Se quiere encontrar un IC para λ_1/λ_2. Se sabe que

$$T_1 = \sum x_i \to \frac{2T_1}{\lambda_1} \sim \chi^2_{2n} \quad T_2 = \sum y_i \to \frac{2T_2}{\lambda_2} \sim \chi^2_{2m} \implies Q = \frac{2T_1/\lambda_1 2n}{2T_2/\lambda_2 2m} \sim F_{2n,2m}$$

$$P(a \leq F_{2n,2m} \leq b) = 1 - \alpha, \quad a = F_{2n,2m}(\frac{\alpha}{2}) \quad F_{2n,2m}(1 - \frac{\alpha}{2})$$

A10

Sea $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$. El test de razón de verosimilitud sobre las hipótesis $H_0 : \mu = \mu_o$ vs $H_1 : \mu \neq \mu_1$ simplificado en la forma $\lambda(x) \leq c$ corresponde a:

$$\Lambda(X) = \frac{\sup_{\mu = \mu_o} L(\mu|X)}{\sup_{\mu \in \mathbb{R}} L(\mu|X)} = \frac{L(\mu_0|X)}{L(\hat{\mu}_{emv}|X)}$$

$$L(\mu|X) = \prod \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2} \implies \hat{\mu} = \bar{X} \implies \Lambda(X) = \frac{e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu_0)^2}}{e^{-\frac{1}{2\sigma^2} \sum (x_i - \bar{X})^2}} = \exp(-\frac{1}{2\sigma^2}(\sum (x_i - \mu_0)^2 - \sum (x_i - \bar{X}^2)))$$

$$\Lambda(X) = \exp(-\frac{n}{2\sigma^2}(\bar{X} - \mu_0)^2)$$

Así, la región crítica corresponde a $\{X : \exp(-\frac{n}{2\sigma^2}(\bar{X} - \mu_0)^2) \leq c\}$

Sea $X_1, \ldots, X_n \overset{iid}{\sim} Pareto(\theta, \nu)$, la distribución está dada por

$$f(x|\theta, \nu) = \frac{\theta \nu^\theta}{x^{\theta+1}} I_{[\nu, \infty)}(x)$$

El EMV de θ y ν

$$L = \prod \frac{\theta \nu^\theta}{x^{\theta+1}} I_{[\nu, \infty)}(x) = \frac{\theta^n \nu^{n\theta}}{(\prod x_i)^{\theta+1}} I_{(0, X_{(1)})}(\nu) \implies \hat{\theta} = \frac{n}{\log(\prod x_i) - n\log(\nu)}, \quad \hat{\nu} = X_{(1)}$$

A12

Sea $X_1, \ldots, X_n$ una muestra aleatoria desde una distribución Binomial$(1,\theta)$. Se quiere probar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Primero se usará el test de razón de verosimilitud y luego los otros test para comparar los resultados.

$$RVG : \Lambda(X) = \frac{\sup_{\theta\in\theta_0} L}{\sup_{\theta\in\theta} L}, \quad C_\Lambda = \{x \in X : \Lambda(x) \leq k_\alpha\}$$

$$L = \prod \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \implies \hat{\theta}_{emv} = \bar{X}$$

$$\Lambda(X) = (\frac{\theta_0}{\bar{X}})^{\sum x_i}(\frac{1-\theta_0}{1-\bar{X}})^{n-\sum x_i} \implies C_\Lambda = \{x \in X : (\frac{\theta_0}{\bar{X}})^{\sum x_i}(\frac{1-\theta_0}{1-\bar{X}})^{n-\sum x_i} \leq k_\alpha\}$$

Para expresarlo en términos de $T = \sum x_i$

$$\Lambda(T) = (\frac{n\theta_0}{T})^T(\frac{n(1-\theta_0)}{n-T})^{n-T}$$

$$\log(\Lambda) = T\log(\frac{n\theta_0}{T}) + (n-T)\log(\frac{n(1-\theta_0)}{n-T})$$

Se llega a

$$\log\Lambda = log(\frac{\theta_0(n-T)}{T(1-\theta_0)}) \text{ , luego si } \begin{cases} \frac{\theta_0(n-T)}{T(1-\theta_0)} > 1 & \log\Lambda > 0 \\ \frac{\theta_0(n-T)}{T(1-\theta_0)} < 1 & \log\Lambda < 0 \end{cases}$$

Nuevamente utilizando RVG. Sabemos que: $T_{RV}(X) = -2\ln\Lambda(X) \xrightarrow{d} \chi^2_r$ con $\Lambda(X)$ calculada anteriormente y $r = dim(\theta = -dim(\theta_0)$.

$-2\ln(\Lambda) = 2\ln(L(\hat{\theta}) - L(\theta_0))$, $C_\Lambda : \{x \in \mathcal{X} : -2\ln\Lambda n(X) \geq k_\alpha\}$. $T_{RV} = -2(\sum X_i\log(\frac{\theta_0}{\bar{X}}) + (n-T)\log(\frac{1-\theta_0}{1-\bar{X}})) = 2(\sum X_i\log(\frac{\bar{X}}{\theta_0}) + (n-T)\log(\frac{1-\bar{X}}{1-\theta_0}))$.

$$C_\Lambda = \{2(\sum X_i\log(\frac{\bar{X}}{\theta_0}) + (n-T)\log(\frac{1-\bar{X}}{1-\theta_0}) \geq k_\alpha\}$$

Como $T_{RV} \overset{approx}{\sim} \chi^2_1$, $k_\alpha = \chi^2_1(1-\alpha)$.

Se prosigue con el test de razón de Wald.

$$T_W(X) = (\hat{\theta}_n - \theta_0)^T\Sigma^{-1}_\theta(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_k$$

Donde $\Sigma_\theta = \mathcal{I}^{-1}_{\underline{X}}(\theta)$, $\hat{\theta}_n = EMV$. $\ell'' = \frac{-\Sigma X_i}{\theta^2} - \frac{n-\Sigma X_i}{(1-\theta)^2}$, $-E(\ell'') = \frac{n}{\theta} + \frac{n}{1-\theta}$, $\mathcal{I}_{\underline{X}}(\theta) = \frac{n-n\theta+n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$

$$T_W(X) = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \overset{approx}{\sim} \chi^2_1 \quad \text{Entonces } C_W = \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_1(1-\alpha)\}$$

Veamos el test Score

$$T_S(X) = \mathcal{U}^T(\theta_0)\mathcal{I}^{-1}_X(\theta_0)\mathcal{U}(\theta_0) \overset{app}{\sim} \chi^2_k \text{ , } C_S = \{x \in \mathcal{X} : T_S(X) \geq \chi^2_k(1-\alpha)\}$$

