INFERENCIA ESTADÍSTICA II

Apuntes de la asignatura

Autores
Victor Elvira Fernández, Tomás Ruiz Rojo, Juan Horrillo Crespo
Universidad de Valladolid

18 de enero de 2025

AVISO

Estos apuntes fueron creados de forma voluntaria por un grupo de estudiantes, invirtiendo tiempo, dedicación y esfuerzo para ofrecer información útil a la comunidad. Apreciamos cualquier apoyo que se nos quiera brindar, ya que nos ayuda a continuar con futuros proyectos de este tipo.

Si deseas colaborar en esta clase de proyectos puedes contactarnos y unirte o invitarnos a unas ricas patatas 5 salsas por el siguiente enlace:

Buy Me a Patatas 5 Salsas

https://www.buymeacoffee.com/ApuntesINdat

- Mail Juan Horrillo
- Mail Victor Elvira
- Mail Tomás Rojo

Si has colaborado de cualquier forma te agradecemos enormemente.

Índice

Tema 1

Propiedades asintóticas del EMV y del estadístico RV

Uso y propiedades asintóticas del estimador de máxima verosimilitud (EMV) y del estadístico de razón de verosimilitud (RV).

1.1. Estimadores en muestras grandes

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes igualmente distribuidas (v.a.i.i.d.) tal que $P_\theta : \theta \in \Theta$. Tenemos nuestra muestra aleatoria simple (m.a.s) con n grande y nos interesa estimar θ (o $g(\theta)$).

Cuando el tamaño de la muestra aumenta, también aumenta la información disponible de θ (o $g(\theta)$), por lo tanto se espera que la estimación sea más precisa. Esto podemos comprobarlo al analizar la distribución de la media muestral ya que como $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, entonces $Var(\overline{X}) \xrightarrow{n\to\infty} 0$.

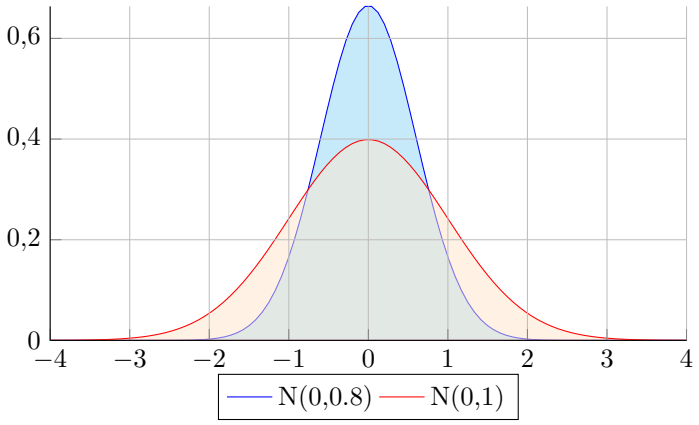

Figura 1: Comparación aumento de n

Consideraremos que un estimador $T(X_1, \ldots, X_n)$ es razonable si cuando n aumenta, el estimador $T(X_1, \ldots, X_n)$ es más preciso. Lo que podemos esperar es que $T(X_1, \ldots, X_n)$ esté próximo a θ con mayor probabilidad.

A continuación, se definirán los criterios que nos permitirán comprobar la calidad de información que nos proporciona un estimador.

1.1.1. Consistencia de un estimador

Se dice que un estimador $T(X_1, \ldots, X_n)$ es consistente si cumple que

$$\forall \varepsilon > 0, \delta > 0 \quad \exists N : n \geq N$$

$$P_\theta(|T(X_1, \ldots, X_n) - \theta| < \varepsilon) \geq 1 - \delta$$

Definición: $T_n(X)$ es un estimador consistente para θ (o $g(\theta)$) si

$$\forall \varepsilon > 0 \quad P_\theta\left(|T(X_1, \ldots, X_n) - \theta| \leq \varepsilon\right) \xrightarrow{n\to\infty} 1$$

es decir, si:

$$T_n(x) \underset{n\to\infty}{\overset{p}{\to}} \theta \quad \text{o} \quad \lim_{n\to\infty} \forall \varepsilon > 0 \quad P_\theta(|T(X_1, \ldots, X_n) - \theta| \leq \varepsilon) = 1$$

Existen algunas estrategias para comprobar si un estimador es consistente:

1. **Convergencia en probabilidad**. Si $T_n(x) \underset{n\to\infty}{\overset{p}{\to}} \theta$ y $G_n(x) \underset{n\to\infty}{\overset{p}{\to}} \theta'$
   - $T_n(x) + G_n(x) \underset{n\to\infty}{\overset{p}{\to}} \theta + \theta'$
   - $T_n(x) \cdot G_n(x) \underset{n\to\infty}{\overset{p}{\to}} \theta \cdot \theta'$
   - Si $\theta' \neq 0$, $\frac{T_n(x)}{G_n(x)} \underset{n\to\infty}{\overset{p}{\to}} \frac{\theta}{\theta'}$
2. **Ley débil de los grandes números**. $X_1, X_2, \ldots, X_n$ i.i.d. con $E(X_i) = \mu < \infty$

   $$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{p}{\to}} \mu$$

   Los momentos muestrales son estimaciones consistentes de los correspondientes momentos poblacionales.
3. **Convergencia en media cuadrática**. Un estimador $T(X_1, \ldots, X_n) \overset{\text{m.c.}}{\xrightarrow[n\to\infty]{}} \theta$ si:
   - $E(T(X_1, \ldots, X_n)) \xrightarrow{n\to\infty} \theta$, sesgo de $T(X_1, \ldots, X_n) \xrightarrow{n\to\infty} 0$
   - $Var(T(X_1, \ldots, X_n)) \xrightarrow{n\to\infty} 0$

   Es decir que si ECM $T(X_1, \ldots, X_n) \xrightarrow{n\to\infty} 0$, el estimador converge en media cuadrática. Un estimador insesgado sería consistente si su varianza converge a 0 con $n \to \infty$. La convergencia en media cuadrática es más fuerte que la convergencia en probabilidad.

Ejercicio 1.1. Sean $X_1, \ldots, X_n$ v.a.i.i.d. con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$, $\forall i = 1, \ldots, n$. ¿Es consistente $T_n(x) = \frac{X_1 + X_2 + \cdots + X_n}{\frac{n}{2}}$?
Usando la ley de los grandes números:

$$\frac{1}{n}\sum X_i \xrightarrow{\text{P}} \mu \quad T_n(x) \xrightarrow{P} 2\mu$$

Resultado: El estimador $T_n(x)$ NO es consistente

La consistencia por si sola no es tan interesante, ya que si T_n es consistente para θ, nos dice que para n grande los errores serán pequeños pero no nos permite conocer el orden del error $\left(\frac{1}{n}, \frac{1}{\sqrt{n}}, \frac{1}{\log(n)}, \dots\right)$

Si $\{K_n\}_{n=1}^{\infty}$ es una sucesión de reales positivos y $\varepsilon > 0$ definimos $P_n(\varepsilon)$ como

$$P_n(\varepsilon) = P_\theta\left(|T_n(x) - \theta| \leq \frac{\varepsilon}{K_n}\right)$$

Habiendo definido $P_n(\varepsilon)$, ¿qué pasará con $P_n(\varepsilon)$ cuando n sea grande?

1. Si K_n crece "lentamente"(por ejemplo: $K_n = \log(n)$), el error disminuye "lentamente.ª medida que aumenta n $P_n(\varepsilon) \xrightarrow{\text{n}\to\infty} 1$. Si K_n crece lentamente, $\frac{\varepsilon}{K_n}$ es más grande, lo que facilitará que el error esté por debajo del umbral.

2. Si K_n crece rápido"(por ejemplo: $K_n = n$), el error disminuye más rápido a medida que aumenta n $P_n(\varepsilon) \xrightarrow{\text{n}\to\infty} 0$. Si K_n crece rápido, $\frac{\varepsilon}{K_n}$ se hace muy pequeño y se hace muy difícil que el error sea tan pequeño.

3. Casos intermedios. Si K_n crece .ªdecuadamente", $P_n(\varepsilon) \xrightarrow{\text{n}\to\infty} H(\varepsilon) \in (0,1)$. Decimos que el error converge a 0 a velocidad $\frac{1}{K_n}$

De manera resumida:

$$P_n(\varepsilon) = P_\theta(K_n|T_n(x) - \theta| \leq \varepsilon) \xrightarrow{n\to\infty} \left\{\begin{array}{c} 0 \text{ si } K_n \xrightarrow{\infty} \text{rápido} \\ 0 \leq P_n(\varepsilon) \leq 1 \text{ si } K_n \xrightarrow{\infty} \text{adecuadamente} \\ 1 \text{ si } K_n \xrightarrow{\infty} \text{lento} \end{array}\right\}$$

La idea es que al multiplicar K_n, se amplifica la velocidad de convergencia de los errores a 0. Si elegimos K_n adecuadamente, de forma que $P_n(\varepsilon)$ sea menor que 1, podemos controlar la velocidad a la que los errores tienden a 0, mejorando la precisión.

1.1.2. Estimador Consistente Asintoticamente Normal (CAN)

Definición: Sean $X_1, \dots, X_n$ v.a.i.i.d. con distribución P_θ donde $\theta \in \Theta$, entonces un estimador $T_n(X) = T(X_1, \dots, X_n)$ con $n \geq 1$ será CAN para θ si

$$\exists \sigma_n(\theta) > 0, \theta \xrightarrow{n\to\infty} 0 \quad \text{tal que} \quad \frac{T_n(X) - \theta}{\sigma_n(\theta)} \xrightarrow{L} N(0,1)$$

En muchas ocasiones, $\sigma_n^2(\theta)$ es de la forma $\frac{Var_\theta(T_n(X))}{n}$ y se llama varianza asintótica del estimador. En este caso $T_n(X)$ es CAN para θ si

$$\sqrt{n}(|T_n(X) - \theta|) \xrightarrow{L} N(0, Var_\theta(T_n(X)))$$

donde los errores convergen a 0 con velocidad $\frac{1}{\sqrt{n}}$.

Al igual que la LGN es clave para obtener estimadores consistentes, el TCL lo es para obtener estimadores CAN.

Sean $X_1, \ldots, X_n$ v.a.i.i.d. tal que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$. Entonces

- $\frac{\sum X_i - E(\sum X_i)}{\sqrt{Var(\sum X_i)}} \xrightarrow{L} N(0,1)$
- $\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \xrightarrow{L} N(0,1)$
- $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Resulta útil aplicar el Δ-método cuando queremos aproximar la distribución de una función no lineal de un estimador que es asintoticamente normal.
Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde su distribución es P_θ tal que $\theta \in \Theta \subseteq \mathbb{R}$ y sea una función $g : \Theta \to \mathbb{R}$ con derivada no nula. Si $T_n(X)$ es CAN para θ tal que $\sqrt{n}(|T_n(X) - \theta|) \xrightarrow{L} N(0, V_t^2(\theta))$ entonces

$$\sqrt{n}(|g(T_n(X)) - g(\theta)|) \xrightarrow{L} N(0, V_t^2(\theta)g'(\theta)^2)$$

Ejercicio 1

Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde $E(X_i) = \mu, \quad Var(X_i) = \sigma^2 \quad \forall i = 1 \ldots n$. Determinar si los siguientes estimadores son consistentes

b) $T(X_1, \ldots, X_n) = \frac{T(X_1 + \cdots + X_{\frac{n}{2}})}{\frac{n}{2}}$
Vemos que

$$T(X_1, \ldots, X_{\frac{n}{2}}) = \frac{\sum_{i=1}^{\frac{n}{2}}}{\frac{n}{2}} = \bar{X}_{\frac{n}{2}} \xrightarrow{P} \mu$$

Por tanto el estimador es consistente

c) $T(X_1, \ldots, X_n) = X_1$
El estimador no es consistente ya que X_1 no depende de n. Lo mínimo para que pueda converger es que dependa de n.

d) $T(X_1, \ldots, X_n) = 2\sum_{i=1}^{n} \frac{iX_i}{n(n+1)}$
No podemos usar la ley de los grandes números porque $X_1, 2X_2, 3X_3, \ldots$ no son v.a.i.i.d. por tanto usaremos la convergencia en media cuadrática.

$$T_n(x) \xrightarrow{m.c.} \mu \begin{cases} E(T_n(X)) \xrightarrow[n\to\infty]{} \mu \\ Var(T_n(X)) \xrightarrow[n\to\infty]{} 0 \end{cases}$$

$$E(T_n(x)) = \frac{2}{n(n+1)}E((\sum_{i=1}^{n} i)X_i) = \frac{2}{n(n+1)}(\sum_{i=1}^{n} i)E(X_i) =$$

$$= \frac{2\mu}{n(n+1)}\sum_{i=1}^{n} i = \frac{2\mu}{n(n+1)}\frac{n(n+1)}{2} = \mu$$

Se cumple que la media tiende a μ cuando n tiende a infinito.

$$Var(T_n(x)) = \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} Var(iX_i) = \frac{4}{n^2(n+1)^2}\sigma^2\sum_{i=1}^{n} i^2$$

$$= \frac{4}{n^2(n+1)^2}\sigma^2\frac{n(n+1)(2n+1)}{6} = \frac{2}{3}\frac{2n+1}{n^2+n}\sigma^2 \xrightarrow[n\to\infty]{} 0$$

Como se cumple que la media tiende a 0 cuando n tiende a infinito el estimador es consistente.

Ejercicio 3 Sean $X_1, \ldots, X_n$ v.a.i.i.d. con distribución uniforme $U(0, \theta)$, ¿$2\bar{X}$ es consistente para θ?

$$E(X_i) = \frac{a+b}{2} = \frac{\theta}{2}; \quad Var(X_i) = \frac{(b-a)^2}{12}$$

Entonces

$$\bar{X} \xrightarrow[n\to\infty]{P} \Longrightarrow \; 2\bar{X} \xrightarrow{P} \theta$$

Por tanto $2\bar{X}$ es consistente para θ

1.1.3. Información de Fisher

La información de Fisher $I_x(\theta)$ es la matriz que mide la cantidad de información que una m.a.s. contiene sobre el estimador.

Definición: Sea $X = (X_1 \ldots X_n)$ i.i.d. de distribución $P_\theta \in P = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ con función de densidad $f(X, \theta)$ y en la que existe $\frac{df(x,\theta)}{d\theta}$, la información de Fisher sobre θ contenida en X es:

$$I_x(\theta) = Var_\theta(S(\theta, X))$$

$$Score = S_x(\theta) = S(\theta, X) = \frac{d}{\mathrm{d}\theta} \log f(x, \theta)$$

1.1.4. Condiciones de regularidad de Cramer-Rao (CRCR)

Llamaremos familias regulares a aquellas familias en las que se verifican las condiciones de regularidad de Cramer-Rao. Estas son las familias con las que trabajaremos.

Condiciones de regularidad de Cramer-Rao:

1. El espacio paramétrico es un intervalo de $\mathbb{R}$.

2. El soporte de la distribución no depende del parámetro θ. Por ejemplo $x = \{x : f(x, \theta) > 0\}$, no depende de θ y sería regular. En cambio U(0,θ) tiene un soporte que depende de θ, por lo que no es regular

3. Se pueden calcular las dos primeras derivadas bajo el signo integral. Además se puede intercambiar la derivada con el signo integral.

$$\frac{d}{d\theta} \int_x f(x, \theta)\, dx = \int_x \frac{d}{d\theta} f(x, \theta)\, dx$$

4. $T_n(x)$ es un estimador insesgado para θ o $g(\theta)$.

Bajo las condiciones de regularidad de Cramer-Rao, podemos definir la cantidad de información esperada como:

$$I_x(\theta) = E_\theta(S(\theta, X)^2) = E_\theta\left(\left(\frac{d}{d\theta} \log f(x, \theta)\right)^2\right)$$

Demostración 1.2. Deberemos probar que $E_\theta(\left(\frac{d}{d\theta}\log f(x,\theta)\right)) = 0$. Si lo conseguimos entonces, $Var(S(\theta,x)) = E(S(\theta,x)^2) - E(S(\theta,x))^2 = E(S(\theta,x)^2)$

$$E_\theta\left(\left(\frac{d}{d\theta}\log f(x,\theta)\right)\right) = \int_x E_\theta\left(\left(\frac{d}{d\theta}\log f(x,\theta)\right)\right) f(x,\theta)\,dx =$$

$$= \int_x \frac{\frac{d}{d\theta}f(x,\theta)}{f(x,\theta)} f(x,\theta)\,dx = \int_x \frac{d}{d\theta}f(x,\theta)\,dx = 0$$

Si se verifican las condiciones de regularidad de Cramer-Rao, otra forma alternativa de calcular la información de Fisher es:

$$I_x(\theta) = -E\left(\frac{d^2}{d\theta^2}\log f(x,\theta)\right)$$

Demostración 1.3. Breve paso previo:

$$\frac{d}{d\theta}\left(\frac{d}{d\theta}\log f(x,\theta)\right) = \frac{d}{d\theta}\left(\frac{\frac{d}{d\theta}f(x,\theta)}{f(x,\theta)}\right) =$$

donde derivando el cociente

$$= \frac{\frac{d^2}{d\theta}f(x,\theta)f(x,\theta) - \left(\frac{d}{d\theta}f(x,\theta)\right)^2}{f(x,\theta)^2} = \frac{\frac{d^2}{d\theta}f(x,\theta)}{f(x,\theta)} - \left(\frac{\frac{d}{d\theta}f(x,\theta)}{f(x,\theta)}\right)^2$$

Demostración:

$$E\left(\frac{d^2}{d\theta}\log f(x,\theta)\right) = \int_x \frac{d}{d\theta}\left(\frac{d}{d\theta}\log f(x,\theta)\right) f(x,\theta)\,dx =$$

$$= \int_x \frac{d^2}{d\theta}f(x,\theta)\,dx - \int_x \left(\frac{\frac{d}{d\theta}f(x,\theta)}{f(x,\theta)}\right)^2 f(x,\theta)\,dx$$

$$= -\int_x \left(\frac{\frac{d}{d\theta}f(x,\theta)}{f(x,\theta)}\right)^2 f(x,\theta)\,dx = -I_x(\theta)$$

Propiedades:

Sean X e Y dos variables independientes de la misma familia de distribuciones

$X \sim P_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}, \quad f(x,\theta), I_x(\theta)$

$Y \sim Q_\theta, \quad \theta \in \Theta \subseteq \mathbb{R}, \quad g(y,\theta), I_y(\theta)$

Entonces

1. Propiedad de la información de Fisher conjunta: $I_{xy}(\theta) = I_x(\theta) + I_y(\theta)$

   Demostración 1.4. Por independencia

   $$f_{xy}(x,y,\theta) = f_x(x,\theta) f_y(y,\theta)$$

   $$I_{xy}(\theta) = Var\left(\frac{d}{d\theta}\log(f_x(x,\theta)f_y(y,\theta))\right) = Var\left(\frac{d}{d\theta}(\log f_x(x,\theta) + \log f_y(y,\theta))\right)$$

   Y por las propiedades de la varianza: $Var(X+Y) = Var(X) + Var(Y)$ si X e Y son independientes

   $$= Var\left(\frac{d}{d\theta}(\log f_x(x,\theta))\right) + Var\left(\frac{d}{d\theta}(\log f_y(y,\theta))\right) = I_x(\theta) + I_y(\theta)$$

2. Sean $X_1, \ldots, X_n$ m.a.s. v.a.i.i.d. $P_\theta, \theta \in \Theta$ con $f(x,\theta)$ de una familia regular:

   $$I_{X_1,\ldots,X_n} = nI_{X_1}(\theta)$$

Ejemplo con la distribucion de Bernoulli:

Sea $X \sim B(p) \to f(x,p) = p^x(1-p)^{1-x}$ donde $x = \{0,1\}$.

Usando la primera definición de $I_x(p)$:

$$I_x(p) = Var((S_x(p))) = Var\left(\frac{d}{dp}\log f(x,p)\right)$$

donde

$$\log f(x,p) = x\log p + (1-x)\log(1-p) \implies S_x(p) = \frac{d}{dp}(x\log(p) + (1-x)\log(1-p))$$

por tanto

$$Var(S_x(p)) = Var\left(\frac{x-p}{p(1-p)}\right) = \frac{Var(x)}{p^2(1-p)^2} = \frac{p(1-p)}{p^2(1-p)^2} = \frac{1}{p(1-p)}$$

Ejemplo con la distribucion de Poisson:

Sea $X \sim Poisson(\lambda) \to f(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$ donde $k \in \mathbb{N}_0$.

