Notas Curso de Estadística (Parte I)

Maikol Solís

Actualizado el 16 septiembre, 2020

Índice general

1. Introducción **7**

2. Inferencia estadística **9**
2.1. Ejemplo 9
2.2. Modelo estadístico 10
2.3. Estadístico 12

3. Densidades previas conjugadas y estimadores de Bayes **15**
3.1. Distribución previa (distribución a priori) 15
3.2. Densidad posterior 17
3.3. Proceso de modelación de parámetros. 19
3.4. Función de verosimilitud 19
3.5. Familias conjugadas 21
3.6. Densidades previas impropias 24
3.7. Funciones de pérdida 25
3.7.1. Función de pérdida cuadrática 26
3.7.2. Función de pérdida absoluta 26
3.7.3. Otras funciones de pérdida 27
3.8. Efecto de muestras grandes 27
3.9. Consistencia 28
3.10. Laboratorio 29
3.10.1. Distribución previa 29
3.10.2. Distribución conjunta 30
3.10.3. Distribución posterior 31
3.10.4. Agregando nuevos datos 32
3.10.5. Familias conjugadas normales 33
3.10.6. Funciones de pérdida 38

3.10.7. Caso concreto 39

4. Estimación por máxima verosimilitud 45
4.1. Propiedades del MLE 52
4.1.1. Propiedad de invarianza 52
4.1.2. Consistencia 53
4.2. Cálculo numérico 54
4.2.1. Método de los momentos 54
4.2.2. Método Delta 56
4.3. Laboratorio 58

5. Estadísticos Suficientes y Criterio de Factorización 61
5.1. Estadísticos suficientes 61
5.2. Teorema de Factorización de Fisher 61
5.3. Estadístico suficiente multivariado. 64
5.4. Estadísticos minimales 66
5.5. Mejorando estimadores 67

6. Distribución muestral de un estadístico 71
6.1. Distribución muestral 71
6.2. Distribución χ^2 74
6.3. Distribución t 78

7. Intervalos de confianza 81
7.1. Intervalos de confianza para la media de una distribución normal 81
7.2. Caso normal. 81
7.3. Intervalos de confianza abiertos 86
7.4. Intervalos de confianza en otros casos 87
7.4.1. Intervalos de confianza aproximados. 90
7.4.2. Transformaciones estabilizadoras de la varianza 92

8. Estimación Bayesiana bajo normalidad 95
8.1. Precisión de una distribución normal 95
8.2. Distribución marginal de μ 98
8.3. Efecto de previas no informativas 101

9. Estimación insesgada 103
9.1. Estimadores insesgados 103
9.2. Estimador insesgado de la varianza 105

9.3. Información de Fisher 106
9.4. Desigualdad de Cramer-Rao 111
9.5. Estimadores eficientes 113
9.6. Comportamiento asintótico del MLE 114

10. Pruebas de hipótesis **117**
10.1. Pruebas de hipótesis 117
10.2. Regiones críticas y estadísticas de prueba 118
10.3. Función de potencia y tipos de error 119
10.4. Valor p 123
10.5. Dualidad entre pruebas de hipótesis y regiones de confianza 124
10.5.1. Dualidad en pruebas unilaterales 126
10.5.2. Pruebas de cociente de verosimilitud (LRT) 128

11. Pruebas con hipótesis simples **131**
11.1. Hipótesis simples 131
11.2. Prueba t 137
11.2.1. Propiedades de las pruebas t 137
11.2.2. Prueba t pareada 140
11.2.3. Pruebas t de dos colas 141

12. Prueba de comparación de medias en 2 poblaciones **143**
12.1. Comparación de medias normales 143
12.2. Prueba t de dos muestras 144
12.2.1. Prueba de 2 colas 146
12.3. Prueba F 146
12.3.1. Prueba de 2 colas (prueba de homocedasticidad) 148

13. Pruebas de hipótesis bayesianas **151**
13.1. Pruebas de hipótesis bayesianas 151
13.2. Hipótesis de una cola 153
13.3. Hipótesis de 2 colas 156

Capítulo 1

Introducción

Capítulo 2

Inferencia estadística

Definición: Hacer afirmaciones probabilísticas respecto a (acerca de) cantidades desconocidas.

2.1. Ejemplo

*__Pregunta__: ¿Será posible modelar cuánto dura un componente electrónico en fallar?

Solución: Podemos responder esta pregunta dividiéndola en dos partes:

1. **Modelo probabilístico:** Asuma que los tiempos de vida del componente son exponenciales (en años).
2. **Parámetro:** Sea $\theta > 0$ la tasa de fallo (unidades: 1/Tiempo(años)).

Es decir, tenemos un modelo (exponencial) y estamos decretando que su información estará concentrada en el parámetro θ.

Nota: El parámetro θ contiene la información del modelo, pero ¿Cómo obtenemos esa información

Muestra: Secuencia (sucesión) de variables aleatorias independientes $X_1, X_2, \ldots, X_n, \ldots$. Tomemos una muestra $X_1, X_2, \ldots, X_n, \ldots \overset{i.i.d}{\sim} \text{Exp}(\theta)$.

Objetivos

- Estimar $X_m, X_{m+1}, \ldots$ si se observa $X_1, X_{m-1}, \ldots$ (Predicción).

- Estimar θ usando información.

Datos: Realizaciones de variables aleatorias $X_1, \ldots, X_m$ pertenecientes a la muestra.

Estimación de θ

Dado que $\mathbb{E}(X) = \dfrac{1}{\theta}$ con $X \sim \text{Exp}(\theta)$, por la ley de grandes números se tiene que

$$\underbrace{\frac{1}{n}\sum_{i=1} X_i}_{\bar{X}_n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}(X) = \frac{1}{\theta}$$

por propiedad de convergencia en probabilidad.

Un posible candidato para estimar θ es $\dfrac{1}{\bar{X}_n}$, bajo el supuesto por Ley de Grandes Números que θ es una constante (frecuentista).

Realidad: θ no necesariamente es determinístico (factores externos, por la naturaleza del fenómeno).

Asumimos un modelo probabilístico para θ (tasa siempre positiva):

$$\theta \sim \Gamma(\alpha_0, \beta_0)$$

Luego, según estudios previos la tasa esperada es 0.5/año

$$\mathbb{E}(\theta) = \frac{1}{2} = \frac{\alpha_0}{\beta_0}.$$

Un primer indicio de que se podría establecer que $\alpha_0 = 1$ y de $\beta_0 = 2$.

2.2. Modelo estadístico

Vamos a definir como típicamente se define un modelo estadístico.

1. Variables aleatorias observables / hipotéticamente observables:

$$\underbrace{X_t}_{\text{Observable}} = \underbrace{Y_t}_{\text{Hip. observable}} + \underbrace{\epsilon}_{\text{Ruido}}$$

En otras palabras Y_t sería la el dato *"verdadero"* que pasó exactamente en el fenómeno analizado. Esta observación es afectada por muchos factores no observables (por ejemplo: errores de medición, cambio de las condiciones de la economía, etc.). La variable ϵ captura toda esa aleatoriedad que no es parte del fénomeno.

Claramente ni Y_t ni ϵ se pueden medir y la mejor representación del nuestro es fenómeno es a partir de X_t.

2. Distribución conjunta de una muestra de variables observables.

   Es decir cuál es el supuesto general que estoy usando para describir mis observaciones.

3. Parámetros que son hipotéticamente observables (desconocidos).

   ¿Cuál sería la mejor calibración de los componentes del modelo anterior de modo que mi modelo se ajuste a los datos?

4. (Opcional) Distribución conjunta de los parámetros.

   En el caso de Bayes, los parámetro dejan de ser simple valores puntuales y se convierten en distribuciones completas.

- **Inferencia estadística**: procedimiento que genera afirmaciones probabilísticas de un modelo estadístico.

Ejemplo de inferencias:

1. Estimar θ a través de $\dfrac{1}{\bar{X}_n}$.

2. ¿Qué tan probable es que el promedio de las siguientes observaciones es al menos 2?

$$\frac{1}{10}\sum_{i=m+1}^{m+10} X_i > 2$$

3. ¿Qué tan cierto es que $\theta \leq 0{,}4$ después de observar la muestra?

- **Parámetro**: característica (s) que determinan la distribución conjunta de las variables aleatorias de interés.
- **Espacio paramétrico** Ω (espacio de parámetros, puede ser de probabilidad)

Ejemplos:

- $\theta > 0$ (ejemplo anterior); $\Omega = (0, +\infty)$.
- $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, (μ, σ^2) parámetros; $\Omega = \mathbb{R} \times [0, +\infty)$.

Ejemplo: Clientes de un banco

¿Qué tan probable es que un cliente no pague su crédito hoy?

- **Datos**: $X_i = \begin{cases} 1 & \text{el cliente } \#i \text{ no pagó} \\ 0 & \text{el cliente } \#i \text{ pagó} \end{cases}$.
- **Muestra**: $X_1, \ldots, X_{10000}$ (realización al día de hoy).
- **Modelos**: $X_1, \ldots, X_{10000} \overset{i.i.d}{\sim} \text{Ber}(p)$ con $p \in [0, 1]$.
- **Parámetro**: p, $\Omega = [0, 1]$.
- **Inferencias:**
  - Estimar p (probabilidad de impago).
  - Suponga que $L(X_i)$ es el saldo en la cuenta del cliente $\#i$.

$$\mathbb{P}\left(\sum_{i=1}^{10000} L(X_i) > u\right) = \text{Probabilidad de ruina}$$

2.3. Estadístico

Definición. Si $X_1, \ldots, X_n$ es una muestra observable. Sea r una función real de n variables:

$$T = r(X_1, \ldots, X_n)$$

es un estadístico.

Nota: T también es aleatorio.

Ejemplos:

- $\hat{p} = \frac{1}{10000} \sum_{i=1}^{10000} X_i = \frac{\# \text{ no pagan}}{\text{Total}} = r(X_1, \ldots, X_{10000})$
- $L_m = \max L(X_i)$ (saldo del cliente más riesgoso).
- $R_m = \max L(X_i) - \min L(X_i), 1 \leq i \leq 10000$

Capítulo 3

Densidades previas conjugadas y estimadores de Bayes

3.1. Distribución previa (distribución a priori)

Suponga que tenemos un modelo estadístico con parámetro θ. Su θ es aleatorio entonces su densidad (antes de observar cualquier muestra) se llama **densidad previa**: π.

Ejemplo: $X_1, \ldots, X_n \sim \text{Exp}(\theta)$ y θ es aleatorio tal que $\theta \sim \Gamma(\overset{\alpha}{1}, \overset{\beta}{2})$ entonces

$$\pi(\theta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \theta^{\alpha-1} e^{\beta\theta} = 2e^{-2\theta}, \quad \theta > 0$$

Ejemplo: Sea θ la probabilidad de obtener cara al tirar una moneda.

En este caso antes de modelar exactamente el θ, lo importante es modelar el tipo de moneda. Es decir, supongamos que tenemos dos opciones

- *Moneda justa:* $\theta = \dfrac{1}{2}$ con probabilidad previa 0,8 ($\pi(\frac{1}{2}) = 0{,}8$).
- *Moneda con solo una cara:* $\theta = 1$ con probabilidad previa 0,2 ($\pi(1) = 0{,}2$).

En este ejemplo si tuviéramos 100 monedas con probabilidad previa π entonces 20 tendrían solo una cara y 80 serían monedas normales.

Notas:

- π está definida en Ω (espacio paramétrico).
- π es definida antes de obtener la muestra.

Ejemplo (Componentes eléctricos) Supoga que se quiere conocer el tiempo de vida de cierto componente eléctrico. Sabemos que este tiempo se puede modelar con una distribución exponencial con parámetro θ desconocido. Este parámetro asumimos que tiene una distribución previa Gamma.

Un experto en componentes eléctricos conoce mucho de su área y sabe que el parámetro θ tiene las siguientes características:

$$\mathbb{E}[\theta] = 0{,}0002, \quad \sqrt{\mathrm{Var}(\theta)} = 0{,}0001.$$

Como sabemos que la previa π es Gamma, podemos deducir lo siguiente:

$$\mathbb{E}[\theta] = \frac{\alpha}{\beta}, \mathrm{Var}(\theta) = \frac{\alpha}{\beta^2}$$

$$\implies \begin{cases} \dfrac{\alpha}{\beta} = 2 \times 10^{-4} \\ \sqrt{\dfrac{\alpha}{\beta^2}} = 1 \times 10^{-4} \end{cases} \implies \beta = 20000, \alpha = 4$$

Notación:

- $X = (X_1, \ldots, X_n)$: vector que contiene la muestra aleatoria.
- Densidad conjunta de X: $f_\theta(x)$.
- Densidad de X condicional en θ: $f_n(x|\theta)$.

Supuesto: X viene de una muestra aleatoria si y solo si X es condicionalmente independiente dado θ.

Consecuencia:

$$f_n(X|\theta) = f(X_1|\theta) \cdot f(X_2|\theta) \cdots f(X_n|\theta)$$

Ejemplo

Si $X = (X_1, \dots, X_n)$ es una muestra tal que $X_i \sim \text{Exp}(\theta)$,

$$f_n(X|\theta) = \begin{cases} \prod_{i=1}^n \theta e^{-\theta X_i} & \text{si } X_i > 0 \\ 0 & \text{si no} \end{cases}$$
$$= \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n X_i} & X_i > 0 \\ 0 & \text{si no} \end{cases}$$

3.2. Densidad posterior

Definición. Considere un modelo estadístico con parámetro θ y muestra aleatoria $X_1, \dots, X_n$. La densidad condicional de θ dado $X_1, \dots, X_n$ se llama *densidad posterior*: $\pi(\theta|X)$

Teorema. Bajo las condiciones anteriores:

$$\pi(\theta|X) = \frac{f(X_1|\theta)\cdots f(X_n|\theta)\pi(\theta)}{g_n(X)}$$

para $\theta \in \Omega$, donde g_n es una constante de normalización.

Prueba:

$$\pi(\theta|X) = \frac{\pi(\theta, X)}{\text{marginal de X}} = \frac{\pi(\theta, X)}{\int \pi(\theta, X)\, d\theta} = \frac{P(X|\theta)\cdot\pi(\theta)}{\int \pi(\theta, X)\, d\theta}$$
$$\frac{f_n(X|\theta)\cdot\pi(\theta)}{g_n(X)} = \frac{f(X_1|\theta)\cdots f(X_n|\theta)\pi(\theta)}{g_n(X)}$$

Del ejemplo anterior,

$$f_n(X|\theta) = \theta^n e^{-\theta y}, y = \sum X_i \text{ (estadístico)}$$

Numerador:

$$f_n(X|\theta)\pi(\theta) = \underbrace{\theta^n e^{-\theta y}}_{f_n(X|\theta)} \cdot \underbrace{\frac{200000^4}{3!}\theta^3 e^{-20000\cdot\theta}}_{\pi(\theta)} = \frac{20000^4}{3!}\theta^{n+3} e^{(20000+y)\theta}$$

Denominador:

$$g_n(x) = \int_0^{+\infty} \theta^{n+3} e^{-(20000+y)\theta}\, d\theta = \frac{\Gamma(n+4)}{(20000+y)^{n+4}}$$

Entonces la posterior corresponde a

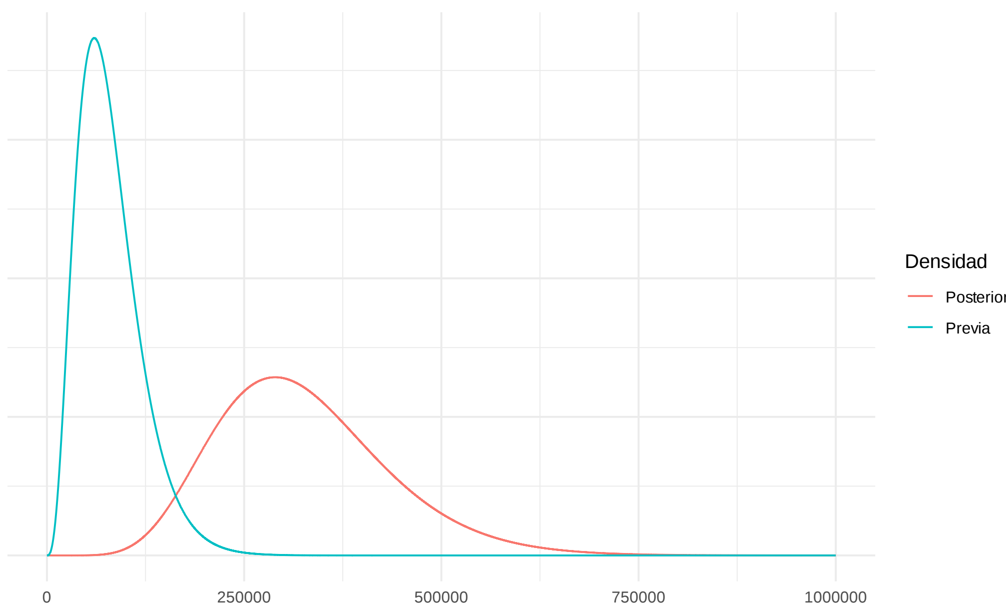
$$\pi(\theta|X) = \frac{\theta^{n+3} e^{-(20000+y)\theta}}{\Gamma(n+4)} (20000+y)^{n+4}$$

que es una $\Gamma(n+4, 20000+y)$.

Con 5 observaciones (horas): 2911, 3403, 3237, 3509, 3118.

$$y = \sum_{i=1}^{5} X_i = 16478, \quad n = 5$$

por lo que $\theta|X \sim \Gamma(9, 36178)$



Es sensible al tamaño de la muestra (una muestra grande implica un efecto de la previa menor).

Hiperparámetros: parámetros de la previa o posterior.

3.3. Proceso de modelación de parámetros.

De ahora en adelante vamos a entender un modelo como el conjunto de los datos $X_1, \ldots, X_n$, la función de densidad f y el parámetro de la densidad θ. Estos dos últimos resumen el comportamiento de los datos.

Ahora para identificar este modelo se hace por partes,

1. La información previa $\pi(\theta)$ es la información extra o basado en la experiencia que tengo del mdoelo.
2. Los datos es la información observada. La función de densidad f filtra y mejora la información de la previa.
3. La densidad posterior es la "mezcla" entre la información y los datos observados. Es una versión más informada de la distribución del parámetro.

3.4. Función de verosimilitud

Bajo el modelo estadístico anterior a $f_n(X|\theta)$ se le llama **verosimilitud** o **función de verosimilitud**.

Observación. En el caso de una función de verosimilitud, el argumento es θ.

Ejemplo.

Sea θ la proporción de aparatos defectuosos, con $\theta \in [0, 1]$

$$X_i = \begin{cases} 0 & \text{falló} \\ 1 & \text{no falló} \end{cases}$$

$\{X_i\}_{i=1}^n$ es una muestra aleatoria y $X_i \sim Ber(\theta)$.

- **Verosimilitud**

$$f_n(X|\theta) = \prod_{i=1}^n f(X_i|\theta) = \begin{cases} \theta^{\sum X_i}(1-\theta)^{n-\sum X_i} & X_i = 0, 1\ \forall i \\ 0 & \text{si no} \end{cases}$$

- **Previa**:

$$\pi(\theta) = 1_{\{0 \leq \theta \leq 1\}}$$

- **Posterior**:

Por el teorema de Bayes,

$\pi(\theta|X) \propto \theta^y(1-\theta)^{n-y} \cdot 1$

$$= \theta^{\overbrace{y+1}^{\alpha}-1}(1-\theta)^{\overbrace{n-y+1}^{\beta}-1} \quad \Longrightarrow \theta|X \sim \text{Beta}(y+1, n-y+1)$$

- **Predicción**.

Supuesto: los datos son secuenciales. Calculamos la distribución posterior secuencialmente:

$$\begin{aligned}
\pi(\theta|X_1) &\propto \pi(\theta)f(X_1|\theta) \\
\pi(\theta|X_1, X_2) &\propto \pi(\theta)f(X_1, X_2|\theta) \\
&= \pi(\theta)f(X_1|\theta)f(X_2|\theta) \text{ (por independencia condicional)} \\
&= \pi(\theta|X_1)f(X_2|\theta) \\
&\vdots \\
\pi(\theta|X_1, \ldots, X_n) &\propto f(X_n|\theta)\pi(\theta|X_1, \ldots, X_{n-1})
\end{aligned}$$

Bajo independencia condicional no hay diferencia en la posterior si los datos son secuenciales.

Luego,

$$\begin{aligned}
g_n(X) &= \int_\Omega f(X_n|\theta)\pi(\theta|X_1, \ldots, X_{n-1})\, d\theta \\
&= P(X_n|X_1, \ldots, X_{n-1}) \text{ (Predicción para } X_n\text{)}
\end{aligned}$$

Continuando con el ejemplo de los artefactos, $P(X_6 > 3000|X_1, X_2, X_3, X_4, X_5)$. Se necesita calcular $f(X_6|X)$. Dado que

$$\pi(\theta|X) = 2{,}6 \times 10^{36}\theta^8 e^{-361788\theta}$$

se tiene

$$f(X_6|X) = 2{,}6 \times 10^{36} \int_0^1 \underbrace{\theta e^{-\theta X_6}}_{\text{Densidad de } X_6} \theta^8 e^{-361178\theta}\, d\theta = \frac{9{,}55 \times 10^{41}}{(X_6 + 36178)^{10}}$$

Entonces,

$$P(X_6 > 3000) = \int_{3000}^{\infty} \frac{9{,}55 \times 10^{41}}{(X_6 + 36178)^{10}}\, dX_6 = 0{,}4882$$

La vida media se calcula como $\frac{1}{2} = P(X_6 > u|X)$.

3.5. Familias conjugadas

Definición. Sea $X_1, \dots, X_n$ i.i.d. condicional dado θ con densidad $f(X|\theta)$. Sea ψ la familia de posibles densidades previas sobre Ω. Si, sin importar los datos, la posterior pertenece a ψ, entonces decimos que ψ es una familia conjugada de previas.

Ejemplos:

- La familia Beta es familia conjugada para muestras según una Bernoulli.
- La familia Gama es familia conjugada para muestras exponenciales.
- Para el caso Poisson, si $X_1, \dots, X_n \sim Poi(\lambda)$,entonces la familia Gamma es familia conjugada.

La función de densidad de una Poisson es $P(X_i = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. La verosimilitud corresponde a

$$f_n(X|\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda_i^X}{X_i!} = \frac{e^{-n\lambda^y}}{\prod_{i=1}^n X_i}.$$

La previa de λ está definida por $\pi(\lambda) \propto \lambda^{\alpha-1}e^{-\beta\lambda}$. Por lo tanto, la posterior es

$$\pi(\lambda|X) \propto \lambda^{y+\alpha-1}e^{-(\beta+n)\lambda} \implies \lambda|X \sim \Gamma(y+\alpha, \beta+n)$$

- En el caso normal, si $X_1, \dots, X_n \sim N(\theta, \sigma^2)$,entonces la familia normal es conjugada si σ^2 es conocido.

Si $\theta \sim N(\mu_0, V_0^2) \implies \theta|X \sim N(\mu_1, V_1^2)$ donde,

$$\mu_1 = \frac{\sigma^2\mu_0 + nV_0^2\bar{X}_n}{\sigma^2 + nV_0^2} = \frac{\sigma^2}{\sigma^2 + nV_0^2}\mu_0 + \frac{nV_0^2}{\sigma^2 + nV_0^2}\bar{X}_n$$

Combina de manera ponderada la previa y la de los datos.

Ejemplo

Considere una verosimilitud Poisson(λ) y una previa

$$\pi(\lambda) = \begin{cases} 2e^{-2\lambda} & \lambda > 0 \\ 0 & \lambda \geq 0 \end{cases} \quad \lambda \sim \Gamma(1,2)$$

Supongamos que es una muestra aleatoria de tamaño n. ¿Cuál es el número de observciones para reducir la varianza, a lo sumo, a 0.01?