$\mathcal{U} = \frac{\sum X_i}{\theta} - \frac{n-\sum X_i}{1-\theta}$, $\mathcal{U}(\theta_0) = \frac{\sum X_i}{\theta_0} - \frac{n-\sum X_i}{1-\theta_0} = \frac{\sum X_i - n\theta}{\theta_0(1-\theta_0)} = \frac{n(\bar{X}-\theta_0)}{\theta_0(1-\theta_0)}$. $\mathcal{I}^{-1}_X(\theta_0) = \frac{\theta_0(1-\theta_0)}{n}$, $(\frac{n\bar{X}-\theta_0}{\theta_0(1-\theta_0)})^2 = \frac{n^2(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)^2}$

$$T_S = \frac{\frac{n^2(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)^2}}{\frac{n}{\theta_0(1-\theta_0)}} = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \text{ , } C_S = \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_i(1-\alpha)\}$$

Finalmente comparemos con el test Gradiente.

$$T_{Gr} : \mathcal{U}^T(\theta_0)(\hat{\theta}_{EMV} - \theta_0) \sim \chi^2_k$$

$$T_{Gr} : \frac{n(\bar{X}-\theta_0)}{\theta_0(1-\theta_0)}(\bar{X}-\theta_0) = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)}$$

$$T_{Gr} : \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_i(1-\alpha)\}$$

A13

Se tiene el siguiente modelo lineal:

$$y_{ij} \sim (\beta_i x_j, \sigma^2), \quad i = 1, 2; j = 1, \dots n.$$

El vector de respuestas y, el vector de parámetros β y la matriz de diseño X corresponden a:

$$y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ \vdots \\ y_{2n} \end{pmatrix} \quad X = \begin{pmatrix} X_1 & 0 \\ \vdots & \vdots \\ X_n & 0 \\ 0 & X_1 \\ \vdots & \vdots \\ 0 & X_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Utilizando el MMC, se tiene que

$$X^T = \begin{pmatrix} X_1 & \dots & X_n & 0 & \dots & 0 \\ 0 & \dots & 0 & X_1 \dots & X_n \end{pmatrix} \quad \rightarrow \quad X^T X = \begin{pmatrix} \sum X_i^2 & 0 \\ 0 & \sum X_i^2 \end{pmatrix} \quad \rightarrow \quad (X^T X)^{-1} = \begin{pmatrix} \frac{1}{\sum X_i^2} & 0 \\ 0 & \frac{1}{\sum X_i^2} \end{pmatrix}$$

$$X^T y = \begin{pmatrix} \sum X_i y_{1i} \\ \sum X_i y_{2i} \end{pmatrix} \quad \rightarrow \quad \hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} \frac{\sum X_i y_{1i}}{\sum X_i^2} \\ \frac{\sum X_i y_{1i}}{\sum X_i^2} \end{pmatrix}$$

En cuanto al estimador de σ^2

$$S^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p}$$

Pauta Examen 2020-2

Muestra de tamaño 1 y estimación de parámetros

Se extrae una muestra de tamaño uno de una distribución con fdp

$$f(x; \theta) = \frac{2}{\theta^2}(\theta - x), \quad 0 < x < \theta.$$

Sea $\hat{\theta}$, el EMV de θ y $\tilde{\theta}$ el estimador de momentos de θ. ¿Qué estimador prefiere, $\hat{\theta}$ o $\tilde{\theta}$? Justifique.

Dado que se tiene una observación $X = x$, luego encontramos la verosimilitud:

$$L(\theta) = f(x; \theta) = \frac{2}{\theta^2}(\theta - x), \quad x < 0$$

Maximizamos θ: $\ell(\theta) = \log L(\theta) = \log 2 - 2\log\theta - \log(\theta - x)$. Luego derivando respecto θ: $\frac{d\ell}{d\theta} = -\frac{2}{\theta} + \frac{1}{\theta - x} = 0$
Luego el EMV es: $\hat{\theta} = 2X$
El estimador de momentos se calcula:

$$E(X) = \frac{\theta}{3}, \ \bar{X} = X \approx E(X) \text{ Luego, } \tilde{\theta} = 3X$$

Para compararlos, recordamos que el EMV suele tener mejores propiedades de varianza, se observa;

- $\hat{\theta} = 2X$ tiene varianza: $V(\hat{\theta}) = 4Var(X)$
- $\tilde{\theta} = 3X$ tiene varianza: $V(\tilde{\theta}) = 9Var(X)$

Por lo que preferimos $\hat{\theta}$

Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria de tamaño n desde una distribución normal $N(\mu, \sigma_0^2)$, con σ_0^2 conocido.**i)** Muestre que la varianza de cualquier estimador insesgado de μ^2 no puede ser menor que

$$\frac{4\mu^2\sigma_0^2}{n}.$$

ii) Muestre que

$$T = \bar{X}^2 - \frac{\sigma_0^2}{n}$$

es un EIVUM de μ^2. ¿Puede usted asegurar que la $\mathrm{Var}(T) = \frac{4\mu^2\sigma_0^2}{n}$? Justifique.

Para cualquier estimador insesgado W de μ^2, por Cramér-Rao:

$$Var(W) \geq \frac{(g'(\mu))^2}{I(\mu)}$$

$$g(\mu) = \mu^2, \quad g'(\mu) = 2\mu, \quad I(\mu) = \frac{n}{\sigma_0^2}$$

Por tanto,

$$Var(W) \geq \frac{4\mu^2\sigma_0^2}{n}$$

Para

$$T = \bar{X}^2 - \frac{\sigma_0^2}{n}$$

$$E(T) = E(\bar{X}^2) - \frac{\sigma_0^2}{n} = \left(\mu^2 + \frac{\sigma_0^2}{n}\right) - \frac{\sigma_0^2}{n} = \mu^2$$

Alcanza la cota, por lo que es EIVUM. No se puede afirmar que

$$Var(T) = \frac{4\mu^2\sigma_0^2}{n}$$

porque depende del sesgo cuadrático.

IDEC y TRV

Suponga que usted ha sido contratado como consultor en un experimento de muestreo que implica medir varillas de acero (en pulgadas). Se tomará una muestra $X_1, X_2, \dots, X_n \sim N(\mu, 30)$. Se requiere un IC del 95 % de longitud 3 pulgadas. ¿Qué tamaño de muestra recomendaría a su cliente?