Usando la primera definición de $I_x(p)$:

$$I_x(\lambda) = Var(S_x(\lambda)); \quad S_x(\lambda) = \frac{d}{d\lambda}\log\left(\frac{\lambda^x e^{-\lambda}}{x!}\right) = \frac{x-\lambda}{\lambda}$$

$$I_x(\lambda) = Var\left(\frac{x-\lambda}{\lambda}\right) = \frac{1}{\lambda^2}Var(x-\lambda) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

Ejemplo con n muestras de la distribucion de Bernoulli:

Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde $X_i \sim B(p)$ entonces la densidad conjunta será

$$f(X_1, \ldots, X_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

como

$$\log f(x_1, ..., x_n) = (\sum_{i=1}^{n} x_i) \log p + (n - \sum_{i=1}^{n} x_i) \log(1-p)$$

entonces

$$S_{x_1,\ldots,x_n}(p) = \frac{d}{dp} \log f(X_1, ..., X_n) = \frac{\sum_{i=1}^n x_i - np}{p(1-p)}$$

y finalmente

$$Var\left(\frac{\sum_{i=1}^n x_i - np}{p(1-p)}\right) = \frac{\sum_{i=1}^n (Var(x_i))}{p^2(1-p)^2} = \frac{n}{p(1-p)} I_{x_1,\ldots,x_n}(p) = nI_{x_1}(p) = \frac{n}{p(1-p)}$$

Ejemplo con la exponencial:

Sea $X \sim exp(\lambda) \to f(x, \lambda) = \lambda e^{-\lambda x}$ simplemente como

$$\log f(x, \theta) = \log \lambda - \lambda X \implies S_x(\lambda) = \frac{1}{\lambda} - X$$

entonces

$$I_x(\lambda) = -E_\lambda\left(\frac{d}{d\lambda} S_x(\lambda)\right) = -E_\lambda\left(\frac{-1}{\lambda^2}\right) = \frac{1}{\lambda^2}$$

Vamos a profundizar un poco en las condiciones de regularidad de Cramer-Rao:

Sea X=$(X_1, \ldots, X_n)$ m.a.s de $P_\theta, \theta \in \Theta \subseteq \mathbb{R}$, con densidad f(x,$\theta$) y con $T_n(x)$ como estimador insesgado de g(θ).

$$E(T_n(x) = g(\theta)) \to g(\theta) = \int_x T_n(x) f(x, \theta)\, dx$$

$g(\theta)$ es diferenciable respecto a θ bajo el signo integral y podemos intercambiar derivada e integral

1.1.5. Cota de Cramer-Rao

Sea $X, \ldots, X_n$ m.a.s $P_\theta, \theta \in \Theta \subseteq \mathbb{R}$ con densidad $f(x,\theta)$ y sea $T_n(x)$ un estimador insesgado de $g(\theta)$ con $\text{Var}(T_n(x)) < \infty$ y que verifica las condiciones de regularidad de Cramer-Rao, entonces:

$$Var(T_n(x)) \geq \frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)}$$

Demostración 1.5. Consideremos $Cov(T_n(x), S_x(\theta)) = E(T_n(x) \cdot S_x(\theta))$

$$Cov(T_n(x)S_x(\theta)) = \int_x T_n(x) \cdot S_x(\theta) \cdot f(x,\theta)\, dx$$

$$\int_x T_n(x) \cdot \frac{d}{d\theta} \log f(x,\theta) \cdot f(x,\theta)\, dx = \int_x T_n(x) \cdot \frac{\frac{d}{d\theta} f(x,\theta)}{f(x,\theta)} \cdot f(x,\theta)\, dx$$

$$\int_x T_n(x) \cdot \frac{d}{d\theta} f(x,\theta)\, dx = \frac{d}{d\theta} \int_x T_n(x) f(x,\theta)\, dx = \frac{d}{d\theta} E(T_n(x))$$

Como $T_n(x)$ es insesgado:

$$= \frac{d}{d\theta} g(\theta) = g'(\theta)$$

Usamos la desigualdad de Cauchy-Schwarz que relaciona varianza y covarianza

$$cov(X,Y) = \sqrt{Var(X) \cdot Var(Y)}$$

entonces,

$$(g'(\theta)^2) = (cov(T_n(x) \cdot S_x(\theta)))^2 \leq Var(T_n(x)) \cdot Var(S_x(\theta)) = Var(T_n(x)) \cdot n \cdot I_{x_1}(\theta)$$

$$Var(T_n(x)) \geq \frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)}$$

La varianza de un estimador insesgado es como mínimo $\frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)}$

Nota: Bajo las mismas condiciones de regularidad de Cramer-Rao puede estimarse en el caso que $T_n(x)$ sea un estimador de θ con $Var(T_n(x)) < \infty$

$$Var(T_n(x)) \geq \frac{1}{n \cdot I_{x_1}(\theta)} = \frac{1}{I_n(\theta)}$$

La demostración es análoga a la anterior.

Ejemplo:

$X_1, \ldots, X_n \sim B(p) \quad p \in (0,1) \quad I_n(\theta) = n \cdot I_{x_1}(\theta) = \frac{n}{p(1-p)}$

Si quiero estimar p, ¿cuál es la cota CR para cualquier estimador insesgado de P?

$Var(T_n(x)) \geq \frac{1}{n \cdot I_{x_1}(p)} = \frac{p(1-p)}{n}$

Supongamos que nuestro estimador para p es $\bar{X}$

$Var(\frac{\sum_{i=1}^n}{n}) = \frac{1}{n^2} Var(\sum_{i=1}^n x_i) = \frac{n \cdot p(1-p)}{n^2} = \frac{p(1-p)}{n}$

Hemos visto que,como consecuencia del Teorema Central del Limite, la velocidad de convergencia de los errores a cero es del orden $\frac{1}{\sqrt{n}}$

Si $X_1, \ldots, X_n$ i.i.d. $P_\theta \in P, \theta \in \Theta \subseteq \mathbb{R}$ y f(x, θ) verifica las condiciones de regularidad de Cramer-Rao, es decir, si la familia es regular.

$\sqrt{n}(T_n(x) - g(\theta)) \xrightarrow{L} N(0, V_T^2(\theta))$

donde $V_T^2(\theta)$ es la varianza asintótica del estimador y $\sqrt{n}$ es la normalización adecuada como consecuencia del TCL.

Es equivalente, $T_n(x) \simeq N(g(\theta), \frac{V_T^2(\theta)}{n})$

Bajo esas condiciones de regularidad, la varianza $\frac{V_T^2(\theta)}{n}$ cumple también la cota de Cramer-Rao

Ejemplo:

$X_1, \ldots, X_n \quad i.i.d \quad U(0, \theta) \quad \theta > 0$

$f(x, \theta) = \frac{1}{\theta}$ (Ejercicio 3 apartado b)

Se demuestra que $X_{(n)}$ es un estimador consistente para θ, pero la velocidad de convergencia no es $\frac{1}{\sqrt{n}}$

1.1.6. Estimador Consistente Asintóticamente Normal (CAN) y Asintóticamente Eficiente (AE)

Sea $X_1, \ldots, X_n$ m.a.s. de $P_\theta, \theta \in \Theta \subseteq \mathbb{R}$ con densidad $f(x, \theta)$ y $T_n(x)$ estimador insestado de g(θ) con una Var$(T_n(x)) < \infty$ y que verifica las condiciones de regularidad de Cramer-Rao, entonces:

$$Var(T_n(x)) \geq \frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)}$$

es decir,

$$\frac{V_T^2(\theta)}{n} \geq \frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)} \implies V_T^2(\theta) \geq \frac{(g'(\theta))^2}{I_{x_1}(\theta)}$$

Concretamente un estimador CAN será mejor cuanto menor sea la varianza asintótica. Si la varianza del estimador original es igual a la cota de Cramer-Rao, decimos que $T_n(x)$ es CAN y asintóticamente eficiente (AE)

Definición: $X_1, \ldots, X_n$ m.a.s.$P_\theta, \theta \in \Theta \subseteq \mathbb{R}, f(x, \theta), T_n(x)$ es un estimador CAN y AE de g(θ) si es CAN y $V_T^2 = \frac{(g'(\theta))^2}{I_{x_1}(\theta)}$, es decir si

$$\sqrt{n}(T_n(x) - g(\theta)) \xrightarrow{L} N\left(0, \frac{(g'(\theta))^2}{I_{x_1}(\theta)}\right)$$

Ejemplo:

$X \sim P(\lambda) \quad \lambda > 0 \quad f(x, \theta) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad x = 0, 1, 2, \ldots$

$E(X) = \lambda \quad Var(X) = \lambda$

$\log F(x, \theta) = -\lambda + x \cdot \log \lambda - \log x!$

Buscamos un estimador CAN y AE para λ.

Por el teorema central del límite, la media muestral es CAN para la media poblacional.

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{L} N(0, \lambda) \quad o \quad \bar{X} \simeq N(\lambda, \frac{\lambda}{n})$$

Para comprobar si $\bar{X}$ es AE, debemos comprobar que $\frac{\lambda}{n}$ es igual a $\frac{1}{n \cdot I_{x_1}(\lambda)}$

$$S_x(\lambda) = \frac{x - \lambda}{\lambda}, \quad I_{x_1} = \frac{1}{\lambda}$$

La varianza de $\bar{X}$ coincide con la cota de Cramer-Rao, por lo que es un estimador asintóticamente eficiente $\left(\frac{1}{n \cdot I_{x_1}(\lambda)} = \frac{1}{n \cdot \frac{1}{\lambda}} = \frac{\lambda}{n}\right)$

Ejemplo:

$X \sim exp(\lambda) \quad f(x, \lambda) = \lambda \cdot e^{-\lambda \cdot x} \quad x > 0, \quad \lambda > 0$

$E(X) = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2}$

$\log f(x, \lambda) = \log \lambda - \lambda \cdot X$

$S(\lambda, x) = \frac{1}{\lambda} - X$

$\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \xrightarrow{L} N(0, \frac{1}{\lambda^2})$

$\bar{X} \simeq N(\frac{1}{\lambda}, \frac{1}{n \cdot \lambda^2})$

Sabemos que $\bar{X}$ es CAN para $\frac{1}{\lambda}$, es decir para $g(\lambda) = \frac{1}{\lambda}; g'(\lambda) = \frac{-1}{\lambda^2}$.

Queremos ver si tambien es AE para la media poblacional. Tenemos que comprobar que $\frac{V_T^2(\lambda)}{n} = \frac{1}{\lambda^2}$ coincida con la cota CR.

$$I_{x_1} = \frac{1}{\lambda^2} \quad \frac{(g'(\theta))^2}{n \cdot I_{x_1}(\theta)} = \frac{(\frac{-1}{\lambda^2})^2}{n \cdot \frac{1}{\lambda^2}} = \frac{1}{n \cdot \lambda^2}$$

¿Y como obtendríamos un estimador AE para λ? Utilizando el delta método.

Si $T_n(x)$ es CAN para $\theta \in \Theta \subseteq \mathbb{R}$ y g es una función con derivada no nula, entonces $g(T_n(x))$ es CAN para $g(\theta)$ con varianza $\frac{V_T^2}{n} \cdot (g'(\theta))^2$

$$\sqrt{n}(T_n(x) - \theta) \xrightarrow{L} N(0, V_T^2(\theta))$$

$$\sqrt{n}(g(T_n(x)) - g(\theta)) \xrightarrow{L} N(0, V_T^2(\theta) \cdot (g'(\theta))^2)$$

Volviendo al ejemplo, tras esta breve parte teórica, ya vimos que $\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \xrightarrow{L} N(0, \frac{1}{\lambda})$ Consideraremos que $\theta = \frac{1}{\lambda^2}$.

Pasos:

1. $\sqrt{n}(\bar{X} - \theta) \xrightarrow{L} N(0, \theta^2)$

2. Delta método:

$$g(\theta) = \frac{1}{\theta} \quad ; \quad g'(\theta) \frac{-1}{\theta^2}$$

$$\sqrt{n}(g(\bar{X}) - g(\theta)) \xrightarrow{L} N(0, \theta^2 (\frac{-1}{\theta^2})^2)$$

$$\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\theta}) \xrightarrow{L} N(0, \frac{1}{\theta^2})$$

3. $\sqrt{n}(\frac{1}{\bar{X}} - \lambda) \xrightarrow{L} N(0, \lambda^2) \quad o \quad \frac{1}{\bar{X}} \approx N(\lambda, \frac{\lambda^2}{n})$

Entonces, $\frac{1}{\bar{X}}$ es CAN para λ y $\frac{V_T^2(\lambda)}{n} = \frac{\lambda^2}{n}$

Ahora nos cuestionamos, ¿es $\frac{1}{\lambda}$ AE?

Teniamos que $I_{x_1}(\lambda) = \frac{1}{\lambda^2}$, por lo que la cota CR será $\frac{1}{n \cdot I_{x_1}(\lambda)} = \frac{\lambda^2}{n}$. Es un estimador asintóticamente eficiente.

Resultado: Todos los estimadores CAN que se obtienen usando el delta método a partir de un estimador asintóticamente eficiente, son asintóticamente eficientes.

Demostración 1.6. Sea $T_n(X)$ estimador CAN para θ

$$\sqrt{n}(T_n(X) - \theta) \xrightarrow{L} N(0, V_T^2(\theta))$$

y $T_n(X)$ es AE entonces

$$V_T^2(\theta) = \frac{1}{I_T(\theta)} \implies I_T(\theta) = \frac{1}{V_T^2(\theta)}$$

Sea g una función con derivada no nula, se tiene

$$\sqrt{n}(g(T_n(X)) - g(\theta)) \xrightarrow{L} N(0, V_T^2(\theta)(g'(\theta))^2)$$

La varianza de $g(T_n(X))$ es

$$Var(g(T_n(X))) = \frac{V_T^2(\theta)(g'(\theta))^2}{n}$$

Y su cota CR es

$$\frac{(g'(\theta))^2}{nI_{X_1}(\theta)} = \frac{(g'(\theta))^2}{\frac{n}{V_T^2(\theta)}} = \frac{V_T^2(\theta)(g'(\theta))^2}{n} \square$$

1.1.7. Estimadores razonables

Los estimadores CAN y AE son estimadores con propiedades razonables, pero podemos tener más de un estimador razonable.

Sean $T_n(x)$ estimador CAN para g(θ) y $G_n(x)$ estimador CAN para g(θ)

- Si uno de los 2 es AE y el otro no, el estimador AE es mejor.
- Si ambos son AE, será mejor el que tenga una menor varianza asintótica(V_T^2).

Para ver cual tiene una menor varianza asintótica, usamos la eficiencia relativa asintótica.

Definición: $X_1, \ldots, X_n$ i.i.d. $\{P_\theta : \theta \in \Theta\}$, sean $T_n(x)$ y $G_n(x)$ dos estimadores CAN de $g(\theta)$

$$ARE_{TG}(\theta) = \frac{V_G^2(\theta)}{V_T^2(\theta)}$$

Si $ARE_{TG}(\theta)$ es menor que 1, $G_n(x)$ es mejor. En cambio si ARE es mayor que 1, $T_n(x)$ es mejor.

1.2. Inferencia Basada en Verosimilitud

1.2.1. Caso Uniparamétrico

Toda información acerca de θ que los datos nos proporcionan está en la función de verosimilitud.

Definición: Sean $X_1, \ldots, X_n$ v.a.i.i.d. Se define como función de verosimilitud a

$$L(\theta; X_1, \ldots, X_n) = \begin{cases} P_\theta(X_1, \ldots, X_n) = \prod_{i=1}^n P_\theta(X_i = x_i) & \text{Caso discreto} \\ f(X_1, \ldots, X_n; \theta) = \prod_{i=1}^n f(x_i; \theta) & \text{Caso continuo} \end{cases}$$

$L(\theta; X_1, \ldots, X_n)$ es una función de θ proporcional a la probabilidad de observar que $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ cuando θ es el verdadero valor del parámetro. Así, un valor del parámetro será más o menos verosimil cuanto mayor o menor sea esa verosimilitud.

Consideremos las siguientes condiciones:

1. El parámetro θ es identificable, es decir, que $\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$
2. Las distribuciones P_θ donde $\theta \in \Theta$ tienen el mismo soporte $\{x : f(x, \theta) > 0\}$ y no depende de θ

Resultado: Sean $X_1, \ldots, X_n$ v.a.i.i.d. con $f(X; \theta)$. Entonces si θ_0 es el verdadero valor del parámetro

$$P_\theta(L(\theta_0; X_1, \ldots, X_n) > L(\theta; X_1, \ldots, X_n)) \xrightarrow{n \to \infty} 1$$

Demostración 1.7. Sea

$$\{x : L(\theta_0; X_1, \ldots, X_n) > L(\theta; X_1, \ldots, X_n)\} = \left\{ x : \prod_{i=1}^n \frac{f(x_i; \theta)}{f(x_i; \theta_0)} < 1 \right\}$$

$$\left\{ x : \frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i; \theta)}{f(x_i; \theta_0)} < 0 \right\}$$

Tenemos que

$$\frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i; \theta)}{f(x_i; \theta_0)} \xrightarrow{P} E_{\theta_0} = \left(\log \frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right)$$

Aplicando la desigualdad de Jensen para funciones convexas: $f(E(X)) \leq E(f(X))$ donde $-\log$ es convexa

$$E_{\theta_0} \left(-\log \frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right) \geq -\log E_{\theta_0} \left(\frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right) \implies$$

$$\implies E_{\theta_0} \left(\log \frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right) \leq \log E_{\theta_0} \left(\frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right)$$

$$\log E_{\theta_0} \left(\frac{f(x_i; \theta)}{f(x_i; \theta_0)} \right) = \log \int \frac{f(x_i; \theta)}{f(x_i; \theta_0)} f(x_i; \theta_0) dx = \log \int f(x_i; \theta_0) dx = \log(1) = 0$$

probamos que $\frac{1}{n} \sum_{i=1}^n \log \frac{f(x_i; \theta)}{f(x_i; \theta_0)}$ converge a una cantidad negativa

Estimación por Máxima Verosimilitud El estadistico EMV es el valor en el que se alcanza el máximo de la verosimilitud o así mismo, puesto que la transformación logaritmica es monótona y creciente, de la log verosimilitud definida como $l(\theta; X_1, \ldots, X_n) = \log L(\theta; X_1, \ldots, X_n)$.

Definición: $\widehat{\theta}(x)$ será EMV de θ si

$$\widehat{\theta}(x) = \underset{\theta \in \Theta}{argmax}(l(\theta; X_1, \ldots, X_n))$$

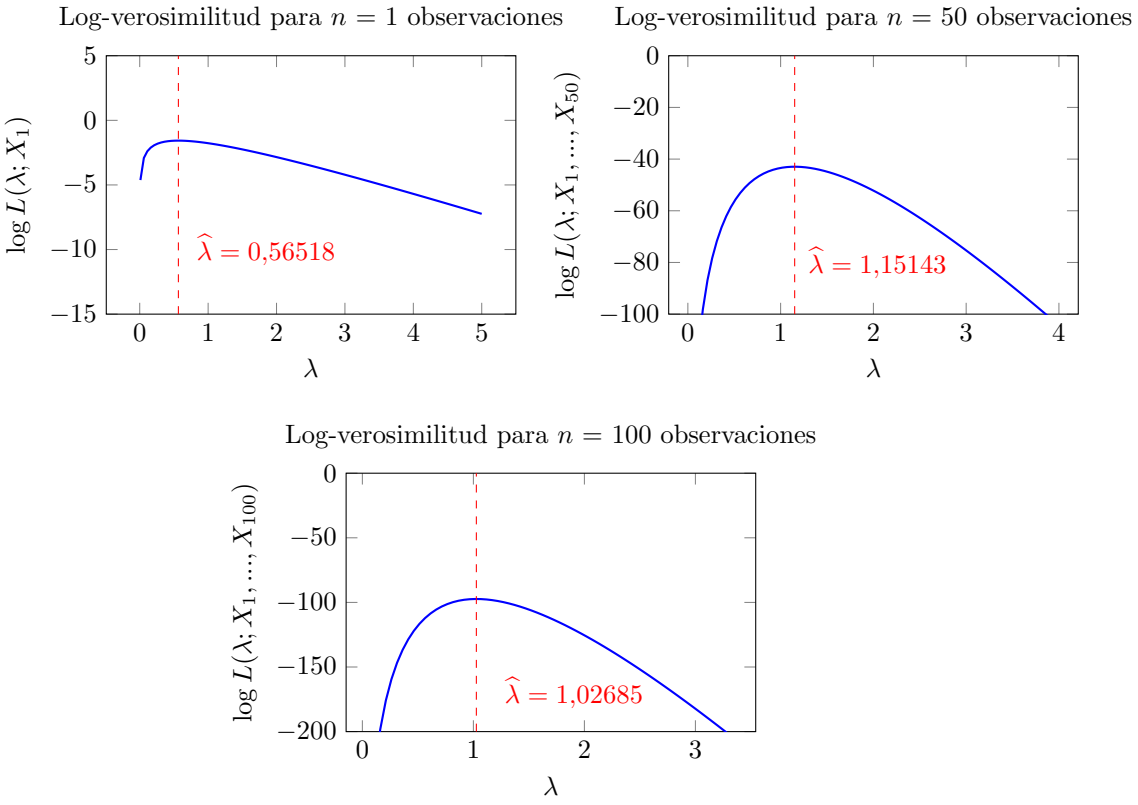

Figura 2: Podemos observar como el EMV aproxima el verdadero valor de λ en función de la cantidad de observaciones (en este caso $\exp(\lambda)$ donde $\lambda = 1$)

En la siguiente página se mostrará un ejemplo de como calcular el estimador máximo verosimil de una distribución Bernoulli de parametro θ.

Ejercicio 1.8. Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde $X_i \sim B(\theta)$ para $\theta \in (0,1)$. Sea la función de distribución Bernoulli $P_\theta(X = x) = \theta^x(1-\theta)^{1-x}$. Estimar θ mediante máxima verosimilitud.
Calculamos la verosimilitud como

$$L(\theta; X) \propto \prod_{i=1}^{n} \theta^x(1-\theta)^{1-x} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$

y la log verosimilitud

$$l(\theta; X_1, \ldots, X_n) = \sum_{i=1}^{n} x_i \log\theta + (n - \sum_{i=1}^{n} x_i)\log(1-\theta)$$

Encontraremos un máximo en donde se cumpla que $\frac{\partial}{\partial\theta} l(\theta; X_1, \ldots, X_n) = 0$

$$\frac{\partial}{\partial\theta} l(\theta; X_1, \ldots, X_n) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{(1-\theta)} \implies$$

$$\implies \sum_{i=1}^{n} x_i(1-\theta) = (n - \sum_{i=1}^{n} x_i)\theta \implies \theta = \frac{\sum_{i=1}^n x_i}{n} = \overline{X}$$

Por tanto concluimos que el EMV de θ es $\widehat{\theta} = \overline{X}$.

Hilando con la figura 1.2.1 donde vemos representada la log verosimilitud de una exponencial, se pondrá como ejemplo también el cálculo del EMV de una distribución exponencial censurada negativa de parámetro λ.

Definición: Sean $X_1, \ldots, X_n$ v.a.i.i.d. que siguen una distribución exponencial. Se considera como censura a la derecha cuando se desconoce el tiempo exacto de fallo para k observaciones, pero si se sabe que ocurrió después de un tiempo t_0. Es decir, que para k observaciones solo sabemos que $x_i > t_0$.

Se resolverá el ejercicio en la siguiente página...