Por teorema de Bayes, la posterior $\lambda|x \sim \Gamma(y+1, n+2)$. Luego, la varianza de la Gamma es

$$\frac{\alpha}{\beta^2} = \frac{\sum x_i + 1}{(n+2)^2} \leq 0{,}01 \implies \frac{1}{(n+2)^2} \leq \frac{\sum x_i + 1}{(n+2)^2} \leq 0{,}01 \implies 100 \leq (n+2)^2 \implies n \geq 8$$

Teorema. Si $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ con σ^2 conocido y la previa es $\theta \sim N(\mu_0, V_0^2)$, entonces $\theta|X \sim N(\mu_1, V_1^2)$ donde

$$\mu_1 = \frac{\sigma^2\mu_0 + nV_0^2\bar{X}_n}{\sigma^2 + nV_0^2}, \quad V_1^2 = \frac{\sigma^2 V_0^2}{\sigma^2 + nV_0^2}$$

Prueba:

- **Verosimilitud**:

$$f_n(X|\theta) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i\theta)^2\right]$$

Luego,

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \theta)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \theta)^2 \\ &= n(\bar{X} + \theta)^2 + \sum_{i=1}^{n}(X_i - \bar{X})^2 + \underbrace{2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \theta)}_{=0 \text{ pues } \sum Xi = n\bar{X})}\end{aligned}$$

Entonces

$$f_n(X|\theta) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2\right].$$

- **Previa**:

$$\pi(\theta) \propto \exp\left[-\frac{1}{2V_0^2}(\theta-\mu_0)^2\right].$$

- **Posterior**:

$$\pi(\theta|X) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2 - \frac{1}{2V_0^2}(\theta-\mu_0)^2\right].$$

Con μ_1 y V_1^2 definidos anteriormente, se puede comprobar la siguiente identidad:

$$-\frac{n}{\sigma^2}(\bar{X}-\theta)^2 - \frac{1}{V_0^2}(\theta-\mu_0)^2 = \frac{1}{V_1^2}(\theta-\mu_1)^2 + \underbrace{\frac{n}{\sigma^2+nV_0^2}(\bar{X}_n-\mu_0)^2}_{\text{Constante con respecto a } \theta}$$

Por lo tanto,

$$\pi(\theta|X) \propto \exp\left[-\frac{n}{2V_1^2}(\theta-\mu_1)^2\right]$$

Media posterior:

$$\mu_1 = \underbrace{\frac{\sigma^2}{\sigma^2+nV_0^2}}_{W_1}\mu_0 + \underbrace{\frac{nV_0^2}{\sigma^2+nV_0^2}}_{W_2}\bar{X}_n$$

Afirmaciones:

1) Si V_0^2 y σ^2 son fijos, entonces $W_1 \xrightarrow[n\to\infty]{} 0$ (la importancia de la media empírica crece conforme aumenta n).
2) Si V_0^2 y n son fijos, entonces $W_2 \xrightarrow[\sigma^2\to\infty]{} 0$ (la importancia de la media empírica decrece conforme la muestra es menos precisa).

3) Si σ^2 y n son fijos, entonces $W_2 \xrightarrow[V_0^2 \to \infty]{} 1$ (la importancia de la media empírica crece conforma la previa es menos precisa).

Ejemplo (determinación de n)

Sean $X_1, \ldots, X_n \sim N(\theta, 1)$ y $\theta \sim N(\mu_0, 4)$. Sabemos que

$$V_1^2 = \frac{\sigma^2 V_0^2}{\sigma^2 + nV_0^2}.$$

Buscamos que $V_1 \leq 0{,}01$, entonces

$$\frac{4}{4n+1} \leq 0{,}01 \implies n \geq 99{,}75 \text{ (al menos 100 observaciones)}$$

3.6. Densidades previas impropias

Definición. Sea π una función positiva cuyo dominio está en Ω. Suponga que $\int \pi(\theta)\, d\theta = \infty$. Entonces decimos que π es una **densidad impropia**.

Ejemplo: $\theta \sim \text{Unif}(\mathbb{R})$, $\lambda \sim \text{Unif}(0, \infty)$.

Una técnica para seleccionar distribuciones impropia es sustituir los hiperparámetros previos por 0.

Ejemplo:

Se presenta el número de soldados prusianos muertos por una patada de caballo (280 conteros, unidades de combate en 20 años).

Unidades	Ocurrencias
144	0
91	1
32	2
11	3
2	4

- Muestra de Poisson: $X_1 = 0, X_2 = 1, X_3 = 1, \ldots, X_{280} = 0 \sim \text{Poi}(\lambda)$.
- Previa: $\lambda \sim \Gamma(\alpha, \beta)$.

- Posterior: $\lambda|X \sim \Gamma(y+\alpha, n+\beta) = \Gamma(196+\alpha, 280+\beta)$.

Sustituyendo, $\alpha = \beta = 0$

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha)}\beta^\alpha \lambda^{\alpha-1} e^{\beta\lambda}$$

$$\propto \lambda^{\alpha-1} e^{-\lambda\beta}$$

$$= \frac{1}{\lambda}$$

donde $\int_0^\infty \frac{1}{\lambda} d\lambda = \infty$.

Por teorema de Bayes,

$$\theta|X \sim \Gamma(196, 280)$$

3.7. Funciones de pérdida

Definición. Sean $X_1, \ldots, X_n$ datos observables cuyo modelo está indexado por $\theta \in \Omega$. Un estimador de θ es cualquier estadístico $\delta(X_1, \ldots, X_n)$.

Notación:

- Estimador $\to \delta(X_1, \ldots, X_n)$.
- Estimación o estimado: $\delta(X_1, \ldots, X_n)(\omega) = \delta(\overbrace{x_1, \ldots, x_n}^{datos})$

Definición. Una **función de pérdida** es una función de dos variables:

$$L(\theta, a), \quad \theta \in \Omega$$

con a un número real.

Interpretación: es lo que pierde un analista cuando el parámetro es θ y el estimador es a.

Asuma que θ tiene una previa. La pérdida esperada es

$$\mathbb{E}[L(\theta, a)] = \int_\Omega L(\theta, a)\pi(\theta)\, d\theta$$

la cual es una función de a, que a su vez es función de $X_1, \ldots, X_n$. Asuma que a se selecciona el minimizar esta esperanza. A ese estimador $a = \delta^*(X_1, \ldots, X_n)$

se le llama **estimador bayesiano**, si ponderamos los parámetros con respecto a la posterior.

$$\mathbb{E}[L(\theta,\delta^*)|X] = \int_\Omega L(\theta,a)\pi(\theta)\ d\theta = \min_a \mathbb{E}[L(\theta|a)X].$$

3.7.1. Función de pérdida cuadrática

$$L(\theta,a) = (\theta-a)^2$$

En el caso en que θ es real y $\mathbb{E}[\theta|X]$ es finita, entonces

$$\delta^*(X_1,\ldots,X_n) = \mathbb{E}[\theta|X] \text{ cuando } L(\theta,a) = (\theta-a)^2.$$

Ejemplo: $X_1,\ldots,X_n \sim \text{Ber}(\theta)$, $\theta \sim \text{Beta}(\alpha,\beta) \implies \theta|X \sim \text{Beta}(\alpha+y,\beta+n-y)$.

El estimador de θ es

$$\delta^*(X_1,\ldots,X_n) = \frac{\alpha+y}{\alpha+\beta+n} = \frac{\overbrace{\alpha}^{\text{Esperanza previa}}}{\alpha+\beta} \cdot \frac{\alpha+\beta}{\alpha+\beta+n} + \frac{\overbrace{y}^{\bar{X}}}{n} \cdot \frac{n}{\alpha+\beta+n}.$$

3.7.2. Función de pérdida absoluta

$$L(\theta,a) = |\theta-a|$$

La pérdida esperada es

$$f(a) = \mathbb{E}[L(\theta,a)|X] = \int_{-\infty}^{+\infty} |\theta-a|\pi(\theta|X)\ d\theta = \int_a^{+\infty} (\theta-a)\pi(\theta|X)\ d\theta + \int_{-\infty}^a (a-\theta)\pi(\theta|X)\ d\theta$$

Usando el teorema fundamental del cálculo,

$$F_\pi(a|X) = \int_{-\infty}^{\hat{a}} \pi(\theta|X)\ d\theta = \frac{1}{2} \Leftrightarrow \hat{a} = \underset{a}{\operatorname{argmin}}\, f(a)$$

La **mediana** es el punto de $X_{0,5}$ tal que $F(X_{0,5}) = \frac{1}{2}$.

Corolario. Bajo la función de pérdida absoluta, el estimador bayesiano es la mediana posterior.

Ejemplo: Bernoulli.

$$\frac{1}{\mathrm{Beta}(\alpha+y,\beta+n-y)}\int_{-\infty}^{X_{0,5}} \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}\,d\theta = \frac{1}{2}$$

Resuelva para $X_{0,5}$.

3.7.3. Otras funciones de pérdida

- $L(\theta,a)=|\theta-a|^k$, $k\neq 1,2$, $0<k<1$.
- $L(\theta,a)=\lambda(\theta)|\theta-a|^2$ ($\lambda(\theta)$ penaliza la magnitud del parámetro).
- $L(\theta,a)=\begin{cases}3(\theta-a)^2 & \theta\leq a \text{ (sobreestima)}\\ (\theta-a)^2 & \theta\geq a \text{ (subestima)}\end{cases}$

3.8. Efecto de muestras grandes

Ejemplo: ítemes malos (proporción: θ), $\theta\in[0,1]$. Función de pérdida cuadrática. El tamaño de muestra son $n=100$ ítemes, de los cuales $y=10$ están malos.

$$X_1,\ldots,X_n\sim \mathrm{Ber}(\theta)$$

- Primer previa. $\alpha=\beta=1$ (Beta). El estimador bayesiano corresponde a

$$\mathbb{E}[\theta|X]=\frac{\alpha+y}{\alpha+\beta+n}=\frac{1+10}{2+100}=0{,}108$$

- Segunda previa. $\alpha=1,\beta=2 \implies \pi(\theta)=2e^{-2\theta},\theta>0$.

$$\mathbb{E}[\theta|X]=\frac{1+10}{1+2+100}=\frac{11}{103}=0{,}107$$

La media es $\bar{X}_n=\dfrac{10}{100}=0{,}1$.

3.9. Consistencia

Definición. Un estimador de θ $\delta(X_1,\dots,X_n)$ es consistente si

$$\delta(X_1,\dots,X_n) \xrightarrow[n\to\infty]{\mathbb{P}} \theta.$$

Bajo pérdida cuadrática, $\mathbb{E}[\theta|X] = W_1\mathbb{E}[\theta] + X_2\bar{X}_n = \delta^*$. Sabemos, por ley de grandes números, que $\bar{X}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta$. Además, $W_1 \xrightarrow[n\to\infty]{} 0$ y $W_2 \xrightarrow[n\to\infty]{} 1$.

En los ejemplos que hemos analizado

$$\delta^* \xrightarrow[n\to\infty]{\mathbb{P}} \theta$$

Teorema. Bajo condiciones generales, los estimadores bayesianos son consistentes.

Estimador. Si $X_1,\dots,X_n$ es una muestra en un modelo indexado por θ, $\theta\in\Omega$ (k-dimensiones), sea

$$h:\Omega\to H\subset\mathbb{R}^d.$$

Sea $\psi = h(\theta)$. Un **estimador** de ψ es un estadístico $\delta^*(X_1,\dots,X_n)\in H$. A $\delta^*(X_1,\dots,X_n)$ estimador de ψ se puede evaluar y construir estimadores nuevos.

Ejemplo. $X_1,\dots,X_n \sim \text{Exp}(\theta)$, $\theta|X\sim\Gamma(\alpha,\beta)=\Gamma(4, 8{,}6)$. La característica de interés es $\psi = \dfrac{1}{\theta}$, el valor esperado del tiempo de fallo.

Es estimador se calcula de la siguiente manera:

$$\begin{aligned}
\delta^*(x) = \mathbb{E}[\psi|x] &= \int_0^\infty \frac{1}{\theta}\pi(\theta|x)\,d\theta\\
&= \int_0^\infty \frac{1}{\theta}\frac{8{,}6^4}{\Gamma(4)}\theta^3 e^{-8{,}6\theta}\,d\theta\\
&= \frac{8{,}6^4}{6}\underbrace{\int_0^\infty \theta^2 e^{-8{,}6\theta}\,d\theta}_{\frac{\Gamma(3)}{8{,}6^3}}\\
&= \frac{8{,}6^4}{6}\frac{2}{8{,}6^3} = 2{,}867 \text{ unidades de tiempo.}
\end{aligned}$$

Por otro lado, vea que $\mathbb{E}(\theta|X) = \dfrac{4}{8{,}6}$. El estimador *plug-in* correspondería a

$$\frac{1}{\mathbb{E}(\theta|X)} = \frac{8{,}6}{4} = 2{,}15.$$

3.10. Laboratorio

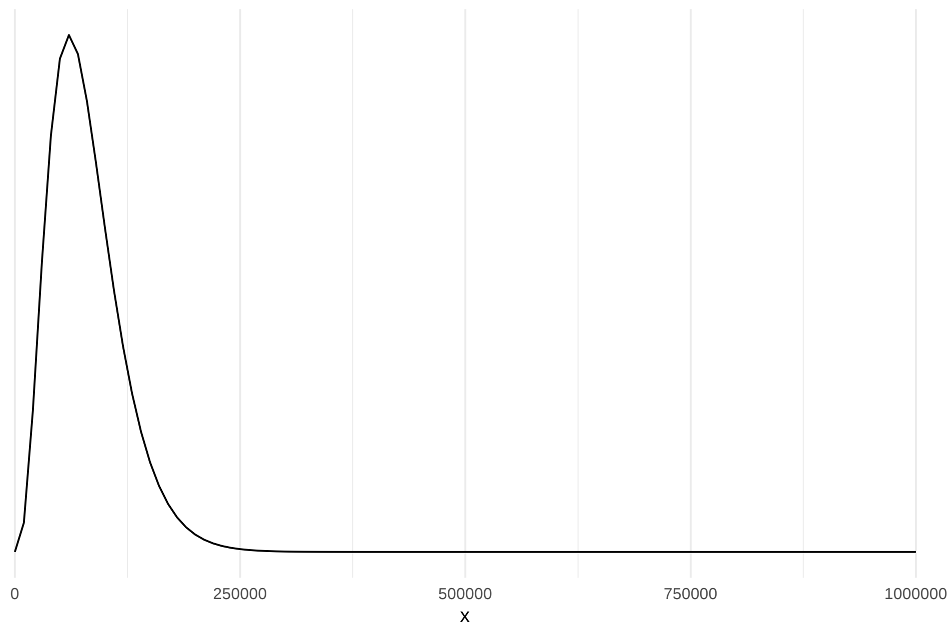
Lo primero es cargar los paquetes necesarios que usaremos en todo el curso

```
library(tidyverse)
```

3.10.1. Distribución previa

En nuestro ejemplo se tenía que $\mathbb{E}[\theta] = 0{,}0002$ y $\mathrm{Var}(\theta) = 0{,}001$. Suponiendo que θ es gamma se puede resolver el sistema de ecuaciones obtenemos que $\beta = 20000$ y $\alpha = 4$.

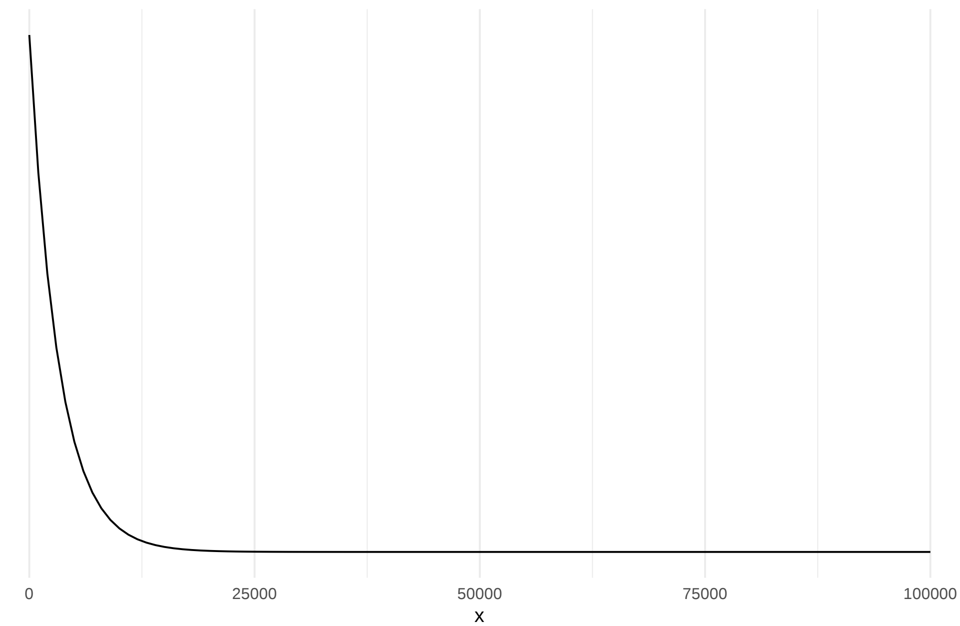
```
alpha_previa <- 4
beta_previa <- 20000

ggplot(data = data.frame(x = c(0, 1e+06)), aes(x)) +
    stat_function(fun = dgamma, args = list(shape = alpha_previa,
        scale = beta_previa)) + ylab("") + scale_y_continuous(breaks = NULL) +
    theme_minimal()
```



3.10.2. Distribución conjunta

Asumiendo que tenemos algunos datos $X_1, ..., X_n$, asumimos que estos son exponencial recordando que $\mathbb{E}[X] = 1/\theta$, entonces una aproximación de esta densidad es

```
x <- c(2911, 3403, 3237, 3509, 3118)

theta <- 1/mean(x)

ggplot(data = data.frame(x = c(0, 1e+05)), aes(x)) +
    stat_function(fun = dexp, args = list(rate = theta)) +
    ylab("") + scale_y_continuous(breaks = NULL) +
    theme_minimal()
```



3.10.3. Distribución posterior

Según los contenidos del curso, se puede estimar los parámetros de la densidad posterior de la forma

```
(y <- sum(x))
```

```
## [1] 16178
```

```
(n <- length(x))
```

```
## [1] 5
```

```
(alpha_posterior <- n + alpha_previa)
```

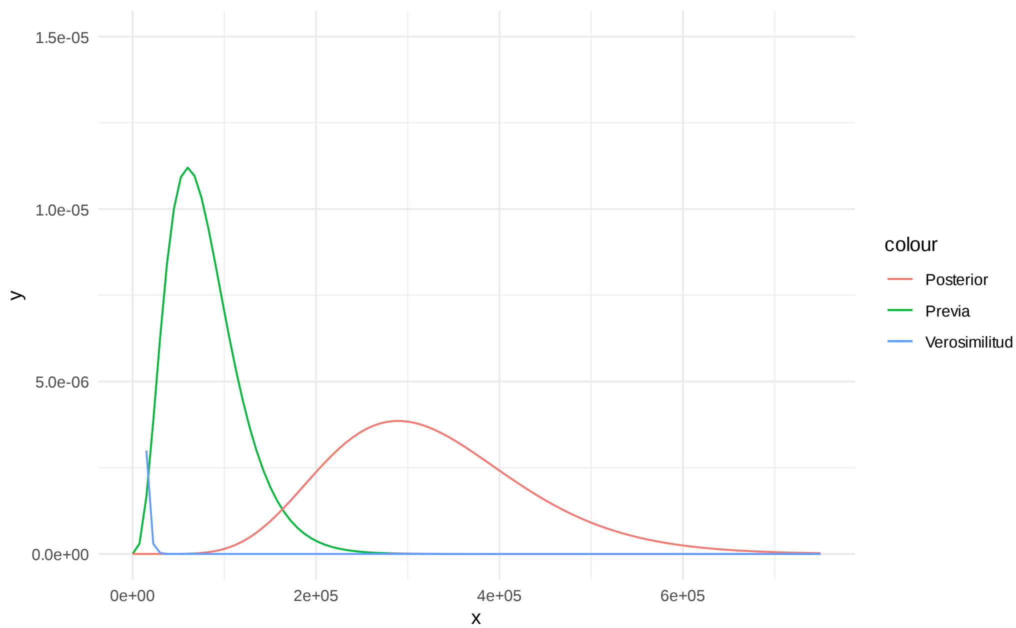
```
## [1] 9
```

```
(beta_posterior <- beta_previa + y)
```

```
## [1] 36178
```

```
ggplot(data = data.frame(x = c(0, 750000)), aes(x)) +
    stat_function(fun = dgamma, args = list(shape = alpha_previa,
```

```
        scale = beta_previa), aes(color = "Previa")) +
    stat_function(fun = dgamma, args = list(shape = alpha_posterior,
        scale = beta_posterior), aes(color = "Posterior")) +
    stat_function(fun = dexp, args = list(rate = theta),
        aes(color = "Verosimilitud")) + ylim(0, 1.5e-05) +
    theme_minimal()
```



3.10.4. Agregando nuevos datos

Si tenemos un 6to dato, y queremos ver cual es su distribución posterior. Lo primero es estimar la densidad posterior de este 6to dato, pero asumiendo que la previa es la densidad que obtuvimos en el caso anterior.

Suponga que $X_6 = 3000$

```
(alpha_previa <- alpha_posterior)
```

```
## [1] 9
```

```
(beta_previa <- beta_posterior)
```

```
## [1] 36178
```

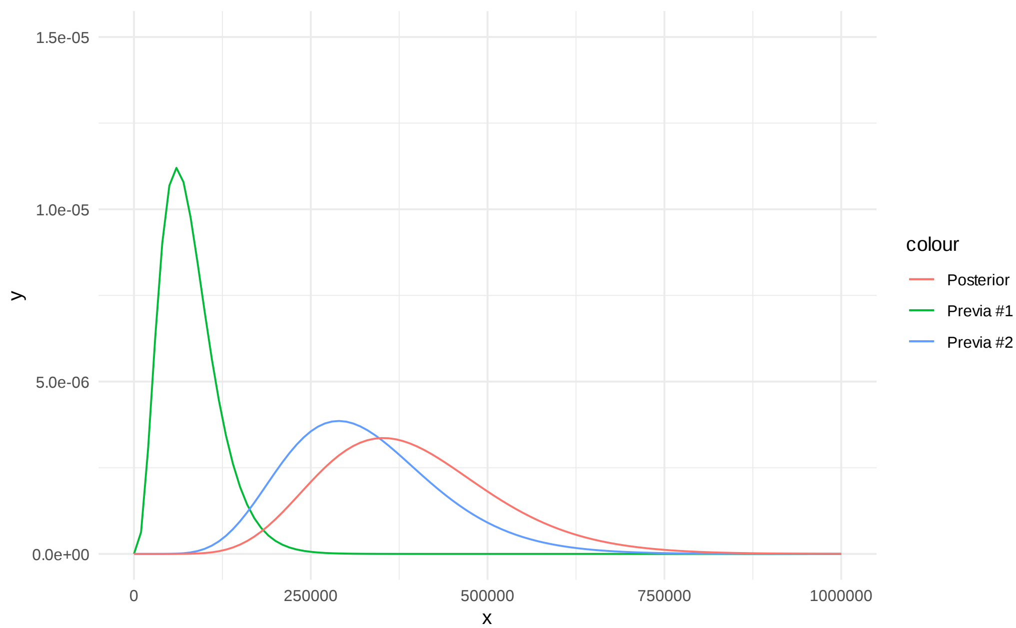
```
(alpha_posterior <- alpha_previa + 1)
```

```
## [1] 10
```

```
(beta_posterior <- beta_previa + 3000)
```

```
## [1] 39178
```

```
ggplot(data = data.frame(x = c(0, 1e+06)), aes(x)) +
    stat_function(fun = dgamma, args = list(shape = 4,
        scale = 20000), aes(color = "Previa #1")) +
    stat_function(fun = dgamma, args = list(shape = alpha_previa,
        scale = beta_previa), aes(color = "Previa #2")) +
    stat_function(fun = dgamma, args = list(shape = alpha_posterior,
        scale = beta_posterior), aes(color = "Posterior")) +
    ylim(0, 1.5e-05) + theme_minimal()
```



3.10.5. Familias conjugadas normales

Si tenemos pocos datos, la información previa es la que "prevalece".

```
x <- rnorm(n = 3, mean = 10, sd = 1)

(mu <- mean(x))
```

```
## [1] 10.64715
```

```
(sigma <- sd(x))
```

```
## [1] 0.5654769
```

```
(n <- length(x))
```

```
## [1] 3
```

```
(mu_previa <- 0)
```

```
## [1] 0
```

```
(sigma_previa <- 1)
```

```
## [1] 1
```