Se tiene $X_i \sim N(\mu, 30)$ y queremos un IDEC al 95 % de longitud 3 para μ, es decir:

$$2z_{0{,}025}\frac{\sigma}{\sqrt{n}} = 3$$

Con $z_{0{,}025} = 1{,}96$, $\sigma = \sqrt{30} \approx 5{,}477$.
Luego,

$$2 \cdot 1{,}96 \cdot \frac{\sqrt{30}}{\sqrt{n}} = 3 \Rightarrow \sqrt{n} = \frac{5{,}477}{0{,}765} \approx 7{,}16 \Rightarrow n \approx 52$$

Sean $X_i \sim b(n_i, \pi_i)$, $i = 1, \dots, k$ variables aleatorias independientes. Use el test de la razón de verosimilitud generalizado para probar la hipótesis

$$H_0 : \pi_1 = \cdots = \pi_k.$$

Para $H_0 :\ \pi_1 = \cdots = \pi_k$ usamos el test LR:

$$\Lambda = -2\log\left(\frac{\text{Máx VL bajo } H_0}{\text{Máx VL general}}\right)$$

Que distribuye asintóticamente χ^2_{k-1}.

$$L(\pi_1,\dots,\pi_k)=\prod_{i=1}^{k}\binom{n_i}{x_i}\pi_i^{x_i}(1-\pi_i)^{n_i-x_i}\ ,\ \ell=\sum_{i=1}^{k}[x_i\log\pi_i+(n_i+x_i)\log(1-\pi_i)]$$

Sin restricciones, para cada i el EMV es:

$$\hat{\pi}_i=\frac{x_i}{n_i},\ \text{log verosimilitud maxima: }\sum_{i=1}^{k}[x_i\log\frac{x_i}{n_i}+(n_i+x_i)\log(1-\frac{x_i}{n_i})$$

Bajo H_0 todos los $\pi_i=\pi$, luego:

$$\ell(\pi,\dots,\pi)=\sum_{i=1}^{k}[x_i\log\pi+(n_i+x_i)\log(1-\pi)]$$

EL EMV se obtiene maximizando o por momentos:$\hat{\pi}=\frac{\sum_{i=1}^{k}x_i}{\sum_{i=1}^{k}n_i}$
El TRV se vería:

$$\Lambda=-2[\ell(\hat{\pi},\dots,\hat{\pi})-\ell(\hat{\pi}_1,\dots,\hat{\pi}_k)]=-2\sum_{i=1}^{k}[x_i\log(\frac{\hat{\pi}}{\hat{\pi}_i}+(n_i-x_i)\log\frac{1-\hat{\pi}}{1-\hat{\pi}_i})]$$

Luego, se rechaza si $\Lambda>\chi^2_{k-1,1-\alpha}$.

EMV, Test Score y Gradiente

Sea $X_1,\dots,X_{200}$ una muestra aleatoria de tamaño 200 desde una población con densidad dada por,

$$f(x;\theta)=\frac{2x}{\theta}\exp\{-x^2/\theta\},\ x>0,\ \theta>0,$$

para la cual se tiene que

$$\mathrm{E}(X)=\frac{\sqrt{\pi\theta}}{2}\text{ y }\mathrm{E}(X^2)=\theta.$$

Encuentre el EMV de θ. ¿Es función de algún estadístico suficiente?
Log-verosimilitud para $X_1,\dots,X_{200}$:

$$\ell(\theta)=\sum\log(2X_i)-n\log(\theta)-\frac{\sum X_i^2}{\theta}=\sum\log 2X_i-n\log\theta-\frac{\sum X_i^2}{\theta}$$

Derivando:

$$\frac{\partial\ell}{\partial\theta}=-\frac{n}{\theta}+\frac{\sum X_i^2}{\theta^2}=0$$
$$\theta=\frac{1}{n}\sum X_i^2=\frac{1}{200}\sum X_i^2$$

Por lo que es función de $\sum X_i^2$, estadístico suficiente $\sum X_i^2$.
Se sabe que para la muestra anterior, el EMV de θ es $\hat{\theta}=2{,}5$. Use los tests **Score** y **Gradiente** para probar la hipótesis $H_0:\theta=3{,}5$ versus $H_1:\theta\neq3{,}5$, a un nivel $\alpha=0{,}01$. Comente los resultados.
Test Score: Se tiene el estadístico: $S=\frac{U(\theta_0)}{\sqrt{\hat{I}(\theta_0)}}\sim N(0,1)$

$$U(\theta)=-\frac{200}{\theta}+\frac{200\hat{\theta}}{\theta^2}$$

Evaluando en $\theta_0=3{,}5$:

$$U(3{,}5)=-\frac{200}{3{,}5}+\frac{200*2{,}5}{3{,}5^2}\approx-57{,}14+40{,}82=-16{,}32$$

La información de Fischer:

$$I(3{,}5) = \frac{200}{3{,}5^2} \approx 16{,}33$$

$$S = \frac{U(3{,}5)}{\sqrt{I(3{,}5)}} \approx -4{,}0$$

Test Gradiente: Similarmente, tenemos el estadístico: $G = \frac{\ell'(3{,}5)}{\sqrt{\hat{I}}}$

$$\hat{I} = \frac{200}{(2{,}5)^2} = 32$$

$$G \approx \frac{-16{,}32}{\sqrt{32}} = -2{,}88$$

Rechazamos H_0 al $\alpha = 0{,}01$ si $|Z| > z_{1-\alpha/2} \approx 2{,}58$, ya que $|S| > 2{,}58$ y $|G > 2{,}58$ se rechaza H_0.

Ejercicios

Como plantear TRV y región crítica

Sea $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$.

a) Plantee un test de razón de verosimilitud para las hipótesis $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. Calcule la región crítica simplificada.