Ejercicio 1.9. Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde X_i sigue una exponencial negativa censurada a la derecha de parámetro λ. Sea la función de distribución exponencial $f(x;\lambda) = \lambda e^{-\lambda x}$. Estimar λ mediante máxima verosimilitud.
La función de verosimilitud tendrá dos tipos de datos, los que se obtienen de la exponencial ($f_\lambda(x;\lambda)$) y los k censurados ($P(X_i > t_0)$).
Entonces la función de verosimilitud queda de la siguiente forma:

$$L(\lambda; x_1, \ldots, x_n) = \left(\prod_{i=1}^{k} \lambda e^{-\lambda x_i} \right) \left(\prod_{i=k+1}^{n} P(X_i > t_0) \right)$$

Sabemos que

$$P(X_i > t_0) = \int_{t_0}^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t_0}$$

Por lo tanto

$$L(\lambda; x_1, \ldots, x_n) = \lambda^k e^{\lambda \sum_{i=1}^{k} x_i} e^{-\lambda(n-k)t_0}$$

Y la log verosimilitud queda

$$l(\lambda; x_1, \ldots, x_n) = k \log \lambda - \lambda \left(\sum_{i=1}^{k} x_i + (n-k)t_0 \right)$$

Que encontrará máximo en $\widehat{\lambda} = \frac{k}{\sum_{i=1}^{k} + (n-k)t_0}$

Como último ejemplo se hará lo mismo con la distribución Poisson

Ejercicio 1.10. Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde X_i sigue una Poisson de parámetro λ. Sea la función de distribución Poisson $P_\lambda(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$. Estimar λ mediante máxima verosimilitud y comprobar si el estimador es CAN y AE.
La verosimilitud quedará como

$$L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

siendo la log verosimilitud

$$l(\lambda; x_1, \ldots, x_n) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda$$

Que encontrará máximo en $\widehat{\lambda} = \overline{X}$
Haciendo uso de este último resultado, es fácil comprobar que $\widehat{\lambda}$ es CAN y AE pues

$$\sqrt{n}(\widehat{\lambda} - \lambda) = \sqrt{n}(\overline{X} - \lambda) \xrightarrow{L} N(0, \frac{1}{I_1(\lambda)})$$

$$I_1(\lambda) = \frac{1}{\lambda} \implies \sqrt{n}(\overline{X} - \lambda) \xrightarrow{L} N(0, \lambda) \square$$

Como generalización del anterior resultado se enuncia el siguiente teorema

Teorema 1.11. Sean $X_1, \ldots, X_n$ v.a.i.i.d. con función de densidad $f(x;\theta)$ donde $\theta \in \Theta$ que verifica CRCR y además que

$$\exists \frac{\partial^3}{\partial\theta^3} \log f(x;\theta) \quad \text{y} \quad \left|\frac{\partial^3}{\partial\theta^3} \log f(x;\theta)\right| \leq M(X); \quad E(M(X)) < \infty$$

Entonces el EMV de θ es CAN y AE. Es decir

$$\sqrt{n}(\widehat{\theta}(X) - \theta) \xrightarrow{L} N(0, \frac{1}{I_1(\theta)})$$

Pongamos un ejemplo

Ejercicio 1.12. Sean $X_1, \ldots, X_n$ v.a.i.i.d. donde X_i sigue una Normal de parámetros $(0, \sigma^2)$ donde $\sigma^2 = \theta$. Sea la función de distribución Normal $f(x;\theta) = \frac{1}{\sqrt{(2\pi\theta)}} e^{\frac{-x^2}{2\theta}}$. Entonces

$$L(\theta; x_1, \ldots, x_n) = \frac{1}{\sqrt{\theta}} e^{\frac{-\sum_{i=1}^n x_i^2}{2\theta}} \implies l(\theta; x_1, \ldots, x_n) = -\frac{n}{2}\log\theta - \frac{\sum_{i=1}^n x_i^2}{2\theta}$$

Que encontrará máximo en $\widehat{\theta} = \frac{\sum_{i=1}^n x_i^2}{n}$. Entonces

$$\sqrt{n}\left(\frac{\sum_{i=1}^n x_i^2}{n} - \theta\right) \xrightarrow{L} N\left(0, \frac{1}{I_1(\theta)}\right)$$

$$I_1(\theta) = Var\left(-\frac{n}{2\theta} + \frac{\sum_{i=1}^n x_i^2}{2\theta^2}\right) = \frac{n}{2\theta^2}$$

por tanto

$$\sqrt{n}\left(\frac{\sum_{i=1}^n x_i^2}{n} - \theta\right) \xrightarrow{L} N(0, 2\theta^2) = N(0, 2\sigma^4)$$

Se podría hacer también para una Normal $N(0,\sigma)$ donde $\widehat{\theta} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$ que es el EMV para σ en modelos $N(0,\sigma)$ Queda como ejercicio para el lector comprobar si

$$\sqrt{n}\left(\sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} - \sigma\right) \xrightarrow{L} N\left(0, \frac{1}{I_1(\sigma)}\right)$$

Intervalos de confianza y contrastes de hipótesis uniparamétricos Ya hemos visto como estimar el EMV en el caso uniparamétrico. Ahora vamos a ver como se hacen los intervalos de confianza y los contrastes de hipótesis.

Ejemplo

Sean $X_1 \dots X_n \sim B(\theta)$ con $0 < \theta < 1$, $f(x,\theta) = \theta^x(1-\theta)^{1-x}$, $E(X) = \theta$ y $Var(X) = \theta(1-\theta)$

Recordemos que el EMV para $X \sim B(\theta)$ es $\widehat{\theta}_n = \overline{X}$

$$I_1(\theta) = -E\Big(\frac{\partial^2}{\partial\theta^2} ln(f(x,\theta))\Big) = \frac{\theta^2 - 2\theta^2 + \theta}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

Por lo tanto:

$$\sqrt{n}(\overline{X} - \theta) \xrightarrow{\mathcal{L}} N\big(0, \theta(1-\theta)\big)$$

Usando esta distribución asintótica podemos construir **intervalos de confianza** para θ.

$$P_\theta\left(\underbrace{\frac{\sqrt{n}|\overline{X} - \theta|}{\sqrt{\theta(1-\theta)}}}_{N(0,1)} < \zeta_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Buscamos el cuantil $\zeta_{1-\frac{\alpha}{2}}$ que hace que la probabilidad sea $1 - \alpha$

$$P_\theta\left(-\zeta_{1-\frac{\alpha}{2}} < \frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta(1-\theta)}} < \zeta_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Utilizamos la información de Fischer $I_n(\theta) = \frac{n}{\theta(1-\theta)}$ para depejar θ

$$P_\theta\left(-\zeta_{1-\frac{\alpha}{2}} < \frac{(\overline{X} - \theta)}{\sqrt{\frac{1}{I_n(\theta)}}} < \zeta_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

$$P_\theta\left(\overline{X} - \zeta_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{I_n(\theta)}} < \theta < \overline{X} + \zeta_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{I_n(\theta)}} \right) = 1 - \alpha$$

Definición: La **Información de Fischer observada** se traduce en sustituir la información de Fischer esperada por su EMV. Como $\widehat{\theta}_n$ es CAN para θ, se tiene que

$$If_{obs} \xrightarrow{p} If_{esp}$$

En nuestro caso particular consiste en reemplazar $I_n(\theta) = \frac{n}{\theta(1-\theta)}$ por $I_n(\widehat{\theta}) = \frac{n}{\overline{X}(1-\overline{X})}$, luego:

$$P_\theta\left(\overline{X} - \zeta_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} < \theta < \overline{X} + \zeta_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right) = 1 - \alpha$$

Este tipo de inferencia basada en la verosimilitud se llama **inferencia de Wald**.

Para los contrastes de hipótesis vamos a definir tres estadísticos que nos sirven para hacer inferencia basada en la verosimilitud.

1. **Estadístico de Razón de Verosimilitud**

   Definamos $\lambda(X)$

   $$\lambda(X) = \lambda(X_1, \ldots, X_n) = \Lambda(X) = \frac{L(\theta_0, X_1, \ldots, X_n)}{\sup\limits_{\theta} L(\theta, X_1, \ldots, X_n)} = \frac{L(\theta_0, X_1, \ldots, X_n)}{L(\widehat{\theta}_n, X_1, \ldots, X_n)} \in [0, 1]$$

   H_0 se rechazará para valores bajos del estadístico. Ahora vamos a hallar la **región crítica del test** $\lambda(X) < k$. Para hacerlo más sencillo, se puede escribir con la log-verosimilitud:

   $$Q_L = -2ln(\lambda(X)) = -2\big(ln(L(\widehat{\theta}_n, X_1 \ldots, X_n)) - ln(L(\theta_0, X_1 \ldots, X_n))\big)$$

   En este caso, se rechazará H_0 para valores grandes. La región crítica del test será $Q_L > C$. Necesitamos conocer la distribución asintótica de Q_L, algo que veremos más adelante.

2. **Estadístico de Wald**

   Bajo las condiciones de regularidad de Cramer-Rao podemos escribir $ln(L(\theta_0, X))$ como **desarrollo de Taylor** en torno a $\widehat{\theta}_n$ como:

   $$ln(L(\theta_0, X)) \approx ln(L(\widehat{\theta}_n, X)) + (\theta - \widehat{\theta}_n)\overbrace{\frac{\partial}{\partial\theta} ln(L(\widehat{\theta}_n, X))}^{0} + \frac{(\theta - \widehat{\theta}_n)^2}{2}\frac{\partial^2}{\partial\theta^2} ln(L(\widehat{\theta}_n, X))$$

   Recordemos que $\widehat{\theta}_n$ es el EMV, y por tanto el valor de su primera derivada es nula, pero el valor de su segunda derivada no tiene por qué serlo. Ahora:

   $$Q_L = -2\big(ln(L(\widehat{\theta}_n, X)) - ln(L(\theta_0, X))\big) = -2(ln(L(\theta_0, X)) + \frac{(\theta - \widehat{\theta}_n)^2}{2}\frac{\partial^2}{\partial\theta^2} ln(L(\widehat{\theta}_n, X)) - ln(L(\theta_0, X))) =$$

   $$= -2\Big(\frac{(\theta - \widehat{\theta}_n)^2}{2}\frac{\partial^2}{\partial\theta^2} ln(L(\widehat{\theta}_n, X))\Big) = (\theta - \widehat{\theta}_n)^2\Big(\underbrace{-\frac{\partial^2}{\partial\theta^2} ln(L(\widehat{\theta}_n, X))}_{\text{I. de Fischer observada}}\Big) =$$

   $$= (\theta - \widehat{\theta}_n)^2 I_n(\theta) = \frac{n(\theta - \widehat{\theta}_n)^2}{Var(\widehat{\theta}_n)} = Q_W \longleftarrow \textbf{ Estadístico de Wald}$$

   Q_L y Q_W son asintóticamente equivalentes.

3. **Estadístico de Rao**

$$Q_R = R = \frac{\left(\overbrace{\frac{\partial}{\partial\theta} ln(L(\theta_0, X))}^{\text{Score}}\right)^2}{I_n(\theta)}$$

Habíamos visto que

$$ln(L(\theta_0, X)) \approx ln(L(\widehat{\theta}_n, X)) + \frac{(\theta - \widehat{\theta}_n)^2}{2} I_n(\theta)$$

Ahora vamos a calcular el desarrollo de Taylor para la función score en un entorno de $\widehat{\theta}_n$

$$S(\theta_0) \approx \underbrace{S(\widehat{\theta}_n)}_{0} + (\theta_0 - \widehat{\theta}_n) \frac{\partial^2}{\partial\theta^2} ln(L(\widehat{\theta}_n, X))$$

(A partir de este punto el profesor hace una demostración que no es correcta, por lo que he preferido no incluirla, aunque las conclusiones que hay a continuación sí son validas)

Este resultado demuestra que Q_R es asintóticamente equivalente a Q_W y Q_L.

Los tres estadísticos rechazarán H_0 para valores grandes. Para ver qué distribución siguen, usamos Q_W.
Sabemos que un EMV es CAN y AE, por lo que:

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N\Big(0, \frac{1}{I_1(\theta)}\Big) \text{ y } \frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sqrt{\frac{1}{I_1(\theta)}}} \xrightarrow{\mathcal{L}} N(0,1)$$

Q_W es $\dfrac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sqrt{\frac{1}{I_1(\theta)}}}$ al cuadrado, por lo que, despejando, se obtiene que:

$$n(\theta - \widehat{\theta}_n)^2 I_1(\theta) \xrightarrow{\mathcal{L}} \chi^2_1$$

En este punto es conveniente recordar que si una variable $Z \sim N(0,1)$ su cuadrado $Z^2 \sim \chi^2_1$

Las distribuciones exactas de Q_L, Q_W y Q_R son diferentes, pero como las tres son **asintóticamente equivalentes**, tenderán a seguir una distribución χ^2_1. Por tanto, para determinar el valor de la región crítica solo habría que calcular el siguiente percentil:

$$P_\theta(Q_i > c) \approx 1 - \alpha \Longrightarrow c = \chi^2_{1,1-\alpha}$$

1.2.2. Caso multiparamétrico

Situacion: $X_1, \ldots, X_n \quad i.i.d. \quad P_\theta, \theta \in \Theta \subseteq \mathbb{R}^s$
El parámetro s-dimensional es $\theta = (\theta_1, \ldots, \theta_s), \quad s \geq 1$ con familia de densidad $\{f(x,\theta), \theta_0 \in \Theta\}$

Igualmente podemos escribir

- Función de verosimilitud: $L(\theta, X_1, \ldots, X_n) = \prod_{i=1}^{n} f(x_i, \theta)$
- Función de log-verosimilitud: $l(\theta, X_1, \ldots, X_n) = \sum_{i=1}^{n} \log f(x_i, \theta)$
- Vector/función score:

$$S(\theta, X_1, \ldots, X_n) = S(X, \theta) = \frac{d}{d\theta} \log L(\theta, X) = \left(\frac{d}{d\theta} \log L(\theta_1, X), \ldots, \frac{d}{d\theta} \log L(\theta, X) \right)$$

Ahora nos preguntamos, ¿cual es la cantidad de información de Fisher que tendremos en el caso multiparamétrico?. Para averiguarlo recurrrimos a la matriz Hessiana, que recordamos que es la matriz s x s de las derivadas parciales de segundo orden

$$I_{ij}(\theta) = E_\theta \left(-\frac{d^2}{d\theta_i d\theta_j} logL(\theta, X) \right) = E \left(\frac{d}{d\theta_i} logL(\theta, X), \frac{d}{d\theta_j} logL(\theta, X) \right)$$

$$H(\theta, X) = \left\{ \frac{d^2}{d\theta_i d\theta_j} logL(\theta, X) \right\}^s = \begin{pmatrix} \frac{d^2}{d\theta_1^2} logL(\theta, X) & \ldots & \frac{d^2}{d\theta_1 d\theta_s} logL(\theta, X) \\ \vdots & \ddots & \vdots \\ \frac{d^2}{d\theta_s d\theta_1} logL(\theta, X) & \ldots & \frac{d^2}{d\theta_s^2} logL(\theta, X) \end{pmatrix}$$

Sabiendo esto, $I(\theta)$,matriz de información esperada(que recordemos que es la matriz de covarianzas del vectir score) será:

$$I(\theta) = E_\theta(-H(\theta, X))$$

Además de esto tendremos tambíen la matriz de información de Fisher observada

EMV en el caso multiparamétrico $\widehat{\theta} = \widehat{\theta_n} = (\widehat{\theta_1}, \ldots, \widehat{\theta_s})'$ es el EMV para $\theta = (\theta_1, \ldots, \theta_s)$ si es la solución al siguiente sistema de ecuaciones:

$$\begin{gathered} \frac{d}{d\theta_1} logL(\theta, X) = 0 \\ \ldots \\ \frac{d}{d\theta_s} logL(\theta, X) = 0 \end{gathered}$$

Estas ecuaciones son las denominadas **ecuaciones de verosimilitud**.

Propiedades asintóticas del EMV multiparamétrico

En la situación $X_1, \ldots, X_n$ i.i.d. $P_\theta, \theta \in \Theta \subseteq \mathbb{R}^s s \geq 1$ y bajo las siguientes condiciones de regularidad:

1. Θ es un intervalo de $\mathbb{R}^s$
2. El soporte de f no depende de θ. $\{x : f(x, \theta) > 0\}$ no depende de θ.
3. $\frac{d}{d\theta_j} f(x, \theta)$ existe y es finita $\forall j = 1, \ldots, s \quad \theta \in \Theta$
4. La matriz de información de Fisher (IF(θ)) es definida positiva

$$(si \quad \forall X \neq A, \quad X^T \cdot A \cdot X > 0)$$

5. $f(x, \theta)\, dx$ se puede derivar bajo el signo integral

Siendo θ_0 el verdadero valor del parámetro, se obtienen los siguientes resultados:

1. $P_\theta(L(\theta_0, X) \geq L(\theta, X), \quad \forall \theta \in \Theta) \xrightarrow[n\to\infty]{L} 1$

   Con probabilidad que tiende a 1, cuando la función de verosimilitud alcanza el máximo en el verdadero valor del parámetro

2. $\widehat{\theta_n} \to \theta_0$. El estimador es consistente.

3. $\widehat{\theta_n} \sim N_s(\theta_0, V(\theta))$ donde $V(\theta) = I^{-1}(\theta)$ (es la inversa de la información de Fisher esperada).

4. Para una componente de $\widehat{\theta_k}$ de $\widehat{\theta}$ (vector) se tiene que $\widehat{\theta_k} \sim N(\theta_{0k}, V_{kk}(\theta))$ donde $V_{kk}(\theta)$ es el k-ésimo elemento de la diagonal de $V(\theta)$

Ejemplo *Para* $X_1, \ldots, X_n \quad N(\mu, \sigma)$ Obtener $I(\theta)$ esperada

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

$$L(\mu, \sigma, X_1, \ldots, X_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{\frac{-1\cdot\sum_{i=1}^n (x_i-\mu)^2}{2\cdot\sigma^2}}$$

$$\log L(\mu, \sigma, X_1, \ldots, X_n) = -n\log\sigma \cdot \frac{-1}{2\cdot\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$$

Sacamos las ecuaciones de verosimilitud:

$$\frac{d}{d\mu}\log L(\mu, \sigma, X_1, \ldots, X_n) = \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2} = 0$$
$$\frac{d}{d\sigma}\log L(\mu, \sigma, X_1, \ldots, X_n) = \frac{-n}{\sigma} + \frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^3} = 0$$

$$I(\mu, \sigma) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix} \quad I^{-1}(\mu, \sigma) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{pmatrix}$$

luego:

$$EMV = \begin{pmatrix} \bar{\mu} \\ \bar{\sigma} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \bar{\mu_0} \\ \bar{\sigma_0} \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{pmatrix}\right)$$

o tambien se puede escribir como

$$EMV = \sqrt{n}\left[\begin{pmatrix} \bar{\mu} \\ \bar{\sigma} \end{pmatrix} - \begin{pmatrix} \bar{\mu_0} \\ \bar{\sigma_0} \end{pmatrix}\right] \to N_2\left(0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix}\right)$$

Ejemplo 2: Sea $X_1, \ldots, X_n$ i.i.d. de una distribución que toma valores en un conjunto finito de valores x=$\{a1, a2, a3\}$ con probabilidad P_i tal que $\sum_{i=1}^3 P_i = 1$, obtener el EMV.

Se trata de un modelo biparamétrico ya que $P_3 = 1 - P_1 - P_2$. $\theta = (P_1, P_2)'$

Función de probabilidad=$f(x;\theta) = P_1\mathbf{1}_{x=a_1} + P_2\mathbf{1}_{x=a_2} + P_3\mathbf{1}_{x=a_3}$

Función de verosimilitud:

$$L(\theta, X_1, \ldots, X_n) = \prod_{i=1}^n f(x_i, P_1, P_2) = P_1^{N1}\cdot P_2^{N2}\cdot(1-P_1-p_2)^{N3} \quad \textit{donde} \quad N_r = \sum_{i=1}^n \mathbf{1}_{x_i=a_r}$$

Log-verosimilitud=$\log L(\theta, X) = N_1\log P_1 + N_2\log P_2 + N_3\log P_3$

Ecuaciones de verosimilitud:

$$\frac{d}{dP_1}\log L(\theta,X)=\frac{N_1}{P_1}-\frac{N_3}{1-P_1-P_2}=0$$
$$\frac{d}{dP_2}\log L(\theta,X)=\frac{N_2}{P_2}-\frac{N_3}{1-P_1-P_2}=0$$

Resolviendo el sistema de ecuaciones obtenemos:

$$\widehat{P_1}=\frac{N_1}{n}\quad \widehat{P_2}=\frac{N_2}{n}\quad \widehat{P_3}=\frac{N_3}{n}$$
$$\begin{pmatrix}\widehat{p_1}\\ \widehat{p_2}\end{pmatrix}\sim N_2\left(\begin{pmatrix}P_1\\ P_2\end{pmatrix},IF^{-1}\right)$$

(Hallar la información de Fisher se deja como tarea para el lector)

Estimación de la varianza a través de la matriz de Información de Fisher observada La matriz de información de Fisher es la negativa de la matriz Hessiana calculada en el EMV, es decir $-H(\widehat{\theta},X)$, de forma que el estimador de $V(\theta)$ es:

$$\widehat{V}=\widehat{V(\theta)}=-[H(\widehat{\theta},X)]^{-1}\quad donde\quad H(\theta,X)=\frac{d^2}{d\theta_i d\theta_j}\log L(\theta,X)$$

esto ocurre ya que: $H(\widehat{\theta},X)\to H(\theta_0,X)$
$\widehat{\theta}\sim N_s(\theta_0,\widehat{V})\widehat{\theta_k}\sim N(\theta_{0_k},\widehat{V_{kk}})$

En esto nos basaremos para hacer inferencia de Wald con intervalos de confianza y contrastes de hipótesis.

$$Z\sim N_s(\theta,\Sigma)\to \text{Tipificando nos queda}$$
$$(Z-\theta)'\Sigma^{-}1(Z-\theta)\sim\chi^2_s$$
$$\frac{(Z-\theta)^2}{\sigma}\sim\chi^2_1$$

1.2.3. Inferencia de Wald en el caso multiparamétrico

Si queremos hacer inferencia para 2 o más parámetros, tenemos:
Sea una particion: $\psi=(\theta_1,\ldots,\theta_r)\quad r\le s\quad ,\quad \psi_0=(\theta_{1_0},\ldots,\theta_{r_0})\quad y\quad \Omega=(\theta_{r+1},\ldots,\theta_s)$
Buscamos contrastar: $H_0:\psi=\psi_0$
El EMV para ψ es $\widehat{\psi}=(\widehat{\theta_1},\ldots,\widehat{\theta_r})$.
Bajo H_0 según lo visto, $\widehat{\psi}\sim N_r(\psi_0,\widehat{V_\psi})$ donde $\widehat{V_\psi}$ es la matriz de covarianzas de $\widehat{\psi}$ que es la matriz rxr superior de $\widehat{V(\theta)}$.
El estadístico de Wald en este caso es:

$$W=(\widehat{\psi}-\psi_0)'[\widehat{V_\psi}]^{-1}(\widehat{\psi}-\psi_0)\sim\chi^2_r$$

O lo que es lo mismo, en el caso uniparamétrico:

$$\frac{(\widehat{\theta}-\theta)^2}{\sigma}\sim\chi^2_1$$

Un test de Wald de nivel α rechazará H_0 si $W > C_k \equiv W > \chi^2_{1-\alpha,r}$, es decir, si α es menor que el p-valor=$P_{\psi=\psi_0}(W > W_{obs}) \backsimeq$1 - pchisq($W_{obs}$,r).