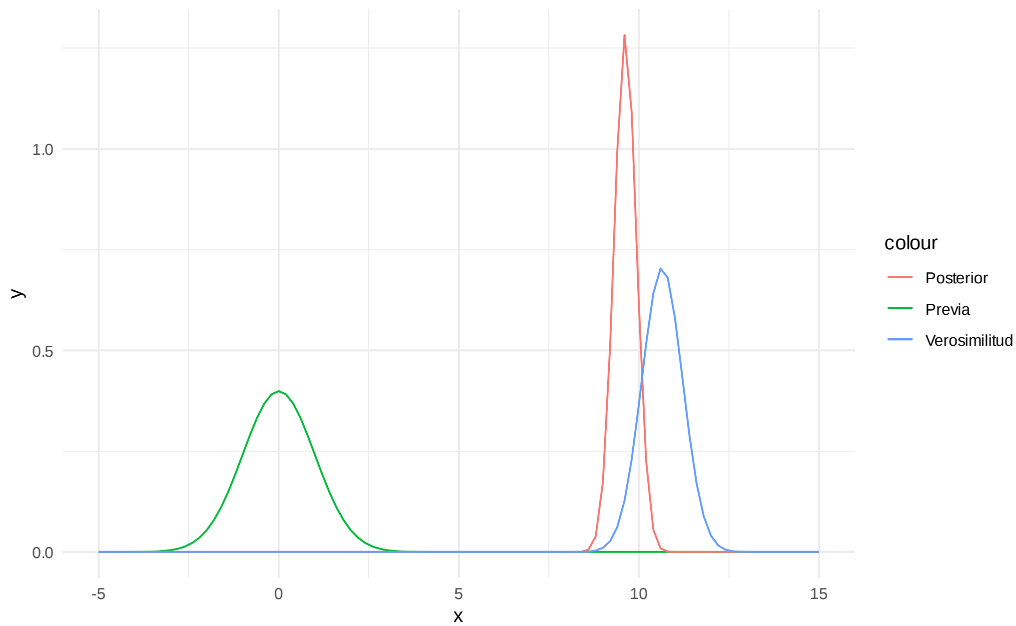
```
(mu_posterior <- ((sigma^2)/(sigma^2 + n * sigma_previa^2)) *
    mu_previa + ((n * sigma_previa^2)/(sigma^2 + n *
    sigma_previa^2)) * mu)
```

```
## [1] 9.621605
```

```
(sigma2_posterior <- (sigma^2 * sigma_previa^2)/(sigma^2 +
    n * sigma_previa^2))
```

```
## [1] 0.09632134
```

```
ggplot(data = data.frame(x = c(-5, 15)), aes(x)) +
    stat_function(fun = dnorm, args = list(mean = mu_previa,
        sd = sigma_previa), aes(color = "Previa")) +
    stat_function(fun = dnorm, args = list(mean = mu_posterior,
        sd = sqrt(sigma2_posterior)), aes(color = "Posterior")) +
    stat_function(fun = dnorm, args = list(mean = mu,
        sd = sigma), aes(color = "Verosimilitud")) +
    theme_minimal()
```



Con más datos, la distribución se ajusta a esto y le quita importancia a la información previa.

```
x <- rnorm(n = 100, mean = 10, sd = 1)

(mu <- mean(x))
```

```
## [1] 9.994564
```

```
(sigma <- sd(x))
```

```
## [1] 1.060066
```

```
(n <- length(x))
```

```
## [1] 100
```

```
(mu_previa <- 0)
```

```
## [1] 0
```

```
(sigma_previa <- 1)
```

```
## [1] 1
```

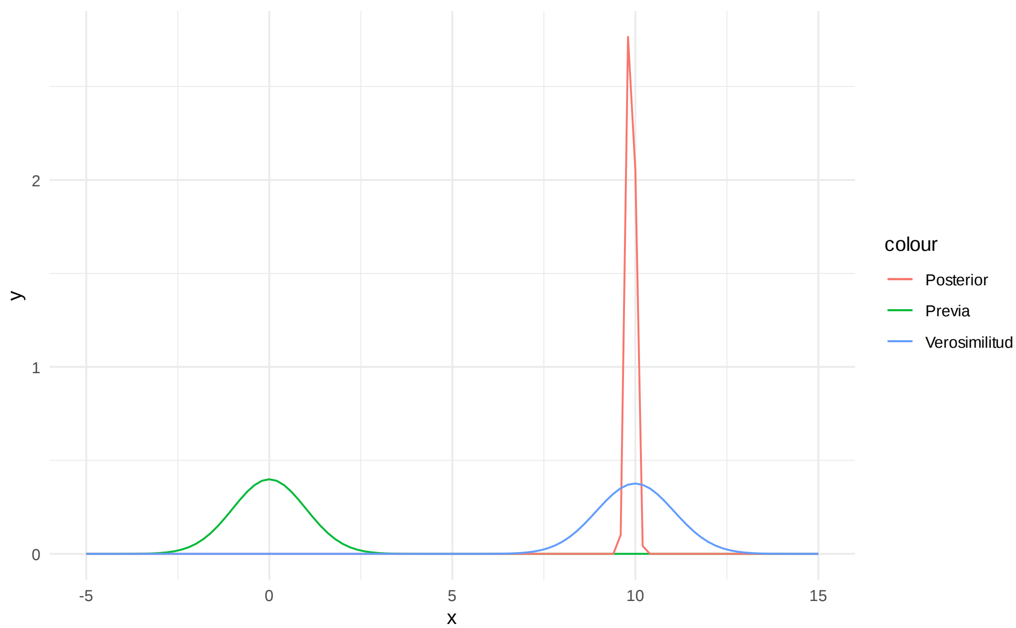
```
(mu_posterior <- ((sigma^2)/(sigma^2 + n * sigma_previa^2)) *
    mu_previa + ((n * sigma_previa^2)/(sigma^2 + n *
    sigma_previa^2)) * mu)
```

```
## [1] 9.883499
```

```
(sigma2_posterior <- (sigma^2 * sigma_previa^2)/(sigma^2 +
    n * sigma_previa^2))
```

```
## [1] 0.01111253
```

```
ggplot(data = data.frame(x = c(-5, 15)), aes(x)) +
    stat_function(fun = dnorm, args = list(mean = mu_previa,
        sd = sigma_previa), aes(color = "Previa")) +
    stat_function(fun = dnorm, args = list(mean = mu_posterior,
        sd = sqrt(sigma2_posterior)), aes(color = "Posterior")) +
    stat_function(fun = dnorm, args = list(mean = mu,
        sd = sigma), aes(color = "Verosimilitud")) +
    theme_minimal()
```



Si los datos por si solo son muy variable, la posterior tiende a parecerse a la distribución previa en lugar que a la verosimilitud.

```
x <- rnorm(n = 10, mean = 10, sd = 5)

(mu <- mean(x))
```

```
## [1] 9.576984
```

```
(sigma <- sd(x))
```

```
## [1] 7.32444
```

```
(n <- length(x))
```

```
## [1] 10
```

```
(mu_previa <- 0)
```

```
## [1] 0
```

```
(sigma_previa <- 1)
```

```
## [1] 1
```

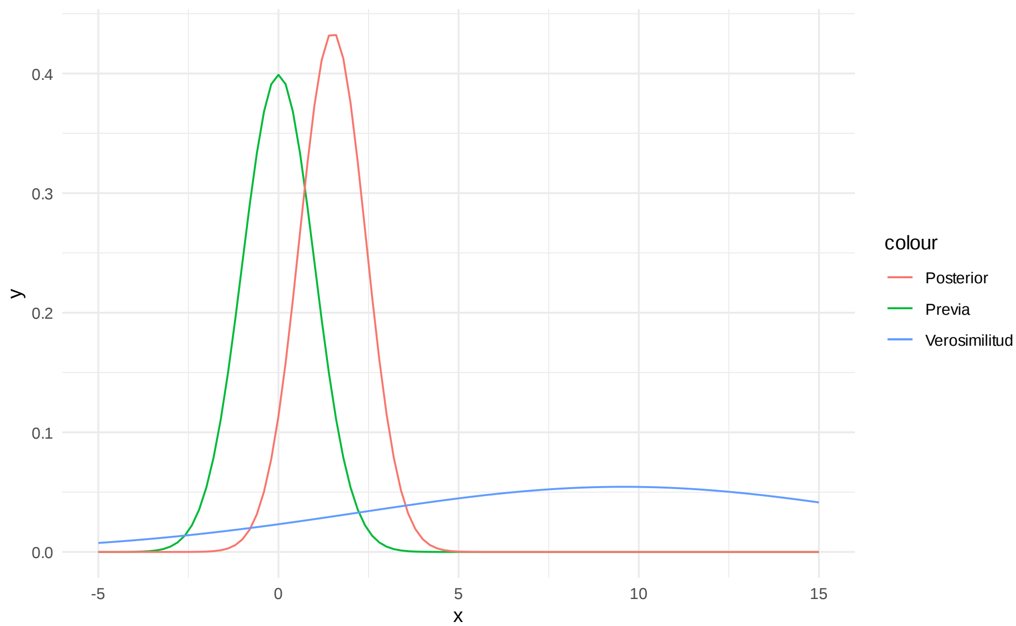
```
(mu_posterior <- ((sigma^2)/(sigma^2 + n * sigma_previa^2)) *
    mu_previa + ((n * sigma_previa^2)/(sigma^2 + n *
    sigma_previa^2)) * mu)
```

```
## [1] 1.504693
```

```
(sigma2_posterior <- (sigma^2 * sigma_previa^2)/(sigma^2 +
    n * sigma_previa^2))
```

```
## [1] 0.8428844
```

```
ggplot(data = data.frame(x = c(-5, 15)), aes(x)) +
    stat_function(fun = dnorm, args = list(mean = mu_previa,
        sd = sigma_previa), aes(color = "Previa")) +
    stat_function(fun = dnorm, args = list(mean = mu_posterior,
        sd = sqrt(sigma2_posterior)), aes(color = "Posterior")) +
    stat_function(fun = dnorm, args = list(mean = mu,
        sd = sigma), aes(color = "Verosimilitud")) +
    theme_minimal()
```



3.10.6. Funciones de pérdida

Lo más importante acá es que dependiendo de la función de pérdida podemos construir una estimador para θ. En el caso de los componentes electrónicos recordemos que la posterior nos daba

```
alpha <- 9
beta <- 36178
```

- **Pérdida cuadrática:** Recoremos que la media de una gamma es α/β entonces

```
(theta <- alpha/beta)
```

```
## [1] 0.00024877
```

Y por lo tanto el tiempo promedio del componente electrónico es $1/\theta$=4019.7777778.

- **Pérdidad absoluta:** La distribución Gamma no tiene una forma cerrada para la mediana, por que se puede aproximar así,

```
m <- rgamma(n = 1000, scale = beta, shape = alpha)
(theta <- median(m))
```

```
## [1] 312498.7
```

Y por lo tanto el tiempo promedio del componente electrónico es $1/\theta$=3,2000131 $\times 10^{-6}$.

OJO: En este caso la pérdida cuadrática ajusta mejor ya que la distribución que la pérdida absoluta ya que la distribución NO es simétrica. En el caso simétrico los resultados serían muy similares.

3.10.7. Caso concreto

Suponga que se que quiere averiguar si los estudiantes de cierto colegio duermen más de 8 horas o menos de 8 horas.

Para esto primero cargaremos el siguiente paquete,

```
library(LearnBayes)
```

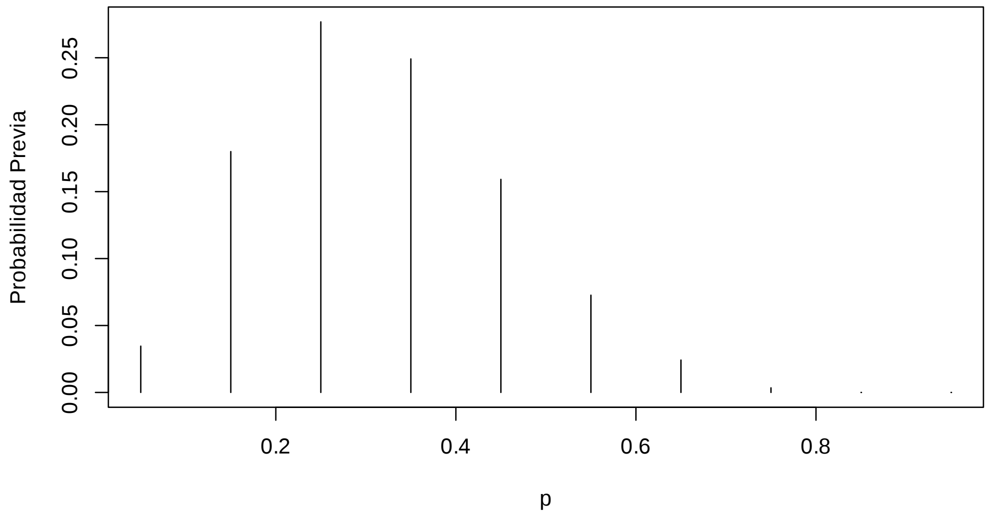
Suponga que se hace una encuesta a 27 estudiantes y se encuentra que 11 dicen que duermen más de 8 horas diarias y el resto no. Nuestro objetivo es encontrar inferencias sobre la proporción p de estudiantes que duermen al menos 8 horas diarias. El modelo más adecuado es

$$f(x|p) \propto p^s(1-p)^f$$

donde s es la cantidad de estudiantes que duermen más de 8 horas y f los que duermen menos de 8 horas.

Una primera aproximación para la previa es usar una distribución discreta. En este caso, el investigador asigna una probabilidad a cierta cantidad de horas de sueño, según su experiencia. Así, por ejemplo:

```
p <- seq(0.05, 0.95, by = 0.1)
prior <- c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
prior <- prior/sum(prior)
plot(p, prior, type = "h", ylab = "Probabilidad Previa")
```



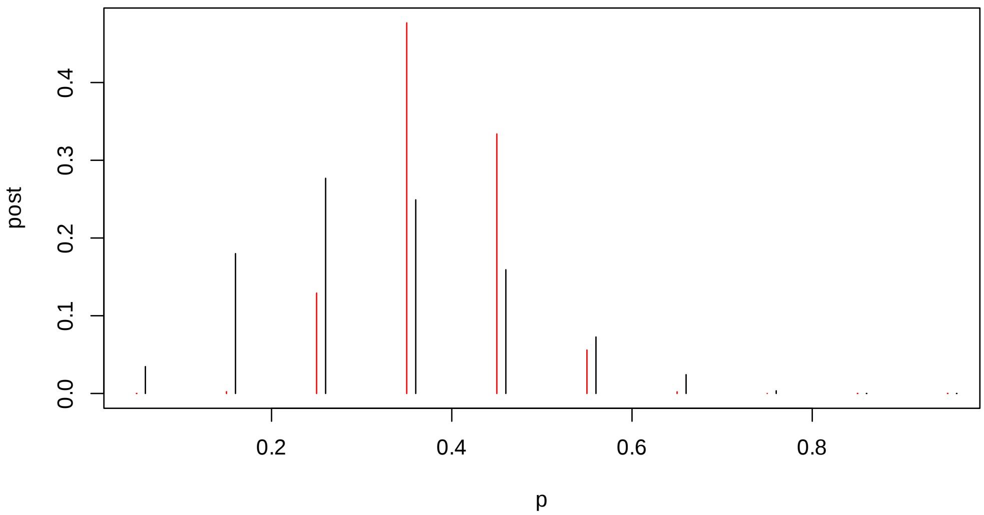
El paquete `LearnBayes` tiene la función `pdisc` que estima la distribución posterior para una previa discreta binomial. Recuerde que el valor 11 representa la cantidad de estudiantes con más de 8 horas de sueño y 16 lo que no duermen esa cantidad.

```
data <- c(11, 16)
post <- pdisc(p, prior, data)
round(cbind(p, prior, post), 2)
```

```
##          p prior post
##  [1,] 0.05  0.03 0.00
##  [2,] 0.15  0.18 0.00
##  [3,] 0.25  0.28 0.13
##  [4,] 0.35  0.25 0.48
##  [5,] 0.45  0.16 0.33
##  [6,] 0.55  0.07 0.06
##  [7,] 0.65  0.02 0.00
##  [8,] 0.75  0.00 0.00
##  [9,] 0.85  0.00 0.00
## [10,] 0.95  0.00 0.00
```

Y podemos ver la diferencia entre la previa (negro) y la posterior (roja),

```
plot(p, post, type = "h", col = "red")
lines(p + 0.01, prior, type = "h")
```



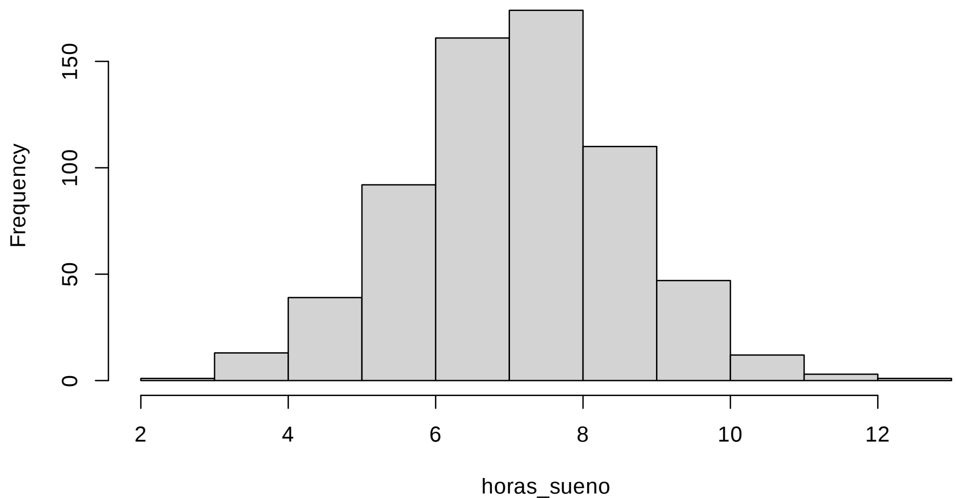
¿Qué se puede deducir de estos resultados?

Ejercicio: Suponga que se tiene la base de datos `studentdata`. Realice los cálculos anteriores con esos datos,

```
data("studentdata")
horas_sueno <- studentdata$WakeUp - studentdata$ToSleep
horas_sueno <- na.omit(horas_sueno)
summary(horas_sueno)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.500   6.500   7.500   7.385   8.500  12.500
```

```
hist(horas_sueno, main = "")
```



Ahora supongamos que se tiene quiere ajustar una previa continua a este modelo. Para esto usaremos una distribución Beta con parámetros α y β, de la forma

$$pi(p|\alpha,\beta) \propto p^{1-\alpha}(1-p)^{1-\beta}.$$

El ajuste de los paramétros de la Beta depende mucho de la información previa que se tenga del modelo. Una forma fácil de estimarlo es a través de cuantiles con los cuales se puede reescribir estos parámetros. Para una explicación detallada revisar https://stats.stackexchange.com/a/237849

En particular, suponga que se cree que el 50 % de las observaciones la proporción será menor que 0.3 y que el 90 % será menor que 0.5.

Para esto ajustaremos los siguientes parámetros

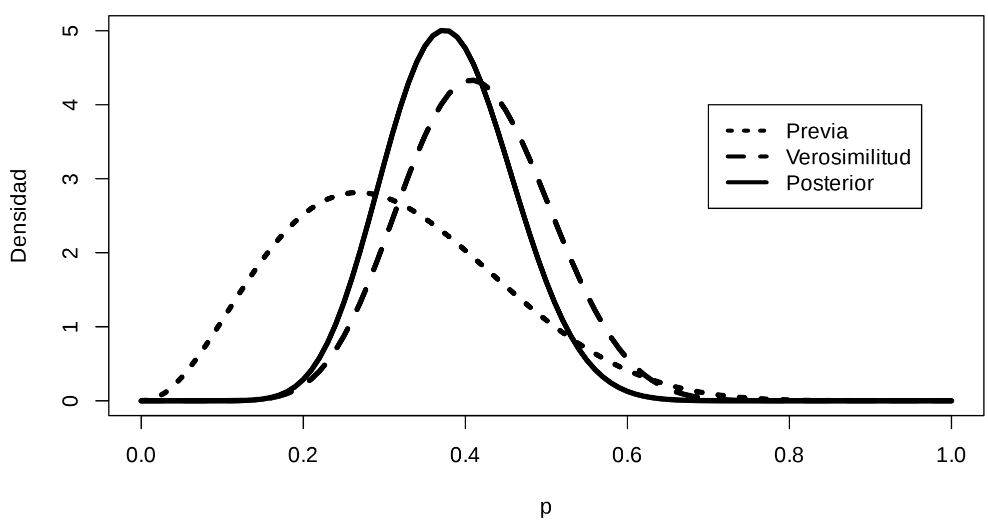
```
quantile2 <- list(p = 0.9, x = 0.5)
quantile1 <- list(p = 0.5, x = 0.3)
(ab <- beta.select(quantile1, quantile2))
```

```
## [1] 3.26 7.19
```

```
a <- ab[1]
b <- ab[2]
s <- 11
f <- 16
```

En este caso se obtendra la distribución posterior Beta con parámetros $\alpha + s$ y $\beta + f$,

```
curve(dbeta(x, a + s, b + f), from = 0, to = 1, xlab = "p",
    ylab = "Densidad", lty = 1, lwd = 4)
curve(dbeta(x, s + 1, f + 1), add = TRUE, lty = 2,
    lwd = 4)
curve(dbeta(x, a, b), add = TRUE, lty = 3, lwd = 4)
legend(0.7, 4, c("Previa", "Verosimilitud", "Posterior"),
    lty = c(3, 2, 1), lwd = c(3, 3, 3))
```



Capítulo 4

Estimación por máxima verosimilitud

¿Será posible estimar sin una densidad previa? Se debería ajustar la noción de muestra a independencia dado el valor de un parámetro.

Recuerde que, para $X_1, \ldots, X_n \overset{i.i.d}{\sim} f(X|\theta)$ con θ fijo, la **función de verosimilitud** se define como

$$f_n(X|\theta) = \pi(X_i|\theta) = G(\theta|X).$$

Si $\theta_1, \theta_2 \in \Omega$, θ es el valor real del parámetro. Si la muestra es fija, evaluamos, para θ_1, $f_n(X|\theta_1) = L(\theta_1|X)$ y, de igual forma para θ_2, $f_n(X|\theta_2) = L(\theta_2|X)$. Supongamos que

$$f_n(X|\theta_1) > f_n(X|\theta_2) \implies L(\theta_1|X) > L(\theta_2|X) \text{ (principio de verosimilitud)}$$

Interpretación. Es más verosímil (realista) que el verdadero parámetro sea θ_1 que θ_2 dada la muestra.

Definición. Para cada $x \in \mathcal{X}$ (espacio muestral), sea $\delta(x) \in \delta$ estimador de θ tal que $f_n(x|\theta)$ es máximo. A $\delta(x)$ se le llama **MLE (estimador de máxima verosimilitud)**.

Ejemplo. Si $X_1, \ldots, X_n \sim \text{Exp}(\theta)$, estime θ.

Determinamos la función de verosimilitud,

$$f_n(X|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-X_i/\theta} = \frac{1}{\theta^n} \exp\left(\frac{1}{\theta}\sum_{i=1}^{n} X_i\right) = \theta^{-n} e^{-y/\theta}.$$

Considere la **log-verosimilitud**

$$\ell(\theta|X) = \ln f_n(X|\theta) = -n\ln\theta - \frac{y}{\theta}$$

Como es una transformación monótona creciente, la función de verosimilitud se maximiza si la log-verosimilitud es máxima. Entonces,

$$\begin{aligned} \frac{\partial}{\partial\theta}\ell(\theta|X) &= \frac{-n}{\theta} + \frac{y}{\theta^2} = 0 \\ \implies \frac{1}{\theta}\left(-n + \frac{y}{\theta}\right) &= 0 \\ \implies \hat{\theta} = \frac{y}{n} &= \bar{X}_n. \end{aligned}$$

Para verificar que es un máximo:

$$\frac{\partial^2 \ell}{\partial\theta^2} = \frac{n}{\theta^2} - \frac{2y}{\theta^3}\bigg|_{\theta=\frac{y}{n}} = \frac{1}{\hat{\theta}^2}\left[n - \frac{2y}{\frac{y}{n}}\right] = \frac{-n}{\hat{\theta}^2} < 0.$$

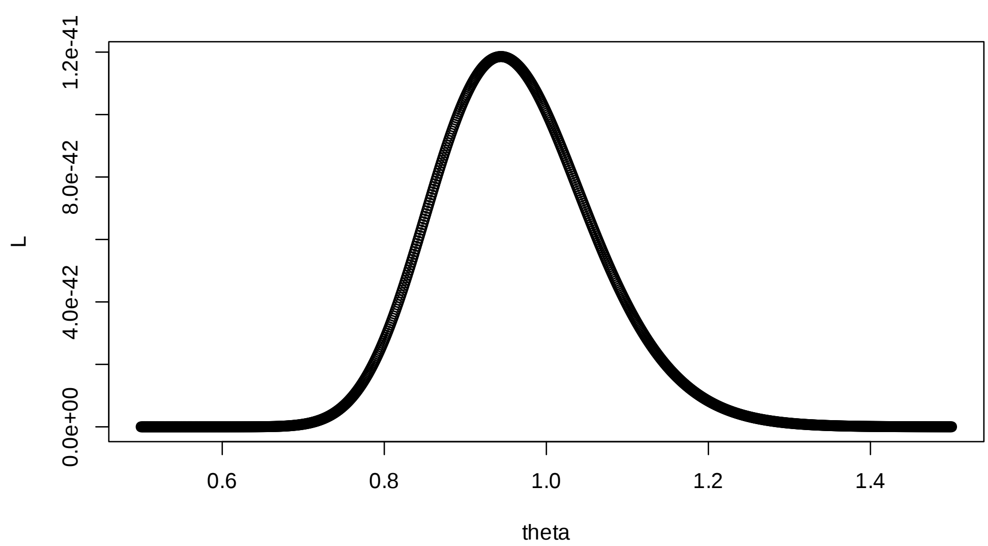
Entonces $\hat{\theta} = \bar{X}_n$ es el MLE de θ.