Solución:

La función de verosimilitud es:

$$L(\mu|x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2}} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum(X_i-\mu)^2}$$

$$\ell(\mu|x) = -\frac{n}{2} log(2\pi) - \frac{1}{2}\sum(X_i - \mu)^2$$

$$\mathcal{U}(\mu|x) = \sum X_i - n\mu = 0\ ,\ \hat{\mu} = \bar{X} = EMV$$

El estimador máximo verosímil (EMV) de μ es $\hat{\mu} = \overline{X}$. La razón de verosimilitud es:

$$\lambda(x) = \frac{\sup_{\mu=\mu_0} L(\mu|x)}{\sup_{\mu\in\mathbb{R}} L(\mu|x)} = \frac{L(\mu_0|x)}{L(\hat{\mu}|x)} =$$

$$\frac{(2\pi)^{-\frac{n}{2}} e^{-\frac{2}{2\sigma^2}} \sum(X_i-\mu_0)^2}{(2\pi)^{-\frac{n}{2}} e^{-\frac{2}{2\sigma^2}} \sum(X_i-\bar{X})^2} = \exp\left[-\frac{1}{2}\left(\sum(X_i-\mu_0)^2 - \sum(X_i-\overline{X})^2\right)\right]$$

Simplificando:

$$\sum(X_i-\mu_0)^2 = \sum(X_i - \bar{X} + \bar{X} - \mu_0)^2 = \sum(X_i-\bar{X})^2 + 2(\bar{X}-\mu_0)\sum(X_i-\bar{X}) + \sum(\bar{X}-\mu_0)^2$$

$$= \sum(X_i-\bar{X})^2 + \sum(\bar{X}-\mu_0)^2 \ldots \exp[-\frac{1}{2}(\sum(X_i-\bar{X})^2 + \sum(\bar{X}-\mu_0)^2 - \sum(X_i-\overline{X})^2]$$

$$\lambda(x) = \exp\left[-\frac{n}{2}(\overline{X}-\mu_0)^2\right]$$

La región crítica es:

$$\left\{x : \exp\left[-\frac{n}{2\sigma^2}(\overline{X}-\mu_0)^2\right] \leq c\right\}$$

b) Transforme la región de rechazo a la forma $\{x : T(x) \leq k_1 \vee T(x) \geq k_2\}$, donde $T(x) = \overline{x}$.

Partiendo de $\lambda(x) \leq c$:

$$\exp\left[-\frac{n}{2}(\overline{X}-\mu_0)^2\right] \leq c \implies (\overline{X}-\mu_0)^2 \geq -\frac{2}{n}\log(c)$$

$$|\overline{X}-\mu_0| \geq \sqrt{-\frac{2}{n}\log(c)}$$

Esto equivale a:

$$\overline{X} \leq \mu_0 - \sqrt{-\frac{2}{n}\log(c)} \quad \text{o} \quad \overline{X} \geq \mu_0 + \sqrt{-\frac{2}{n}\log(c)}$$

Rechazamos H_O si se cumple una de las anteriores. Por lo tanto, la región de rechazo es:

$$\{x : \overline{X} \leq k_1 \vee \overline{X} \geq k_2\}$$

donde $k_1 = \mu_0 - \sqrt{-\frac{2}{n}\log(c)}$ y $k_2 = \mu_0 + \sqrt{-\frac{2}{n}\log(c)}$.

c) Suponga $\overline{x} = 2$ y $\mu_0 = 0$, encuentre valores de c que no rechacen H_0.

Para rechazar H_0, se debe cumplir:

$$\exp\left[-\frac{n}{2}(2-0)^2\right] \leq c \implies \exp(-2n) \leq c$$

Por lo tanto, los valores de c que no rechazan H_0 son aquellos con $c < e^{-2n}$.

Test de razón de verosimilitud

Sea $X_1, \ldots, X_n \overset{iid}{\sim}$ Pareto(θ, ν), con densidad:

$$f(x|\theta,\nu) = \frac{\theta\nu^\theta}{x^{\theta+1}} I_{[\nu,\infty)}(x), \quad \theta > 0, \nu > 0.$$

a) Encuentre el EMV de θ y ν.

La función de verosimilitud es:

$$L(\theta,\nu|x) = \prod_{i=1}^{n} \frac{\theta\nu^\theta}{x_i^{\theta+1}} I_{\nu,\infty}(x_i) = \frac{\theta^n\nu^{\theta n}}{(\prod x_i)^{\theta+1}} I_{(0,x_{(1)}}(\nu)$$

$$\ell(\theta,\nu|x) = n\log(\theta) + \theta n\log(\nu) - \theta\log(\prod X_i) + \log(\prod X_i)$$

donde la función indicadora en la productoria $I_{\nu,\infty}(x_i)$ depende de ν, para cualquier $x_i \leq \nu$ la función vale 0, por lo que $\nu \leq \text{mín}(X_1, \ldots, X_n) = X_{(1)}$. Luego, maximizando la verosimilitud, se tiene que para $\theta > 0$ el EMV de $\hat{\nu}$ es $\hat{\nu} = x_{(1)}$. Derivando para encontrar el EMV de θ:

$$\ell(\theta,\nu|x) = n\log(\theta) + \theta n\log(\nu) - \theta\sum\log(X_i) + \sum\log(X_i) \backslash \frac{\partial}{\partial\theta}$$

$$\mathcal{U}(\theta,\nu|x) = \frac{n}{\theta} + n\log(\nu) - \sum\log(X_i) = 0 \implies \frac{n}{\log(\frac{\prod X_i}{\nu^n})} = \frac{n}{\log(\frac{\prod X_i}{X_{(1)}^n})} = \hat{\theta}$$

b) Muestre que el TRV para $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$ tiene región crítica:

$$\{x : T(x) \leq c_1 \vee T(x) \geq c_2\}, \quad \text{donde} \quad T = \log\left[\frac{\prod_{i=1}^n x_i}{(x_{(1)})^n}\right] \text{ y tambien } 0 \leq c_1 \leq c_2.$$

$$\sup_{\theta=1} L(\theta,\nu|X) = L(1, X_{(1)}|X)$$

$$\sup L(\theta,\nu|X) = L(\hat{\theta}, X_{(1)}|X)$$

Entonces la RVM es:

$$\lambda(x) = \frac{\frac{X_{(1)}^n}{(\prod X_i)^2}}{\frac{(\frac{n}{T})^n X_{(1)}^{\frac{n^2}{T}}}{(\prod X_i)^{\frac{n}{T}+1}}} = \frac{X_{(1)}^n (\prod X_i)^{\frac{n}{T}+1}}{(\frac{n}{T})^n X_{(1)}^{\frac{n^2}{T}} (\prod X_i)^2} = \dots$$

$$\lambda(x) = (\frac{T}{n})^n (e^T)^{n-T}$$

La región crítica $\lambda(x) \leq c$ corresponde a valores extremos de T, es decir:

$$T \leq c_1 \quad \text{o} \quad T \geq c_2$$

donde c_1 y c_2 se determinan según el nivel de significancia.

Tamaño del Test y error tipo I

Como encontrar el tamaño del test:

Sea X una variable aleatoria con función de densidad:

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 \leq x \leq 1, \quad \theta > 0.$$

Se desea probar las hipótesis:

$$H_0 : \theta = 1 \quad \text{vs} \quad H_1 : \theta \neq 1.$$

El test propuesto D rechaza H_0 si $X < 0{,}025$ o $X > 0{,}975$.