La región de confianza para ψ con $1-\alpha$ es una elipse r-dimensional para todos los valores $(W \leq \chi^2_{r,1-\alpha})$ en los que el test no rechaza H_0

$$\text{Región de confianza}(1-\alpha) = \{\psi_0/(\widehat{\psi}-\psi_0)'[\widehat{V_\psi}]^{-1}(\widehat{\psi}-\psi_0) = \frac{(\widehat{\theta}-\theta_s)^2}{\sigma} \leq \chi^2_r\}$$

¿Cómo hacemos inferencia para una función del parámetro? Utilizando el delta método s-variante.

Partimos de la situacion $H_0 : g(\theta) = 0$.

Si g es una función de θ tal que g:$\theta \to \mathbb{R}^p \quad p \leq S, \quad g(\theta) = (g_1(\theta), \ldots, g_p(\theta))$ y sea $G(\theta)$ la matriz p x s de las primeras derivadas respecto a θ:

$$G(\theta) = \begin{pmatrix} \frac{d}{d\theta_1}g_1(\theta) & \ldots & \frac{d}{d\theta_s}g_1(\theta) \\ \ldots & \ldots & \ldots \\ \frac{d}{d\theta_1}g_p(\theta) & \ldots & \frac{d}{d\theta_s}g_p(\theta) \end{pmatrix}$$

Si $\widehat{\theta}$ tiene distribución asintótica $(N_s(\theta_s, V))$ y $g(\widehat{\theta})$ también tiene distribución asintótica $(N_p(g(\theta_0)), G_{pxs}(\theta_0) \cdot V \cdot G(\theta)')$

$$W = (g(\widehat{\theta}) - g(\theta_0))'(G(\widehat{\theta}) \cdot \widehat{V} G(\widehat{\theta})')^{-1}(g(\widehat{\theta}) - g(\theta_0)) \sim \chi^2_p$$

Ejercicio 1 Se analizan 100 lotes con 10 muestras cada uno $\sum_{i=1}^{100} Y_i = 12 \quad Y_1, \ldots, Y_{100} \sim B(\theta)$.

θ="probabilidad de que la toxina esté presente en el lote"

p="probabilidad de que la toxina esté presente en una muestra individual"

Apartado a.

Nos piden hacer inferencia sobre p, sabiendo que p es función de θ.

$$P(Y_1 = 0) = 1 - \theta = (1-p)^{10} \quad P(Y_i = 1) = \theta$$

Ya que ya tenemos la relación, hacemos inferencia sobre p.

$$L(\theta, Y_1, \ldots, Y_{100}) = \prod_{i=1}^{100} \theta^{Y_i}(1-\theta)^{1-Y_i} = \theta^{\sum_{i=1}^{100} Y_i}(1-\theta)^{100-\sum_{i=1}^{100} Y_i}$$

$$\log L(\theta, Y_1, \ldots, Y_{100}) = (\sum_{i=1}^{100} Y_i)\log\theta + (100 - \sum_{i=1}^{100} Y_i)\log(1-\theta)$$

$$\frac{d}{d\theta} logL(\theta, Y_1, \ldots, Y_{100}) = \frac{\sum_{i=1}^{100} Y_i}{\theta} - \frac{100 - \sum_{i=1}^{100} Y_i}{1-\theta} = 0$$

El EMV es invariante por transformación: $EMV \to \widehat{\theta} = \bar{Y}$

$$p = g(\theta) = 1 - (1-\theta)^{0,1} \Longrightarrow \widehat{p} = g(\widehat{\theta}) = 1 - (1-\bar{Y})^{0,1}$$

Apartado b.

Sabemos que $\widehat{\theta} = \bar{Y}$ y que tiene distribución asintóticamente normal.

$$I_n(\theta) = \frac{n}{\theta(1-\theta)} \quad \widehat{\theta} \simeq N(\theta, \frac{1}{I_n(\theta)}) = N(\theta, \frac{\theta(1-\theta)}{n})$$

Sabiendo esto, ¿cuál será la distribución de p?

$$\widehat{p} = g(\widehat{\theta}) \simeq N(g(\theta), \frac{\widehat{\theta}(1-\widehat{\theta})}{n} \cdot (g'(\theta))^2)$$

¿Y cual sería un intervalo de confianza de Wald con confianza del 95 % para p?

$$\widehat{p} \pm qnorm(0{,}975)\sqrt{Var(\widehat{p})}$$

Apartado c. Calcular el ICRV con confianza 0.95 para p.

$$Q_L \approx Q_W \sim \chi^2_1$$

Usamos la fórmula que sabemos para calcular este intervalo para θ.

$$ICRV = \{\theta_0 : 2[\log L(\widehat{\theta}, X) - logL(\theta_0, X)] \leq qchisq(0{,}95, 1)\}$$
$$= \{\theta_0 : \log L(\theta_0, X) \geq \log L(\bar{Y}, X) - \frac{qchisq(0{,}95, 1)}{2}\}$$

Como los intervalos de confianza basados en RV son invariantes por transformación, el intervalo de confianza RV para p es:

$$[1-(1-L)^{0,1}, 1-(1-M)^{0,1}]$$

(L y M son los puntos entre los que se cumple que $logL(\theta_0, X) \geq \log L(\bar{Y}, X) - \frac{qchisq(0{,}95,1)}{2}$)

Ejercicio 9

$X_1, \ldots, X_n$ i.i.d. con distribución discreta ($P_1 = P(ab), P_2 = P(Ab), P_3 = P(aB), P_4 = P(AB)$).

$$n = 3839, \quad N_1 = 1997, N_2 = 904, N_3 = 906, N_4 = n - N_1 - N_2 - N_3$$

$$L(x) = \prod_{i=1}^{n} P(X_i = x_i) = P_1^{N_1} \cdot P_2^{N_2} \cdot P_3^{N_3} \cdot (1 - P_1 - P_2 - P_3)^{n-N_1-N_2-N_3}$$

$$logL(p, X_1, \ldots, X_n) = 1997 \log P_1 + 904 \log P_2 + 906 \log P_3 + 32 \log(1 - P_1 - P_2 - P_3)$$

Ecuaciones de verosimilitud:

$$\frac{d}{dP_1}\log L(P_1, X_1, \ldots, X_n) = \frac{1997}{P_1} - \frac{32}{1-P_1-P_2-P_3} = 0$$

$$\ldots$$

$$\frac{d}{dP_3}\log L(P_3, X_1, \ldots, X_n) = \frac{906}{P_3} - \frac{32}{1-P_1-P_2-P_3} = 0$$

$$\text{Resultado:} \quad \widehat{P_1} = \frac{N_1}{n} = \frac{1997}{3834} \quad \widehat{P_2} = \frac{N_2}{n} = \frac{904}{3834} \quad \widehat{P_3} = \frac{N_3}{n} = \frac{906}{3834}$$

Apartado b.

Calcular el p-valor para el test de Wald.

¿Como relacionamos nuestros parámetros a θ para contrastar H_0? Tenemos que escribir H_0 como una función de $g(\theta)$.

El modelo global tiene 3 parámetros. Bajo H_0 depende de 1 parámetro y tiene que haber 2 relaciones (2 funciones).

$$H_0: \quad P_1 = \frac{2+\theta}{4}, \quad P_2 = \frac{1-\theta}{4} = P_3, \quad P_4 = \frac{\theta}{4}$$

$$g_1(P) = P_2 - P_3 = 0, \quad g_2(P) = P_1 + P_2 - \frac{3}{4} = 0$$

$$H_0: \begin{pmatrix} g_1(P) = 0 \\ g_2(p) = 0 \end{pmatrix} \qquad (P = 2, s = 3) \quad H_0 = (g(\theta) = (g_1(\theta), \ldots, g_P(\theta)))$$

Utilizando el delta-método multivariante:

$$g(\widehat{p}) \sim N_2(g(P), G_{2x3}(\widehat{P}) \cdot \widehat{Var(P)}_{3x3} \cdot G_{3x2}(\widehat{P})')$$

donde $G(\widehat{P})$ es la matriz de derivadas parciales evaluadas en $\widehat{P}$

$$G(\widehat{P}) = \begin{pmatrix} \frac{d}{dP_1}g_1(P) & \frac{d}{dP_2}g_1(P) & \frac{d}{dP_3}g_1(P) \\ \frac{d}{dP_1}g_2(P) & \frac{d}{dP_2}g_2(P) & \frac{d}{dP_3}g_2(P) \end{pmatrix}$$

$$W = \left(g_1(\widehat{P}), g_2(\widehat{P})\right)' \cdot \left(G(\widehat{P})\,\widehat{\mathrm{Var}}(P)\,G(\widehat{P})'\right)^{-1} \cdot \left(g_1(\widehat{P}), g_2(\widehat{P})\right) \sim \chi^2_2 \quad (\text{bajo } H_0)$$

Recordando:

$$\begin{matrix} g_1(P) = P_2 - P_3 = 0 \\ g_2(P) = P_1 + P_2 - \frac{3}{4} = 0 \end{matrix} \quad G(\widehat{P}) = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$W = \left(\widehat{P_2} - \widehat{P_3}, \widehat{P_1} + \widehat{P_2} - \frac{3}{4}\right)' \cdot \left(G(\widehat{P})\,\widehat{\mathrm{Var}}(P)\,G(\widehat{P})'\right)^{-1} \cdot \left(\widehat{P_2} - \widehat{P_3}, \widehat{P_1} + \widehat{P_2} - \frac{3}{4}\right) \sim \chi^2_2$$

$$\text{p-valor} = P_{H_0}(W \geq W_{obs}) = 1 - pchisq(W_{obs}, 2) = 0{,}36$$

Volviendo al caso multiparamétrico.

Situación: $\theta = (\theta_1, \ldots, \theta_s) \subseteq \mathbb{R}^s$ y sea $g: \mathbb{R}^{\sim} \to \mathbb{R}^r$

$$g(\theta) = (g_1(\theta), \ldots, g(\theta))' \quad G(\theta) = \begin{pmatrix} \frac{d}{d\theta_1}g_1(\theta) & \ldots & \frac{d}{d\theta_s}g_1(\theta) \\ \ldots & \ldots & \ldots \\ \frac{d}{d\theta_1}g_r(\theta) & \ldots & \frac{d}{d\theta_s}g_r(\theta) \end{pmatrix}$$

Se requiere contrastar: $H_0 : g(\theta) = 0 \quad H_1 : g(\theta) \neq 0$.
H_0 depende de s-r parámetros libres. Con el delta método podemos llegar a calcular un p-valor basado en el test de Wald

Bajo H_0:

$$g(\widehat{\theta}) - g(\theta) \sim N_r(0, G(\widehat{\theta}) \cdot \widehat{V}(\widehat{\theta}) \cdot G(\widehat{\theta})')$$

$$W = (g(\widehat{\theta}) \cdot [G(\widehat{\theta}) \cdot \widehat{V}(\widehat{\theta}) \cdot G(\widehat{\theta})']^{-1} \cdot g(\widehat{\theta})) \sim \chi^2_r$$

$$\text{p-valor} = P_{H_0}(\chi^2_r > W_{obs})$$

Aunque este test es potente, el test de razón de verosimilitud (RV) es más potente.

1.2.4. Test de razón de verosimilitud (RV)

Situación: $X_1, \ldots, X_n$ i.i.d. $P_\theta : \theta \in \Theta \subseteq \mathbb{R}^s$
$H_0 : \theta \in \Theta_0 \subseteq \Theta \quad H_1 : \theta \notin \Theta_0$
donde $\Theta_0 = \{\theta \in \Theta : \theta = (\theta_{1_0}, \ldots, \theta_{r_0}, \theta_{r+1}, \theta_s)\}$
Θ depende de s parámetros libres. Θ_0 depende de r-s parámetros libres.

El test de razón de verosimilitud (TRV) compara el máximo de la verosimilitud en H_0 con el mínimo en el EMV

$$\Delta(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta, X)}{\sup_{\theta \in \Theta} L(\theta, X)}$$

El estadístivo test de razón de verosimilitud se escribe habitualmente como $-2 \cdot \log \Delta(x)$.

$$Q_L(x) = 2 \cdot [\log L(\widehat{\theta}, X) - \log L(\widehat{\theta_0}, X)]$$

($\log L(\widehat{\theta_0}, X)$ es el EMV de θ restringido a Θ_0 y que depende de s-r parámetros libres)

$Q_L(x)$ bajo H_0 tiene una distribución asintótica χ^2_r.

- Si $r = s \to H_0 : \theta \in \Theta$ es una hipótesis simple, es decir, no hay parámetros libres bajo H_0.$(dim(\Theta_0) = 0)$
- Si $r < s \to H_0 : \theta \in \Theta_0$ es una hipótesis compuesta $(dim(\Theta_0) = s - r)$ con s-r parámetros que pueden tomar varios valores posibles.

Vamos a usar la notación de partición que vimos antes.

$$\theta = (\phi, \lambda) \text{ donde } \phi = (\theta_1, \ldots, \theta_r) \quad y \quad \lambda = (\theta_{r+1}, \ldots, \theta_s)$$

El contraste es: $H_0 : \phi = \phi_0 \quad H_1 : \phi \neq \phi_0$

Entonces para maximizar sobre el espacio paramétrico restringido a Θ_0 es una maximización sobre λ porque $\Theta_0 = \phi_0 = (\theta_{1_0}, \ldots, \theta_{r_0})$ y ϕ_0 están restringidos.

$$\sup_{\theta \in \Theta} L(\theta, X) = \sup_{\lambda} L(\phi_0, \lambda, X)$$

Bajo las condiciones de regularidad de Cramer-Rao multiparamétrico, $Q_L(X) \sim \chi^2_r$ (bajo H_0) y rechazamos H_0 para valores grandes de $Q_L(x)$. El test de razón de verosimilitud rechaza H_0 a nivel α si $Q_L(x) > \chi^2_{2-\alpha,r} = qchisq(1-\alpha, r)$.

$$\text{p-valor} \to P_{H_0}(\chi^2_r > Q_{obs}) = 1 - pchisq(Q_{obs}, r)$$

Región de confianza de la razón de verosimilitud La región de confianza (1-α) para $\phi = (\theta_1, \ldots, \theta_r) \in \mathbb{R}^r$ es la colección de valores $\phi_0 = (\theta_{1_0}, \ldots, \theta_{r_0})$ para los que $H_0 : \phi = \phi_0$ no se rechaza a nuvel α.

$$RCRV(1-\alpha) = \{\phi_0 \in \mathbb{R}^r : H_0 : \phi = \phi_0\}$$

$$= \{\phi_0 \in \mathbb{R}^r : 2[\log L(\widehat{\theta}, X) - \sup \log L(\phi_0, \lambda, X)] \leq qchisq(1-\alpha, r)\}$$

La región de confianza de la razón de verosimilitud será un elipsoide r-dimensional. Con s=2 se puede representar.

Caso particular para r=1:

Intervalo de confianza para θ_1 basado en RV con $\theta = (\theta_1, \ldots, \theta_s)$ y $\phi = \phi_0$.

$$ICRV(\theta_1, 1-\alpha) =$$

$$\{\theta_{1_0} : H_0 : \theta_1 = \theta_{1_0}\} = \left\{\theta_{1,0} : 2\left[\log L(\widehat{\theta}, X) - \sup \log L(\phi_0, \lambda, X)\right] \leq qchisq(1-\alpha, r)\right\}$$

Buscamos el intervalo de confianza para el caso s=2, es decir, intervalo de confianza para θ_1 con $\phi_0 = \theta_1$ y $\lambda = \theta_2$.

$$ICRV(\theta_1, 1-\alpha) =$$

$$= \left\{\theta_{1_0} : h(\theta_{1_0}) = \sup_{\theta_2} \log L(\theta_{1_0}, \theta_2, X) \geq \log(\widehat{\theta}, X) - \frac{qchisq(1-\alpha, 1)}{2} = d1\right\}$$

Podemos representarlo fácilmente si conocemos la función $h(\theta_{1_0})$.

El resultado con s=2 es:

$$\theta_{1_0} \in ICRV(\theta_1, 1-\alpha) \Longleftrightarrow \exists \alpha_2 / (\theta_1, \theta_2) \in B$$

$$B = \{\theta = (\theta_1, \theta_2) : \log L(\theta, X) > d_1\}$$

$$d1 = \log L(\widehat{\theta}, X) - \frac{qchisq(1-\alpha, 1)}{2}$$

En general, incluso con tamaños de muestra relativamente grandes, el p-valor del test de razón de verosimilitud no coincide con el de Wald. Siempre es mejor aproximación el test de razón de verosimilitud, especialmente en r>1.

Ejercicio 6 (Script de R en el campus)
$X_1, \ldots, X_n$ i.i.d. discreta $Y_i = \sum_{j=1}^{50} \mathbf{1}_{x_j=i}$ $i = 1, 2, 3$ $Y_1 = 8$ $Y_2 = 14$ $Y_3 = 28$
Apartado a.
Contrastar $H_0 : p_2 - 2 \cdot p_1 = 0$ $H_1 : p_2 - 2 \cdot p_1 \neq 0$ con TRV.
Pasos:

1. Escribimos la función de verosimilitud
2. Calculamos el EMV
3. Calculamos la log-verosimilitud en el EMV
4. Repetimos los 3 primeros pasos bajo H_0

Tenemos $p = (p_1, p_2)$ como vector de parámetros.

$$L(p, x) = \prod_{i=1}^{50} P_p(X = X_i) = P_1^{N_1} \cdot P_2^{N_2} \cdot (1 - P_1 - P_2)^{50-P_1-P_2} = P_1^8 \cdot P_2^{14} \cdot (1 - P_1 - P_2)^{28}$$

$$l(p, x) = 8 \log P_1 + 14 \log P_2 + 28 \log(1 - P_1 - P_2)$$

$$EMV = \begin{cases} \frac{d}{dp_1} \log L(p, x) = 0 & \widehat{p_i} = \dfrac{N_i}{n} \quad \widehat{p_1} = \frac{8}{50} \\ \frac{d}{dp_2} \log L(p, x) = 0 & \quad\quad\quad\quad \widehat{p_2} = \frac{14}{50} \end{cases}$$

$$Q_L(x) = 2[\log L(\widehat{p}, x) - \sup_{p_2=2 \cdot p_1} \log L(p, x)]$$

EMV bajo H_0:

$$sup_{p_1}(8 \log p_1 + 14 \log(2 \cdot p_1) + 28 \log(1 - 3 \cdot P_1)) \implies \widehat{p_{1_0} = 0{,}15}$$

$$Q_l(x) = Q_{obs} \qquad \text{p-valor}=P_\theta(\chi_1^2 \geq Q_{obs}) = 0{,}76$$

No se rechaza H_0

1.3. Maximización de la verosimilitud

En la mayoría de casos el EMV se obtiene de forma explícita. Sin embargo, en muchos casos prácticos la ecuación (o ecuaciones) de verosimilitud son muy complejas. Por ejemplo, para obtener el EMV en R podemos utilizar la función *optim*. En teoría veremos dos algoritmos distintos: el de **Newton-Raphson** y el **algoritmo EM**.

1.3.1. Algoritmo de Newton-Raphson (NR)

Está basado en la aproximación analítica de la función objetivo vía la aproximación lineal de su derivada.

Sean $X_1, X_2, ..., X_n$ v.a.i.i.d, con P_θ $\theta \in \Theta \subseteq \mathbb{R}^s$ y $\theta = (\theta_1, ..., \theta_s)$ donde la función objetivo es $l(\theta, x) = ln(L(\theta, x))$. El algoritmo NR busca el máximo de la log-verosimilitud a través de la aproximación de su derivada:

$$\frac{\partial}{\partial \theta} ln(L(\theta, x)) = \Big(\frac{\partial}{\partial \theta_1} ln(L(\theta, x)), ..., \frac{\partial}{\partial \theta_s} ln(L(\theta, x)) \Big)$$

basada en el desarrollo de Taylor.

Partimos de un valor inicial razonable"del parámetro: θ_0 y aproximamos la función por el desarrollo de Taylor.

$$\frac{\partial}{\partial\theta} ln(L(\theta, x)) \approx \frac{\partial}{\partial\theta} ln(L(\theta_0, x)) + H(\theta_0, x)(\theta - \theta_0)$$

Siendo $H(\theta_0, x) = \left\{ \frac{d^2}{d\theta_i d\theta_j} ln(L(\theta, x)) \right\}_{i,j}$ la matriz Hessiana de la función de log-verosimilitud.

Este resultado es válido para todo θ y se deduce:

$$\frac{\partial}{\partial\theta} ln(L(\widehat{\theta}, x)) \approx \frac{\partial}{\partial\theta} ln(L(\theta_0, x)) + H(\theta_0, x)(\widehat{\theta} - \theta_0)$$

Despejando obtenemos lo siguiente:

$$\widehat{\theta} = \theta_0 - H(\theta_0, x)^{-1} \frac{\partial}{\partial\theta} ln(L(\theta_0, x)) = \theta^{(1)}$$

Si repetimos k veces el paso anterior llegaremos a $\theta^{(k)}$, por lo que la fórmula de actualización del algoritmo en la siguiente iteración será:

$$\theta^{(k+1)} = \theta^{(k)} - H(\theta^{(k)}, x)^{-1} \frac{\partial}{\partial\theta} ln(L(\theta^{(k)}, x))$$

Es importante la elección del punto inicial para evitar convergencia a **óptimos locales**.