Laboratorio:

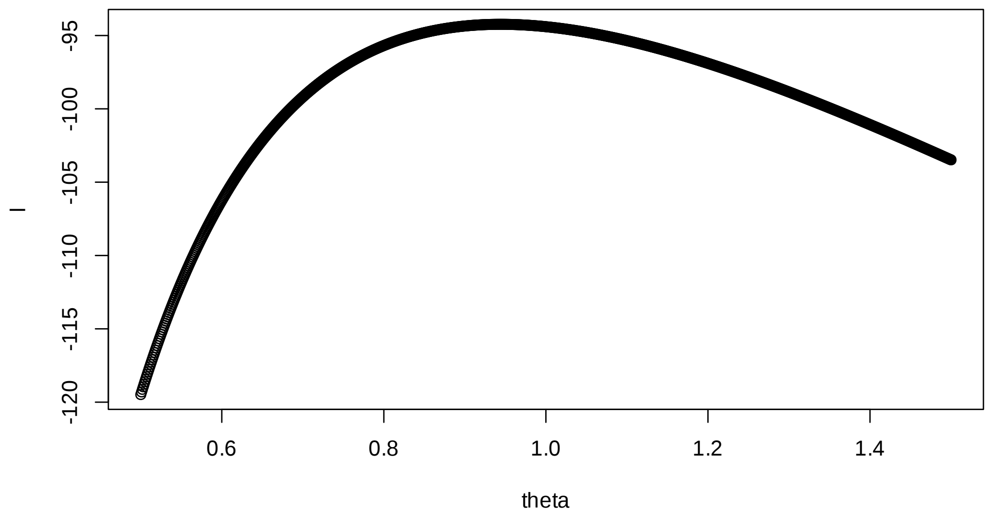
Suponga que se tiene 100 valores con distribución exponencial con parámetro $\theta = 1$.

```
x <- rexp(n = 100, rate = 1)
n <- length(x)
y <- sum(x)
theta <- seq(0.5, 1.5, length.out = 1000)
```

```
L <- theta^(-n) * exp(-y/theta)
plot(theta, L)
```




```
l <- -n * log(theta) - y/theta
plot(theta, l)
```



Ejemplo. En una prueba sobre alguna enfermedad, en un 90 % da la verdadera condición (enfermo) y en un 10 % la prueba se equivoca (que diga que la persona esté enferma cuando está sana). Considere una variable aleatoria Bernoulli(θ),$\theta \in \{0{,}9, 0{,}1\}$ Una muestra sería

$$x = \begin{cases} 1 & \text{si la prueba es positiva} \\ 0 & \text{si no} \end{cases}$$

Si $x = 0$, entonces $f(0|\theta) = \begin{cases} 0{,}9 & \text{si } \theta = 0{,}1 \\ 0{,}1 & \text{si } \theta = 0{,}9 \end{cases}.$

Si $x = 1$, entonces $f(1|\theta) = \begin{cases} 0{,}1 & \text{si } \theta = 0{,}1 \\ 0{,}9 & \text{si } \theta = 0{,}9 \end{cases}.$

El MLE corresponde a

$$\hat{\theta} = \begin{cases} 0{,}1 & \text{si } x = 0 \\ 0{,}9 & \text{si } x = 1 \end{cases}$$

Ejemplo. Para el caso normal, $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, σ^2 conocida, estime μ.

$$f_n(x|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

La log-verosimilitud es de la forma

$$\ell(\mu|x) = \frac{-n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Basta con minimizar $Q(\mu) = \sum_{i=1}^{n}(x_i-\mu)^2$.

$$\frac{\partial Q}{\partial \mu} = -2\sum_{i=1}^{n}(x_i-\mu) \implies n\mu = \sum_{i=1}^{n} x_i \implies \hat{\mu} = \bar{x}_n.$$

No hace falta verificar la condición de segundo orden, pues Q es una función cuadrática de μ y tiene un único máximo.

$$\hat{\mu}_{MLE} = \bar{x}_n \quad (*)$$

Ahora, si $X_1,\dots,X_n \sim N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$ desconocido, por $(*)$,

$$\ell(\sigma^2|X_1,\dots,X_n) = \frac{n}{2}\ln(2\pi\sigma^2) - -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{2\pi\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 = 0$$

Entonces

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \text{ (varianza muestral)}$$

Las condiciones de segundo orden quedan como ejercicio.

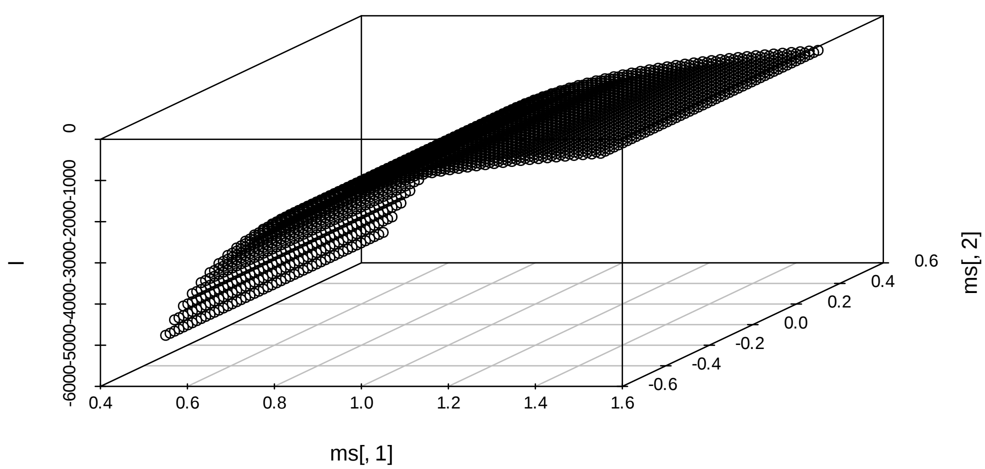
Nota. Si θ_{MLE} de θ, entonces $h(\theta_{MLE})$ es el MLE de $h(\theta)$.

Sea $h(x,y) = \sqrt{y}$ (es inyectiva). $h(\bar{x}_n, \hat{\sigma}^2) = \sqrt{\hat{\sigma}^2} = \hat{\sigma}$.

El MLE de $\dfrac{\sigma}{\mu} = \dfrac{\hat{\sigma}}{\bar{x}_n}$.

Laboratorio:

```
library(scatterplot3d)

x <- rnorm(100)
n <- length(x)

mu <- seq(-0.5, 0.5, length.out = 50)
sigma <- seq(0.5, 1.5, length.out = 50)

ms <- expand.grid(sigma, mu)

l <- -(n/2) * log(2 * pi/ms[, 1]^2) - (1/(2 * ms[,
    1]^2) * sum((x - ms[, 2])^2))

scatterplot3d(ms[, 1], ms[, 2], l, angle = 45)
```



Ejemplo. $X_1, \ldots, X_n \overset{i.i.d}{\sim}$ Unif$(0, \theta)$. Estime θ ($\theta > 0$). Suponga que $x_i > 0 \forall i$.

$$f(X|\theta) = \frac{1}{\theta} \cdot 1_{[0,\theta]}(x)$$

La verosimilitud es

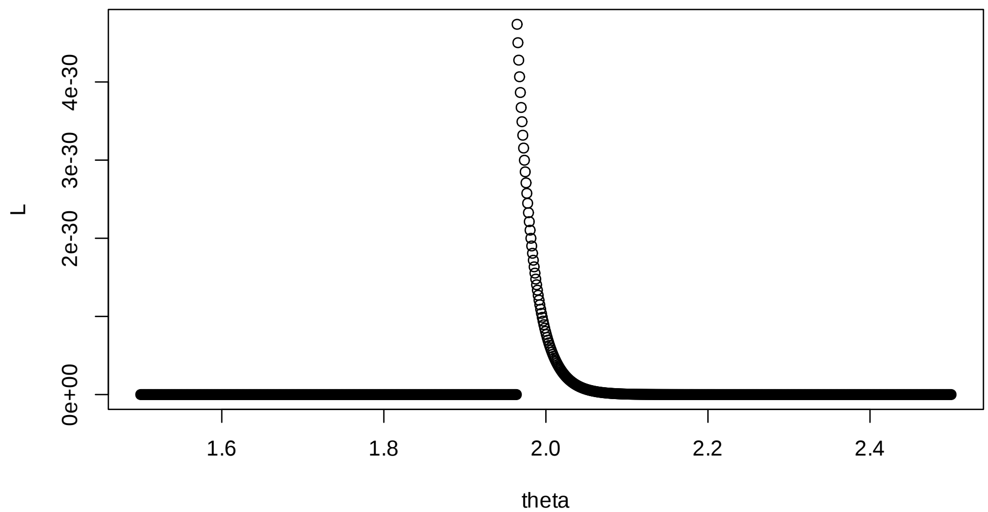
$$f_n(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} 1_{\{0 \leq X_i \leq \theta\}} \quad 0 \leq X_i \leq \theta \; \forall i$$

Vea que $f_n(x|\theta)$ es válido si y solo si $0 \leq X_{(n)} \leq \theta$.

El valor de la muestra $\{X_1, \ldots, X_n\}$ en la i-ésima posición cuando los datos se ordenan de menor a mayor se denota $X_{(i)}$ (estadístico de orden). En este caso, $X_{(n)} = \text{máx}\{X_1, \ldots, X_n\}$. Entonces $\hat{\theta}_{MLE} = x_{(n)}$.

Laboratorio:

```
x <- runif(100, 0, 2)
n <- length(x)

theta <- seq(1.5, 2.5, length.out = 1000)

L <- numeric(1000)
for (k in 1:1000) {
    L[k] <- 1/theta[k]^n * prod(x < theta[k])
}

plot(theta, L)
```



4.1. Propiedades del MLE

4.1.1. Propiedad de invarianza

Teorema. Si $\hat{\theta}$ es el MLE de $\hat{\theta}$ y si g es biyectiva, entonces $g(\theta)$ es el MLE de $g(\theta)$.

Prueba:

Sea Γ el espacio paramétrico $g(\Omega)$. Como g es biyectiva entonces defina h la inversa de g: $\theta = h(\psi), \psi \in \Gamma$.

Reparametrizando la verosimilitud,

$$f_n(x|\theta) = f_n(x|h(\psi)).$$

El MLE de $\psi : \hat{\psi}$ satisface que $f_n(x|h(\hat{\psi}))$ es máximo.

Como $f_n(x|\theta)$ se maximiza cuando $\theta = \hat{\theta}$, entonces $f_n(x|h(\psi))$ se maximiza cuando $\hat{\theta} = h(\psi)$ para algún ψ.

Se concluye que $\hat{\theta} = h(\hat{\psi}) \implies \hat{\psi} = g(\hat{\theta})$.

Ejemplo: $g(\theta) = \dfrac{1}{\theta}$ es biyectiva si $\theta > 0$. Así,

$$\frac{1}{\hat{\theta}} = \frac{1}{\frac{1}{\bar{X}_n}} = \bar{X}_n \text{ es parámetro de la tasa.}$$

¿Qué pasa si h no es biyectiva?

Definicion (Generalización del MLE). Si g es una función de θ y G la imagen de Ω bajo g. Para cada $t \in G$ defina

$$G_t = \{\theta : g(\theta) = t\}$$

Defina $L^*(t) = \max\limits_{\theta \in G_t} \ln f_n(x|\theta)$. El MLE de $g(\theta)(= \hat{t})$ satisface $L^*(\hat{t}) = \max\limits_{t \in G} L^*(t)$.

Teorema. Si $\hat{\theta}$ es el MLE de θ entonces $g(\hat{\theta})$ es el MLE de $g(\theta)$ (g es arbitraria).

Prueba. Basta probar $L^*(\hat{t}) = \ln f_n(x|\hat{\theta})$. Se cumple que $\hat{\theta} \in G_{\hat{t}}$. Como $\hat{\theta}$ maximiza $f_n(x|\theta)$ $\forall \theta$, también lo hace si $\theta \in G_{\hat{t}}$. Entonces $\hat{t} = g(\hat{\theta})$ (no pueden existir 2 máximos en un conjunto con la misma imagen).

Ejemplos.$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.

- Si $h(\mu, \sigma^2) = \sigma$ (no es biyectiva) $\implies h(\hat{X}_n, \hat{\sigma}^2) = \sqrt{\hat{\sigma}^2}$ es el MLE de σ.
- $h(\mu, \sigma^2) = \dfrac{\sigma^*}{\mu}$ (coeficiente de variación). $\dfrac{\hat{\sigma}}{\bar{X}_n}$ es el MLE de CV.
- $h(\mu, \sigma^2) = \mu^2 + \sigma^2$. $\mathbb{E}[X^2] - \mu^2 = \sigma^2 \implies \mathbb{E}[X^2] = \mu^2 + \sigma^2$. El MLE de $\mathbb{E}[X^2]$ es $\bar{X}_n^2 + \hat{\sigma}^2$.

4.1.2. Consistencia

Los estimadores bayesianos son de la forma

$$EB = W_1 \mathbb{E}[\text{Previa}] + W_2 \hat{X}_n.$$

El estimador bayesiano "combina" la esperanza de la previa y el $\hat{\theta}_{MLE}$. El $\hat{\theta}_{MLE}$ "hereda la consistencia del estimador bayesiano".

$$EB = W_1\mathbb{E}[\text{Previa}] + W_2\hat{\theta}_{MLE}.$$

Afirmación. Bajo “condiciones usuales”,

$$\hat{\theta}_{MLE} \xrightarrow[n\to\infty]{\mathbb{P}} \theta.$$

4.2. Cálculo numérico

4.2.1. Método de los momentos

Ejemplo. $X_1, \dots, X_n \sim \Gamma(\alpha, 1)$. Estime α.

$$f_n(x|\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1}e^{-x}.$$

Verosimilitud: $f_n(x|\alpha) = \dfrac{1}{\Gamma(\alpha)^n}(\prod x_i)e^{\sum x_i}$.

$$\begin{aligned}\frac{\partial}{\partial\alpha}L(\alpha|x) &= \frac{\partial}{\partial\alpha}\Big[-n\ln\Gamma(\alpha) + (\alpha-1)\ln(\pi x_i) - \sum x_i \Big] \\ &= -n\frac{1}{\Gamma(\alpha)}\frac{d}{d\alpha}\Gamma(\alpha) + \ln(\prod x_i) = 0\end{aligned}$$

Definición. Asumimos que $X_1, \dots, X_n \sim F$ indexada con un parámetro $\theta \in \mathbb{R}^k$ y que al menos tiene k momentos finitos. Para $j = 1, \dots, k$ sea $\mu_j(\theta) = \mathbb{E}[X_1^j|\theta]$. Suponga que $\mu(\theta) = (\mu_1(\theta), \dots, \mu_2(\theta))$ es biyectiva. Sea M la inversa de μ,

$$M(\mu(\theta)) = \theta = M(\mu_1(\theta), \dots, \mu_2(\theta))$$

y defina los momentos empíricos

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

El estimador según el método de los momentos es

$$\hat{\theta} = M(m_1, \dots, m_k).$$

Del ejemplo anterior, $\mu_1(\alpha) = \mathbb{E}[x_1|\alpha] = \alpha$.Dado que $m_1 = \hat{x}_n$, el sistema por resolver es

$$\mu_1(\alpha) = m_1 \iff \alpha = \bar{x}_n$$

El estimador por método de momentos es $\hat{\alpha} = \bar{X}_n$.

Ejemplo. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \Gamma(\alpha, \beta)$. La varianza de X es

$$\frac{\alpha}{\beta^2} = \mathrm{Var}X = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \frac{\alpha^2}{\beta^2}.$$

Se debe resolver el sistema

$$\begin{cases} \mu_1(\theta) = \dfrac{\alpha}{\beta} = \bar{X}_n = m_1 & (1) \\ \mu_2(\theta) = \dfrac{\alpha(\alpha+1)}{\beta^2} = m_2 & (2) \end{cases}$$

De (1), $\alpha = m_1\beta$. Sustituyendo en (2),

$$m_2 = \frac{m_1\beta(m_1\beta + 1)}{\beta^2} = m_1^2 + \frac{m_1}{\beta} = m_2 \implies m_2 - m_1^2 = \frac{m_1}{\beta}.$$

De esta manera,

$$\hat{\beta} = \frac{m_1}{m_2 - m_1^2}, \quad \hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}$$

Teorema. Si $X_1, X_2, \ldots$ i.i.d con distribución indexada por $\theta \in \mathbb{R}^k$. Suponga que los k momoentos teóricos son finitos $\forall\theta$ y suponga que M es continua. Entonces el estimador por el método de momentos es consistente.

¿Cuál es el comportamiento en la distribución de $\hat{\theta}$ cuando la muestra es grande?

Del teorema del límite central,

$$\frac{\bar{X}_n - \theta}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma} \xrightarrow{d} N(0,1)$$

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\sum \mathrm{Var}(X_1) = \frac{\sigma^2}{n}$$

Implica que se debe multiplicar la media muestral por una constante para hacer la desviación visibile y, con ello, hacer inferencia del parámetro.

Caso general. Si $f(X|\theta)$ es "suficientemente suave" como función de θ, es puede comprobar que la verosimilitud tiende a una normal conforme $n \to \infty$. Es decir,

$$f(X|\theta) \propto \exp\left[\frac{-1}{2\frac{V_n(\theta)}{n}}(\theta - \hat{\theta})\right], \quad n \to \infty \quad (*)$$

donde $\hat{\theta}$ es el MLE de θ.

$$V_n(\theta) \xrightarrow[n\to\infty]{} V_\infty(\theta) < \infty$$

Notas:

1) Si $n \to \infty$ la normal en $(*)$ tiene muchísima precisión y es concentrada alrededor de $\hat{\theta}$.

2) En el caso bayesiano, ninguna previa en θ puede anular el efecto en la verosimilitud cuando $n \to \infty$.

3) Por $(*)$ el MLE se distribute asintóticamente como

$$N\left(\theta, \frac{V_\infty(\theta)}{n}\right),$$

$Var(X_n) \xrightarrow[n\to\infty]{} 0$ y $\mathbb{E}[X_n] = X \implies X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ (confirma que el MLE es consistente).

4.2.2. Método Delta

Si $Y_1, Y_2, \ldots$ es una sucesión de variables aleatorias y sea F^* su c.d.f. continua. Sea $\theta \in \mathbb{R}$ y $\{a_n\}$ sucesión de números positivos tal que $a_n \nearrow \infty$. Suponga que $a_n(Y_n - \theta) \xrightarrow{d} F^*$. Si α es una función tal que $\alpha'(\theta) \neq 0$, entonces

$$\frac{a_n}{\alpha'(\theta)}[\alpha(Y_n) - \alpha(\theta)] \xrightarrow{d} F^*$$

Ejemplo. $X_1, X_2, \ldots$ i.i.d. de variables con media μ y varianza σ^2. Sea α una función tal que $\alpha'(\mu) \neq 0$. Por el T.L.C,

$$\frac{\sqrt{n}}{\sigma}(X_n - \mu) \xrightarrow{d} N(0,1)$$

Entonces por el método Delta

$$\frac{\sqrt{n}}{\sigma\alpha'(\mu)}[\alpha(\bar{X}_n) - \alpha(\mu)] \xrightarrow{d} N(0,1)$$

Si $\alpha(\mu) = \dfrac{1}{\mu}$ $(\mu \neq 0) \implies -\dfrac{1}{\mu^2} = \alpha'(\mu)$. Entonces por el método Delta

$$\frac{\sqrt{n}}{\sigma}\mu^2\left[\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right] \xrightarrow{d} N(0,1)$$

Ejemplo

Si $X_1, X_2 \ldots \overset{i.i.d}{\sim} \text{Exp}(\theta)$. Sea $T_n = \sum X_i \implies \hat{\theta} = \dfrac{1}{\bar{X}_n} = \dfrac{n}{T_n}$.

Note que $\dfrac{1}{\hat{\theta}} = \bar{X}_n$ y

$$\frac{\sqrt{n}}{\sigma}\left[\bar{X}_n - \frac{1}{\theta}\right] \xrightarrow[n\to\infty]{d} N(0,1).$$

La varianza de una exponencial es $\sigma^2 = \text{Var}(X_1) = \dfrac{1}{\theta^2}$, entonces

$$\theta\sqrt{n}\left[\bar{X}_n - \frac{1}{\theta}\right] \xrightarrow[n\to\infty]{d} N(0,1).$$

El método Delta nos dice, con $\alpha(\mu) = \dfrac{1}{\mu}$, $\alpha'(\mu) = -\dfrac{1}{\mu^2}$, el comportamiento asintótico de MLE:

$$
\begin{aligned}
\frac{\theta\sqrt{n}}{\alpha'(1/\theta)}\left[\bar{\alpha}(X_n) - \alpha\left(\frac{1}{\theta}\right)\right] &= \frac{\theta\sqrt{n}}{\frac{1}{1/\theta}}\left[\frac{1}{\bar{X}_n} - \theta\right] \xrightarrow[n\to\infty]{d} N(0,1) \\
&= \frac{\sqrt{n}}{\theta}\left[\frac{1}{\bar{X}_n} - \theta\right] \xrightarrow[n\to\infty]{d} N(0,1)
\end{aligned}
$$

El MLE $\hat{\theta} = \dfrac{1}{\bar{X}_n}$ es asintóticamente normal con media θ y varianza $\dfrac{V_n(\theta)}{n} = \dfrac{\theta^2}{n}$.