1. **Tamaño del test:**
   El tamaño del test (α) es la probabilidad de rechazar H_0 cuando es verdadera. Bajo H_0 ($\theta = 1$), la densidad es uniforme: $f(x;1) = 1$.

   $$\alpha = P_{H_0}(X < 0{,}025) + P_{H_0}(X > 0{,}975) = \int_0^{0{,}025} 1\, dx + \int_{0{,}975}^{1} 1\, dx = 0{,}025 + 0{,}025 = 0{,}05.$$

   $\therefore$ El tamaño del test es $\alpha = 0{,}05$.

2. **Función de potencia y test inesgado:**
   La función de potencia $\beta_D(\theta)$ es la probabilidad de rechazar H_0 para un valor dado de θ:

   $$\beta_D(\theta) = P_\theta(X < 0{,}025) + P_\theta(X > 0{,}975) = \int_0^{0{,}025} \theta x^{\theta-1} dx + \int_{0{,}975}^{1} \theta x^{\theta-1} dx.$$

   Resolviendo las integrales:

   $$\beta_D(\theta) = x^\theta \Big|_0^{0{,}025} + x^\theta \Big|_{0{,}975}^{1} = 0{,}025^\theta + (1 - 0{,}975^\theta).$$

   Para $\theta = 1$ (bajo H_0): $\beta_D(1) = 0{,}05$.

   Test inesgado:
   Un test es inesgado si $\beta_D(\theta) \geq \alpha$ para todo $\theta \in H_1$. Evaluemos en $\theta = 2$:

   $$\beta_D(2) = 0{,}025^2 + (1 - 0{,}975^2) \approx 0{,}000625 + 0{,}0494 \approx 0{,}0500.$$

   Para $\theta = 0{,}5$:

   $$\beta_D(0{,}5) = \sqrt{0{,}025} + (1 - \sqrt{0{,}975}) \approx 0{,}158 + 0{,}012 \approx 0{,}170 > 0{,}05.$$

   Sin embargo, para $\theta = 15$:

   $$\beta_D(15) = 0{,}025^{15} + (1 - 0{,}975^{15}) \approx 0 + (1 - 0{,}683) \approx 0{,}317 > 0{,}05.$$

   Conclusión: El test es inesgado porque $\beta_D(\theta) \geq \alpha$ para todo $\theta \neq 1$.

Error tipo I

Se tiene una muestra $X_1, \dots, X_8$ con densidad:

$$f(x;\theta) = (\theta+1)x^{\theta}, \quad 0 \le x \le 1, \quad \theta > 0.$$

Hipótesis:

$$H_0 : \theta = 2 \quad \text{vs} \quad H_1 : \theta > 2.$$

Definimos $N = \#\{X_i \ge 0{,}88\}$ y rechazamos H_0 si $N \ge 5$.
Error tipo I (α): Probabilidad de rechazar H_0 cuando es verdadera. Esto también, es lo mismo que el tamaño del test $P_{\theta_0}(X \in C_D)$ Bajo H_0 $(\theta = 2)$:

$$P(X_i \ge 0{,}88) = \int_{0{,}88}^{1} 3x^2 dx = x^3\Big|_{0{,}88}^{1} = 1 - 0{,}88^3 \approx 0{,}3185.$$

Entonces, $N \sim \text{Binomial}(8, 0{,}3185)$. Calculamos:

$$\alpha = P(N \ge 5) = \sum_{k=5}^{8} \binom{8}{k} (0{,}3185)^k (0{,}6815)^{8-k}.$$

Test UMP y tamaño de la muestra

Como derivar test más potente

Derivar el test más potente (MP) de tamaño α para:

$$H_0 : f = f_0 \quad \text{vs} \quad H_1 : f = f_1,$$

Tenemos que pensar en f_0 y f_1 como la misma función pero con un θ distinto, en este caso tenemos que conseguir un MP para:

$$f_0(x) = 1 \quad \text{y} \quad f_1(x) = 2x, \quad 0 < x < 1.$$

La función puede ser entonces: $\theta x^{\theta-1}$ y $\theta_0 = 1$, $\theta_1 = 2$.
Aplicamos el Lema de Neyman-Pearson. La razón de verosimilitud es:

$$\Lambda(x) = \frac{f_0(x)}{f_1(x)} = \frac{1}{2x}.$$

El test rechaza H_0 si $\Lambda(x) \le k$, es decir:

$$\frac{1}{2x} \le k \implies x \ge \frac{1}{2k}.$$

Para tamaño α, elegimos k tal que:

$$P_{H_0}\left(X \ge \frac{1}{2k}\right) = \int_{\frac{1}{2k}}^{1} 1\, dx = 1 - \frac{1}{2k} = \alpha \implies \frac{1}{2k} = 1 - \alpha \implies k = \frac{1}{2(1-\alpha)}.$$

Reemplazando en k, la región crítica es:

$$C^* = \{x : x \ge 1 - \alpha\}.$$

$\therefore$ El test MP rechaza H_0 si $X \ge 1 - \alpha$.

Tamaño de la muestra y región crítica (A11 p2b)

Densidad:

$$f(x;\theta) = \frac{3}{\theta} x^2 \exp\left(-\frac{x^3}{\theta}\right), \quad x > 0, \quad \theta > 0.$$

Se sabe que $\frac{2}{\theta}\sum_{i=1}^{n} X_i^3 \sim \chi^2(2n)$.

1. **Test UMP para $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$:**
   Si $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta) = a(\theta)h(x)e^{b(\theta)t(x)}$ (familia exponencial) $\implies$ hay RVM con $\sum\limits_{i=1}^{n} t(X_i)$

- $b(\theta)$ creciente → RVM decreciente.
- $b(\theta)$ decreiente → RVM creciente.

¿Qué nos dice Karlin-Rubin?:

- Si $L(\theta)$ tiene RVM no decreciente: $C^* = \{\ T(X) \leq k\}$ es UMP para $H_0 : \theta \leq \theta_0\ H_1 : \theta > \theta_0$
- Si $L(\theta)$ tiene RVM no creciente: $C^* = \{\ T(X) \geq k\}$ es UMP para $H_0 : \theta \leq \theta_0\ H_1 : \theta > \theta_0$ (incluye $\theta = \theta_0\ \theta > \theta_1$).