1.3.2. Algoritmo EM (*Expectation-maximization*)

Utilizado en situaciones donde los datos que tenemos se pueden considerar **incompletos** (censurados). Es decir, cuando el experimento tiene dos variables aleatorias Y y B pero solamente se observa $Y = y$. B no se observa y los datos completos serán $X = (Y, B)$.

Sean $X_1, X_2, ..., X_n$ v.a.i.i.d, con P_θ $\theta \in \Theta \subseteq \mathbb{R}^s$ con $X_i = (Y_i, B_i)$ y función de densidad $f(x, \theta) = f(y, b, \theta)$ donde solo $Y_1, ..., Y_n$ se observan con función de densidad $g(Y, \theta)$. En este contexto:

- f es la función de densidad de los datos **completos**.
- g es la función de densidad de los datos **observados**.

Ambas densidades dependen de θ pero la densidad de X depende también de B, datos que no hemos observado. En estos casos, el EMV lo obtendremos **siempre** de los **datos observados**.

La función de verosimilitud de los datos observados es:

$$L(Y_1, ..., Y_n, \theta) = \prod_{i=1}^{n} g(Y_i, \theta)$$

$$ln(L(Y, \theta)) = \sum_{i=1}^{n} ln(g(Y_i, \theta))$$

Y el EMV será (como siempre): $\widehat{\theta} = arg \underset{\theta}{max}[ln(L(Y,\theta))]$

Sin embargo, hay situaciones en las que es difícil resolver y plantear las ecuaciones de verosimilitud. El algoritmo EM nos permite calcular la función de verosimilitud a partir de los datos completos cuando es complicado hacerlo a partir de los observados.

$$L(X_1,...,X_n,\theta) = L((Y_1,B_1),...,(Y_n,B_n),\theta) = \prod_{i=i}^{n} f(Y_i,B_i,\theta)$$

El algoritmo EM obtiene un valor aproximado del EMV que es el máximo de los datos observados a partir del EMV de los datos completos.

Desarrollo del algoritmo EM Se puede calcular la esperanza de $ln(L(X,\theta))$ condicionada a $Y = y$ dado un valor de θ. A partir de un **valor inicial** θ_0, el algoritmo establece dos pasos:

1. **Esperanza (expectation):** Calcular $E_{\theta_0}\Big(ln(L(X,\theta))/Y\Big)$ y sustituir la probabilidad de los valores no observados por el valor esperado condicionado a lo observado.
2. **Maximización (maximization):** Obtener $\theta^{(n)}$ en el punto en el que se alcanza el máximo en la log-verosimilitud habiendo sustituido la expresión anterior.

Este proceso se repite **iterativamente** k veces.

Ejemplo:(*Este ejemplo es el de la mixtura de la práctica 3 del laboratorio.*)

Sean:

$$f_1(Y) = P_p(Y = y) = \begin{cases} 1 & \text{si } y = 0, \\ 0 & \text{si } y \neq 0 \end{cases}, \quad y = 0, 1, \dots$$

$$f_2(Y) = P_\lambda(Y = y) = e^{-\lambda}\frac{\lambda^y}{y!}, \quad y = 0, 1, \dots$$

Primero obtenemos las fórmulas de adecuación del algoritmo para p y λ.

- Datos observados: Y_i
- Datos completos: $X_i = (Y_i, B_i)$
- B_i son datos NO observados

Función de densidad conjunta:

$$f(x,p,\lambda) = f(y,b,p,\lambda) = \begin{cases} p & \text{si } b_i = 1 \ y \ y_i = 0 \\ (1-p)e^{-\lambda}\frac{\lambda^y}{y!} & \text{si } b_i = 0 \ y \ y_i \neq 0 \end{cases}$$

$$L(p,\lambda,b,y) \propto \prod_{i=1}^{n} p^{b_i}\Big[(1-p)e^{-\lambda}\frac{\lambda^y}{y!}\Big]^{1-b_i} = p^{\sum b_i}(1-p)^{n-\sum b_i}e^{-\lambda(n-\sum b_i)}\lambda^{\sum y_i(1-b_i)}$$

$$ln(L(p,\lambda,b,y)) = ln(p)\sum b_i + (n-\sum b_i)ln(1-p) - \lambda(n-\sum b_i) + ln(\lambda)\sum y_i(1-b_i)$$

Paso 1, Esperanza: Dados p_0 y λ_0 sustituir b_i por $E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)$:

$$E_{p_0,\lambda_0}\Big(ln(L(p_0,\lambda_0,b,y))\Big/Y_1,...,Y_n\Big)$$

$$= ln(p)\left[\sum E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)\right] + ln(1-p)\left[n-\sum E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)\right]$$

$$-\lambda\left[n-\sum E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)\right] + ln(\lambda)\left[\sum y_i\Big(1-E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)\Big)\right]$$

Si tenemos que $b_i^* = E_{p_0,\lambda_0}\Big(B_i\Big/Y_i=y\Big)$ entonces, haciendo uso de la definición de esperanza y Teorema de Bayes:

$$b_i^* = P\Big(B_i=1\Big/Y=y\Big)(1) + P\Big(B_i=0\Big/Y=y\Big)(0) = P\Big(B_i=1\Big/Y=y\Big)$$

$$= \frac{P_{p_0,\lambda_0}(B_i=1)P\Big(Y_i=y_i\Big/B_i=1\Big)}{P_{p_0,\lambda_0}(B_i=1)P\Big(Y_i=y_i\Big/B_i=1\Big) + P_{p_0,\lambda_0}(B_i=0)P\Big(Y_i=y_i\Big/B_i=0\Big)}$$

$$= \begin{cases} \frac{p_0}{p_0+(1-p_0)e^{-\lambda_0}} & \text{si } y_i=0 \\ 0 & \text{si } y_i\neq 0 \end{cases}$$

Sea $b^* = \sum_{i=0}^{n} b_i^*$, al ser p_0 y λ_0 constantes, b^* será **una constante** y tenemos que:

$$E_{p_0,\lambda_0}\Big(ln(L(p_0,\lambda_0,b,y))\Big/Y_1,...,Y_n\Big) = b^*ln(p)+(n-b^*)ln(1-p)-\lambda(n-b^*)+\sum y_i(1-b^*)ln(\lambda)$$

Paso 2, Maximización: Una vez tenemos la expresión para la esperanza, derivamos y buscamos el máximo en función de cada una de las variables:

$$\begin{cases} \frac{\partial}{\partial p}\ E_{p_0,\lambda_0}\Big(ln(L(p_0,\lambda_0,b,y))\Big/Y_1,...,Y_n\Big) = (...) = \frac{b^*}{p} - \frac{n-b^*}{1-p} = 0 \\ \frac{\partial}{\partial \lambda}\ E_{p_0,\lambda_0}\Big(ln(L(p_0,\lambda_0,b,y))\Big/Y_1,...,Y_n\Big) = (...) = (n-b^*) + \frac{\sum y_i(1-b_i^*)}{\lambda} = 0 \end{cases}$$

La solución de este sistema serán los valores $p^{(1)}$ y $\lambda^{(1)}$, a partir de los cuales podemos repetir el proceso de forma iterativa.

Ejercicio 1.Practica 4:

Tenemos 2 distribuciones N(0,1) y N(1,0.8). Observamos Y_i:

$$Y_i \sim \begin{cases} N(0,1) & \text{con probabilidad P} \\ N(1,0{,}8) & \text{con probabilidad 1-P} \end{cases} \qquad \text{con P desconocida}$$

Como los datos son incompletos, no observamos $B_i \sim B(p)$. Los datos completos serían $X = (Y_i, B_i)$ con i=1, . . . , n. Función de densidad para los datos completos:

$$f(y_i, b_i, p) = \begin{cases} f(y_i, 2, p) = P_p(B = 1) \cdot f_{Y/B=1}(y_i \cdot p) = p \cdot N(0,1) & \text{si b=1} \\ f(y_i, 2, p) = P_p(B = 0) \cdot f_{Y/B=0}(y_i \cdot p) = (1-p) \cdot N(1,0{,}8) & \text{si b=0} \end{cases}$$

$$= [p \cdot dnorm(y_i, 0, 1)]^{b_i} \cdot [(1-p) \cdot dnorm(y_i, 1, 0{,}8)]^{1-b_i}$$

Densidad de los datos observados Y_i.

$$g(Y_i, p) = \sum_{b=0}^{1} f(Y_i, b_i, p) = (p \cdot dnorm(y_i, 0, 1)) + ((1-p) \cdot dnorm(y_i, 1, 0{,}8))$$

$$L(p, y) = \prod_{i=1}^{n} g(Y_i, p) = \prod_{i=1}^{n} [(p \cdot dnorm(y_i, 0, 1)) + ((1-p) \cdot dnorm(y_i, 1, 0{,}8))]$$

$$logL(p, y) = \sum_{i=1}^{n} \log[(p \cdot dnorm(y_i, 0, 1)) + ((1-p) \cdot dnorm(y_i, 1, 0{,}8))]$$

$$\frac{d}{dp} logL(p, y) = \sum_{i=1}^{n} \frac{dnorm(y_i, 0, 1) - dnorm(y_i, 1, 0{,}8)}{(p \cdot dnorm(y_i, 0, 1)) + ((1-p) \cdot dnorm(y_i, 1, 0{,}8))}$$

$$\frac{d^2}{dp^2} logL(p, y) = \sum_{i=1}^{n} \frac{-(dnorm(y_i, 0, 1) - dnorm(y_i, 1, 0{,}8))^2}{((p \cdot dnorm(y_i, 0, 1)) + ((1-p) \cdot dnorm(y_i, 1, 0{,}8)))^2}$$

En general:

$$\theta_{n+1} = \theta_n - \frac{f'(\theta_n)}{f''(\theta_n)}$$

Dado un valor inicial P_0, obtenemos P.

$$p^{(1)} = p_0 - \frac{\frac{d}{dp} \log L(p_0, y)}{\frac{d^2}{dp^2} \log L(p_0, y)}$$

$$p^{(k+1)} = p_0 - \frac{\frac{d}{dp} \log L(p^{(k)}, y)}{\frac{d^2}{dp^2} \log L(p^{(k)}, y)}$$

Aplicamos el Algoritmo EM:

Verosimilitud datos completos:

Paso Esperanza: (Dado un P_0 inicial) Solo reemplazamos b_i (desconocido) por su esoeranza condicionada a las observaciones.

$$E_{p_0}(\log L(p, Y_i, B_i)/Y_1, \ldots, Y_n) = (\sum_{i=1}^{n} E_{p_0}(B_i/Y_i = y) \log p)$$

$$+(n - \sum_{i=1}^{n} E_{p_0} E_{p_0}(B_i/Y_i = y) \log(1-p))$$

$$+ \sum_{i=1}^{n} E_{p_0}(B_i/Y_i = y) \log(dnorm(y_i, 0, 1)) + \sum_{i=1}^{n} ((1 - E_{p_0}(B_i/Y_i = y)) \cdot \log(dnorm(y_i, 1, 0{,}8)))$$

Necesitamos calcular $E_{p_0}(B_i/Y_i = y)$

$$B_i = E_{p_0}(B_i/Y_i = y) = 1 \cdot E_{p_0}(B_i = 1/Y_i = y) + 0 \cdot E_{p_0}(B_i = 0/Y_i = y)$$
$$= \frac{p_0 \cdot dnorm(y_i, 0, 1)}{p_0 \cdot dnorm(y_i, 0, 1) + (1 - p_0) dnorm(y_i, 1, 0{,}8)}$$

Sea $B^* = \sum_{i=1}^{n} B_i^*$:

$$E_{p_0}(\log L(p, y_i, B_i)/y_i = y_i) = B^* \log p + (n - B^*) \log(1 - p)$$

$$+ \sum_{i=1}^{n} b_i^* \log dnorm(y_i, 0, 1) + \sum_{i=1}^{n} (1 - b_i^*) \log dnorm(y_i, 1, 0{,}8)$$

$$\frac{d}{dp} E_{p_0}(\log L(p, Y_i, B_i)/Y_1, \ldots, Y_n) = \frac{B^*}{p} - \frac{n - B^*}{1 - p} = 0$$

Despejando

$$p^{(1)} = \frac{B^*}{n} - \frac{1}{n} \sum_{i=1}^{n} \frac{p_0 \cdot dnorm(y_i, 0, 1)}{p_0 \cdot dnorm(y_i) + (1 - p_0) dnorm(y_i, 1, 0{,}8)}$$

Tema 2

Simulación y Bootstrap

Introducción a técnicas de simulación y el método bootstrap para estimación estadística.

2.1. Introducción al Bootstrap

El bootstrap es un mecanismo generador de datos. Hasta ahora hemos trabajado en una situación en la que tenemos una muestra $X = (X_1, \ldots, X_n)$ v.a.i.i.d. de una distribución P_θ, $\theta = (\theta_1, \ldots, \theta_s)$ con el interés de obtener un estimador $T(\theta)$ razonable para θ o $g(\theta)$.

Todo ello en el concepto de **inferencia frecuentista**; se tiene un estimador del parámetro en base al que queremos hacer inferencia respecto a θ. Para esto es necesario conocer la distribución del estadístico (distribución exacta o asintótica). Supongamos que:

- No conocemos la distribución de los datos
- No se cumplen las condiciones de regularidad de Cramer-Rao

Cuando se da uno de los casos anteriores, el bootstrap puede ser una buena opción. Puede resultar interesante poder repetir un mecanismo generador de datos con el que se obtuvo la muestra original de forma que podamos obtener tales muestras como se quiera, y cada una de ellas obtendrá sus estadísticos correspondientes. Es decir, a partir de P_θ obtendremos:

$$\left.\begin{array}{ccccccc} X_{11} & X_{12} & \cdots & X_{1n} & \longrightarrow & T_n^1(X) \\ X_{21} & X_{22} & \cdots & X_{2n} & \longrightarrow & T_n^2(X) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ X_{5000,1} & X_{5000,2} & \cdots & X_{5000,n} & \longrightarrow & T_n^{5000}(X) \end{array}\right\} \text{Simulaciones si } P_\theta \text{ es conocida}$$

Podemos usar bootstrap para **aproximar cualquier característica** de la distribución y hacer inferencia a partir de los datos simulados.

Sin embargo, no siempre se conoce P_θ, si no que solo se dispone de los datos observados. En estos casos no es posible simpular a partir de P_θ. Podremos simularlos si somos capaces de estimar $F_\theta(\cdot)$, la verdadera función de distribución, La estimaremos a partir de la distribución empírica:

$$\hat{F}(X) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(x_i \leq x)}$$

Donde $\nVdash$ se refiere a la función indicadora. *En los apuntes del tema 3 de probabilidad (2^o curso) se ve otra forma de definir la función de distribución empírica.*

Definición: **Principio plug-in:** cualquier característica de una distribución pueede ser aproximada. El principio plug-in está apoyado por el **Teorema de Glivenko-Cantelli**:

$$\underset{x\in\mathbb{R}}{sup}\ |\hat{F}(X) - F_0(X)| \xrightarrow{c.s.} 0$$

La idea es simular por remuestreo el experimento y, a continuación, reajustar el modelo y recalcular estimadores con los datos simulados. Estos serían los pasos:

1. Estimar $F_0(X)$ a partir de la muestra
2. Simular $\hat{F}(X)$

Con el bootstrap podemos obtener también estimadores sobre el sesgo, intervalos de confianza y contrastes de hipótesis.

2.1.1. Aproximación bootstrap de la distribución EMV

Sean $X_1, \ldots, X_n$ con $F(\dot{})$, $\hat{\theta}$ es el EMV de θ. El bootstrapsimula la distribución de $\hat{\theta}$.

1. Se estima $\hat{F}(X)$ $\begin{cases} \text{En el caso no paramétrico, a partir de la función de distribución empírica} \\ \text{En el caso paramétrico, estimando los parámetros necesarios} \end{cases}$
2. Generamos datos artificiales: las muestras bootstrap:
   - $X_1^*, \ldots, X_n^*$ con función de densidad $\hat{F}$ estimada de F
   - Se obtiene el EMV $\hat{\theta}^*$ basado en la muestra bootstrap

La idea del procedimiento es la siguiente: la distribución $\hat{\theta}^* - \hat{\theta}$ aproxima la distribución de $\hat{\theta} - \theta$. Al repetir los pasos anteriores en un proceso B veces se obtiene una versión bootstrap del EMV.

Existen dos tipos de bootstrap:

- **Bootstrap paramétrico:** si el estimador de F en el paso 1 es un estimador paramétrico
- **Bootstrap no paramétrico:** si usamos la función de distribución empírica para estimar F en el paso 1

Ejemplo

Sean $X_1, \ldots, X_n$ v.a.i.i.d de una $N(\mu, \sigma^2)$. Según lo visto en temas anteriores, sabemos que el EMV para $\theta = (\mu, \sigma)$ sigue la siguiente distribución:

$$\sqrt{(n)} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \end{pmatrix} \xrightarrow{\mathcal{L}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 4\sigma^2 \end{pmatrix} \right)$$

En este ejercicio supondremos que sabemos que los datos vienen de una distribución normal (aunque desconocemos los valores de μ y σ). Para conseguir una muestra bootstrap tenemos que seguir los pasos ya mencionados anteriormente:

1. Estimar μ y σ a partir del EMV

2. Simular desde nuestra nueva $\hat{F}$ (que sabemos que sigue una distribución normal)

(*El algoritmo se ejecuta de forma iterativa, y es un proceso muy laborioso a mano. Por ello, se deja como ejercicio propuesto al lector hacer un script que siga dichos pasos (se ve en clases prácticas.*)
La distribución de $\begin{pmatrix} \mu_i^* \\ \sigma_i^{2*} \end{pmatrix} - \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} \forall i = 1, 2, \ldots, B$ aproxima la distribución de $\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$
De igual forma, el histograma de $\mu_1^*, \ldots, \mu_B^*$ aproxima el de $\hat{\mu}$. Pasa lo mismo para σ^2

En el caso no paramétrico estimaríamos F a partir de la muestra original y su función de distribución empírica.

Estimador bootstrap de la varianza del EMV Estamos estudiando la distribución de θ, por lo que tenemos que poder estudiar cualquier característica que dependa de θ. Estimaremos la varianza del EMV usando las B muestras bootstrap.

Justificación $\Longrightarrow$ La distribución de $\hat{\theta}^* - \hat{\theta}$ aproxima la de $\hat{\theta} - \theta$; por lo que la distribución de $Var(\hat{\theta}^* - \hat{\theta})$ aproxima la de $Var(\hat{\theta} - \theta)$, y, por tanto $Var^*(\hat{\theta}^*) \approx Var^*(\hat{\theta})$

2.1.2. Intervalos de confianza bootstrap (Método percentil)

Consideremos $X_1, \ldots, X_n$, con función de distribución $F_0(\dot{})$, dependiente del parámetro s-dimensional $\theta = (\theta_1, .s, \theta_s)$. Para obtener un intervalo de confianza de nivel 1-α para la k-ésima componente de θ (θ_k) utilizaremos el **método percentil**.

Sean $\hat{\theta}_{k1}^*, \ldots, \hat{\theta}_{kB}^*$ las B versiones bootstrap del estimador $\hat{\theta}_k$, y sean $\hat{\theta}_{k,\frac{\alpha}{2}}^*$ y $\hat{\theta}_{k,1-\frac{\alpha}{2}}^*$ los cuantiles $\alpha/2$ y $1-\alpha/2$ respectivamente. El intervalo de confianza bootstrap percentil para θ_k será $\left(\hat{\theta}_{k,\frac{\alpha}{2}}^*, \hat{\theta}_{k,1-\frac{\alpha}{2}}^*\right)$. Es decir, el método percentil consiste en sustituir los extremos del intervalo por los percentiles correspondientes para nuestro nivel α.

$$P_\theta\left(\hat{\theta}_{k,\frac{\alpha}{2}}^* \leq \theta \leq \hat{\theta}_{k,1-\frac{\alpha}{2}}^*\right) \approx 1 - \alpha$$

La justificación viene dada por la suposición de que, bajo las condiciones de regularidad apropiadas, el comportamiento de $\hat{\theta}$ como estimador de θ sea parecido al comportamiento de $\hat{\theta}^*$ como estimador de $\hat{\theta}$. En otras palabras, un intervalo que contenga a $\hat{\theta}_k^*$ con probabilidad aproximada $1-\alpha$ es también un intervalo que contiene a θ_k con probabilidad aproximada $1-\alpha$.

En cuanto a una transformación $g(.)$, si la función g es monótona creciente, el método percentil es **invariante a transformaciones**.

2.1.3. Contrastes de hipótesis bootstrap

Sean $X_1, \ldots, X_n$ i.i.d. con $f(., \theta), \theta \in \Theta$. Vamos a contrastar $H_0 : \theta \in \Theta_0$; $H_1 : \theta \notin \Theta_0$.

Sea T el estadístico de contraste y $\{T \geq C_\alpha\}$ la región crítica de nivel α. Existen situaciones en las que es posible calcular la distribución asintótica o exacta del estadístico T bajo la hipótesis nula. En ese caso, podemos determinar directamente C_α y calcular el p-valor del test. En el caso en el que calcular la distribución de T no sea posible podemos aproximarla mediante bootstrap.

- H_0 **es simple**: el bootstrap no es necesario, basta con simulación. La idea es simular un número grande de muestras (B) de tamaño n de la distribución; sean $T_1, \ldots, T_B$ los valores del estadístico T observados en las B muestras, podemos aproximar C_α por el cuantil $1 - \alpha$ de estos B valores y el p-valor del test por la proporción de estos valores mayores que el observado para los datos originales.

- H_0 **es compuesta**: bootstrap paramétrico. Si $\hat{\theta}_0$ es el EMV de θ bajo H_0, se generan B muestras **bootstrap** de la distribución. Al igual que antes se calcula el p-valor del test a partir de la proporción de los valores mayores que el valor observado para los datos; pero esta vez con muestreo bootstrap.