Caso bayesiano. Tome una previa conjugada $\theta \sim \Gamma(\alpha, \beta)$, posterior $\theta \sim \Gamma(\alpha + n, \beta + y)$, $y = \sum X_i$. Supongamos que es entero positivo.

$$
\Gamma(\alpha + n, \beta + y) \sim \sum_{i=1}^{\alpha+n} e^{\beta+y}
$$

Por el T.L.C., la distribución posterior $\theta|X$ se distribuye como una normal con media $\dfrac{\alpha + n}{\beta + y}$ y varianza $\dfrac{\alpha + n}{(\beta + y)^2}$. Tomando una previa poco informativa, (α, β son pequeños), la media es

$$
\frac{n}{y} = \frac{1}{\bar{X}_1} = \hat{\theta}_{MLE}
$$

y la varianza

$$
\frac{1}{y^2/n} = \frac{\theta^2}{n} = \frac{V_n(\hat{\theta})}{n}.
$$

4.3. Laboratorio

Suponga que tenemos una tabla con los siguientes datos, los cuales representan la cantidad de giros hacia la derecha en cierta intersección.

```r
(X <- c(rep(0, 14), rep(1, 30), rep(2, 36), rep(3,
    68), rep(4, 43), rep(5, 43), rep(6, 30), rep(7,
    14), rep(8, 10), rep(9, 6), rep(10, 4), rep(11,
    1), rep(12, 1)))
```

```
##   [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  1  1  1  1  1  1  1  1  1  1
##  [26]  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2
##  [51]  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2
##  [76]  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
## [101]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
## [126]  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  4  4
## [151]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
## [176]  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  5  5  5  5  5  5  5  5  5
## [201]  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
## [226]  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6  6
## [251]  6  6  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7  7  7  7  7  7  7  7
## [276]  7  7  7  8  8  8  8  8  8  8  8  8  8  9  9  9  9  9  9 10 10 10 10 11 12
```

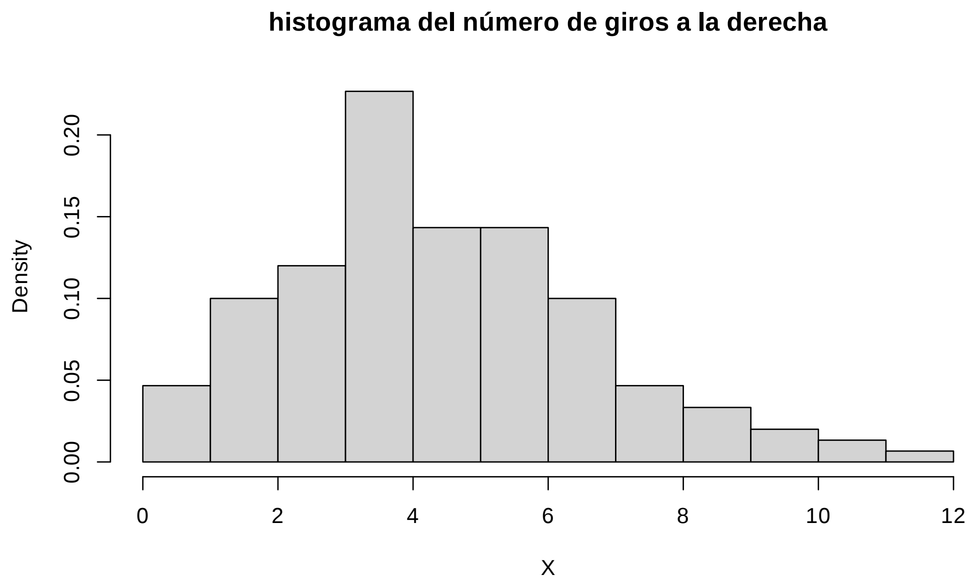
Queremos ajustar esta tabla a una distribución Poisson con función de densidad

$$\mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Se puede probar que teórico de máxima verosimilitud para λ es $\overline{X}$ (Tarea). Queremos estimar este parámetro alternativamente maximizando la función de verosimilitud.

Primero veamos la forma de los datos,

```r
hist(X, main = "histograma del número de giros a la derecha",
    right = FALSE, prob = TRUE)
```



Definamos la función correspondiente a $-\log(\mathbb{P}(X = x))$

```
n <- length(X)
negloglike <- function(lambda) {
    n * lambda - sum(X) * log(lambda) + sum(log(factorial(X)))
}
```

Para encontrar el parámetro deseado, basta minimizar la función `negloglike` usando el la instrucción de optimización no lineal `nlm`.

```
lambda.hat <- nlm(negloglike, p = c(0.5), hessian = TRUE)
```

Aquí el valor `p = c(0.5)` representa un valor inicial de búsqueda y `hessian = TRUE` determina el cálculo explícito de la segunda derivada.

Compare el resultado de `lambda.hat$estimate` con `mean(X)`.

```
lambda.hat$estimate
```

```
## [1] 3.893331
```

```
mean(X)
```

```
## [1] 3.893333
```

Capítulo 5

Estadísticos Suficientes y Criterio de Factorización

5.1. Estadísticos suficientes

Una función de verosimilitud se va a describir a través de un número. El objetivo es buscar un estadístico $T = r(X_1, \ldots, X_n)$ que resuma de manera óptima la información de $X_1, \ldots, X_n$

Definición. Sea $X_1, \ldots, X_n$ una muestra indexada por θ. Sea T un estadístico, suponga que para cada $\theta \in \Omega$ y para cada t en la imagen de T, $X_1 \cdots X_n | T = t$ depende solamente de t y no de θ. Entonces T es suficiente.

5.2. Teorema de Factorización de Fisher

Teorema. Si $X_1, \ldots, X_n$ es una muestra aleatoria de $f(X|\theta)$, el parámetro θ es desconocido. Un estadístico $T = r(X_1, \ldots, X_n)$ es suficiente si y solo si

$$f_n(x|\theta) = u(x)v(r(x), \theta)\ \forall x \in \mathbb{R},\ \forall \theta \in \mathbb{R}.$$

Prueba (Discreta). $f_n(x|\theta) = \mathbb{P}(X = x|\theta)$

"$\Leftarrow$" Sea $A(t) = \{x \in \mathbb{R} | r(x) = t\}$. Para $\theta \in \mathbb{R}$, $x \in A(t)$,

$$
\begin{aligned}
\mathbb{P}(X=x|T=t) &= \frac{\mathbb{P}(X=x\cap T=t)}{\mathbb{P}(T=t)} \\
&= \frac{f_n(x|\theta,T=t)}{\sum_{y\in A(t)} f_n(y|\theta)} \\
&= \frac{u(x)v(r(x),\theta)}{\sum_{y\in A(t)} u(y)v(r(y),\theta)} \\
&= \frac{u(x)v(t,\theta)}{v(t,\theta)\sum_{y\in A(t)} u(y)} (\text{Como } y\in A(t) \text{ entonces } r(y)=t \text{ que es constante.}) \\
&= \frac{u(x)}{\sum_{y\in A(t)} u(y)}
\end{aligned}
$$

no depende de θ.

Si $x\notin A(t) \implies \mathbb{P}(X=x|T=t)=0$ no depende de θ.

"$\Rightarrow$" Si T es un estadístico suficiente, $u(x)=\mathbb{P}(X=x|T=t)$ no depende de θ. Sea $v(t,\theta)=\mathbb{P}_\theta(T=t)$. Entonces

$$f_n(x|\theta)=\mathbb{P}(X=x|\theta)=\frac{\mathbb{P}(X=x|\theta)}{\mathbb{P}(T=t)}\mathbb{P}(T=t)=u(x)v(t,\theta).$$

Consecuencia: $f_n(x|\theta)\propto v(r(x),\theta)$ ($u(x)$ es una constante con respecto a θ). Aplicando el teorema de Bayes,

$$\pi(\theta|x)\propto \pi(\theta)v(r(x),\theta).$$

Corolario. Un estadístico $r(x)$ es suficiente si y solo si no importa cuál previa de θ se use, la posterior depende solamente de $r(x)$ a través de los datos.

Ejemplo. $X_1,\ldots,X_n \sim \text{Poi}(\lambda)$,

$$f_n(x|\theta)=\prod_{i=1}^n \frac{e^{-\lambda}}{x_i!}=\frac{e^{-\lambda n}\lambda^{\sum x_i(\ =r(x))}}{\prod x_i!}=\underbrace{\frac{1}{\prod_{i=1}^n x_i!}}_{u(x)}\underbrace{e^{-\lambda n}\lambda^{r(x)}}_{v(r(x),\lambda)}$$

Si $x_i < 0$ para al menos un i, entonces $f_n(x|\theta) = 0$. Tome $u(x) = 0$. Por el teorema de factorización, $r(x) = \sum x_i$ es un estadístico suficiente para λ.

Ejemplo. $X_1, \ldots, X_n \sim f(x|\theta)$

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otro caso} \end{cases}$$

Verosimilitud: $(0 < x_i < 1 \ \forall i)$

$$f_n(x|\theta) = \theta^n \Big[\underbrace{\prod(x_i)}_{r(x)}\Big]^{\theta-1} = \underbrace{\theta^n (r(x))^{\theta-1}}_{v(r(x),\theta)} \cdot \underbrace{1}_{u(x)}$$

Por el teorema de factorización $r(x) = \prod x_i$ es un estadístico suficiente.,

Ejemplo. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ (σ^2 conocido).

$$\begin{aligned} f_n(x|\theta) &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n} X_i^2}_{r_2(x)} + \frac{\mu}{\sigma^2}\underbrace{\sum_{i=1}^{n} X_i}_{r_1(x)} - \frac{\mu^2 n}{2\sigma^2}\right] \end{aligned}$$

Tome

$$\begin{aligned} u(x) &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} X_i^2\right], \\ v(r_1(x),\mu) &= \exp\left[\frac{\mu}{\sigma^2} r_1(x) - \frac{n\mu^2}{2\sigma^2}\right]. \end{aligned}$$

Por teorema de factorización, $r_1(x) = \sum X_i$ es un estadístico suficiente para μ.

Con σ^2 desconocido, $\theta = (\mu, \sigma^2)$, tome $u(x) = 1$,

$$v(r_1(x), r_2(x), \theta) = (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-r_2(x)}{2\sigma^2} + \frac{\mu r_1(x)}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right]$$

Entonces

$$(r_1(x), r_2(x)) = \left(\sum x_i, \sum x_i^2\right)$$

es un estadístico suficiente para (μ, σ^2).

Ejemplo. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \text{Unif}(0, \theta)$, $\theta > 0$, $f(x|\theta) = 1_{[0,\theta]}(x)\dfrac{1}{\theta}$.

$$f_n(x|\theta) = \prod_{i=1}^{n} 1_{[0,\theta]}(x_i)\left(\frac{1}{\theta}\right)$$

Nota: si al menos uno de los $x_i < 0$ o $x_i > \theta$, $u(x) = 0$ $(f(x|\theta) = 0)$ (Trivial).

Si $0 < x_i < \theta \ \forall i \implies f_n(x|\theta) = 1_{[0,\theta]}(\text{máx}\{x_i\})\left(\dfrac{1}{\theta}\right)^n$.

Si $T = r(x) = X_{(n)} \implies f_n(x|\theta) = u(x)v(r(x), \theta)$, $u(x) = 1$. Por teorema de factorización, $r(x) = x_{(n)}$ es un estadístico suficiente para θ.

5.3. Estadístico suficiente multivariado.

Si $\theta \in \mathbb{R}^k$, $k \geq 1$ se necesita al menos k estadísticos $(T_1, \ldots, T_k)$ para cada $i = 1, \ldots, k$, $T_i = r_i(X_1, \ldots, X_n)$.

Definición. Suponga que para cada $\theta \in \Omega$ y $(t_1, \ldots, t_k) \in \mathbb{R}^k$ valor del estadístico $(T_1, \ldots, T_k)$, la distribución condicional de $X_1, \ldots, X_n$ dado $(T_1, \ldots, T_k) = (t_1, \ldots, t_k)$ no depende de θ, entonces $(T_1, \ldots, T_k)$ es un **estadístico suficiente** para θ.

Criterio de factorización:

$$f_n(x|\theta) = u(x)v(r_1(x), \ldots, r_k(x), \theta) \Leftrightarrow T = (r_1(x), \ldots, r_k(x)) \text{ es suficiente}$$

Si $(T_1, \ldots, T_k)$ es suficiente para θ y si $(T_1', \ldots, T_k') = g(T_1, \ldots, T_k)$ donde g es biyectiva, entonces $(T_1', \ldots, T_k')$ es suficiente para θ.

$$u(x)v(r(x)|\theta) = u(x)v(g^{-1}(g(r(x))), \theta).$$

Ejemplo. Considere los siguiente

$$\begin{aligned}
T_1 &= \sum_{i=1}^{n} X_i \\
T_2 &= \sum_{i=1}^{n} X_i^2 \\
T_1' &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
T_2' &= \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X}_n)^2
\end{aligned}$$

Entonces defina la siguiente función

$$(T_1', T_2') = g(T_1, T_2) = \left(\frac{1}{n}T_1, \frac{1}{n}T_2 - \frac{1}{n^2}T_1^2\right).$$

De la primera entrada,

$$T_1' = \frac{1}{n}T_1 \implies T_1 = nT_1'.$$

De la segunda,

$$\begin{aligned}
T_2' = \frac{1}{n}T_2 - \frac{1}{n^2} &= \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2 \\
&= \frac{1}{n}\sum X_i^2 - 2X_i\bar{X}_n^2 + \bar{X}_n \\
&= \frac{1}{n}\sum (X_i - \bar{X}_n)^2 = \hat{\sigma}_n^2
\end{aligned}$$

Como g es biyectiva entonces $(\bar{X}_n, \sigma_n^2)$ es un estadístico suficiente para (μ, σ^2).

Ejemplo. $X_1, \ldots, X_n \sim \text{Unif}(a, b)$, $a < b$. Encuentre un estadístico suficiente.

1. Si $x_i \leq a$ o $x_i > b$, tome $u(x) = 0$.
2. Si $a < x_i < b \ \forall i$,
   a. $x_i > a \ \forall i \Leftrightarrow x_{(1)} > a$.
   b. $x_i < b \ \forall i \Leftrightarrow x_{(n)} < b$.

La verosimilitud es de la forma

$$f_n(x|(a,b)) = \prod_{i=1}^{n} 1_{[a,b]}(x_i) = \underbrace{\frac{1}{(b-a)^n} 1_{\{(z,w):z>a,w<b\}}(X_{(1)}, X_{(n)})}_{v(r_1(x),r_2(x),(a,b))} \cdot \underbrace{1}_{u(x)}$$

Por teorema de factorización $(r_1(x), r_2(x)) = (X_{(1)}, X_{(n)})$ es un estadístico suficiente para (a,b).

5.4. Estadísticos minimales

Idea: un estadístico suficiente que garantice una partición de $\mathcal{X}$ (espacio muestral) de la manera más simple posible.

Definición (Estadístico de orden). Sean $X_1, \ldots, X_n \overset{i.i.d}{\sim} f$. Al ordenar los datos

$$(Y_1, \ldots, Y_n) = (X_{(1)}, \ldots, X_{(n)}) \text{ tal que } Y_1 < \ldots < Y_n$$

Nota: $(X_{(1)}, \ldots, X_{(n)})$ es un estadístico suficiente de θ.

Ejemplo. $X_1, \ldots, X_n \sim \text{Cauchy}(\alpha)$.

$$f(x) = \frac{1}{\pi}[1 + (x-\alpha)^2]^{-1}, x \in \mathbb{R}$$

Busque un estimador suficiente para $\alpha \in \mathbb{R}$.

$$f_n(x|\alpha) = \prod (x|\alpha) = \frac{1}{\pi}[1 + (x_i - \alpha)^2]^{-1} = \underbrace{\frac{1}{\pi^n}}_{u(x)} \underbrace{\prod_{i=1}^{n} [1 + (x_i - \alpha)^2]^{-1}}_{v(y,\alpha)}$$

donde $y = (X_{(1)}, \ldots, X_{(n)})$ es suficiente para α.

Ejercicio: estime α usando R o usando método de momentos.

Definición. Un estadístico T es **suficiente minimal** si T es suficiente y es función de cualquier otro estadístico suficiente.

Teorema. Si $T = r(X_1, \ldots, X_n)$ es un estadístico suficiente para θ, entonces el MLE $\hat{\theta}$ de θ depende de $X_1, \ldots, X_n$ solamente a través de T. Además, si $\hat{\theta}$ es suficiente entonces $\hat{\theta}$ es minimal.

Prueba. Por teorema de factorización, $f_n(x|\theta) = u(x)v(r(x), \theta)$ de $T = r(x)$ es suficiente y

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, f_n(x|\theta) = \underset{\theta}{\operatorname{argmax}}\, v(r(x), \theta), \quad (\Delta)$$

Como $\hat{\theta} = g(T)$ para cualquier T estadístico suficiente, entonces $\hat{\theta}$ es minimal.

Teorema. Si $T = r(X_1, \ldots, X_n)$ es un estadístico suficiente para θ entonces el estimador bayesiano (bajo una escogencia de L) depende de $X_1, \ldots, X_n$ solamente a través de T (el estimador bayesiano es minimal).

Prueba. Sustituya (Δ) por $\pi(\theta|x) \propto v(r(x), \theta) \cdot \pi(\theta)$. Como cualquier estimador bayesiano depende de $\pi(\theta|x)$, cualquier estimador bayesiano depende e los datos a través de $r(x)$.

5.5. Mejorando estimadores

Idea: ¿Será posible mejorar un estimar que no es suficiente?

¿Existirá otra medida de comparación entre estimadores?

Considere una **función de riesgo o pérdida**

$$R(\theta, \delta) = \mathbb{E}[(\delta(x) - \theta)^2]$$

Si $\delta(x)$ estima una característica de F:

$$R(\theta, \delta) = \mathbb{E}[(\delta(x) - h(\theta))^2] \quad (\Delta\Delta)$$

donde h es la característica.

Nota: la función de riesgo puede ser calculada con una posterior $\pi(\theta|X)$.

Definición.

- Decimos que δ es **inadmisible** si $\exists \delta_0$ (otro estimador) tal que $R(\theta, \delta_0) \leq R(\theta, \delta)\ \forall \theta \in \Omega$. deltadelta

- Decimos que δ_0 **"domina"** a δ en el caso anterior.
- Decimos que δ_0 es admisible si no existe otro estimador que domine a δ_0.
- A $(\Delta\Delta)$ se le llama **MSE** o **error cuadrático medio**.

Teorema (Rao-Blackwell). Sea $\delta(X)$ un estimador y T un estadístico suficiente para θ y sea $\delta_0 = \mathbb{E}[\delta(X)|T]$. Entonces

$$R(\theta, \delta_0) \leq R(\theta, \delta)\ \forall \theta \in \Omega$$

Prueba. Por la desigualdad de Jensen,

$$\mathbb{E}_\theta[(\delta(x) - \theta)^2] \geq (E_\theta[(\delta(x) - \theta)])^2.$$

También,

$$\mathbb{E}[(\delta(x) - \theta)^2 | T] \geq (E[(\delta(x)|T)] - \theta)^2 = (\delta_0(T) - \theta)^2.$$

Entonces,

$$\mathbb{E}[(\delta(x) - \theta)^2] \leq \mathbb{E}[\mathbb{E}[(\delta(x) - \theta)^2 | T]] = \mathbb{E}[(\delta(x) - \theta)^2] = R(\theta, \delta).$$

Nota. Si cambiamos a $R(\theta, \delta) = \mathbb{E}[|\delta(x) - \theta|]$ (error medio absoluto), el resultado anterior es cierto.

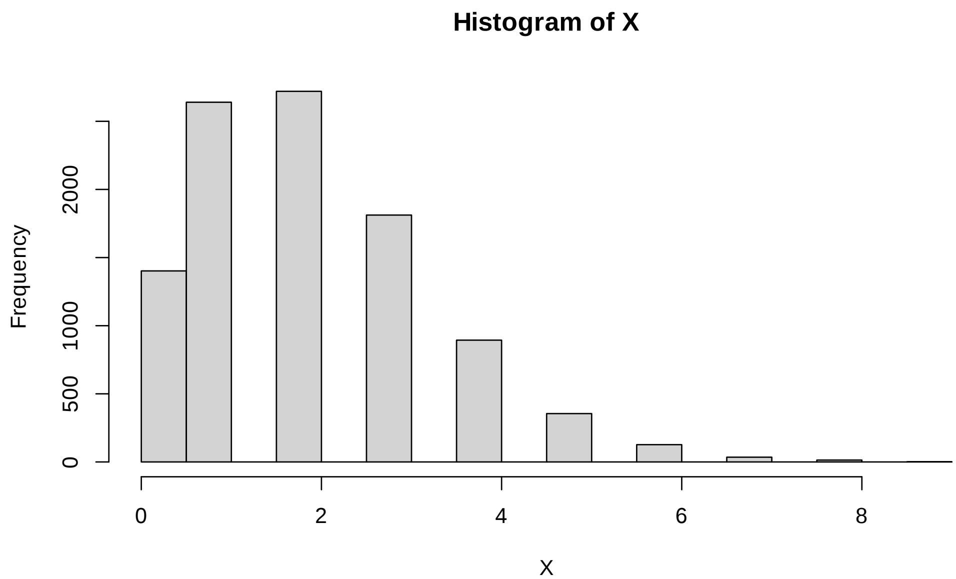
Ejemplo. Sean $X_1, \ldots, X_n \overset{i.i.d}{\sim}$ Poisson(θ) donde θ es la tasa de "visitas" de clientes por hora.

Numericamente podemos hacer el ejemplo con $\theta = 2$ y una muestra de $n = 10000$,

```r
X <- rpois(n = 10000, lambda = 2)
head(X, 20)
```

```
##  [1] 3 2 3 4 1 5 3 0 3 0 0 0 5 0 2 5 0 3 2 2
```

```r
hist(X)
```



A partir de la verosimilitud,

$$f_n(X|\theta) = \frac{e^{-\theta n}\theta^{\sum X_i}}{\prod X_i!}$$

se tiene que $T = \sum X_i$ es un estadístico suficiente para θ.

Sea $Y_i = \begin{cases} 1 & \text{si } X_i = 1 \\ 0 & \text{si } X_i \neq 1 \end{cases}$.

Esta Y se calcula de la forma

```
Y <- X == 1
head(Y, 10)
```

```
##  [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
```

El objetivo es estimar p donde p es la probabilidad de que $X_i = 1$ (solo llegue un cliente por hora). Un estimador de p (MLE) es

$$\delta(x) = \frac{\sum Y_i}{n}$$

```r
(delta <- mean(Y))
```

```
## [1] 0.264
```

¿Es el óptimo?

Calculamos

$$\mathbb{E}[\delta(x)|T] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i|T)$$

Vea que

$$\begin{aligned}\mathbb{E}[Y_i|T=t] = \mathbb{P}(X_i=1|T=t) &= \frac{\mathbb{P}(X_i=1, T=t)}{\mathbb{P}(T=t)}\\ &= \frac{\mathbb{P}(X_i=1, \sum_{j\neq i} X_j = t-1)}{\mathbb{P}(T=t)}\\ &= \frac{\mathbb{P}(X_i=1)\mathbb{P}(\sum_{j\neq i} X_j = t-1)}{\mathbb{P}(T=t)} = \Delta\end{aligned}$$

- $\mathbb{P}(X_i=1) = \theta e^{-\theta}$
- $\mathbb{P}(\sum_{j\neq i} X_j = t-1) = e^{-(n-1)\theta}\frac{((n-1)\theta)^{t-1}}{(t-1)!}$
- $\mathbb{P}(T=t) = e^{-n\theta}\frac{(n\theta)^t}{t!}$

Entonces,

$$\Delta = \frac{\theta e^{-n\theta}\frac{((n-1)\theta)^{t-1}}{(t-1)!}}{e^{-n\theta}\frac{(n\theta)^t}{t!}} = \frac{t}{n}\left(1-\frac{1}{n}\right)^{t-1} = G\left(\frac{t}{n}\right)$$

es el estadístico con MSE mínimo.

```r
T <- sum(X)
n <- length(X)
(delta_0 <- (T/n) * (1 - 1/n)^(T - 1))
```

```
## [1] 0.2707382
```

En este caso δ_0 es mejor que δ bajo una pérdida cuadrática.

Capítulo 6

Distribución muestral de un estadístico

6.1. Distribución muestral

Definición. Suponga que $X_1, \ldots, X_n$ es una muestra con parámetro θ con parámetro θ (desconocido). Sea $T = r(X_1, \ldots, X_n, \theta)$. La distribución de T dado θ se llama **distribución muestral**.

Ejemplo. Si $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. El MLE de μ es

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

La distribución muestral del estadístico $\bar{X}_n$ es

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- $\mathbb{E}[\bar{X}_n] = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbb{E}[X_i] = \dfrac{1}{n}\cdot n\mathbb{E}[X_1] = \mu.$
- $\operatorname{Var}(\bar{X}_n) = \operatorname{Var}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n^2}\cdot n\cdot \operatorname{Var}(X_1) = \dfrac{\sigma^2}{n}.$

Ejemplo. X_i : tiempo de vida de un aparato. $X_1, \ldots, X_n \overset{i.i.d}{\sim} \text{Exp}(\theta)$. La previa de θ es $\Gamma(1,2)$. Solamente observamos $n = 3$. La posterior sería

$$\theta|X \sim \Gamma(1+3, 2+\sum_{i=1}^{3} X_i).$$

El estimador bayesiano, bajo pérdida cuadrática, es

$$\mathbb{E}[\theta|X] = \frac{4}{2+\sum X_i} = \hat{\theta}$$

Problema: estimar $\mathbb{P}(|\hat{\theta}-\theta| < 0{,}1)$.