En este caso $f(x;\theta) = \frac{3}{\theta}x^2 \exp(-\frac{x^3}{\theta})$

$$a(\theta) = \frac{3}{\theta}\ ,\ h(x) = x^2\ ,\ b(\theta) = -\frac{1}{\theta}\ ,\ t(x) = x^3$$

$$b(\theta) = -\frac{1}{\theta} : b'(\theta) = \frac{1}{\theta^2} > 0 \text{ creciente} = \text{RVM decreciente.}$$

$$t(x) = X^3\ ,\ T = \sum_{i=1}^{n} X^3$$

por Karlin-Rubin $C^* = \{T(X) \geq k\}$ es UMP.
Usando la distribución chi-cuadrado la región crítica es (multiplicando por $\frac{2}{\theta}$:

$$C^* = \{\frac{2T(X)}{\theta} \geq \frac{2k}{\theta}\} \to \alpha = P(\frac{2T}{\theta} \geq \frac{2k}{\theta}) \to \frac{2k}{\theta_0} = \chi^2_{2n}(1-\alpha)\ (\star)$$

$$C^* = \{\sum X^3 \geq \frac{\theta_0}{2}\chi^2_{2n}(1-\alpha)\}$$

2. **Tamaño de muestra y región crítica:**
Dado $\theta_0 = 100$, probabilidad de error tipo I, $\alpha = 0{,}05$, y $\beta_D(400) = 0{,}95$, buscamos n tal que:

$$P_{\theta=100}(T \geq k) = 0{,}05 \quad \text{y} \quad P_{\theta=400}(T \geq k) = 0{,}95.$$

Usando ($\star$):

$$\frac{2k}{100} = \chi^2_{2n}(0{,}95) \quad \text{y} \quad \frac{2k}{400} = \chi^2_{2n}(0{,}05).$$

Esto implica:

$$\chi^2_{2n}(0{,}95) = 4\chi^2_{2n}(0{,}05).$$

Resolviendo numéricamente (e.g., con Python):

$$n = 6 \quad \text{cumple} \quad \chi^2_{12}(0{,}95) \approx 21{,}03 \quad \text{y} \quad \chi^2_{12}(0{,}05) \approx 5{,}23 \quad (21{,}03 \approx 4 \times 5{,}23).$$

$n = 6$ y la región crítica es $\left\{\sum X_i^3 \geq 50 \times \chi^2_{12}(0{,}95) \approx 1051{,}5\right\}$.

Derivación de tests

Sea $X_1, \ldots, X_n \overset{iid}{\sim} Binomial(1,\theta)$ Realice test de razón de verosilimitud para probar $H_0 : \theta = \theta_0\ vs\ H_1 : \theta \neq \theta_0$

1. **Test de razón de verosimilitud** $\Lambda(x) = \frac{\sup_{\theta\in\theta_0} L(\theta)}{\sup_{\theta\in\theta_0} L(\theta)}$, $C_\Lambda = \{x \in \mathcal{X} : \Lambda(x) \leq k_\alpha\}$.

$$L(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$$

El EMV de θ: $\ell = \sum X_i log(\theta) + (n - \sum X_i) log(1-\theta)$ derivando e igualando a 0 se obtiene $\hat{\theta} = \bar{X}$.

Luego, $\Lambda(X) = \frac{\theta_0^{\sum X_i}(1-\theta_0)^{n-\sum X_i}}{\bar{X}^{\sum X_i}(1-\bar{X})^{n-\sum X_i}} = (\frac{\theta_0}{\bar{X}})^{\sum X_i}(\frac{1-\theta}{1-\bar{X}})^{n-\sum X_i}$. Entonces $C_\Lambda = \{x \in \mathcal{X} : (\frac{\theta_0}{\bar{X}})^{\sum X_i}(\frac{1-\theta}{1-\bar{X}})^{n-\sum X_i} \leq k_\alpha\}$
Para dejarlo expresado en términos de T, sea $T = \sum X_i$. Luego tomando el logaritmo de la RVG y analizando su derivada queda: $\log(\Lambda) = \log(\frac{\theta_0(n-T)}{T(1-\theta_0)})$. Con este análisis podemos llegar a la conclusión de que para $\theta_0 n > T$ RVG es creciente. Para $\theta_0 n < T$ la RVG es decreciente. ∴ la región de rechazo también puede ser descrita como $C = \{x \in \mathcal{X} : T \leq C_1 \lor T \geq C_2\}$

2. **Test RVG** Sabemos que: $T_{RV}(X) = -2\ln\Lambda(X) \xrightarrow{d} \chi^2_r$ con $\Lambda(X)$ calculada anteriormente y $r = dim(\theta = -dim(\theta_0)$.

   $-2\ln(\Lambda) = 2\ln(L(\hat{\theta}) - L(\theta_0))$, $C_\Lambda : \{x \in \mathcal{X} : -2\ln\Lambda n(X) \geq k_\alpha\}$. $T_{RV} = -2(\sum X_i \log(\frac{\theta_0}{\bar{X}}) + (n-T)\log(\frac{1-\theta_0}{1-\bar{X}})) = 2(\sum X_i \log(\frac{\bar{X}}{\theta_0}) + (n-T)\log(\frac{1-\bar{X}}{1-\theta_0}))$.

$$C_\Lambda = \{2(\sum X_i \log(\frac{\bar{X}}{\theta_0}) + (n-T)\log(\frac{1-\bar{X}}{1-\theta_0}) \geq k_\alpha\}$$

   Como $T_{RV} \overset{approx}{\sim} \chi^2_1$, $k_\alpha = \chi^2_1(1-\alpha)$.

3. **Test de razón de Wald**

$$T_W(X) = (\hat{\theta}_n - \theta_0)^T \Sigma_\theta^{-1} (\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_k$$

   Donde $\Sigma_\theta = \mathcal{I}_{\underline{X}}^{-1}(\theta)$, $\hat{\theta}_n = EMV$. $\ell'' = \frac{-\Sigma X_i}{\theta^2} - \frac{n-\Sigma X_i}{(1-\theta)^2}$, $-E(\ell'') = \frac{n}{\theta} + \frac{n}{1-\theta}$, $\mathcal{I}_{\underline{X}}(\theta) = \frac{n-n\theta+n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$

$$T_W(X) = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \overset{approx}{\sim} \chi^2_1 \quad \text{Entonces } C_W = \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_1(1-\alpha)\}$$

4. **Test Score**

$$T_S(X) = \mathcal{U}^T(\theta_0)\mathcal{I}_X^{-1}(\theta_0)\mathcal{U}(\theta_0) \overset{app}{\sim} \chi^2_k \ , \ C_S = \{x \in \mathcal{X} : T_S(X) \geq \chi^2_k(1-\alpha)\}$$