Tema 4

Técnicas de Bondad de Ajuste

Explicación de las técnicas de bondad de ajuste para evaluar la adecuación de modelos estadísticos.

3.1. Introducción al test de Bondad de Ajuste

Hasta ahora los problemas planteados parten de unos datos obtenidos en un experimento aleatorio del que se conoce el mecanismo con el que se genera (Familia de distribución P_θ).

Definición: El Test de Bondad de Ajuste es un test para comprobar si una familia de distribuciones, representa correctamente el mecanismo con el que se generaron los datos.

Planteamiento:

$X_1, \ldots, X_n$ con funcion de distribución F. Si F_0 es una función de distribución conocida, (por ejemplo, una Poisson con $\lambda = 3$), el problema se reduce a:

$$H_0 : F = F_0 \quad H_1 : F \neq F_0$$

F_0 está completamente especificada, por lo que es una hipótesis simple.

La hipótesis nula sería compuesta si F_0 depende de parámetros desconocidos, por ejemplo, una P(λ)

Empezaremos con el caso de H_0 simple. Existen dos tipos de Test de Bondad de Ajuste si F_0 es una función de distribución conocida.

1. F_0 discreta: Se comparan las frecuencias observadas con las esperadas bajo H_0 (Test χ^2). También se puede hacer si F_0 es continua agrupando, pero hay tests más potentes para esos casos.

2. F_0 continua: Comparamos la funcion de distribución empírica con la teórica.

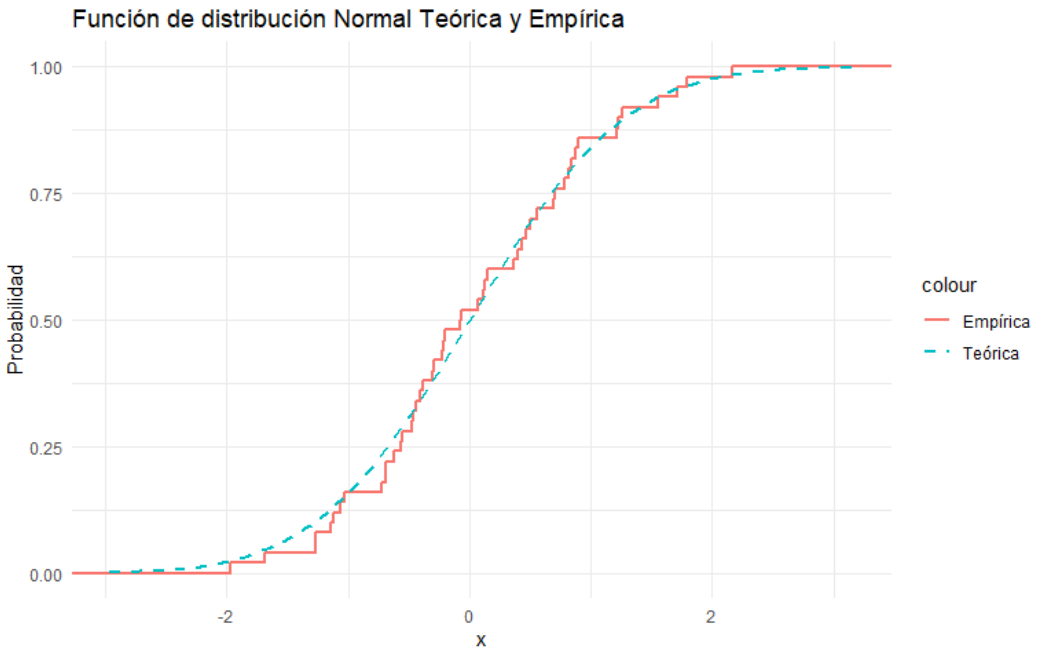

Figura 3: Visualización de una distribución teórica y empírica

3.1.1. Test Chi-Cuadrado de Bondad de Ajuste

Sea $X_1, \ldots, X_n$ con distribución F discreta:

$$\begin{array}{l} C_1 \to P_1 \\ \quad \ldots \qquad p_j \geq 0, \quad \sum_{j=1}^{n} p_j = 1 \\ C_k \to P_k \end{array}$$

Caso en el que F_0 completamente especificada bajo H_0:

$$H_0 : p = p^0 \quad p = (p_1, \ldots, p_k)'$$

$$H_1 : p \neq p^0 \quad p^0 = (p_1^0, \ldots, p_k^0)'$$

Este problema ya lo sabemos resolver: es un contraste para el parámetro p de una distribución multinomial.

Frecuencia observada en $C_j : f_j = \sum_{i=1}^{n} \mathbb{1}_{(x_i = c_j)}, \quad j = 1, \ldots, k \quad \sum_{j=1}^{k} f_j = n$

Frecuencia esperada bajo H_0 en C_j:

$$e_j = n \cdot P_0(x = c_j) = n \cdot p_j^0$$

Ejemplo:

Sea una variable aleatoria X cuya función de masa de probabilidad es la siguiente

$$P(X=x)=\begin{cases}\frac{1}{3} & \text{si } x=0\\ \frac{1}{3} & \text{si } x=1\\ \frac{1}{3} & \text{si } x=2\end{cases}$$

Entonces la hipotesis nula para el test será que

$$H_0: \quad p_1=\frac{1}{3} \quad p_2=\frac{1}{3} \quad p_3=\frac{1}{3}$$

Tomando una muestra de n=10 quedan 7 ceros, 2 unos y 1 dos.

A simple vista no parece que siga esa distribución. Usaremos un estadístico para medir que tan diferente es de nuestra distribución pues lo que esperabamos obtener es:

$$e_1=3,33$$
$$e_2=3,33$$
$$e_3=3,33$$

El test χ^2 de ajuste proporciona unas bases probabilisticas para decidir si las diferencias son suficientemente grandes tal que no hayan ocurrido por puro azar. Se define como

$$\chi^2=\sum_{j=1}^{k}\frac{(f_j-e_j)^2}{e_j}$$

Este estadístico es lo que se conoce como distancia χ^2 entre f_j y e_j. Para valores grandes en χ^2 implica frecuencias observadas y esperadas muy diferentes.

Fijado α:

- Región critica: $(\chi^2 > C_\alpha)$
- p-valor: $p_0(\chi^2 > X_{obs}) = p_0(\chi^2 > t_{obs})$

Necesitamos conocer la distribución del estadístico bajo H_0. Se podría conocer de forma exacta aunque es muy complejo, por ello, nos interesará la distribución asintótica para n grande.

Resultado:

$$\lim_{n\to\infty} P_0(\chi^2 \leq Z) = P(\chi^2_{k-1} \leq Z)$$

La distribución asintótica del estadístico χ^2 bajo H_0 es χ^2_{k-1} g'.

Ya hemos visto que en este contexto F_0 es una distribución multinomial. Vamos a ver que el estadístico T bajo H_0 para el modelo multinomial con $k-1$ parámetros libres es asintóticamente equivalente al estadístico χ^2_1.

Demostración 3.1. Estadístico RV para:

$$H_0 : F = F_0$$

$$H_1 : F \neq F_0$$

Recordemos:

$$L(p, x) = \prod_{j=1}^{k} p_j^{f_j} \quad \Longrightarrow \quad \log L(p, x) = \sum_{j=1}^{k} f_j \log p_j$$

El estadístico es:

$$T = 2\left[\log L(\widehat{p}, x) - \log L(p_0, x)\right] = 2\left[\sum_{j=1}^{k} f_j \log \widehat{p_j} - \sum_{j=1}^{k} f_j \log p_j^0\right]$$

$$= -2\sum_{j=1}^{k}\left(\log p_j^0 - \log \widehat{p_j}\right)$$

Aproximamos $\log p_j^0$ por un desarrollo de Taylor en torno a $\log \widehat{p_j}$:

$$\log p_j^0 \approx \log \widehat{p_j} + (p_j^0 - \widehat{p_j})\frac{1}{\widehat{p_j}} - \frac{(p_j^0 - \widehat{p_j})^2}{2 \cdot \widehat{p_j}^2} + \dots$$

$$\rightarrow \log p_j^0 - \log \widehat{p_j} \approx (p_j^0 - \widehat{p_j})\frac{1}{\widehat{p_j}} - \frac{(p_j^0 - \widehat{p_j})^2}{2}\frac{1}{\widehat{p_j^2}}$$

Sabiendo que $\widehat{p_j} = \frac{f_j}{n}$

$$= \left(p_j^0 - \frac{f_j}{n}\right) \cdot \frac{n}{f_j} - \frac{\left(p_j^0 - \frac{f_j}{n}\right)^2}{2} \cdot \left(\frac{n}{f_j}\right)^2 \xrightarrow{c.s.} 0$$

Esto converge a 0, ya que, por la Ley de los Grandes Números (LGN), $\frac{F_j}{n}$ es un estimador consistente de p_j:

$$\lim_{n\to\infty} P\left(\left|\frac{F_j}{n} - p_j\right| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0$$

Por lo tanto, el estadístico T queda:

$$T = -2\sum_{j=1}^{k}\left(n \cdot p_j^0 - f_j\right) + \frac{\sum_{j=1}^{k}(f_j - n \cdot p_j^0)^2}{f_j}$$

El test χ^2 y el TRV son asintóticamente equivalentes y como TRV converge a χ^2_{k-1}, el estadístico χ^2 también. Fijado un α.

- Región critica: $P_0(\chi^2 \geq C_\alpha) \quad C_\alpha = qchisq(1-\alpha, k-1)$
- p-valor: $P_0(\chi^2 \geq t_{obs}) = 1 - qchisq(t_{obs}, k-1)$

Nota:

Esta aproximación es válida para frecuencias mayores o iguales a 5

Ejercicio 1: Script R

Ejercicio 4: En una fábrica con 220 empleados, el número de trabajadores que tuvieron accidentes se recoge en latabla siguiente:

Nº accidentes	0	1	2	3	4	5	6+
Nº trabajadores	181	9	4	10	7	4	5

Queremos contrastar si los datos observados podrían atribuirse a un proceso de Poisson. Lo habitual es desconocer la información poblacional de dicha distribución, lo que significaría tener una hipótesis compuesta.

Para simplificar el ejercicio en este caso nos preguntaremos... ¿Son estos datos consistentes con la distribución de Poisson para $\lambda = 1$?

$$f(x,1) = \frac{e^{-1} \cdot \lambda^x}{x!} = \frac{e^{-1}}{x!}$$

$$p_0^0 = e^{-1}, \quad p_1^0 = e^{-1}, \quad p_2^0 = \frac{e^{-1}}{2}, \quad \ldots, \quad p_6^0 = 1 - \sum_{j=0}^{b} p_j^0$$

Calculamos las frecuencias esperadas

$$e_0 = 220 \cdot e^{-1}, \quad e_1 = 220 \cdot e^{-1}, \quad \ldots$$

$$\chi^2 = \frac{(181 - 220 \cdot e^{-1})^2}{220 \cdot e^{-1}} + \ldots$$

Por tanto concluimos que es muy probable que no siga una $P(1)$

Ejercicio 10: Se lanza una moneda hasta que aparece la primera cara. Este experimento se repite 100 veces. Las frecuencias observadas del numero de ensayos necesarios hasta que aparece la primera cara son:

Nº ensayos	1	2	3	4	5+
Frecuencia	40	32	15	7	6

¿Se puede concluir que la moneda es perfecta?

Distribución geométrica:

$$P_p(X = k) = (1-p)^{k-1}p \quad o \quad P_p(X = k) = (1-p)^k p$$

Probabilidad bajo H_0.

$$p_1^0 = \frac{1}{2} \quad p_2^0 = \frac{1}{2^2} \quad \ldots \quad p_5 = 1 - \sum_{k=1}^{4} p_k$$

Como H_0 es "Los datos provienen de una distribución geométrica $\left(\frac{1}{2}\right)$".

Según el test, no rechazamos H_0.

Si en realidad los datos vienen de una distribución geométrica $\left(\frac{1}{3}\right)$, ¿cuál es la probabilidad de que el test lo detecte? ¿Y si vienen de una distribución geométrica (0,52)?

3.1.2. Distribución del test bajo H_1

Podemos aumentar la potencia sacrificando α o aumentando el tamaño de la muestra. Si queremos asegurar una potencia de, por ejemplo, 0,8 o 0,9, necesitamos calcular cuánto debe valer n.

Para esto, necesitamos conocer la distribución del test bajo H_1.

Resultado:

$$\lim_{n\to\infty} P_{\theta_1}(\chi^2_{k-1} \leq t) = P(\chi^2_{k-1}(\delta) \leq t)$$

Donde δ es el parámetro de descentralidad para la χ^2_{k-1} descentrada.

Definición: La distribución asintótica bajo H_1 del estadístico χ^2 de prueba de ajuste es una χ^2 descentrada con $k-1$ grados de libertad y parámetro de descentralidad δ.

La demostración no está incluida en este curso. Sin embargo, hacemos algunos comentarios sobre esta definición:

Comentarios:

1. - Si X es una variable aleatoria tal que $X \sim N(0,1)$, entonces $X^2 \sim \chi^2_1$.
   - Si $X_1, \ldots, X_n$ son v.a.i.i.d. $N(0,1)$, entonces $\sum_{i=1}^n X_i^2 \sim \chi^2_n$.
   - Si $X_1, \ldots, X_n$ son v.a.i.i.d. $N(\mu, \sigma^2)$, entonces $\sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2_n$ descentrada con parámetro δ.

2. Para contrastar
$$H_0 : p = (p_1^0, \ldots, p_k^0)$$
$$H_1 : p \neq p^0$$
si bajo H_1 la alternativa es $p = (p_1^1, \ldots, p_k^1)$, el test χ^2 sigue una distribución $\chi^2_{k-1}(\delta)$, donde δ mide la distancia entre los dos vectores:
$$\Delta = \sum_{j=1}^k \frac{(p_j^1 - p_j^0)^2}{p_j^0}; \quad \delta = n \cdot \Delta$$
O equivalentemente:
$$\delta = \sum_{j=1}^k \frac{\left(n \cdot p_j^1 - n \cdot p_j^0\right)^2}{n \cdot p_j^0}.$$

3. Existen tablas para $\chi^2_{k-1}(\delta)$ (por ejemplo, las tablas estándar de χ^2 descentrada).

Ejercicio 11: Durante 60 meses consecutivos se observó la variable aleatoria X="nº de accidentes al mes en un cruce de carreteras". Los resultados fueron:

X	0	1	2	3+
Nº de meses	23	18	12	7

a) Contrastar la hipótesis de que la distribución de X es geométrica:

$$G(\theta); \quad p_\theta(X = x) = \theta \cdot (1-\theta)^x, \quad X = 0, 1, 2, \ldots$$

$H_0 : X \sim G(\theta) = \text{EMV para } X \in \{0,1,2,3+\}$

$$P_0^0 = \theta, \quad P_1^0 = \theta \cdot (1-\theta), \quad P_2^0 = \theta \cdot (1-\theta)^2, \quad P_3^0 = (1-\theta)^3$$

$$L(\theta, f) = \prod_{j=1}^{4} [p_j^0]^{f_i} = \theta^{23} \cdot (\theta \cdot (1-\theta))^{18} \cdot (\theta \cdot (1-\theta)^2)^{12} \cdot ((1-\theta)^3)^7$$

$$logL(\theta, f) = 53 \log \theta + 63 \log(1-\theta)$$

$$\frac{\partial \log(\theta, f)}{\partial \theta} = \frac{53}{\theta} - \frac{63}{1-\theta} \Longrightarrow \widehat{\theta} = \frac{53}{53+63}$$

Test χ^2:

$$\widehat{\chi^2} = \sum_{j=0}^{3} \frac{(f_j - 60 \cdot P_l^0(\widehat{\theta}))^2}{60 \cdot p_j^0(\widehat{\theta})} \sim \chi_2^2$$

b) Potencia de $H_0 : G(0{,}4) \quad B_n(2, 0{,}6)$:

$$P_1(\chi^2 > C_{0,05}) = P(\chi_3^2(\delta) > C_{0,05})$$

$$S = n \cdot \Delta, \quad \Delta = \sum_{j=0}^{3} \left(\frac{(p_j^0 - p_j^1)^2}{p_j^0} \right)$$

3.2. Test de Kolmogorov-Smirnov

3.2.1. Test de Kolmogorov-Smirnov para hipótesis simples

$X_1, \ldots, X_n$ i.i.d. con función de distribución F continua y se quiere contrastar

$$H_0 : F = F_0 \quad H_1 : F \neq F_0$$

El test de Kolmogorov-Smirnov es más potente que el χ^2 en el caso de F continua. Tenemos:

- $F_0 \rightarrow$ Función de distribución bajo H_0.
- $\widehat{F_n} \rightarrow$ Función de distribución muestral o empírica.

$$\widehat{F_n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(x_i \leq x)}, \quad x \in \mathbb{R}$$

Tambien se puede definir a partir de los estadísticos de orden:

$$\widehat{F_n}(x) = \begin{cases} 0, & \text{si} \quad x \leq X_{(1)} \\ \frac{i}{n} & \text{si} \quad X_{(i)} \leq x \leq X_{(i+1)}, \quad i = 1, \ldots, n-1 \\ 1 & \text{si} \quad X_n \leq x \end{cases}$$

Propiedades:

1. $n \cdot \widehat{F_n}(x)$ es el total de valores de la muestra menores o iguales a x y sigue una distribución $B(n, F(x))$, $\forall x \in \mathbb{R}$

2. Según el resultado anterior, junto con las propiedades de la distribución binomial, se tiene $\widehat{F}(x)$ es un estimador consistente de $F(x)$

$$\lim_{n\to\infty} P(|\widehat{F_n}(x) - F(x)| < \varepsilon) \to 0$$

y es CAN: $\sqrt{n} \cdot (\widehat{F_n}(x) - F(x)) \backsimeq N(0, F(x)(1 - F(x)))$

3. Por el teorema de Glivenko-Cantelli, $\widehat{F_n}(x)$ converge uniformemente y casi seguro a $F(x)$.

A medida que n crece, al función escalonada de $\widehat{F_n}(x)$ con saltos en los valores de los estadísticos de orden $X_{(1)}, \ldots, X_{(n)}$ aproxima la distribución $F(x)'$.

Por lo tanto cuando n es grande, la mayor diferencia entre $\widehat{F_n}(x)$ y $F(x)$ converge a 0.

Este resultado nos sugiere el estadístico $D_n = sup_x[\widehat{F_n}(x) - F(x)]$ el cual es una medida razonable de la precisión de la estimación.

Llamaremos estadístico de Kolmogorov-Smirnov de una muestra a

$$D_n = sup_x|\widehat{F_n}(x) - F_0(x)|$$

El test rechaza la hipótesis para valores grandes del estadísitico ($D_n > C$). Como siempre debemos conocer la distribución de D_n para obtener el valor crítico.

1. Es importante ver que la máxima diferencia en $|\widehat{F_n}(x) - F_0(x)|$ es la máxima diferencia en los puntos en los que $\widehat{F_n}(x) > F_0(x)$ y la máxima en los punto donde $F_0(x) > \widehat{F_n}(x)$. Así podemos definir los estadísticos Kolmogorov-Smirnov de un lado.

$$\left.\begin{array}{l} D_n^+ = sup_x(\widehat{F_n}(x) - F_0(x)) \\ D_n^- = sup_x(F_0(x) - \widehat{F_n}(x)) \end{array}\right\} D_n = max\{D_n^+, D_n^-\}$$

   D_n^+ y D_n^- son útiles para conocer la distribución.

2. La distribución de D_n^+, D_n^- y D_n no depende de F_0, es decir, c_α será el mismo sin importar si estamos contrastando diferentes hipótesis simples (Como por ejemplo $N(3, 1)$, $exp(2)$, $\ldots$). Se dice que el estadístivo es de distribución libre ya que no depende de F_0.

Definimos los estadísticos de orden como $X_{(0)} = -\infty$ y $X_{(n)} = +\infty$, por lo tanto podemos definir $\widehat{F_n}(x) = \frac{i}{n} \quad \forall X_{(i)} \leq x < X_{(i+1)}$.

D_n^+ lo podemos ir calculando a trozos como:

$$D_n^+ = sup_x(\widehat{F_n}(x) - F_0(x))$$

Tomará un valor máximo unicamente en los extremos del intervalo $[X_{(i)}, X_{(i+1)})$

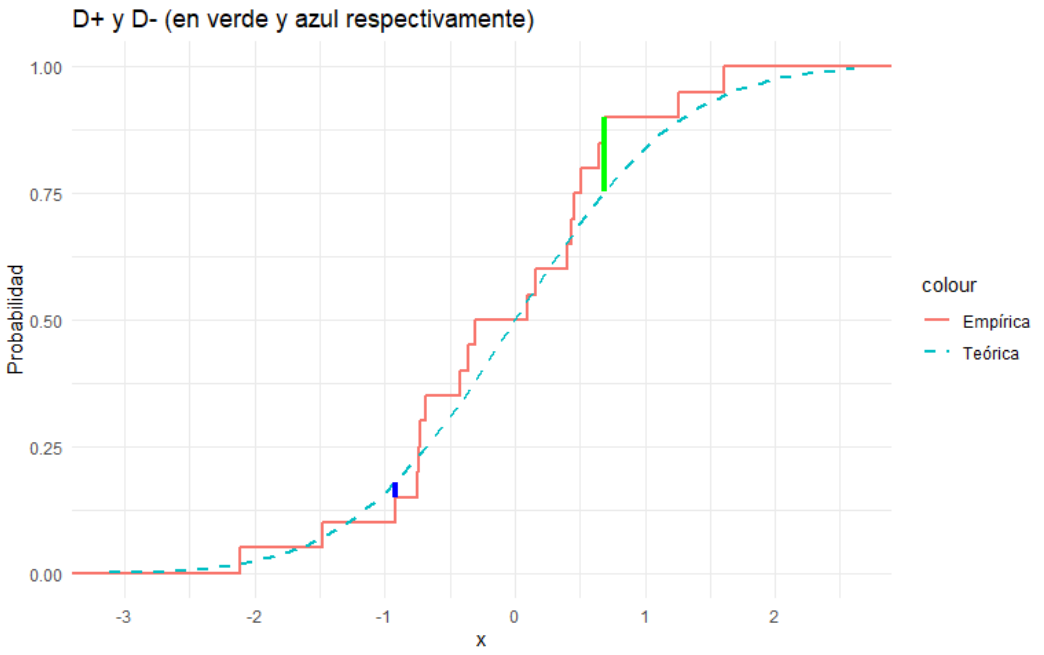

Figura 4: Visualización del estadísitico D+ y D-

$$D_n^+ = max_{1\leq i\leq n}\left(\frac{i}{n} - F_0(x_{(i)})\right)$$

Análogamente:

$$D_n^- = max_{1\leq i\leq n}\left(F_0(x_{(i)}) - \frac{i-1}{n}\right)$$

D_n^+, D_n^-(por lo tanto D_n) dependen de las variables aleatorias.

$$F_0(X_{(1)}), \ldots, F_0(X_{(n)}) \sim U(0,1)$$

Esto ocurre por la trasnformada integral de probabilidad, siempre y cuando las variables aleatorias sean continuas.