Note que

$$\begin{aligned}
\mathbb{P}(|\hat{\theta}-\theta| < 0{,}1) &= \mathbb{E}[1_{|\hat{\theta}-\theta|<0{,}1|\theta)}] \\
&= \mathbb{E}[\mathbb{E}[1_{|\hat{\theta}-\theta|<0{,}1|\theta)}|\theta]] \\
&= \mathbb{E}[\mathbb{P}(|\hat{\theta}-\theta| < 0{,}1|\theta)]
\end{aligned}$$

Debemos definir primero cuál es la función de distribución de $\hat{\theta}$.

$$\begin{aligned}
F_{\hat{\theta}}(t|\theta) = \mathbb{P}(\hat{\theta} \le t|\theta) &= \mathbb{P}\left(\frac{4}{2+T} \le t \middle| \theta\right) \\
&= \mathbb{P}\left(2+T \ge \frac{4}{t} \middle| \theta\right) \\
&= \mathbb{P}\left(T \ge \frac{4}{t} - 2 \middle| \theta\right)
\end{aligned}$$

nn

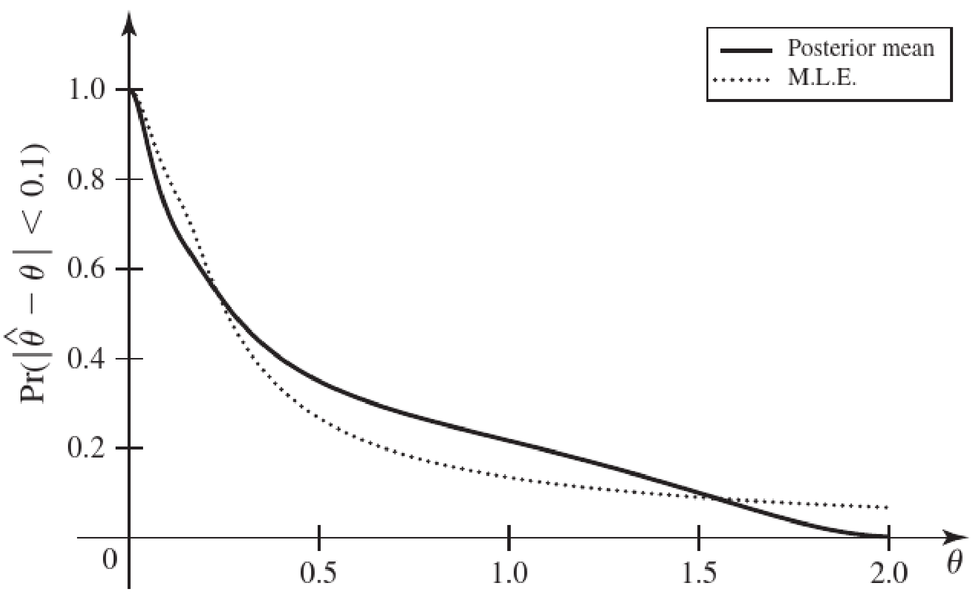
Nota: Recuerde que sumas de exponenciales es una gamma. (Ver teorema 5.7.7)

Entonces $T = \sum_{i=1}^{3} X_i \sim \Gamma(3,\theta)$, por lo que $F(t|\theta) = 1 - G_{\Gamma(3,0)}\left(\frac{4}{t} - 2\right)$. Aqui denotamos como G a la distribución de T.

De esta manera,

$$\begin{aligned}\mathbb{P}[|\hat{\theta}-\theta|<0{,}1|\theta] &= \mathbb{P}[-0{,}1+\theta<\hat{\theta}<0{,}1+\theta|\theta] \\ &= G_{\Gamma(3,\theta)}\left(\frac{4}{-0{,}1+\theta}-2\right)-G_{\Gamma(3,\theta)}\left(\frac{4}{0{,}1+\theta}-2\right)\end{aligned}$$

y se toma la esperanza para estimar la esperanza. Este valor no se puede estimar de forma cerrada, sino que se podría aproximar mediante una simulación



Otra solución es estimar θ usando el MLE $\hat{\theta}=\frac{3}{T}$. Se podría construir esa probabilidad de forma que no dependa de θ.

$$\mathbb{P}\left(\left|\underbrace{\frac{\hat{\theta}_{MLE}}{\theta}-1}_{\text{Cambio relativo}}\right|<0{,}1\middle|\theta\right)=\mathbb{P}\left(\left|\frac{3}{\theta T}-1\right|<0{,}1\middle|\theta\right)=\Delta$$

Si $T\sim\Gamma(3,\theta)\implies\theta T\sim\Gamma(3,1)$.

Por lo tanto,

$$\Delta = \mathbb{P}\left(0{,}9 < \frac{3}{\theta T} < 1{,}1 \middle| \theta\right) = \mathbb{P}\left(\frac{3}{1{,}1} < \theta T < \frac{3}{0{,}9}\right) = 13,4\,\%$$

6.2. Distribución χ^2

Definición. Para $m > 0$ definimos

$$\chi^2_m \sim \Gamma\left(\frac{m}{2}, \frac{1}{2}\right)$$

la distribución **chi-cuadrado** con m grados de libertad.

Propiedades:

- $\mathbb{E}[X] = m$.
- $\mathrm{Var}(X) = 2m$.
- Para $X_i \sim \chi^2_{m_i}$, $i = 1, \ldots, k$, independientes, entonces

$$\sum_{i=1}^{k} X_i \sim \chi^2_{\sum m_i}$$

- Si $X \sim N(0,1) \implies Y = X^2 \sim \chi^2_1$.
- Si $X_i \overset{i.i.d}{\sim} N(0,1) \implies \sum_{i=1}^{m} X_i^2 = \chi^2_m$.

Ejemplo. Si $X_1, \ldots, X_n \sim N(\mu, \sigma^2) \implies Z = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)\ \forall i$.

Entonces

$$\sum Z_i^2 \sim \chi^2_n \implies \sum \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n \quad (*)$$

Además, si μ es conocido y σ^2 desconocido, entonces el MLE de σ^2 es

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

De esta manera, observe que, de $(*)$,

i

$$\frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = n\frac{\hat{\sigma}_0^2}{\sigma^2} \sim \chi_n^2$$

La principal limitación es que μ es conocida. Asuma que también es desconocida. ¿Cuál es la distribución muestral de $(\bar{X}_n, \hat{\sigma}^2)$?

Teorema. Bajo las condiciones anteriores,

1) $\bar{X}_n$ y $\hat{\sigma}_n$ son independientes aunque $\hat{\sigma}_n$ es función de $\bar{X}_n$.

2) La distribución muestral de $\bar{X}_n$ es $N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

3) $n\dfrac{\hat{\sigma}_0^2}{\sigma^2} = \sum_{i=1}^{n} \dfrac{(X_i-\mu)^2}{\sigma^2} \sim \chi_{n-1}^2$.

Nota: De álgebra lineal, recuerde que una matriz $A_{n\times n}$ es ortogonal si cumple que $A^{-1} = A$, $\det(A) = 1$. Si $X, Y \in \mathbb{R}^n$, $AX = Y$, A ortogonal, entonces

$$\|Y\|_2^2 = \|X\|_2^2 \quad (\Delta\Delta)$$

Teorema. Si $X_1, \ldots, X_n \sim N(0,1)$, A es ortogonal $n \times n$ y $Y = AX$ donde $X = (X_1, \ldots, X_n)^T$ entonces $Y_1, \ldots, Y_n \sim N(0,1)$.

Prueba. Ver 8.3.1.

Si $X_1, \ldots, X_n \sim N(0,1)$, use Gram-Schmidt con vector inicial

$$u = \left[\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right]$$

Generamos $A = \begin{bmatrix} u \\ \vdots \end{bmatrix}$. Defina $Y = AX$. Entonces

$$Y_1 = uX = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{n}\bar{X}_n.$$

Por la propiedad $(\Delta\Delta)$, $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2$. Entonces,

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2 \sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Como Y_1^2 y $\sum_{i=2}^n Y_i^2$ son independientes, entonces $\bar{X}_n$ y $\dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ son independientes.

Note que $\sum_{i=2}^n Y_i^2 \sim \chi^2_{n-1}$ ya que $Y_i \overset{i.i.d}{\sim} N(0,1)$.

Si $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, tome $Z_i = \dfrac{X_i - \mu}{\sigma}$ y repita todo lo anterior.

Ejemplo. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ (μ, σ desconocidos). Los MLE son

$$\hat{\mu} = \bar{X}_n, \quad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2\right]^{\frac{1}{2}}.$$

Encuentre n tal que

$$p = \mathbb{P}\left[|\hat{\mu} - \mu| < \frac{6}{5}, |\hat{\sigma} - \sigma| < \frac{6}{5}\right] \geq \frac{1}{2}.$$

Por independencia de $\bar{X}_n$ y $\hat{\sigma}_n^2$,

$$p = \mathbb{P}\left[|\hat{\mu} - \mu| < \frac{\sigma}{5}\right]\mathbb{P}\left[|\hat{\sigma} - \sigma| < \frac{\sigma}{5}\right]$$

Por un lado,

$$\mathbb{P}\left[|\hat{\mu} - \mu| < \frac{6}{5}\right] = \mathbb{P}\left[-\frac{\sqrt{n}}{5} \leq \underbrace{\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma}}_{N(0,1)} < \frac{\sqrt{n}}{5}\right] = \Phi\left(\frac{\sqrt{n}}{5}\right) - \Phi\left(-\frac{\sqrt{n}}{5}\right).$$

Además,

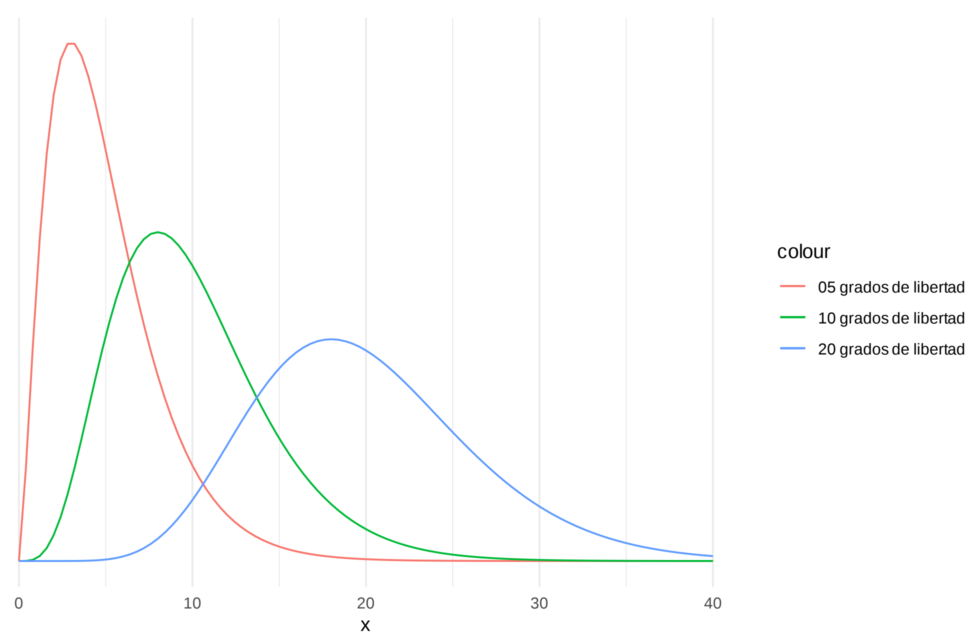
$$\begin{aligned}
\mathbb{P}\left[|\hat{\sigma} - \sigma| < \frac{\sigma}{5}\right] &= \mathbb{P}\left[\frac{4}{5}\frac{\hat{\sigma}}{\sigma} < \frac{6}{5}\right] \\
&= \mathbb{P}\left[0{,}64n\frac{n\hat{\sigma}}{\sigma} < 1{,}44n\right] \\
&= F_{\chi^2_{n-1}}(1{,}44n) - F_{\chi^2_{n-1}}(0{,}64n)
\end{aligned}$$

Estime n de manera que

$$\left[1 - 2\Phi\left(-\frac{\sqrt{n}}{5}\right)\right]\left[F_{\chi^2_{n-1}}(1{,}44n) - F_{\chi^2_{n-1}}(0{,}64n)\right] \geq \frac{1}{2}.$$

Se resuelve numéricamente, y si $n = 21$ se cumple.

```
ggplot(data = data.frame(x = seq(0, 40, length.out = 1000)),
    aes(x)) + stat_function(fun = dchisq, args = list(df = 5),
    aes(color = "05 grados de libertad")) + stat_function(fun = dchisq,
    args = list(df = 10), aes(color = "10 grados de libertad")) +
    stat_function(fun = dchisq, args = list(df = 20),
        aes(color = "20 grados de libertad")) + ylab("") +
    scale_y_continuous(breaks = NULL) + theme_minimal()
```



6.3. Distribución t

Definición. Sea Y y Z dos variables independientes tal que $Y \sim \chi^2_m$ y $Z \sim N(0,1)$. Si

$$X := \frac{Z}{\sqrt{\frac{Y}{m}}},$$

tiene una distribución t **de Student** con m grados de libertad. Tiene como densidad

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad x \in \mathbb{R}.$$

Propiedades:

1) f_X es simétrica.
2) La media de X no existe si $m \leq 1$. Si la media existe, es 0.
3) Las colas de una t de Student son más pesadas que una $N(0,1)$.
4) Si m es entero, los primeros $m-1$ momentos de X existen y no hay momentos de orden superior.
5) Si $m > 2$, $\mathrm{Var}\,(X) = \dfrac{m}{m-2}$.
6) Si $m = 1$, $X \sim$ Cauchy.
7) **Ejercicio**: $f_x(x) \xrightarrow[m\to\infty]{} \Phi(x)$ (sirve como aproximación). La discrepancia de ambas está en la cola y se disipa cuando m es grande.

Recuerde que, por el teorema 8.3.1, $\bar{X}_n$ y $Y = \dfrac{n\hat{\sigma}^2}{\sigma}$ son independientes, con $\bar{X}_n \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ y $Y \sim \chi^2_{n-1}$. Además,

$$Z = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1).$$

Sea

$$T = \frac{Z}{\sqrt{\dfrac{Y}{n-1}}} = \frac{\sqrt{n}\dfrac{\bar{X}_n-\mu}{\sigma}}{\sqrt{\dfrac{\dfrac{n\hat{\sigma}^2}{\sigma^2}}{n-1}}} = \frac{\bar{X}_n-\mu}{\sqrt{\dfrac{\hat{\sigma}}{n-1}}}$$

el cual no depende de σ.

Teorema. Si $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, defina

$$\sigma' = \left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right]^{\frac{1}{2}}.$$

Entonces

$$\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma'} \sim t_{n-1}$$

Nota. $\sigma' = \left(\dfrac{n}{n-1}\right)^{\frac{1}{2}}\hat{\sigma}$ (si n es grande, $\sigma' = \hat{\sigma}$).

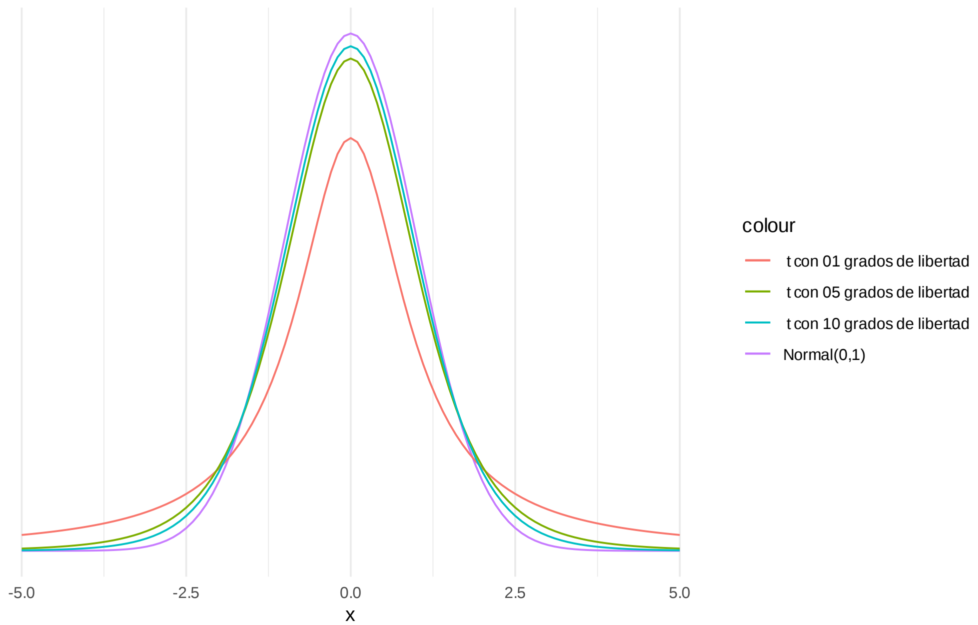
Prueba. Sean

$$S_n^2 = \sum_{i=1}^{n}(X_i-\bar{X}_n)^2, \quad Z = \sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}.$$

Dado que $Y = \dfrac{S_n^2}{\sigma^2} \sim \chi^2_{n-1}$, entonces

$$\begin{aligned} U = \frac{Z}{\sqrt{\dfrac{Y}{n-1}}} &= \frac{\dfrac{\sqrt{n}}{\sigma}(\bar{X}_n-\mu)}{\sqrt{\dfrac{S_n^2}{\sigma^2(n-1)}}} \\ &= \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sqrt{\dfrac{S_n^2}{n-1}}} \\ &= \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma'} \sim t_{n-1}. \end{aligned}$$

```
ggplot(data = data.frame(x = seq(-5, 5, length.out = 1000)),
    aes(x)) + stat_function(fun = dnorm, args = list(mean = 0,
    sd = 1), aes(color = "Normal(0,1)")) + stat_function(fun = dt,
    args = list(df = 1), aes(color = " t con 01 grados de libertad")) +
    stat_function(fun = dt, args = list(df = 5), aes(color = " t con 05 grados
    stat_function(fun = dt, args = list(df = 10), aes(color = " t con 10 grado
    ylab("") + scale_y_continuous(breaks = NULL) +
    theme_minimal()
```



Capítulo 7

Intervalos de confianza

7.1. Intervalos de confianza para la media de una distribución normal

Dado θ un parámetro en $\mathbb{R}$ hemos estudiado procedimientos para encontrar estadísticos $T \in \mathbb{R}$ para estimarlo. La limitación que tenemos acá es que no sabemos que tan aleatorio es T. Entonces podemos sustituir este estadístico T con otros dos estadísticos T_1 y T_2 de modo que sepamos que

$$T_1 \leq \theta \leq T_2$$

En caso que $\theta \in \mathbb{R}^k$ se puede construir un conjunto de estadísticos $T_1, \ldots, T_{k'}$ con $k' = 2k$ tal que

$$\theta \in [T_1, T_2] \times \cdots \times [T_{k'-1}, T_{k'}]$$

7.2. Caso normal.

En el caso normal, $\bar{X}_n$ es un estimador puntual de μ. ¿Será posible encontrar un estimador por intervalos?

Para efectos didácticos, vamos a empezar al revés de lo que usualmente se acostumbra.

Defina $U = \dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma'} \sim t_{n-1}$. Si $c > 0$,

$$\begin{aligned}\mathbb{P}[-c < U < c] &= \mathbb{P}\left[-c < \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma'} < c\right] \\ &= \mathbb{P}\left[-\frac{c\sigma'}{\sqrt{n}} < \bar{X}_n - \mu < \frac{c\sigma'}{\sqrt{n}}\right] \\ &= \mathbb{P}\left[\bar{X}_n - \frac{c\sigma'}{\sqrt{n}} < \mu < \bar{X}_n + \frac{c\sigma'}{\sqrt{n}}\right]\end{aligned}$$

El intervalo

$$T = \left[\bar{X}_n - \frac{c\sigma'}{\sqrt{n}}, \bar{X}_n + \frac{c\sigma'}{\sqrt{n}}\right]$$

es un intervalo aleatorio que "contiene" a μ. Si queremos restringir la probabilidad anterior, tome $\gamma \in (0,1)$:

$$\mathbb{P}(\mu \in T) = \gamma.$$

Para que se cumpla lo anterior, seleccione c tal que

$$\begin{aligned}\gamma = \mathbb{P}(\mu \in T) &= F_{t_{n-1}}(c) - F_{t_{n-1}}(-c) \\ &= F_{t_{n-1}}(c) - [1 - F_{t_{n-1}}(c)] \\ &= 2F_{t_{n-1}} - 1\end{aligned}$$

Entonces

$$\frac{\gamma+1}{2} = F_{t_{n-1}}(c) \implies c = F_{t_{n-1}}^{-1}\left(\frac{\gamma+1}{2}\right).$$

Definición. Si X es una variable aleatoria continua con distribución F (monótona creciente), entonces $x = F^{-1}(p)$ es el **cuantil** de orden p de F (p-cuantil).

El intervalo aleatorio

$$\left[\bar{X}_n - {Ft_{n-1}}^{-1}\left(\frac{\gamma+1}{2}\right)\frac{\sigma'}{\sqrt{n}}, \bar{X}_n - F_{t_{n-1}}^{-1}\left(\frac{\gamma+1}{2}\right)\frac{\sigma'}{\sqrt{n}}\right]$$

contiene a μ con probabilidad γ.

***Definición**. Sea $X = (X_1, \ldots, X_n)$ una muestra con parámetro θ. Sea $g(\theta)$ una característica de la distribución que genera la muestra. Sea $A < B$ dos estadísticos que cumplen $(\forall\theta)$:

$$\mathbb{P}[A < g(\theta) < B] \geq \gamma. \quad (*)$$

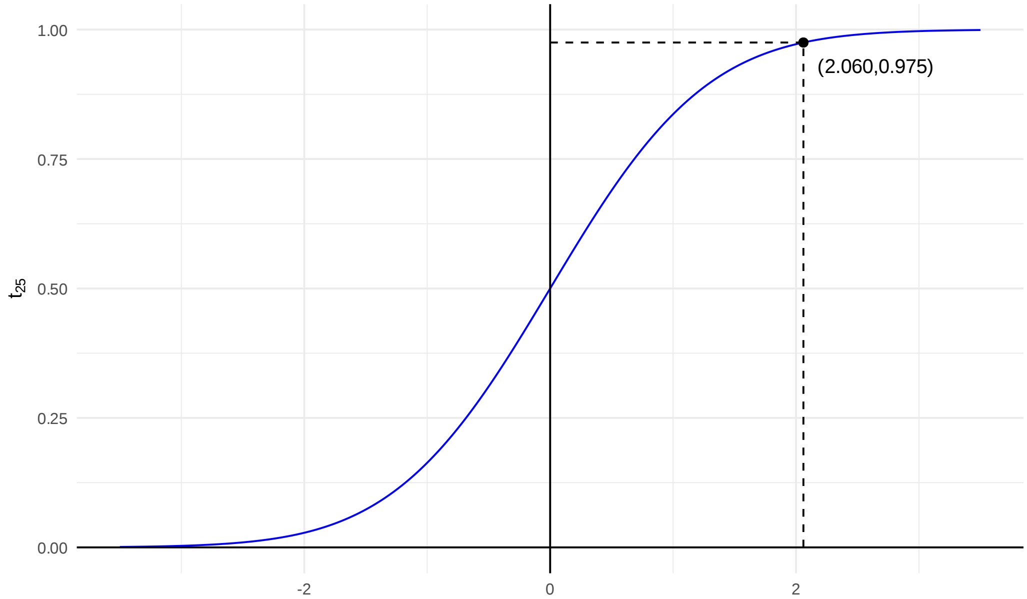
Al intervalo (A, B) le llamamos **intervalo de confianza con coeficiente** γ **para** $g(\theta)$ (intervalo de confianza al 100γ para $g(\theta)$). En el caso que $(*)$ tenga una igualdad, el intervalo es exacto.

Nota. Si observamos X, calculamos $A = a$, $B = b$. Entonces (a, b) es el valor observado de un intervalo de confianza.

Ejemplo. Se mide la lluvia con nubes inyectadas con "sulfato de plata" con $n = 26$ observaciones. Se desea hacer inferencia sobre μ, la cantidad de lluvia media (escala logarítmica). Para $\gamma = 0{,}95$, se calcula

$$c = F_{t_{25}}^{-1}\left(\frac{1+\gamma}{2}\right) = F_{t_{25}}^{-1}(0{,}975) = 2{,}060$$

Note que $\frac{1+\gamma}{2} = 0.975$ y el segundo valor se obtiene de una tabla de valor de la t-student o de la expresión `qt(p = 0.975, df = 26-1)` $= 2.06$



El intervalo de confianza para μ al 95 % es

$$\bar{X}_n \pm \underbrace{0{,}404}_{\frac{2{,}060}{\sqrt{26}}} \sigma'$$

Si $\bar{X}_n = 5{,}134$ y $\sigma' = 1{,}6$ el valor observado del intervalo de confianza al 95 % para μ corresponde a

$$[5{,}134 - 0{,}404 \cdot 1{,}6, 5{,}134 + 0{,}404 \cdot 1{,}6] = [4{,}47, 5{,}78]$$

Interpretación. El intervalo observado $[4{,}48, 5{,}78]$ contiene a μ con un nivel de confianza del 95. Usualmente a $\dfrac{c\sigma'}{\sqrt{n}}$ se le llama **margen de error** (MOE).