   $\mathcal{U} = \frac{\sum X_i}{\theta} - \frac{n-\sum X_i}{1-\theta}$, $\mathcal{U}(\theta_0) = \frac{\sum X_i}{\theta_0} - \frac{n-\sum X_i}{1-\theta_0} = \frac{\sum X_i - n\theta}{\theta_0(1-\theta_0)} = \frac{n(\bar{X}-\theta_0)}{\theta_0(1-\theta_0)}$. $\mathcal{I}_X^{-1}(\theta_0) = \frac{\theta_0(1-\theta_0)}{n}$, $(\frac{n\bar{X}-\theta_0}{\theta_0(1-\theta_0)})^2 = \frac{n^2(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)^2}$

$$T_S = \frac{\frac{n^2(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)^2}}{\frac{n}{\theta_0(1-\theta_0)}} = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \ , \ C_S = \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_i(1-\alpha)\}$$

5. **Test Gradiente**

$$T_{Gr} : \mathcal{U}^T(\theta_0)(\hat{\theta}_{EMV} - \theta_0) \sim \chi^2_k$$

$$T_{Gr} : \frac{n(\bar{X}-\theta_0)}{\theta_0(1-\theta_0)}(\bar{X}-\theta_0) = \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)}$$

$$T_{Gr} : \{x \in \mathcal{X} : \frac{n(\bar{X}-\theta_0)^2}{\theta_0(1-\theta_0)} \geq \chi^2_i(1-\alpha)\}$$

Como afrontar regresión lineal

Sea $y_{ij} \sim (\beta_i x_j, \sigma^2)$, $\quad i = 1, 2, j = 1, \ldots, n$,

Especificación del modelo

- Vector de respuestas y: $\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \end{pmatrix}$
- Vector de parámetros β: $\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$

- Matriz de diseño X: $\begin{pmatrix} X_1 & 0 \\ \vdots & \vdots \\ X_n & 0 \\ 0 & X_1 \\ \vdots & \vdots \\ 0 & X_n \end{pmatrix}$

Luego el modelo lineal se reescribe como:

$$y = X\beta + \epsilon \longrightarrow \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \end{pmatrix} = \begin{pmatrix} X_1\beta_1 \\ \vdots \\ X_n\beta_1 \\ X_1\beta_2 \\ \vdots \\ X_n\beta_2 \end{pmatrix} + \epsilon$$

Estimador de mínimos cuadrados y estimador insesgado de σ^2

El EMC es:

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

Luego, $(X^\top X) = \sum_{j=1}^{n} 2x_j^2 = 2\sum x_j^2$. Por otro lado $X^\top y = \sum_{j=1}^{n} x_j(y_{1j} + y_{2j})$. Por lo tanto:

$$\hat{\beta} = \frac{\sum_{j=1}^{n} x_j(y_{1j} + y_{2j})}{2\sum_{j=1}^{n} x_j^2}$$

El estimador insesgado de σ^2 es:

$$\hat{\sigma^2} = S^2 = \frac{SCE}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Donde n = número de observaciones (en este caso 2n) y p = número de parámetros $p = 1$. Luego:

$$S^2 = \frac{1}{2n-1}\sum_{j=1}^{n}(y_{1j} - \hat{\beta}x_j)^2 + (y_{2j} - \hat{\beta}x_j)^2$$

EMC de otro modelo

SUponga que $n+1$ observaciones $y_0, \dots, y_n$ siguen el modelo lineal simple:

$$y_i \sim (\beta_0 + \beta_1, i, \sigma^2), \quad i = 0, 1, \dots, n$$

En forma matricial:

$$\begin{pmatrix} y_0 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + 0\beta_1 \\ \vdots \\ \beta_0 + n\beta_1 \end{pmatrix} + \epsilon$$

Luego:

$$X^\top X = \begin{pmatrix} n+1 & \sum_{i=0}^{n} i \\ \sum_{i=0}^{n} i & \sum_{i=0}^{n} i^2 \end{pmatrix} \quad X^\top y = \begin{pmatrix} \sum_{i=0}^{n} Y_i \\ \sum_{i=0}^{n} iY_i \end{pmatrix}$$

De las slides tenemos:
CASO:

$$X^\top X = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \quad X^\top y = \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

y

$$(X^\top X)^{-1} = \begin{pmatrix} (\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}) & -\frac{\bar{x}}{S_{xx}} \\ -\frac{\bar{x}}{S_{xx}} & \frac{1}{S_{xx}} \end{pmatrix}$$

Donde $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$. Luego,

$$\hat{\beta} = (X^\top X)^{-1}X^\top y = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \text{ , con } \hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x} \text{ y } \hat{\beta}_2 = S_{xy}/S_{xx}, \text{ donde } S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

En nuestro caso, entonces tenemos: $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=0}^{n}(i-\bar{i})(y_i-\bar{y})}{\sum_{i=0}^{n}(i-\bar{i})^2}$, con $\bar{i} = \frac{n}{2}$.

Probar que estimador es insesgado

Se considera el estimador alternativo para $\beta_1 = \tilde{\beta}_1 = \frac{y_n - y_0}{n}$.

Tenemos que tomar la esperanza:

$$E(\tilde{\beta}_1) = \frac{E(y_n) - E(y_0)}{n} = \frac{\beta_0 + n\beta_1 - \beta_0}{n} = \beta_1$$

Es insesgado.

Calculamos su varianza:
Como $y_o \sim (\beta_0, \sigma^2)$ y $y_o \sim (\beta_0 + \beta_1 n, \sigma^2)$ son independientes, se tiene:

$$Var(\tilde{\beta}_1) = \frac{Var(y_n) + Var(y_0)}{n^2} = \frac{2\sigma^2}{n^2}$$