3.2.2. Test de Kolmogorov-Smirnov para una hipótesis compuesta

Lo anterior nos sirve para hipótesis simples. Sin embargo en la mayoría de los escenarios reales, tenemos hipótesis compuestas. Por ejemplo, utlizar Kolmogorov-Smirnov en ANOVA.

Situación:

$$H_0 : F = F_0(\theta) \quad H_1 : F \neq F_0(\theta)$$

Lo haremos como siempre:

1. Estimamos θ a partir de los datos(con el Estimador Máximo Verosimil).
2. Se calcula el estadístico Kolmogorov-Smirnov como en el caso de la hipótesis simple:

$$\widehat{D_n} = sup_x|\widehat{F_n}(x) - F(x)|$$

Se rechaza H_0 con valores grandes de $\widehat{D_n}$ ($\widehat{D_n} > C$).

Es importante recalcar que $\widehat{D_n}$ no sigue la misma distribución que D_n. La distribución no es libre, depende de la familia que estemos evaluando.

Tenemos tablas obtenidas por Lilliefors por simulación para la normal y para la exponencial. Pasos a seguir:

1. Se estiman los parámetros $\hat{\mu}$ y $\widehat{\lambda}$
2. Se obtiene la muestra estandarizada $z_1, \ldots, z_n$.
3. Hacer el test ($H_0 : F_0 \equiv N(0,1)$ en el caso normal y $H_0 : F_0 \equiv exp(1)$ en el caso exponencial).
4. Se rechaza H_0 si $\widehat{D_n} > C$ para nivel α. Se usan las tablas de Lilliefors para definir C_α.

Caso exponencial:

$$H_0 : F \equiv exp(\lambda)$$
$$H_1 : F \not\equiv exp(\lambda)$$

$$\widehat{\lambda} = \frac{1}{\overline{X}}, \quad z_i = \frac{X_i}{\overline{X}} \to z_i \sim exp(1) = \gamma(1, \lambda)$$

El contraste es equivalente a:

$$H_0 : X \sim \gamma\left(1, \frac{1}{\overline{X}}\right) \longleftrightarrow H_0 : Z \sim \gamma(1,1)$$

Podrás descubrir más estadísticos en el campus virtual

3.3. Test de Shapiro-Wilk

(Este test es más potente que el de Kolmogorov-Smirnov)

Mide como de bien se ajustan los datos a una distribución normal esperada. Se basa en combinación lineal de estadísticos ordenados.

$$W = \frac{\left(\sum_{i=1}^{n} a_i \cdot X_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

donde

- $X_{(i)}$ es el estadístico de orden *i-ésimo*.
- a_i son los coeficientes calculados a partir de los cuartiles esperados de una distribución normal estandar.

$$(a_1, a_2, \ldots, a_n) = \frac{m^T \cdot v^{-1}}{\sqrt{m^T \cdot V^{-1} \cdot V^{-1} \cdot m}}$$

tal que

- $m = (m_{(1)}, \ldots, m_{(n)})$
- V es la matriz de las covarianzas de $m_{(i)}$

El numerador nos indica como de bien se alinean los datos con la normalidad esperada. Si los datos son normales, el numerador será grande.

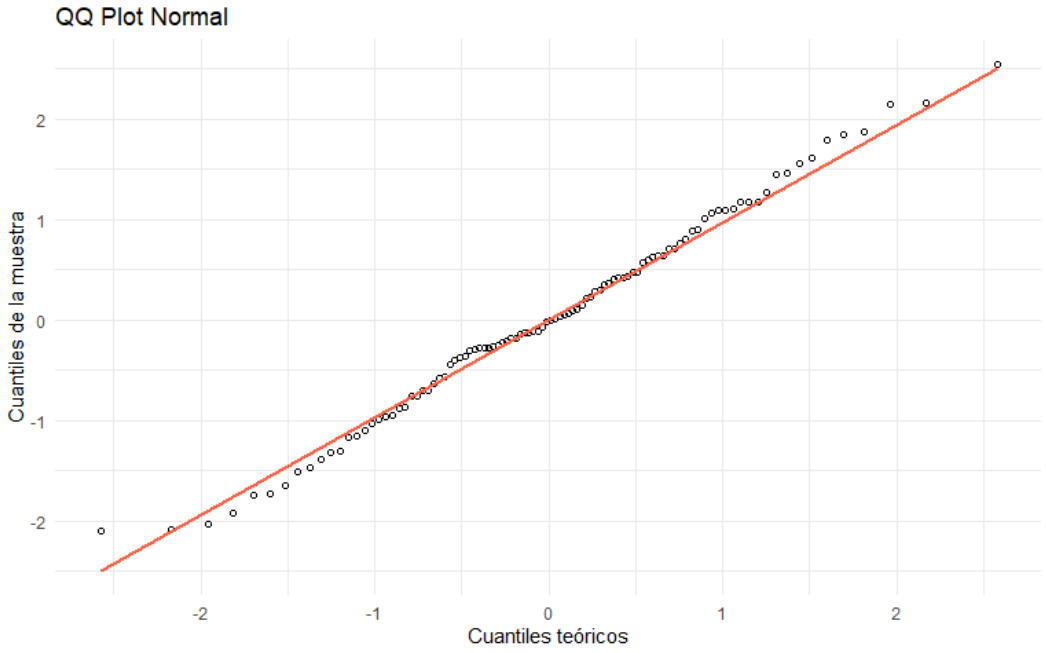

Figura 5: Visualización del QQ plot entre una distribución normal y la muestra

Tema 5

Contrastes basados en estadísticos de rangos

Introducción a los contrastes estadísticos basados en rangos para la comparación de muestras.

4.1. Test de rangos

Caso paramétrico normal.
$X_1, \ldots, X_n$ i.i.d. $N(\mu_1, \sigma_1)$
$Y_1, \ldots, Y_n$ i.i.d. $N(\mu_2, \sigma_2)$

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

En el caso no paramétrico:
$X_1, \ldots, X_n$ i.i.d. con distribución F
$Y_1, \ldots, Y_n$ i.i.d. con distribución G

$$H_0 : F = G$$

$$H_1 : F \neq G$$

En ambos casos el objetico es el mismo, comparar tratamientos o resultados.

Ejemplo: Digamos que se quiere contrastar la eficacia de un nuevo medicamento para una enfermedad. Lo primero que se tiene que hacer es diseñar un experimento para obtener los datos.

Kowalski, opciones (para diseñar el experimento xd):

Tenemos dos opciones...

1. **Modelo de aleatorización**: Los datos vienen de un diseño controlado en el que los individuos de análisis han sido asignados aleatoriamente a diferentes grupos.

2. **Modelo poblacional:** Los datos son extraidos de una población y se asume que esa población tiene ciertas propiedades. Por ejemplo: una distribución

4.1.1. Modelo de aleatorización

Se dispone de N individuos. Se eligen n individuos a los que se asigna un tratamiento (una medición). Caso donde

$$X_1, \ldots, X_m \quad H_0 : \text{El tratamiento no tiene efecto}$$

$$Y_1, \ldots, Y_n \quad H_1 : \text{El tratamiento tiene efecto}$$

El estadístico que utilicemos, rechazará H_0 cuando los valores de la variable considerada en los tratados sean mayores a los de control. Para esto, usaremos **el estadístico basado en rangos**. El estadístico basado en rangos, no depende de unidades de medida. Tomaremos toda la muestra y asignaremos rangos.

Definición: Un rango es el lugar que ocupa la observación en la muestra ordenada. Sean $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ nuestra muestra completa. Se ordenan los valores y se asignan rangos $(i_1, \ldots, i_m, i_{m_1}, i_n)$ Esto es la permutación donde i_k es el valor que ocupa la observación k-ésima en la muestra ordenada.

Ejemplo:

Tenemos $X = \{5, 8, 9\}$ e $Y = \{6, 7, 10\}$

Por tanto la muestra completa será $\{5, 8, 9, 6, 7, 10\}$

Que ordenado $\{5, 6, 7, 8, 9, 10\}$, y se le asigna un rango a cada elemento $(\{1, 2, 3, 4, 5, 6\})$

Usando notación de tuplas, (a, b) donde a es un elemento de la muestra y b el rango asociado, la anterior asignación quedaría de la siguiente manera $\{(5, 1), (6, 2), (7, 3), (8, 4), (9, 5), (10, 6)\}$

Por tanto quedarían asignados...

Rangos de X: $\{1, 4, 5\}$.

Rangos de Y: $\{2, 3, 6\}$.

Notación formal:

$R_1, \ldots, R_m \longrightarrow$ rangos correspondientes a las observaciones $X_1, \ldots, X_m$.

$S_1, \ldots, S_n \longrightarrow$ rangos correspondientes a las observaciones $Y_1, \ldots, Y_n$.

¿Cuando las Y son mayores que las X?. Cuando s_i sean más grandes, es decir, la suma de rangos sea más grande.

Estadístico de suma de rangos ***Definición:*** La suma

$$W_s = s_1 + \cdots + s_n$$

es conocida como el estadístico de Wilcoxon de suma de rangos ($W_r = R_1 + \cdots + R_m$). Se rechazará H_0 para valores de W_s grandes ($W_s > c_\alpha$).

Como siempre, debemos conocer la distribución del estadístico de W_s bajo H_0. Como la distribución es discreta, y podemos calcular lo siguiente...

$$P_{H_0}(W_s = k) = \sum_{s_1+\cdots+s_n=k} P_{H_0}((S_1, \ldots, S_n) = (s_1, \ldots, s_n))$$

... encontrar la distribución de W_s bajo H_0 se reduce a encontrar la distribución de $(S_1, \ldots, S_n)$.

$$P_{H_0}((S_1, \ldots, S_n) = (s_1, \ldots, s_n)) = \frac{1}{\binom{N}{n}}$$

Cada resultado es igual de probable bajo H_0.

Ejemplo: Dados $N = 5$, $m = 2$, $n = 3$

$$\binom{N}{n} = \binom{5}{3} = \frac{5!}{3! \cdot 2!} = 10$$

Habrá 10 posibles resultados de s_1, s_2, s_3.

Tratados	$(1,2,3)$	$(1,2,4)$	$(1,2,5)$	$\ldots$	$(3,4,5)$
$P_{H_0}((S_1, \ldots, S_n) = (s_1, \ldots, s_n))$	0,1	0,1	0,1	$\ldots$	0,1
W_s	6	7	8	$\ldots$	12

Distribución de W_s bajo H_0:

k	6	7	8	9	10	11	12
$P_{H_0}(W_s = k)$	0,1	0,1	0,2	0,2	0,2	0,1	0,1

Rechazaremos H_0 cuando el valor de W_s sea poco probable.

Observaciones:

La distribución de W_s no es la misma si hubiéramos asignado n a tratamientos y m a control. Lo que es lo mismo, W_R no sigue la misma distribución que W_s bajo H_0. La distribución dependerá de n y m.

4.1.2. Estadístico de Mann-Whitney

En el caso anterior, el valor mínimo de W_s corresponde a la situación a la que los individuos con tratamiento toman los valores más pequeños. Donde en el ejemplo anterior, $W_s = 6 \quad 6 = \frac{n \cdot (n+1)}{2}$. Si consideramos el estadístico de Mann-Whitney,

$$W_{XY} = W_S - \frac{n(n+1)}{2}$$

se nos facilitará hacer la tabla porque toma valores $0, 1, \ldots, (n \cdot m)$. Tomará el valor 0 cuando todos los valores de Y son los más pequeños y el valor $n \cdot m$ cuando todos los Y tomen los valores más grandes.

Una ventaja que tiene el estadístico es que toma los mismos valores sin importar la decisión de cuántos asignar a n y cuántos a m.

Del mismo modo:

$$W_{YX} = W_R - \frac{m(m+1)}{2}$$

W_{YX} también toma valores 0,1,$\ldots, n \cdot m$.
W_{XY} y W_{YX} siguen la misma distribución bajo H_0.
Existen tablas para esta distribución.

Observaciones:

La distribución bajo H_0 de W_S (o W_R) es simétrica respecto a $\frac{n\cdot(N+1)}{2}$.

$$\forall k \quad P_{H_0}\left(W_S = \frac{n\cdot(N+1)}{2} + k\right) = P_{H_0}\left(W_S = \frac{n\cdot(N+1)}{2} - k\right)$$

Bajo H_0 X e Y están igualmente distribuidas, por lo que todos los elementos deberían ser indistinguibles en término de rangos. Por esto W_{XY} y W_{YX} están igualmente distribuidas bajo H_0.

Ejemplo del uso de las tablas:

Tenemos $W = 10$, $n = 6$, $m = 4$
Queremos saber $P_{H_0}(W_s \geq 35)$
Tenemos las tablas para $W_{XY} = W_S - \frac{n\cdot(n+1)}{2} = W_S - 21$
Debemos escribir $P_{H_0}(W_s \geq 35)$ como $P_{H_0}(W_{XY} \leq a)$
Sabemos que W_S es simétrico a $\frac{n\cdot(N+1)}{2} = 33$. Entonces...

$$P_{H_0}(W_S \geq 35) = P_{H_0}(W_S \geq 33+2) = P_{H_0}(W_S \leq 33-2) = P_{H_0}(W_S \leq 31)$$

$$= P_{H_0}\left(W_S - \frac{n\cdot(n+1)}{2} \leq 31 - 21\right) = P_{H_0}(W_{XY} \leq 10) = 0{,}3810 \text{(Usando las tablas)}$$

Llegaríamos al mismo resultado usando $W_S + W_R = 55$.

$$P_{H_0}(W_S \geq 35) = P_{H_0}(W_R \leq 20) = P_{H_0}\left(W_R - \frac{m\cdot(m+1)}{2} \leq 20 - 10\right)$$

$$= P_{H_0}(W_{YX} \leq 10) = 0{,}3810$$

Hay tablas hasta $n = m = 10$. A partir de así usariamos la distribución asintótica.

Forma alternativa del estadístico de Mann-Whitney Una manera alternativa de ver el estadístico de Mann-Whitney que va a ser útil para calcular la potencia y también intervalos de confianza es la siguiente

Definición: Si $X_1, \ldots, X_m$ son valores para individuos sin tratamientos y $Y_1, \ldots, Y_n$ son valores para individuos con tratamiento, y $W_{XY} = W_S - \frac{n\cdot(n+1)}{2}$, W_{XY} es también el número de pares (X_i, Y_j) $i = 1, \ldots, m$, $j = 1, \ldots, n$ para los que $X_i < Y_i$

$$W_{XY} = \#[(X_i, Y_j) | X_i < Y_j]$$

Demostración 4.1. Sean $Y_{(1)}, \ldots, Y_{(n)}$ valores ordenados de $Y_1, \ldots, Y_n$ y sean $s_1, \ldots, s_n$ los rangos correpondientes también ordenados,

- Hay $s_1 - 1$ observaciones menores que $Y_{(1)}$ y todas son X.

$$\#[(X_i, Y_{(1)})|X_i < Y_{(1)}] = S_1 - 1$$

- Hay $s_2 - 1$ observaciones menores que $Y_{(2)}$, de ellas una es $Y_{(1)}$ y el resto son X.

$$\#[(X_i, Y_{(2)})|X_i < Y_{(2)}] = S_2 - 1 - 1 = S_2 - 2$$

- Hay $s_n - 1$ observaciones menores que $Y_{(n)}$, de ellas $n-1$ son Y y el resto son X.

$$\#[(X_i, Y_{(n)})|X_i < Y_{(n)}] = S_n - n$$

Por lo tanto:

$$\#[(X_i, Y_j)|X_i < Y_j] = (S_1 - 1) + (S_2 - 2) + \cdots + (S_n - n)$$

$$= (S_1 + \cdots + S_n) - (1 + 2 + \cdots + n) = W_S - \frac{n \cdot (n+1)}{2} = W_{XY}$$

Distribución asintótica de W_s Si los valores n y m son grandes (mayores que 10), se considera que la distribución asintótica para W_S bajo H_0 es, por el Teorema Central del Limite...

$$\frac{W_S - E_0(W_S)}{\sqrt{Var_0(W_S)}} \xrightarrow[H_0]{L} N(0,1)$$

$$E_0(W_S) = \frac{n \cdot (N+1)}{2} \quad Var_0(W_S) = \frac{n \cdot (N-n) \cdot (N+1)}{12}$$

4.2. Test de rangos con observaciones coincidentes

Hemos visto contrastes para 2 muestras independientes, donde se eligen n individuos para un tratamiento y m para un grupo de control (u otro tratamiento), siendo $N = n + m$.

$$H_0 : \text{Tratamiento no tiene efecto}$$
$$H_1 : \text{Tratamiento tiene efecto}$$

Mediamos la variable de interés en los N individuos.

$$X_1, \ldots, X_m \text{ para los individuos de control}$$
$$Y_1, \ldots, Y_n \text{ para los individuos del tratamiento}$$

Obteniamos los rangos de la muestra ($R_1, \ldots, R_m$ y $S_1, \ldots, S_n$). Si la alternativa es mayor, se rechaza H_0 para $W_S = S_1 + \cdots + S_n > C$. Si la alternativa es menor, se rechaza H_0 para $W_S = S_1 + \cdots + S_n < C$

¿Que pasa si hay coincidencias?

Ejemplo:

Muestra:

$$1.2,\ 1.7,\ 1.7,\ 1.7,\ 2,\ 3.1,\ 3.1,\ 5$$

Los rangos serían:

$$1,\ 2,\ 2,\ 2,\ 5,\ 6,\ 6,\ 8$$

Pero estos rangos no serían correctos. Para aquellos valores en los que coincida el rango, se les dan distintos y el rango de todos los que coincidan se calculara como la media de sus rangos. Por tanto, los rangos serían: 1, 3, 3, 3, 5, 6.5, 6.5, 8

A esto lo vamos a llamar **semi-rangos**. Siempre que tengamos coincidencias, calcularemos los semi-rangos

4.2.1. Semirangos

Definición: Los semi-rangos correspondientes a observaciones coinicidentes, se calculan como la media de los rangos que les corresponderían si no tuviéramos empates.

Notación: Cuando haya coincidencias, los semi-rangos se asignan y representan como:

$$R_1^*, \ldots, R_m^* \text{ semi-rango individual de control}$$
$$S_1^*, \ldots, S_n^* \text{ semi-rango individual de tratamiento}$$

Para controlar H_0, se rechaza si:

$$W_S^* = S_1^* + \cdots + S_n^* > C$$

Como siempre, necesitamos la distribución de W_S^* bajo H_0 que no es la misma que cuando no hay coincidencias, aunque llegaremos a la distribución de la misma forma.

Ejemplo:

$n = m = 3$

$X_1 = 5, \quad X_2 = 5, \quad X_3 = 9, \quad Y_1 = 5, \quad Y_2 = 10, \quad Y_3 = 10$

Los semi-rangos son 2,2,4 2,5.5,5.5 (Nota: si los empates estan en el mismo grupo no nos afectan en nada)

$$(S_1^*, S_2^*, S_3^*) = (2, 5{,}5, 5{,}5) \quad W_S^* = 13$$

Después de conocer los semi-rangos, calculamos la distribución de W_S^* bajo H_0 de la misma forma que anteriormente.

Bajo H_0, (hipótesis de que el tratamiento no tiene efecto), los 6 individuos recibirían los semi-rangos independientemente de que fueran asignados al grupo de tratamiento o de control. Por lo tanto, para la selección de los 3 individuos a tratamiento, hay $\binom{6}{3}$ =20 posibles ekecciones de 3 individuos a tratamiento y 3 a control. Pero no todas diferentes, porque hay repetidos.

S_1^*, S_2^*, S_3^*	$(2,2,2)$	$(2,2,4)$	$(2,2,5{,}5)$	$(2,4,5{,}5)$	$(2,5{,}5,5{,}5)$	$(4,5{,}5,5{,}5)$
W_S^*	6	8	9,5	11,5	13	15
P_{H_0}	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{6}{20}$	$\frac{3}{20}$	$\frac{1}{20}$

La distribución depende de la configuración de las coincidencias. No se tienen tablas ya que habría que considerar cada caso. Para n grande, se tiene al distribución asintótica bajo H_0.

Configuración de las coincidencias ***Definición:*** Configuración de las coincidencias.

- Sea $N = n + m$ el número de individuos tal que n sea el numero de individuos de control y m el numero de individuos del tratamiento
- Sea e el número de observaciones distintas entre los tratamientos
- Sea d_1, el número de observaciones iguales a la más pequeña
- Sea d_2, el número de observaciones iguales a la siguiente más pequeña
- Sea d_e el número de observaciones iguales a la más grande

Al vector $(e, d_1, \ldots, d_e)$ se le conoce como configuración de las coincidencias.
Ejemplo
$n = m = 3$
$X_1 = 5, \quad , X_2 = 5, \quad X_3 = 9, \quad Y_1 = 5, \quad Y_2 = 10, \quad Y_3 = 10$
Los semi-rangos son 2,2,4;2,5.5,5.5
En este caso:

$$(e, d_1, \ldots, d_e) = (3, 3, 1, 2)$$

- El semirango de las d_1:

$$\frac{1 + \cdots + d_1}{d_1} = \frac{d_1 + 1}{2}$$

- El semirango de las d_2:

$$\frac{(d_1 + 1) + \cdots + (d_1 + d_2)}{d_2} = d_1 + \frac{d_2 + 1}{2}$$

- El semirango i-ésimo

$$\frac{(d_{i-1} + 1) + \cdots + (d_{i-1} + d_i)}{d_i} = d_1 + \cdots + d_{i-1} + \frac{d_i + 1}{2}$$

$$d_1 = \frac{3 + 1}{2} = 2 \quad d_2 = 3 + \frac{1 + 1}{2} = 4 \quad d_3 = 3 + 1 + \frac{2 + 1}{2} = 5{,}5$$

Estos conteos se pueden hacer solo para n y m pequeños. En este caso, podemos relacionar W_S^* con el estadístico de Mann-Whitney análogo para el caso sin coincidencias.