Interpretación gráfica. El proceso de construir un intervalo de confianza, quiere decir que si usted repitiera ese experimento muchas veces, el 100γ % (e.g, 95 % o 99 %) de la veces, el intervalo escogido tendría el parámetro real de la población θ.

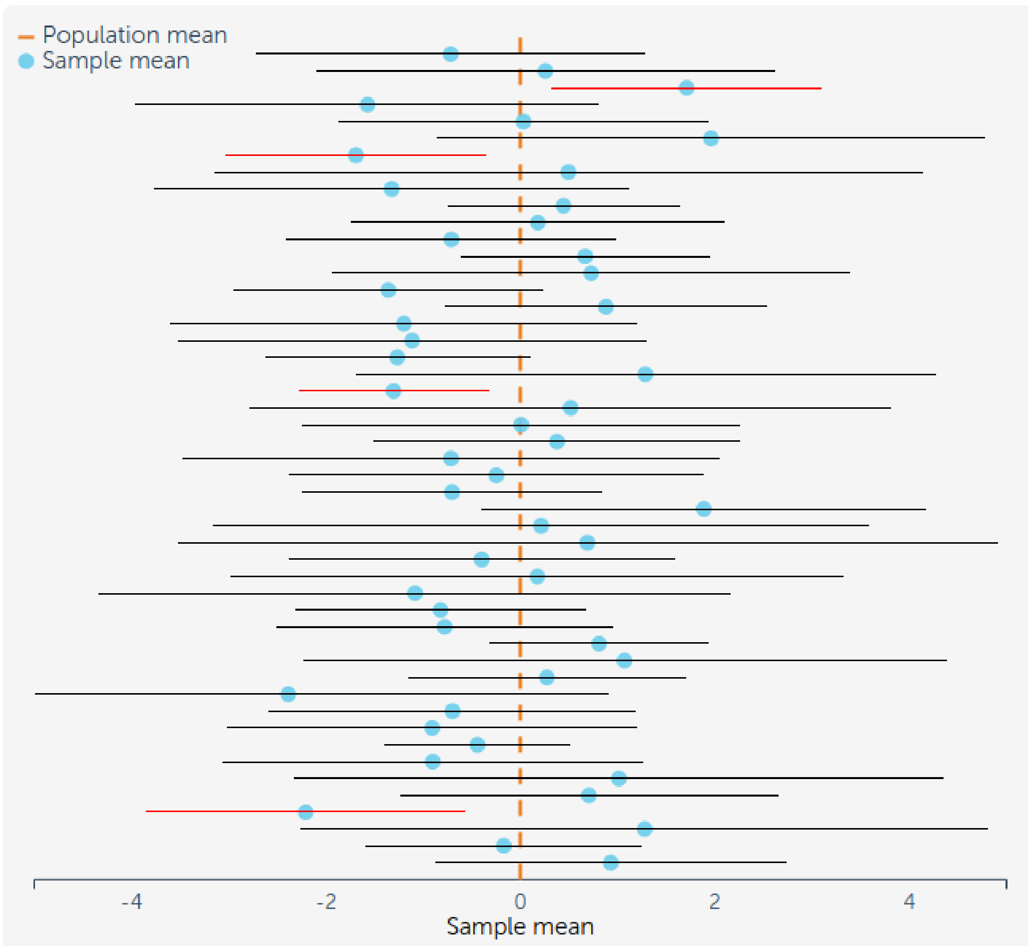


Figura 7.1: Ejemplo interactivo sobre intervalos de confianza. Tomado de (R Psycologist)[https://rpsychologist.com/d3/ci/]

7.3. Intervalos de confianza abiertos

Si γ es el nivel de confianza dado, sea $\gamma_1 < \gamma_2$ tal que $\gamma_2 - \gamma_1 = \gamma$. Sea $U = \frac{\sqrt{n}}{\sigma'}(\bar{X}_n - \mu)$.

Si

$$A = \bar{X}_n + T_{n-1}^{-1}(\gamma_1)\frac{\sigma'}{\sqrt{n}} \text{ y } B = \bar{X}_n + T_{n-1}^{-1}(\gamma_2)\frac{\sigma'}{\sqrt{n}},$$

se cumple que (A, B) es un intervalo de confianza al 100γ ya que

$$\mathbb{P}[\mu \in (A, B)] = \mathbb{P}[T_{n-1}^{-1}(\gamma_1) < U < T_{n-1}^{-1}(\gamma_2)] = \gamma_2 - \gamma_1 = \gamma.$$

Definición (Intervalos de confianza abiertos). Bajo las condiciones anteriores, si A es un estadístico que satisface $\forall\theta$:

$$\mathbb{P}[A < g(\theta)] \geq \gamma,$$

A A se le llama **límite inferior de confianza al** 100γ y al intervalo (A, ∞) es el **intervalo de confianza inferior al** 100γ.

De forma análoga, si B satisface:

$$\mathbb{P}[g(\theta) < B] \geq \gamma,$$

a $(-\infty, B)$ se le llama **intervalo de confianza superior** para $g(\theta)$, con nivel γ. Si hay igualdad, el intervalo es exacto.

Ejemplo. En el caso normal, encuentra B tal que $\mathbb{P}(\mu < B) = \gamma$. Se sabe que

$$F_{t_{n-1}}(c) = \mathbb{P}(U > -c) = \mathbb{P}\left(\frac{\sqrt{n}(\mu - \bar{X}_n)}{\sigma'} < c\right).$$

Entonces

$$\gamma = \mathbb{P}\left(\mu < \bar{X}_n \frac{\sigma'}{\sqrt{n}} c\right).$$

Tome c tal que

$$F_{t_{n-1}}(-c) = \gamma \implies c = -F_{t_{n-1}}(\gamma)$$

Por lo tanto

$$B = \bar{X}_n - \frac{\sigma'}{\sqrt{n}} F_{t_{n-1}}^{-1}(\gamma).$$

7.4. Intervalos de confianza en otros casos

Ejemplo. Tiempos de vida, $n = 3$, $X_i \sim \text{Exp}(\theta)$.

Si $T = \sum_{i=1}^{3} X_i$, $\theta T \sim \Gamma(3, 1)$.

Queremos calcular un intervalo de confianza superior para θ al 100δ (exacto): $\mathbb{P}[\theta < B] = \gamma$.

Si G es la función de distribución de la gamma, sabemos que

$$\gamma = \mathbb{P}[\theta T < G^{-1}(\gamma)] = \mathbb{P}\left[\theta < \frac{G^{-1}(\gamma)}{T}\right].$$

El límite superior es $\dfrac{G^{-1}(\gamma)}{T}$.

```
theta <- 2
X <- rexp(3, rate = theta)
T <- sum(X)
G_inv <- qgamma(p = 0.98, shape = 3, rate = 1)
```

Entonces el intervalo de confianza para este caso es

```
c(0, G_inv/T)
```

```
## [1] 0.000000 5.702774
```

Definición. Sea $X = (X_1, \ldots, X_n)$ una muestra de una distribución F_θ. Sea $V(X, \theta)$ una variable aleatoria cuya distribución no depende de θ. Decimos que V es una **cantidad pivotal**.

Los intervalos de confianza se determinan a partir de un proceso de inversión de la cantidad pivotal.

Encuentre $r(v, x)$ tal que

$$r(V(X,\theta)) = g(\theta) \quad (*)$$

y g es una función cualquiera.

Del ejemplo anterior, $V(X,\theta) = \theta T$,

$$r(V(X,\theta),X) = \frac{V(X,\theta)}{T} = g(\theta) = \theta.$$

Teorema. Bajo las condiciones anteriores, si V existe sea G su c.d.f. y asume que G es continua. Asuma que r en es cierta y asuma que $r(v,x) \nearrow$ en v para cada x. Sea $0 < \gamma < 1$ y $\gamma_2 > \gamma_1$ tal que $\gamma_2 - \gamma_1 = \gamma$. Entonces los extremos del intervalo de confianza para $g(\theta)$ al 100γ son

$$A = r(G^{-1}(\gamma_1), X), \quad B = r(G^{-1}(\gamma_2), X).$$

Ejemplo. $X_1,\ldots,X_n \overset{i.i.d}{\sim} N(\mu,\sigma^2)$. Encuentra A, B tales que $\mathbb{P}[A < \sigma^2 < B] = \gamma$.

Se sabe que

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Tome $V(X,\sigma^2) = \dfrac{n\hat{\sigma}^2}{\sigma^2}$. Entonces

$$\gamma = \mathbb{P}[\chi^2_{n-1,\gamma_1} < V(X,\sigma^2) < \chi^2_{n-1,\gamma_2}]$$

donde $\gamma = \gamma_2 - \gamma_1$. Tome

$$r(v,X) = \frac{\sum(X_i - \bar{X}_n)^2}{v} = \frac{n\hat{\sigma}}{v}.$$

Invirtiendo el intervalo,

$$\gamma = \mathbb{P}\left[\underbrace{\frac{\sum(X_i - \bar{X}_n)^2}{\chi^2_{n-1,\gamma_2}}}_{A} < \sigma^2 < \underbrace{\frac{\sum(X_i - \bar{X}_n)^2}{\chi^2_{n-1,\gamma_1}}}_{B}\right]$$

El IC para σ^2 al 100δ es

$$\left[\frac{\sum(X_i - \bar{X}_n)^2}{\chi^2_{n-1,\gamma_2}}, \frac{\sum(X_i - \bar{X}_n)^2}{\chi^2_{n-1,\gamma_1}}\right].$$

Por ejemplo

```
X <- rnorm(n = 1000, 0, 2)

gamma1 <- 0.025
gamma2 <- 0.975

gamma2 - gamma1
```

```
## [1] 0.95
```

```
(chi2_gamma1 <- qchisq(p = gamma1, df = 1000 - 1))
```

```
## [1] 913.301
```

```
(chi2_gamma2 <- qchisq(p = gamma2, df = 1000 - 1))
```

```
## [1] 1088.487
```

```
(diferencias <- sum((X - mean(X))^2))
```

```
## [1] 4158.789
```

Finalmente el intervalo es

```
c(diferencias/chi2_gamma2, diferencias/chi2_gamma1)
```

```
## [1] 3.820706 4.553580
```

NOTA: Las cantidades pivotales no siempre existen. Esto ocurre principalemente con las distribuciones discretas

7.4.1. Intervalos de confianza aproximados.

Sean $X_1, \ldots, X_n \overset{i.i.d}{\sim} F_\mu$ donde $\mathbb{E}[X_i] = \mu$ y $\mathrm{Var}(X_i) = \sigma^2$ (conocida). Note que

$$D = \mathbb{P}[A < \mu < B] = \mathbb{P}\left[- z_{\frac{1+\gamma}{2}} < \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} < z_{\frac{1+\gamma}{2}} \right] \overset{TLC}{\approx} \gamma.$$

Así,

$$D \underset{n\to\infty}{\approx} \Phi\left(z_{\frac{1+\gamma}{2}}\right) - \Phi\left(-z_{\frac{1+\gamma}{2}}\right) = \gamma.$$

Ejercicio. El intervalo de confianza correspondiente para μ es

$$\bar{X}_n \pm z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}.$$

Considere $U = \dfrac{\bar{X}_n - \mu}{\sigma'/\sqrt{n}}$. U es pivotal, pero no necesariamente una t_{n-1}.

Considere que $(\sigma')^2 = \dfrac{n}{n-1}\hat{\sigma}^2$ y además $\hat{\sigma}^2$ es el MLE de σ^2 y por lo tanto consistente.

$$\hat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^2$$
$$((\sigma')^2 \xrightarrow{\mathbb{P}} \hat{\sigma}^2).$$

Recuerde que si $X_n \xrightarrow{d} Z$ y $Y_n \xrightarrow{\mathbb{P}} a$, entonces $X_n Y_n \xrightarrow{d} aZ$.

Por lo tanto,

$$\underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1)} \cdot \underbrace{\frac{\sigma/\sqrt{n}}{\sigma'/\sqrt{n}}}_{\xrightarrow{\mathbb{P}} 1} \xrightarrow{d} N(0,1)$$

Entonces $U \xrightarrow{d} N(0,1)$.

Como consecuencia

$$\mathbb{P}\left[-z_{\frac{1+\gamma}{2}} < \frac{\bar{X}_n - \mu}{\sigma'/\sqrt{n}} < z_{\frac{1+\gamma}{2}} \right] \overset{TLC}{\approx} \gamma.$$

y el IC aproximado para μ al 100γ

$$\bar{X}_n \pm z_{\frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}}.$$

Ejemplo. Si $X_1, \ldots, X_n \sim \text{Poi}(\theta)$, $\mu = \sigma^2 = \theta$. Por TLC,

$$\sqrt{n} \frac{\bar{X}_n - \theta}{\sqrt{\theta}} \xrightarrow{d} N(0,1).$$

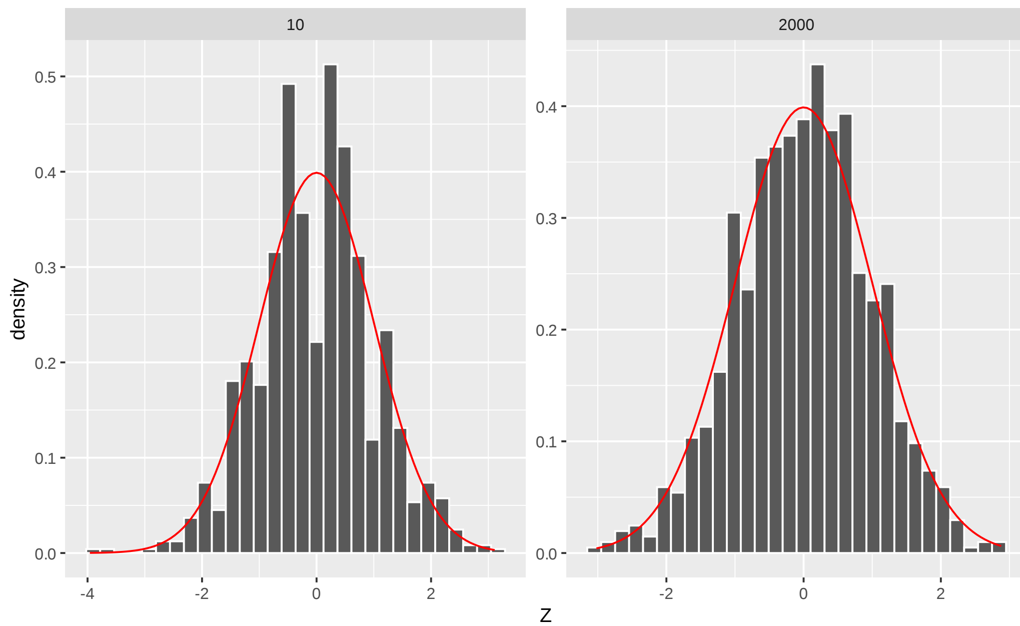
Entonces

$$\mathbb{P}[|\bar{X}_n - \theta| < c] = \mathbb{P}\left[\frac{\sqrt{n}|\bar{X}_n - \theta|}{\sqrt{\theta}} < \frac{c\sqrt{n}}{\sqrt{\theta}} \right] \approx 2\Phi\left(\frac{c\sqrt{n}}{\sqrt{\theta}} \right) - 1.$$

Explicación del fenómeno: En este caso recuerden que $\bar{X}_n$ es una variable aleatoria. Lo que dice el teorema del límite central es que conforme n es grande, la distribución de $\bar{X}_n$ (centrada y escalada apropiadamente) converge a una normal estándar.

```r
Xbar <- data.frame(n = numeric(), Z = numeric())

idx <- rep(x = c(10, 2000), times = 1000)
for (k in 1:length(idx)) {
    muestra <- rpois(n = idx[k], lambda = 5)
    Xbar[k, "Z"] <- sqrt(idx[k]) * (mean(muestra) -
        5)/sqrt(5)
    Xbar[k, "n"] <- idx[k]
}



ggplot(Xbar) + geom_histogram(mapping = aes(x = Z,
    y = ..density..), color = "white") + stat_function(fun = dnorm,
```

```
    args = list(mean = 0, sd = 1), color = "red") +
    facet_wrap(. ~ n, scales = "free")
```



7.4.2. Transformaciones estabilizadoras de la varianza

[Ver página 365 del libro de texto. }

¿Cómo transformar X_n para que tenga varianza constante? Note que en el caso anterior se necesitó saber explícitamente el valor exacto de θ para hacer el ejercicio.

Por el método Delta, la varianza "aproximada" de $\alpha(\bar{X}_n)$ es

$$\left(\frac{\alpha'(\mu)}{a_n}\right)^2 = \left(\frac{\alpha'(\mu)\sigma}{\sqrt{n}}\right)^2 = \frac{\alpha'(\mu)^2\sigma^2(\mu)}{n}.$$

Si se desea que la varianza sea constante con respecto a μ,

$$
\begin{aligned}
&\alpha'(u)^2\sigma^2(\mu) = 1\\
&\implies \alpha'(\mu) = \frac{1}{\sigma(\mu)} \quad (\sigma(\mu) > 0)\\
&\implies \alpha(\mu) = \int_a^\mu \frac{dx}{\sigma(x)} dx
\end{aligned}
$$

donde a es una constante arbitraria que hace la integral finita (y fácil de calcular).

Del ejemplo anterior (Poisson), recuerde que $\sigma^2 = \theta = \mu$, entonces se podría tomar que $\sigma(\mu) = \sqrt{\mu}$ y por lo tanto definimos

$$\alpha(\mu) = \int_0^\mu \frac{dx}{\sqrt{x}} = 2\sqrt{\mu}$$

Por el método Delta,

$$2\bar{X}_n^{\frac{1}{2}} \underset{n \text{ grande}}{\sim} N\left(2\theta^{\frac{1}{2}}, \frac{1}{n}\right)$$

De esta manera

$$\mathbb{P}[|2\bar{X}_n^{\frac{1}{2}} - 2\theta^{\frac{1}{2}}| < c] = \mathbb{P}\left[\frac{|2\bar{X}_n^{\frac{1}{2}} - 2\theta^{\frac{1}{2}}|}{\sqrt{1/n}} < \sqrt{n}c\right] \approx 2\Phi(\sqrt{n}c) - 1$$

Desarrollando,

$$\mathbb{P}[-c + 2\bar{X}_n^{\frac{1}{2}} < 2\theta^{\frac{1}{2}} < c + 2\bar{X}_n^{\frac{1}{2}}] \approx 2\Phi(\sqrt{n}c) - 1$$

Se despeja c tal que

$$\Phi(\sqrt{n}c) = \frac{1+\gamma}{2} \implies c = \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}.$$

El intervalo para $2\theta^{\frac{1}{2}}$ es

$$\left[2\bar{X}_n^{\frac{1}{2}} - \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, 2\bar{X}_n^{\frac{1}{2}} + \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right]$$

```
set.seed(42)
X <- rpois(n = 1000, lambda = 5)
Xbar <- mean(X)
z <- qnorm(p = 0.975)

c(2 * sqrt(Xbar) - 1/sqrt(1000) * z, 2 * sqrt(Xbar) +
    1/sqrt(1000) * z)
```

```
## [1] 4.371529 4.495488
```

Para estimar el IC para θ, vea que si $y = 2x^{\frac{1}{2}} \implies x = \dfrac{y^2}{4}$. Aplicando esta transformación al intervalo anterior, se obtiene

$$\left[\frac{1}{4}\left(2\bar{X}_n^{\frac{1}{2}} - \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)^2, \frac{1}{4}\left(2\bar{X}_n^{\frac{1}{2}} + \frac{1}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)^2\right].$$

```
c((1/4) * (2 * sqrt(Xbar) - 1/sqrt(1000) * z)^2, (1/4) *
    (2 * sqrt(Xbar) + 1/sqrt(1000) * z)^2)
```

```
## [1] 4.777567 5.052354
```

Capítulo 8

Estimación Bayesiana bajo normalidad

8.1. Precisión de una distribución normal

Definición. La precisión τ de una normal se define como $\tau = \dfrac{1}{\sigma^2}$.

Sean $X_1, \ldots, X_n \sim N(\mu, \sigma^2) = N(\mu, \tau)$. Su densidad corresponde a

$$f(x|\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right) \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] = \left(\frac{\tau}{2\pi}\right) \exp\left[-\frac{\tau}{2}(x-\mu)^2\right] = f(x|\mu,\tau).$$

La verosimilitud es

$$f_n(x|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left[-\frac{\tau}{2}\sum_{i=1}^{2}(x_i-\mu)^2\right].$$

La previa de la densidad conjunta es $[\mu,\tau|x] \propto [\mu|\tau]\cdot[\tau]$ y la posterior $[\mu,\tau|x] \propto [\mu|\tau,x]\cdot[\tau|x]$.

Las previas por seleccionar son $[\mu|\tau] \sim$ Normal y $[\tau] \sim$ Gamma.

Teorema. Si $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu,\tau)$, $\mu \in \mathbb{R}$, $\tau > 0$ (precisión) y $\mu \sim N(\mu_0, \lambda_0\tau)$, $\mu \in \mathbb{R}$, $\lambda_0 > 0$, $\tau \sim \Gamma(\alpha_0, \beta_0)$, $\alpha_0, \beta_0 > 0$.

Entonces
$$[\mu, \tau|x] \propto [\mu|\tau, x] \cdot [\tau|x]$$
donde $[\mu|\tau, x] \sim N(\mu_1, \lambda_1\tau)$ con

$$\lambda_1 = \lambda_0 + n, \quad \mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}_n}{\lambda_0 + n},$$

y $[\tau] \sim \Gamma(\alpha_1, \beta_1)$,

$$\alpha_1 = \alpha_0 + \frac{n}{2}, \quad \beta_1 = \beta_0 \frac{1}{2} s_n^2 + \frac{n\lambda_0(\bar{X}_n - \mu_0)^2}{2(\lambda_0 + n)}.$$

Prueba.

- Previa:

$$\begin{aligned}
[\mu, \tau] &\propto [\mu|\tau] \cdot [\tau] \\
&= \tau^{\frac{1}{2}} \exp\left[-\frac{\lambda_0\tau}{2}(\mu - \mu_0) \cdot \tau^{\alpha_0 - 1} e^{-\beta_0\tau} \right] \\
&= \tau^{\alpha_0 - \frac{1}{2}} \exp\left[-\frac{\lambda_0\tau}{2}(\mu - \mu_0)^2 - \beta_0\tau \right]
\end{aligned}$$

- Por Bayes:

$$\begin{aligned}
[\mu, \tau|x] &\propto [\mu, \tau] \cdot [x|\mu, \tau] \\
&\propto [\mu, \tau] \cdot \tau^{\frac{1}{2}} \exp\left[-\frac{\tau}{2} \sum (x_i - \mu)^2 \right] \\
&\propto \tau^{\alpha_0 + \frac{n+1}{2} - 1} \exp\left[-\frac{\tau}{2}(\lambda_0[\mu - \mu_0]^2 + \sum (x_i - \mu)^2 - \beta_0\tau \right]
\end{aligned}$$

Además

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x}_n + \bar{x}_n - \mu)^2 = s_n^2 + n(\bar{x}_n - \mu)^2.$$

Completando cuadrados (queda como ejercicio) se obtiene

$$n(\bar{x}_n-\mu)^2+\lambda_0(\mu-\mu_0)^2=(\lambda_0+n)(\mu-\mu_1)^2+\frac{n\lambda_0(\bar{x}_n-\mu_0)}{\lambda_0+n}.$$

Entonces

$$\sum(x_i-\mu)^2+\lambda_0(\mu-\mu_0)^2=(\underbrace{\lambda_0+n}_{\lambda_1})(\mu-\mu_1)+\underbrace{s_n^2+\frac{n\lambda_0(\bar{x}_n-\mu_0)}{\lambda_0+n}}_{\beta_1}$$

Entonces

$$[\mu,\tau|x]\propto\underbrace{\tau^{\overbrace{\alpha_0+\frac{n}{2}-1}^{\alpha_1}}\exp[-\beta_1\tau]}_{[\tau|x]}\cdot\underbrace{\tau^{\frac{1}{2}}\exp\left[-\frac{\lambda_1\tau}{2}(\mu-\mu_1)^2\right]}_{[\mu|\tau,x]}$$

Por lo que $[\tau|x]\sim\Gamma(\alpha_1,\beta_1)$ y $[\mu|\tau,x]\sim N(\mu_1,\lambda_1\tau)$.