Comparación de estimadores

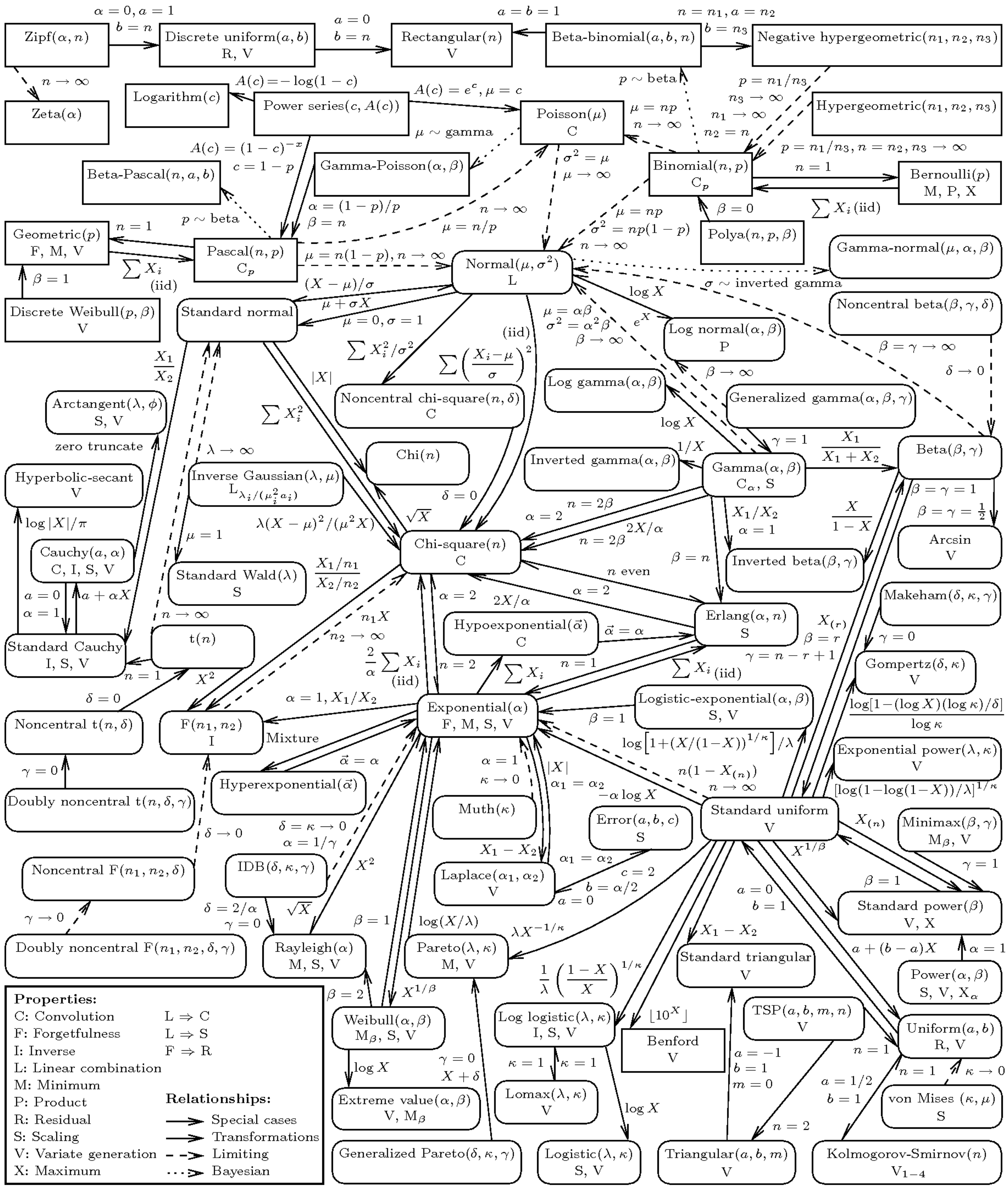
Calculamos ahora la varianza de $\hat{\beta}_1 = \frac{\sigma^2}{\sum_{i=0}^{n}(i-\frac{n}{2})^2}$. Luego tenemos que comparar (haciendo un poco de algebrita y recordando $\sum_{i=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}$):

$$\frac{12\sigma^2}{n(n+1)(n+2)} \text{ con } \frac{2\sigma^2}{n^2}$$

$$\rightarrow (n^2 + n)(n+2) \ vs \ 6n^2 \rightarrow n^3 + 3n^2 + 2n \ vs \ 6n^2 \rightarrow n^3 - 3n^2 + 2n > 0?$$

$$\text{con } n = 0 : 8 - 12 + 4 = 0 \text{ , con } n > 2 : n^3 - 3n^2 + 2n > 0$$

Elegir $\hat{\beta}_1$.



Zipf(α, n)
Discrete uniform(a, b) R, V
Rectangular(n) V
Beta-binomial(a, b, n)
Negative hypergeometric(n1, n2, n3)
Zeta(α)
Logarithm(c)
Power series(c, A(c))
Poisson(μ) C
Hypergeometric(n1, n2, n3)
Beta-Pascal(n, a, b)
Gamma-Poisson(α, β)
Binomial(n, p) Cp
Bernoulli(p) M, P, X
Geometric(p) F, M, V
Pascal(n, p) Cp
Polya(n, p, β)
Normal(μ, σ2) L
Gamma-normal(μ, α, β)
Discrete Weibull(p, β) V
Standard normal
Noncentral beta(β, γ, δ)
Log normal(α, β) P
Arctangent(λ, φ) S, V
Noncentral chi-square(n, δ) C
Log gamma(α, β)
Generalized gamma(α, β, γ)
Hyperbolic-secant V
Chi(n)
Inverted gamma(α, β)
Gamma(α, β) Cα, S
Beta(β, γ)
Inverse Gaussian(λ, μ)
Cauchy(a, α) C, I, S, V
Standard Wald(λ) S
Chi-square(n) C
Inverted beta(β, γ)
Arcsin V
Makeham(δ, κ, γ)
Standard Cauchy I, S, V
t(n)
Hypoexponential(α⃗) C
Erlang(α, n) S
Gompertz(δ, κ) V
Noncentral t(n, δ)
F(n1, n2) I
Exponential(α) F, M, S, V
Logistic-exponential(α, β) S, V
Exponential power(λ, κ) V
Hyperexponential(α⃗)
Doubly noncentral t(n, δ, γ)
Muth(κ)
Error(a, b, c) S
Standard uniform V
Minimax(β, γ) Mβ, V
Noncentral F(n1, n2, δ)
IDB(δ, κ, γ)
Laplace(α1, α2) V
Standard power(β) V, X
Doubly noncentral F(n1, n2, δ, γ)
Rayleigh(α) M, S, V
Pareto(λ, κ) M, V
Standard triangular V
Power(α, β) S, V, Xα
Weibull(α, β) Mβ, S, V
Log logistic(λ, κ) I, S, V
Benford V
TSP(a, b, m, n) V
Uniform(a, b) R, V
Extreme value(α, β) V, Mβ
Lomax(λ, κ) V
von Mises (κ, μ) S
Generalized Pareto(δ, κ, γ)
Logistic(λ, κ) S, V
Triangular(a, b, m) V
Kolmogorov-Smirnov(n) V1−4
Properties:
C: Convolution
F: Forgetfulness
I: Inverse
L: Linear combination
M: Minimum
P: Product
R: Residual
S: Scaling
V: Variate generation
X: Maximum
L ⇒ C
L ⇒ S
F ⇒ R
Relationships:
Special cases
Transformations
Limiting
Bayesian