4.2.2. Estadístico de Mann-Whitney con observaciones no distintas

El estadístico de W_S^* es una generalización de W_S cuando no todas las observaciones son distintas. Del mismo modo, se puede generalizar el estadístico de Mann-Whitney.

Sea $X_1, \ldots, X_m$ valores de la variable de interés para control y $Y_1, \ldots, Y_m$ valores de la variable de interés del tratamiento, si todas las observaciones son distintas, definiamos el

estadístico de Mann-Whitney como:

$$W_{XY} = \#[(X_i, Y_i)|X_i < Y_i]$$

(# = numero de casos en que:)

En caso de tener coincidencias, se puede definir para cada par (X_i, Y_j)

$$\phi(X_i, Y_j) = \begin{cases} 1 & si \quad X_i < Y_j \\ \frac{1}{2} & si \quad X_i = Y_j \\ 0 & si \quad X_i > Y_j \end{cases}$$

Si definimios $W^*_{XY} = \sum \phi(X_i, Y_j)$, es decir:

$$W^*_{XY} = \#[(X_i, Y_i)|X_i < Y_i] + \frac{1}{2} \cdot \#[(X_i, Y_i)|X_i = Y_i]$$

Resultado: Los tests basados en W^*_S y en W^*_{XY} son equivalentes y además

$$W^*_{XY} = W^*_S - \frac{n \cdot (n+1)}{2}$$

Demostración en el campus

Nota: se puede usar para categorías

Distribución asintótica de W^*_S Si n y m son grandes y la proporción máxima de observaciones coincidentes no es próxima a 1, es decir, si:

$$max_{i=1,\ldots,e} \left\{ \frac{d_i}{N} \right\} \ll 1$$

es decir, no hay un grupo en el que estén casi todas las observaciones.

$$\frac{W^*_S - E_\theta(W^*_S)}{\sqrt{Var_\theta(W^*_S)}} \xrightarrow{L} N(0,1)$$

$$E_\theta(W^*_S) = \frac{n \cdot (N+1)}{2}$$

$$Var_\theta(W^*_S) = \frac{n \cdot m \cdot (N-1)}{12} - n \cdot m \cdot \sum_{i=1}^{e} \frac{d_i \cdot (d_i^2 - 1)}{12 - N \cdot (N+1)}$$

Ejercicio 9 En un estudio sobre la efectividad de los consejos psicológicos, 80 jovenes se dividen aleatoriamente en un grupo control de 40 jóvenes, a quienes se aconseja de un modo tradicional, y un grupo de 40 que recibe un tratamiento especial. El cambio en el comportamiento de los jóvenes se califica como pobre, medianamente pobre, medianamente bueno y bueno. Obtenemos los siguientes resultados:

	Pobre	Medianamente pobre	Medianamente bueno	Bueno
Tratamiento	5	7	16	12
Control	7	9	15	9

Contrastar si el efecto del tratamiento es positivo.
Nos piden contrastar:

$$H_0\text{: no hay diferencias entre control y tratamiento}$$
$$H_1\text{: El tratamiento aumenta la respuesta}$$

Hay 4 grupos, por lo tanto $e = 4$.

- En el primer grupo hay 12 individuos

(e, d_1, d_2, d_3, d_4)=(4,12,16,31,21)

$$\text{Semi-rangos: } \begin{cases} d_1 = \frac{12+1}{2} = 6{,}5 \\ d_2 = 12 + \frac{16+1}{2} = 20{,}5 \\ d_3 = 12 + 16 + \frac{31+1}{2} = 44 \\ d_4 = 12 + 16 + 31 + \frac{21+1}{2} = 70 \end{cases}$$

$$W_S^* = \sum_{i=1}^{4} B_i(\text{semirangos}) = 5 \cdot (65) + 7 \cdot (205) + \cdots + 12 \cdot (70) = 1720$$

Vemos si el valor es grande con su distribución asintótica
El p-valor sería:

$$E_0(W_S^*) = \frac{n \cdot (N+1)}{2} = \frac{40 \cdot 81}{2} = 1620$$
$$Var(W_S^*) = 9854{,}937$$
$$P_{H_0}(W_S^* \geq 1720) = P\left(\frac{W_S^* - E(W_S^*)}{\sqrt{\text{Var}(W_S^*)}} \geq \frac{1720 - E(W_S^*)}{\sqrt{\text{Var}(W_S^*)}}\right)$$
$$= P\left(Z \geq \frac{1720 - 1620}{\sqrt{9854{,}937}}\right) = 1 - \Phi(1{,}01) = 0{,}16$$

4.3. Modelo poblacional

El precio que pagamos usando un modelo de aleatorización es que los resultados solo son válidos para los N individuos de estudio y no se pueden extrapolar a una población más amplia. Para que eso sea posible, será necesario que los N individuos representen a toda la población. Dicho de otra forma, necesitamos una **muestra aleatoria simple** de la población.

La situación es la siguiente: Tenemos $N = n + m$ individuos al azar de la población,

$$n \longrightarrow \text{elegidos al azar} \longrightarrow \text{grupo de tratamiento}$$
$$m \longrightarrow \text{restantes al grupo de control}$$

$$Y : \text{Variable respuesta de individuos que reciben el tratamiento}$$
$$X : \text{Variable respuesta de individuos que son del grupo de control}$$

X e Y son dos variables aleatorias con funciones de distribución $X \sim F$ y $Y \sim G$

Queremos contrastar la hipótesis de que el tratamiento NO es efectivo

H_0 : El tratamiento no tiene efecto $(F = G)$

H_1 : El tratamiento aumenta/disminuye la respuesta$(F > G/F < G)$

El modelo poblacional tiene dos ventajas fundamentales:

1. Los resultados son extrapolables
2. Podemos estudiar la potencia del test

Si tenemos un modelo poblacional sin coincidencias podemos utilizar el estadístico W_s y el test de Wilcoxon $(W_s > C_\alpha)$; bajo H_0, W_s sigue la misma distribución que en el modelo de aleatorización.

$$X_1, \ldots, X_m \quad Y_1, \ldots, Y_n$$
$$R_1, \ldots, R_m \quad S_1, \ldots, S_n$$
$$W_s = S_1 + \cdots + S_n$$

Si hay coincidencias, tenemos que encontrar la distribución de W_s^* bajo H_0.
En este caso el estadístico $W_s^* = S_1^* + \cdots + S_n^*$ no es de distribución libre. La distribución bajo H_0 de los semi-rangos de los n individuos depende de F. Esto se debe (al igual que en el modelo de aleatorización) a que la distribución depende de la configuración de las coincidencias $(e, d_1, \ldots, d_e)$, que en el modelo de aleatorización son un número pero aquí son variables aleatorias cuya distribución depende de F.

Ejemplo

Supongamos F discreta de tal forma que

$$F : \begin{cases} a & \text{Con probabilidad } p \\ b & \text{Con probabilidad } 1-p \end{cases}$$

Si $a < b$, y con $m = 2$ y $n = 1$, entonces los posibles resultados son:

$X_1X_2Y_1$	Probabilidad	Semi-rangos
$a\ a\ a$	p^3	2 2 2
$a\ a\ b$	$p^2(1-p)$	1,5 1,5 2
$a\ b\ a$	$p^2(1-p)$	1,5 2 1,5
$b\ a\ a$	$p^2(1-p)$	2 1,5 1,5
$a\ b\ b$	$p(1-p)^2$	1 2 2
$b\ b\ a$	$p(1-p)^2$	2 2 1
$b\ a\ b$	$p(1-p)^2$	1,5 1 1,5
$b\ b\ b$	$(1-p)^3$	1 1 1

La distribución de W_s^* bajo H_0 será:

S_n^*	1	1,5	2	2,5	3
$P_0(S_1^* = s_1^*)$	$p(1-p)^2$	$2p^2(1-p)$	$p^3 + (1-p)^3$	$2p(1-p)^2$	$p^2(1-p)$

Evidentemente la distribución de W_s^* depende de p; es decir, de F.

Al igual que en el modelo de aleatorización, la distribución de W_s^* depende de la configuración de las coincidencias, solo que esta vez esas coincidencias son v.a. que dependen de F.

4.3.1. Potencia del test

Una ventaja del modelo poblacional es que podemos calcular la potencia del test. Para ello debemos especificar la hipótesis alternativa. Sean F y G las distribuciones de las variables respuesta en individuos de control y tratamiento respectivamente,

$$H_0: \ F = G$$
$$H_1: \ \text{El tratamiento aumenta la respuesta, } F > G$$

¿Qué significa en términos de F y G que el tratamiento aumente la respuesta?

$$\forall z \in \mathbb{R} \qquad P(Y > z) \geq P(X > z) \iff 1 - G(z) \geq 1 - F(z) \iff F(z) \geq G(z)$$

Teorema 4.2. Sean X e Y v.a. tales que $X \sim F$ y $Y \sim G$ con F y G funciones de distribución. Se dice que Y (respecto a X) es estocásticamente mayor que X (respecto a Y) cuando los valores que toma la v.a. Y son mayores que los que toma la v.a. X, es decir:

$$G(z) \leq F(z) \qquad \forall z \in \mathbb{R}$$
$$H_0: \ F(x) = G(x)$$
$$H_1: \ F(x) \geq G(x)$$

El cálculo de la potencia requiere la distribución de los rangos. En el caso de F y G continuas es muy complicado, por lo que aproximaremos con la distribución asintótica y usaremos el estadístico de Mann-Whitney.

Potencia asintótica

$$\Pi(F, G): \ X \sim F, \ Y \sim G \text{ si n y m son suficientemente grandes}$$
$$\Pi(F, G) = \underset{\text{Bajo } H_1}{P_{F,G}} (W_{XY} \geq C_\alpha) = P_{F,G}\left(\frac{W_{XY} - E_{FG}(W_{XY})}{\sqrt{Var_{FG}(W_{XY})}} \geq \frac{C_\alpha - E_{FG}(W_{XY})}{\sqrt{Var_{FG}(W_{XY})}}\right) =$$
$$= 1 - \Phi\left(\frac{C_\alpha - E_{FG}(W_{XY})}{\sqrt{Var_{FG}(W_{XY})}}\right)$$

$$E(W_{XY}) = mnp_1$$
$$Var(W_{XY}) = mnp_1(1 - p_1) + mn(n-1)(p_2 - p_1^2) + mn(m-1)(p_3 - p_1^2)$$

Siendo:

$$p_1 = P_{FG}(X < Y)$$
$$p_2 = P(X < Y, X < Y')$$
$$p_3 = P(X < Y, X' < Y)$$

El problema viene porque p_1, p_2 y p_3 son difíciles de calcular, por lo que usaremos una aproximación de la potencia.

4.3.2. Modelo Shift de aproximación de la potencia

Teorema 4.3. F y G se agrupan en un modelo Shift si

$$\exists \Delta > 0,\ \forall x \qquad G(x) = F(x - \Delta)$$

El modelo Shift queda

$$H_0 : F(x) = G(x) \iff \Delta = 0 \qquad H_1 : G(x) = F(x - \Delta) \iff \Delta > 0$$

La potencia se escribe como:

$$\Pi_F(\Delta) = P_\Delta(W_{XY} > C_\alpha),\ \Delta > 0$$

En particular, $\Pi_F(0) = \alpha$

Teorema 4.4. Sea F^* la función de distribución de la diferencia de las dos v.a. independientes con distribución F y sea $f^*(0)$ su densidad en el 0. Entonces

$$\Pi(\Delta) \approx \Phi\left[\sqrt{\frac{12mn}{N+1}} f^*(0)\Delta - \mu_\alpha\right]$$

Donde μ_α / $\Phi(\mu_\alpha) = 1 - \alpha$

Supongamos que N es suficientemente grande como para poder usar la aproximación normal para encontrar C_α

$$\alpha = P_0(W_{XY} > C_\alpha) = P_0\left(\frac{W_{XY} - E_0(W_{XY})}{\sqrt{Var_0(W_{XY})}} \geq \frac{C_\alpha - E_0(W_{XY})}{\sqrt{Var_0(W_{XY})}}\right)$$

$$W_{XY} = W_s - \tfrac{n(n+1)}{2}$$
$$E(W_s) = \tfrac{n(N+1)}{2}$$
$$E(W_{XY}) = \tfrac{n(N+1)}{2} - \tfrac{n(n+1)}{2} = \cdots = \tfrac{nm}{2}$$
$$Var(W_{XY}) = \tfrac{1}{12}mn(N+1)$$

$$\alpha = P_0\left(\frac{W_{XY} - \frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(N+1)}} \geq \frac{C_\alpha - \frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(N+1)}}\right)$$

Por lo que

$$\mu_\alpha = \frac{C_\alpha - \frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(N+1)}} \Longrightarrow C_\alpha = \frac{1}{2}mn + \sqrt{\frac{1}{12}mn(N+1)}\mu_\alpha$$

Calculando la potencia:

$$\Pi_F(\Delta) = P_\Delta\left(W_{XY} \geq \frac{1}{2}mn + \sqrt{\frac{1}{12}mn(N+1)}\mu_\alpha\right) =$$

$$= P_\Delta\left(\frac{W_{XY} - E_\Delta(W_{XY})}{\sqrt{(Var_\Delta(W_{XY}))}} \geq \frac{\frac{1}{2}mn + \sqrt{\frac{1}{12}mn(N+1)}\mu_\alpha - mnp_1}{\sqrt{Var_\Delta(W_{XY})}}\right) =$$

$$= 1 - \Phi\left(\frac{\left(\frac{1}{2} - p_1\right)mn + \mu_\alpha\sqrt{\frac{1}{12}mn(N+1)}}{\sqrt{Var_\Delta(W_{XY})}}\right)$$

Sustituyendo $p_1 = P_{FG}(X < Y)$

$$p_1 = P_\Delta(X < Y) = P_0(X < Y - \Delta) = P_0(Y - X > \Delta) = P_0(\underbrace{Y - X}_{\text{Misma dist. bajo } H_0} - \Delta > 0) = 1 - F^*(\Delta)$$

$F^*(\Delta)$ será la función de distribución de la diferencia de las dos v.a. independientes con distribución F. Si desarrollamos $F^*(\Delta)$ en torno al 0 con el polinomio de Taylor, sabiendo que $(F^*(x))' = f^*(x)$ y por simetría respecto al 0:

$$F^*(\Delta) \approx F^*(0) + (\Delta - 0)f^*(0) = \frac{1}{2} + \Delta f^*(0)$$

Supongamos $D = X - Y$, $X \sim F$, $Y \sim F$ y $D \sim F^*$, entonces:

$$F^*(0) = P(D \leq 0) = P(X - Y \leq 0) = P(X \leq Y) = \frac{1}{2}$$

Por lo tanto,

$$p_1 = 1 - F^*(0) \approx \frac{1}{2} + \Delta f^*(0) \Longrightarrow p_1 - \frac{1}{2} \approx \Delta f^*(0)$$

Entonces, ya podemos hacer una primera aproximación para el cálculo de la potencia:

$$\Pi_F(\Delta) \approx \Phi\left(\frac{mn\Delta f^*(0) - \mu_\alpha\sqrt{\frac{1}{12}mn(N+1)}}{\sqrt{Var_\Delta(W_{XY})}}\right)$$

Nos faltaría calcular $Var_\Delta(W_{XY})$. Podemos hallar una aproximación cuando Δ es pequeño, ya que

$$Var_\Delta(W_{XY}) \approx Var_0(W_{XY}) = \frac{mn(N+1)}{12}$$

Por lo que la expresión para la potencia quedaría como

$$\Pi_F(\Delta) \approx \Phi\left(\frac{mn\Delta f^*(0)}{\sqrt{\frac{mn(N+1)}{12}}} - \mu_\alpha\right) = \Phi\left(\sqrt{\frac{12mn}{N+1}}\Delta f^*(0) - \mu_\alpha\right)$$

$$= \Phi\left(\sqrt{\frac{12\,m}{N+1}}\,\Phi^*(0)\Delta - \mu_\alpha\right) = \Pi$$

Inverso

$$\sqrt{\frac{12\,m}{N+1}}\,\Phi^*(0)\Delta - \mu_\alpha = \Phi^{-1}(\Pi) \Longrightarrow \frac{12\,m}{N+1} = \frac{\left(\Phi^{-1}(\Pi) + \mu_\alpha\right)^2}{\left(\Phi^*(0)\Delta\right)^2}$$

Aproximamos asumiendo $m \simeq n$ y asumimos también N suficientemente grande para que $N \simeq N+1$:

$$n \simeq \frac{\left(\Phi^*(0)\Delta + \mu_\alpha\right)^2}{6\Delta^2\Phi^*(0)^2}$$

Intervalos de confianza para pares Calculamos diferencias de nuestros dos:

$$D_{ij} = Y_i - X_j \quad \text{para todas las pares } i = 1, \ldots, m \quad j = 1, \ldots, n$$

Tomamos como estimador $\hat{\Delta} = \text{mediana}(D_{ij})$ ya que la mediana es robusta y menos sesgada.

Test de signos para muestras pareadas Antes del tratamiento:

$$X = \{2, 4, 5, 6, 8\}$$

Después del tratamiento:

$$Y = \{3, 5, 7, 4, 10\}$$

Se calculan las diferencias:

$$D = \{3-2, 5-4, \ldots\} = \{1, 1, 2, -2, 2\}$$

Si no hubiera diferencias, se deberían distribuir las diferencias positivas y negativas (y viceversa). El estadístico:

$$S \sim b(n, 0{,}5)$$

En un test bilateral, el p-valor:

$$p = 2 \cdot P(5 \leq \text{mín}(S^+, S^-))$$

Parcial 2024-2025

Resolución e indicaciones

Resolución del examen parcial y recomendaciones para su preparación.

Ejercicio 1:

Un artículo, producto de un tipo de componente mecánico que puede deteriorarse a dos velocidades distintas, depende de factores de fabricación no controlados. Se observa que el tiempo de vida (X) de cada componente sigue una mezcla de distribuciones exponenciales con las siguientes características:

- Con probabilidad p, el componente tiene un tiempo de vida X que sigue una distribución exponencial con parámetro λ_1, lo cual corresponde a componentes con alta resistencia.
- Con probabilidad $1-p$, el componente tiene un tiempo de vida X que sigue una distribución exponencial con parámetro λ_2, lo cual corresponde a componentes con menor resistencia.

El fichero `Tiempos.RData` corresponde a un conjunto de observaciones independientes $x_1, x_2, \ldots$, que representan los tiempos de vida medidos en horas de varios componentes.

1. Obtener el EMV y concretar su distribución asintótica para los parámetros del modelo que subyace a partir de estos datos. **Nota:** Planteadas las ecuaciones de verosimilitud, utilice la función `optim` para el cálculo del EMV.

$$f(x, p, \lambda_1, \lambda_2) = p \cdot e^{-\lambda_1 x} + (1-p) \cdot e^{-\lambda_2 x}$$

$$L(x, p, \lambda_1, \lambda_2) = \prod_{i=1}^{n} f(x_i, p, \lambda_1, \lambda_2) = \prod_{i=1}^{n} \left(p \cdot e^{-\lambda_1 x_i} + (1-p) \cdot e^{-\lambda_2 x_i}\right)$$

$$\log L(x, p, \lambda_1, \lambda_2) = \sum_{i=1}^{n} \log f(x_i, p, \lambda_1, \lambda_2)$$

$$\begin{pmatrix} \widehat{p} \\ \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} p \\ \lambda_1 \\ \lambda_2 \end{pmatrix}, \mathrm{Var(EMV)}_{3\times 3} \right)$$

2. Obtener el IC de Wald con confianza aproximada de 95 % para cada parámetro del modelo.

$$\widehat{p} \pm \texttt{qnorm}(0{,}975) \cdot \sqrt{\text{Var}(\widehat{p})}$$

3. Obtener el *p-valor* basado en el estadístico de Wald para contrastar la hipótesis $H_0 : \lambda_1 = 0{,}5$ vs. $H_a : \lambda_1 \neq 0{,}5$.

$$W = \frac{(\widehat{\lambda}_1 - 0{,}5)^2}{\text{Var}(\widehat{\lambda}_1)} \sim \chi^2_1$$

4. Obtener el *p-valor* basado en el estadístico de RV para contrastar la hipótesis $H_0 : \lambda_1 = 0{,}5$ vs. $H_a : \lambda_1 \neq 0{,}5$.

$$Q_L = 2 \cdot \left[\log L(\widehat{p}, \widehat{\lambda}_1, \widehat{\lambda}_2, x) - \sup_{p, \lambda_2} \log L(p, \lambda_1 = 0{,}5, \lambda_2, x) \right] \sim \chi^2_1$$

Referencias

[1] Juan Camilo Yepes Borrero,
Apuntes Manuscritos Tema 1.
Universidad de Valladolid 2024.

[2] Yolanda Larriba González,
Apuntes INFE2 Tema 1.
Universidad de Valladolid 2023.

[3] Juan Camilo Yepes Borrero,
Apuntes Manuscritos Tema 2.
Universidad de Valladolid 2024.

[4] Yolanda Larriba González,
Apuntes INFE2 Tema 2.
Universidad de Valladolid 2023.

[5] Juan Camilo Yepes Borrero,
Apuntes Manuscritos Tema 4.
Universidad de Valladolid 2024.

[6] Yolanda Larriba González,
Apuntes INFE2 Tema 4.
Universidad de Valladolid 2023.

[7] Juan Camilo Yepes Borrero,
Apuntes Manuscritos Tema 5.
Universidad de Valladolid 2024.

[8] Yolanda Larriba González,
Apuntes INFE2 Tema 5.
Universidad de Valladolid 2023.

[9] Juan Camilo Yepes Borrero,
Resolución examen en el aula.
Universidad de Valladolid 2024.