Definición Sean μ,τ dos variables aleatorias. Si $\mu|\tau\sim N(\mu_0,\lambda_0\tau)$, $\tau\sim\Gamma(\alpha_0,\beta_0)$; decimos que

$$[\mu,\tau]\sim\text{Normal - Gamma}(\mu_0,\lambda_0,\alpha_0,\beta_0).$$

- *Conclusión*: la Normal-Gamma conjuga con una verosimilitud normal.
- *Limitación*: μ y τ son independientes. Al combinar con la verosimilitud, cualquier tipo de independencia a nivel de previas se pierde.

Ejemplo. Concentraciones de ácido en queso $X_1,\dots,X_n\sim N(\mu,\tau)$,

$$\mu,\tau\sim\text{Normal-Gamma}(\mu_0=1,\lambda_0=1,\alpha_0=1/2,\beta_0=1/2)$$

Los datos de este experimento son $n=10$, $\bar{x}_n=1{,}379$, $s_n^2=0{,}9663$. Aplicando las fórmulas del teorema anterior:

- $\mu_1=\dfrac{1\cdot1+10\cdot1{,}379}{1+10}=1{,}345.$

- $\lambda_1 = \lambda_0 + n = 1 + 10 = 11.$
- $\alpha_1 = \alpha_0 + \frac{n}{2} = \frac{1}{2} + \frac{10}{2} = 5{,}5.$
- $\beta_1 = \frac{1}{2} + \frac{1}{2} \cdot 0{,}9663 + \frac{10 \cdot 1 \cdot (1{,}379 - 1)^2}{2(1+10)} = 1{,}0484.$

La posterior es

$$[\mu, \tau] \sim \text{Normal - Gamma}(\mu_1, \lambda_1, \alpha_1, \beta_1).$$

Calculamos

$$\begin{aligned}
\mathbb{P}[\sigma > 0{,}3|x] &= \mathbb{P}\left[\sqrt{\frac{1}{\tau}} > 0{,}3 \middle| x\right] \\
&= \mathbb{P}\left[\frac{1}{\tau} > 0{,}3^2 \middle| x\right] \\
&= \mathbb{P}\left[\tau > \frac{1}{0{,}3^2} \middle| x\right] = 0{,}984
\end{aligned}$$

dado que $[\tau|x] \sim \Gamma(\alpha_1, \beta_1) = \Gamma(5{,}5, 1{,}0484)$.

8.2. Distribución marginal de μ

Teorema. Suponga que $[\mu, \tau] \sim$ Normal-Gamma$(\mu_0, \lambda_0, \alpha_0, \beta_0)$. Entonces

$$\left(\frac{\lambda_0 \alpha_0}{\beta}\right)^{\frac{1}{2}} (\mu - \mu_0) \sim t_{2\alpha_0}.$$

Prueba. Vea que $\mu|\tau \sim N(\mu_0, \lambda_0 \tau)$. Se despeja la desviación estándar,

$$\lambda_0 \tau = \frac{1}{\sigma^2} \implies \sigma = (\lambda_0 \tau)^{-\frac{1}{2}}.$$

Entonces

$$Z = \frac{\mu - \mu_0}{\sqrt{\lambda_0 \tau}} \bigg| \tau \sim N(0, 1).$$

La densidad conjunta de (Z, τ) es

$$f(z,\tau) = \pi_2(\tau) \cdot \pi_1(z|\tau)$$

Si $g_1(\mu|\tau)$ es la densidad de $\mu|\tau$, por teorema de cambio de variable

$$f(z,\tau) = \pi_2(\tau) \cdot \underbrace{g_1((\lambda_0\tau)^{-\frac{1}{2}}z + \mu_0|\tau)(\lambda_0\tau)^{-\frac{1}{2}}}_{\phi(z)} = \pi_2\phi(z)$$

Entonces Z y τ son independientes y $Z \sim N(0,1)$.

Sea $Y = 2\beta_0\tau$ y $\tau \sim \Gamma(\alpha_0, \beta_0)$, entonces

$$Y \sim \Gamma\left(\frac{2\alpha_0}{2}, \frac{1}{2} \implies Y \sim \chi^2_{2\alpha_0}\right)$$

y Y es independiente de Z.

Por lo tanto,

$$U = \frac{Z}{\left(\frac{Y}{2\alpha_0}\right)^{\frac{1}{2}}} \sim t_{2\alpha_0}.$$

Observe que

$$U = \frac{(\lambda_0\tau)^{\frac{1}{2}}(\mu - \mu_0)}{\left(\frac{2\beta_0\tau}{2\alpha_0}\right)^{\frac{1}{2}}} = \left(\frac{\lambda_0\alpha_0}{\beta_0}\right)^{\frac{1}{2}}(\mu - \mu_0).$$

Consecuencia:

$$\mu = \left(\frac{\beta_0}{\lambda_0\alpha_0}\right)^{\frac{1}{2}} U + \mu_0, \quad U \sim t_{2\alpha_0}.$$

Propiedades:

- $\mathbb{E}(\mu) = \mu_0 + 0 = \mu_0$.
- $\mathrm{Var}(\mu) = \dfrac{\beta_0}{\alpha_0\lambda_0} \cdot \dfrac{\alpha_0}{\alpha_0 - 1} = \dfrac{\beta_0}{\lambda_0(\alpha_0 - 1)}$.

Ejemplo. $X_1, \ldots, X_{18}$ días de hospitalización en 18 centros de salud.

$$[\mu, \tau] \sim \text{Normal-Gamma}(\mu_0 = 200, \lambda_0 = 2, \alpha_0 = 2, \beta_0 = 6300).$$

Encuentre un intervalo que contenga μ_1 centrado en μ_0 tal que la probabilidad que eso pase sea 0,95.

$$\left(\frac{\alpha_0 \lambda_0}{\beta_0}\right)^{\frac{1}{2}} (\mu - \mu_0) = 0{,}025(\mu - 200) \sim t_{2 \cdot 2} = t_4.$$

Entonces

$$0{,}95 = \mathbb{P}[l < 0{,}025(\mu - 200) < u] = 2F_{t_4}(u) - 1 \implies u = t_{4,0{,}975} = 2{,}776.$$

Así,

$$\mathbb{P}[-2{,}776 < 0{,}025(\mu - 200) < 2{,}776] = 0{,}95$$

y el intervalo es $[89, 311]$.

Con datos: $\bar{X}_n = 182{,}17$ y $s_n^2 = 88678{,}5$. Los hiperparámetros posteriores son $\mu_1 = 183{,}95$, $\lambda_1 = 20$, $\alpha_1 = 11$, $\beta_1 = 50925{,}37$.

Resolvemos el mismo problema:

$$\left(\frac{\alpha_1 \lambda_1}{\beta_1}\right)^{\frac{1}{2}} (\mu - \mu_0) = 0{,}0657(\mu - 183{,}95) \sim t_{2\alpha_1 = 22}.$$

Se busca u:

$$F_{t_{22}}(u|x) = \frac{0{,}95 + 1}{2} \implies u = t_{22,0{,}975} = 2{,}074$$

y

$$0{,}95 = \mathbb{P}[-2{,}074 < 0{,}0657(\mu - 183{,}95) < 2{,}074|x].$$

El **intervalo de credibilidad o predicción** es $[152{,}38, 215{,}52]$.

Si $X_1, \ldots, X_{18} \sim N(\mu, \sigma^2)$, μ, σ^2 fijos y desconocidos.

$$\bar{X}_n + t_{17,0{,}975} \frac{\sigma'}{\sqrt{18}} \text{ al } 95\,\%.$$

El intervalo de confianza es $[146{,}25, 218{,}09]$.

8.3. Efecto de previas no informativas

Considere una **previa no informativa**: $[\mu, \tau] \propto [\mu] \cdot [\tau]$ (supuesto de independencia), con $[\mu] \propto 1$, $\tau = \frac{1}{\sigma^2}$ y $[\sigma] \propto \frac{1}{\sigma}$.

Dado que $\sigma = (\tau)^{-\frac{1}{2}}$, usando el teorema de cambio de variables,

$$\frac{d\theta}{d\tau} = -\frac{1}{2}\tau^{-\frac{3}{2}} \implies \left|\frac{1}{2}\tau^{-\frac{3}{2}}\right| f_\sigma\left(\frac{1}{\tau^{\frac{1}{2}}}\right) = \frac{1}{2}\tau^{-1}.$$

Entonces $[\mu, \tau] \propto \tau^{-1}$.

Ejercicio. Verifique que $[\mu, \tau] \sim$ Normal-Gamma$(\mu_0 = 0, \lambda_0 = 0, \alpha_0 = -1/2, \beta_0 = 0)$.

Usando Bayes, $X_1, \ldots, X_n \sim N(\mu, \tau)$.

$$\begin{aligned}
\pi(\mu, \tau|x) &\propto [\mu, \tau] \cdot [x|\mu, \tau] \\
&= \tau^{-1}(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum(X_i - \mu)^2\right] \\
&\propto \tau^{-1}\tau^{n/2} \exp\left[-\frac{\tau}{2}s_n^2 - \frac{n\tau}{2}(\mu - \bar{X}_n)^2\right] \\
&= \tau^{1/2} \exp\left[-\frac{n\tau}{2}(\mu - \bar{X}_n)^2\right] \cdot \tau^{\frac{n-1}{2}-1} \exp\left[-\frac{s_n^2}{2}\tau\right]
\end{aligned}$$

Entonces

$$\begin{aligned}
\mu|\tau &\sim N(\bar{X}_n, n\tau) \\
\tau|x &\sim \Gamma\left(\frac{n-1}{2}, \frac{s_n^2}{2}\right)
\end{aligned}$$

.

Por lo tanto,

$$\mu, \tau|x \sim \text{Normal-Gamma}(\mu_1 = \bar{X}_n, \lambda_1 = n, \alpha_1 = (n-1)/2, \beta_0 = s_n^2/2).$$

Ejemplo. Tomando $\bar{X}_n = 5{,}134$, $s_n^2 = 63{,}96$ con una previa no informativa para μ, τ. Entonces la posterior es Normal-Gamma con hiperparámetros: $\mu_1 = 5{,}134$, $\lambda_1 = 26$, $\alpha = \dfrac{25}{2}$, $\beta_1 = 31{,}98$. Queremos hacer inferencia sobre μ:

$$
\begin{aligned}
0{,}95 &= \mathbb{P}[-t_{25,0{,}975} < U < t_{25,0{,}975}] \\
&= \mathbb{P}\left[-t_{25,0{,}975} < \left(\frac{26 \cdot 12{,}5}{31{,}98}\right)^{\frac{1}{2}} (\mu - 5{,}134) < t_{25,0{,}975}\right]
\end{aligned}
$$

El intervalo es $[4{,}488, 5{,}78]$.

Calculemos $\mathbb{P}[\mu > 4|x]$. Sea $w = \left(\dfrac{\alpha_1\lambda_1}{\beta_1}\right)^{\frac{1}{2}} = 3{,}188$.

$$
\mathbb{P}[\mu > 4|x] = P[w(\mu - \bar{X}_n) > w(4 - \bar{X}_n)] = 1 - T_{t_{25}}(-3{,}615) = 0{,}9993.
$$

Generalizando:

$$
w = \left(\frac{n(n-1)/2}{s_n^2/2}\right)^{\frac{1}{2}} = \left(\frac{n(n-1)}{s_n^2}\right)^{\frac{1}{2}} = \left(\frac{n}{(\sigma')^2}\right)^{\frac{1}{2}}.
$$

Entonces

$$
\begin{aligned}
\gamma &= \mathbb{P}\left[-t_{n-1,\frac{1+\gamma}{2}} < \left(\frac{n}{(\sigma')^2}\right)^{\frac{1}{2}} (\mu - \bar{X}_n) < t_{n-1,\frac{1+\gamma}{2}}\right] \\
&= \mathbb{P}\left[\bar{X}_n - t_{n-1,\frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1,\frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}}\right].
\end{aligned}
$$

Capítulo 9

Estimación insesgada

9.1. Estimadores insesgados

Definición. Un estimador $\delta(x)$ es un **estimador insesgado** de $g(\theta)$ si $\mathbb{E}_\theta[\delta(x)] = g(\theta)$, $\forall\theta$. A $\mathbb{E}_\theta[\delta(x)] - g(\theta)$ se le denomina **sesgo**.

Ejemplo. Si $X_1, \ldots, X_n \overset{i.i.d}{\sim} F_\theta$, $\mu = \mathbb{E}[X_1]$, entonces

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \mu$$

$\bar{X}_n$ es estimador insesgado de μ.

Ejemplo. $X_1, X_2, X_3 \overset{i.i.d}{\sim} \mathrm{Exp}(\theta)$. El MLE de θ es

$$\hat{\theta} = \frac{3}{T} = \frac{3}{\sum_{i=1}^{3} X_i}$$

¿Será $\hat{\theta}$ un estimador insesgado?

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{3}{T}\right] = 3\mathbb{E}\left[\frac{1}{T}\right], \quad T \sim \Gamma(3, \theta)$$

Como $\dfrac{1}{T} \sim$ Gamma Inversa, se tiene que

$$\mathbb{E}\left[\frac{1}{T}\right] = \frac{\theta}{2} \implies \mathbb{E}[\hat{\theta}] = \frac{3\theta}{2} \neq \theta$$

Por lo que $\hat{\theta}$ es un estimador sesgado, con sesgo

$$\text{sesgo}(\hat{\theta}) = \frac{3\theta}{2} - \theta = \frac{\theta}{2}.$$

Si $U = \dfrac{2\hat{\theta}}{3} = \dfrac{2}{3} \cdot \dfrac{3}{T} = \dfrac{2}{T}$,

$$\mathbb{E}[U] = \frac{2}{3}\mathbb{E}(\hat{\theta}) = \frac{2}{3} \cdot \frac{3}{2}\theta.$$

Entonces U es un estimador insesgado.

Necesitamos encontrar estimadores en donde $\text{Var}(\delta(x)) \to 0$ insesgados. ¿Cómo controlar sesgo y varianza?

$$\begin{aligned}
\text{sesgo}^2(\delta(x)) + \text{Var}(\delta(x)) &= (\mathbb{E}_\theta[\delta(x)] - \theta)^2 + \mathbb{E}[[\delta(x) - \mathbb{E}[\delta(x)]]^2] \\
&= \mathbb{E}[\underbrace{(\mathbb{E}_\theta[\delta(x)] - \theta)^2}_{A^2} + \underbrace{[\delta(x) - \mathbb{E}[\delta(x)]]^2}_{B^2}] \\
&= \mathbb{E}[A^2 + B^2 - 2(\mathbb{E}[\delta(x)] - \theta)\underbrace{(\delta(x) - \mathbb{E}[\delta(x)])}_{=0}] \\
&= \mathbb{E}[(\mathbb{E}[\delta(x)] - \theta - \mathbb{E}[\delta(x)] + \delta(x))^2] \\
&= \mathbb{E}[(\delta(x) - \theta)^2] = MSE(\delta(x))
\end{aligned}$$

Corolario. Si δ tiene varianza finita, entonces

$$MSE_\theta(\delta(x)) = \text{sesgo}^2(\delta(x)) + \text{Var}(\delta(x)).$$

Ejemplo. Comparar $\hat{\theta}$ y $\delta(x) = \dfrac{2}{T}$ en términos del MSE.

Dado que $\text{Var}\left(\dfrac{1}{T}\right) = \dfrac{\theta^2}{4}$, se tiene

- $MSE(\delta(x)) = \text{Var}\left(\dfrac{2}{T}\right) = 4\dfrac{\theta^2}{4} = \theta^2.$
- $MSE(\hat{\theta}) = (\text{sesgo}(\hat{\theta}))^2 + \text{Var}\left(\dfrac{3}{T}\right) = \dfrac{\theta^2}{4} + \dfrac{9\theta^2}{4} = \dfrac{5\theta}{2}.$

$\delta(x)$ es mejor estimador en términos de MSE que el $\hat{\theta}$.

9.2. Estimador insesgado de la varianza

Teorema. Si $X_1, \dots, X_n \sim F_\theta$ con varianza finita y $g(\theta) = \mathrm{Var}(X_1)$ entonces

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$$

es un estimador insesgado de σ^2.

Prueba. Considere que

$$\sum (X_i - \mu)^2 = s_n^2 + n(\bar{X}_n - \mu)^2$$

Entonces

$$\mathbb{E}[\hat{\sigma}_0^2] = \mathbb{E}\left[\frac{s_n^2}{n}\right] = \mathbb{E}\left[\frac{1}{n}\sum (X_i - \mu)^2\right] - \mathbb{E}[(\bar{X}_n - \mu)^2] = \sigma^2 - \frac{\sigma^2}{n} = \left(\frac{n-1}{n}\right)\sigma^2.$$

Para que $\hat{\sigma}_0^2$ sea insesgado,

$$\mathbb{E}\left[\frac{n}{n-1}\hat{\sigma}_0^2\right] = \mathbb{E}[\hat{\sigma}_1] = \sigma^2.$$

Entonces $\hat{\sigma}_1$ es estimador insesgado de σ^2.

Ejemplo. Sean $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathrm{Poi}(\theta)$. $\mathbb{E}(X_i) = \mathrm{Var}(X_i) = \theta$. Estimadores insesgados de θ son:

1) $\bar{X}_n$.
2) $\hat{\sigma}_1^2$.
3) Si $\alpha \in (0,1)$, $T = \alpha \bar{X}_n + (1-\alpha)\hat{\sigma}_1^2$ también es un estimador insesgado (corrige otros problemas).

Ejemplo. (Normal) ¿Cuál estimador tiene menor MSE, $\hat{\sigma}_0^2$ o $\hat{\sigma}_1^2$?

Defina $T_c = c s_n^2$. Si $c = 1/n$, $T_c = \hat{\sigma}_0$ y si $c = 1/(n-1)$, $T_c = \hat{\sigma}_1$. De esta manera,

$$MSE_{\sigma^2}(T_c) = \mathbb{E}[(T_c - \sigma^2)^2] = (\mathbb{E}(T_c) - \sigma^2)^2 + \mathrm{Var}(T_c).$$

- $\mathbb{E}[T_c] = c\mathbb{E}[s_n^2] = c(n-1)\mathbb{E}\left[\frac{s_n^2}{n-1}\right] = c(n-1)\sigma^2.$
- $\text{Var}(T_c) = c^2\text{Var}(s_n) = c^2\text{Var}\left(\sigma^2 \underbrace{\sum \frac{(X_i - \bar{X}_n)}{\sigma^2}}_{\sim \chi^2_{n-1}}\right) = 2c^2\sigma^4(n-1).$

Entonces

$$MSE_{\sigma^2}(T_c) = [c(n-1)\sigma^2 - \sigma^2]^2 + 2c^2\sigma^4(n-1) = [[c(n-1)-1]^2 + 2c^2(n-1)]\sigma^4.$$

Optimizando,

$$\min_c MSE(T_c) = \min_c [(n^2-1)c^2 - 2(n-1)c + 1],$$

se encuentra que $\hat{c} = \dfrac{1}{n+1}$. Así, $T_{\hat{c}} = \dfrac{s_n^2}{n+1}$ es el mejor estimador de σ^2 en el sentido de MSE.

Ejercicio. Compare $\hat{\sigma}_0^2$ y $\hat{\sigma}_1^2$.

9.3. Información de Fisher

¿Cómo cuantificar la información de un estadístico?

Sea $X \sim f(x|\theta)$, $\theta \in \Omega \subset \mathbb{R}$ parámetro fijo.

- *Supuesto 1*: para cada $x \in \mathcal{X}$ (espacio muestral de X) $f(x|\theta) > 0$ $\forall \theta \in \Omega$.
- *Restricción*: la imagen de la variable aleatoria no puede depender de θ.

Ejemplo. Unif$[0,\theta]$, $f(x|\theta) = 1_{(0,\theta)}(x)$. No aplica el supuesto, ya que si $x > \theta$, $f(x|\theta) = 0$.

Definición. Se define la **función Score**:

$$\lambda(x|\theta) := \ln f(x|\theta)$$

cuyas derivadas son

$$\lambda'(x|\theta) = \frac{\partial}{\partial\theta}\ln f(x|\theta)$$
$$\lambda''(x|\theta) = \frac{\partial^2}{\partial\theta^2}\ln f(x|\theta)$$

- *Supuesto 2*: $f(x|\theta)$ es dos veces diferenciable.

Definición. Si X y $f(x|\theta)$ satisfacen los supuestos anteriores, la **información de Fisher** ($I(\theta)$) de X es

$$I(\theta) := \mathbb{E}[(\lambda'(x|\theta))^2]$$

donde la esperanza es integral o suma, dependiendo de X.

Teorema. Bajo las condiciones anteriores, y suponiendo que las dos derivadas de $\int_{\mathcal{X}} f(x|\theta)dx$ con respecto a θ (*Supuesto 3*) se pueden calcular al intercambiar el orden de integración y derivación. Entonces

$$I(\theta) = -\mathbb{E}_\theta[\lambda''(x|\theta)] = \mathrm{Var}[\lambda'(x|\theta)].$$

Prueba:

$$\begin{aligned}
\mathbb{E}[\lambda'(x|\theta)] & \int_{\mathcal{X}} \lambda'(x|\theta)f(x|\theta)dx \\
&= \int_{\mathcal{X}} \frac{f'(x|\theta)}{f(x|\theta)} f(x|\theta)dx \\
&= \int_{\mathcal{X}} f'(x|\theta)dx \\
&= \frac{d}{d\theta}\int_{\mathcal{X}} f(x|\theta)dx \quad \text{por el supuesto} \\
&= \frac{d}{d\theta}1 = 0
\end{aligned}$$

En consecuencia,

$$\mathrm{Var}(\lambda'(x|\theta)) = \mathbb{E}[(\lambda'(x|\theta))^2] - 0 = I(\theta).$$

Además,

$$\lambda''(x|\theta) = \left(\frac{f'(x|\theta)}{f(x|\theta)}\right)' = \frac{f(x|\theta)f''(x|\theta) - f'(x|\theta)^2}{f^2(x|\theta)} = \frac{f''(x|\theta)}{f(x|\theta)} - (\lambda'(x|\theta))^2$$

Note que

$$\begin{aligned}\mathbb{E}\left[\frac{f''(x|\theta)}{f(x|\theta)}\right] &= \int_{\mathcal{X}} \frac{f''(x|\theta)}{f(x|\theta)} f(x|\theta)dx \\ &= \frac{d}{d\theta}\left[\frac{d}{d\theta}\int_{\mathcal{X}} f(x|\theta)dx\right] \\ &= \frac{d}{d\theta}\left[\frac{d}{d\theta}1\right] = 0\end{aligned}$$

Entonces,

$$\mathbb{E}[\lambda''(x|\theta)] = \mathbb{E}\left[\frac{f''(x|\theta)}{f(x|\theta)}\right] - \mathbb{E}[(\lambda'(x|\theta))^2] = -I(\theta).$$

Se concluye, además, que $\lambda'(x|\theta)$ es centrada y su varianza es $I(\theta)$.

Ejemplo. $X \sim \text{Ber}(p)$.

- $f(x|p) = p^x(1-p)^{1-x}$, $x = 0, 1$ satisface supuesto 1.
- $\int_{\mathcal{X}} f(x|p)dx$ " $=$ " $f(0|p) + f(1|p)$ satisface el supuesto 3.

Entonces,

- $\lambda(x|p) = \ln[p^x(1-p)^x] = x\ln p + (1-x)\ln(1-p)$.
- $\lambda'(x|p) = \dfrac{x}{p} - \dfrac{1-x}{1-p}$.
- $\lambda''(x|p) = -\dfrac{x}{p^2} - \dfrac{1-x}{(1-p)^2}$.

De esta manera,

$$I(p) = \mathbb{E}\left[\frac{x}{p} + \frac{1-x}{(1-p)^2}\right] = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)} = \frac{1}{\text{Var}(X)}.$$

Ejemplo. $X \sim N(\mu, \sigma^2)$, μ desconocida, σ^2 conocida.

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

Vea que

$$\begin{aligned}
\frac{d}{du}\int_{\mathbb{R}} f(x|\mu)dx &= \int_{\mathbb{R}} f'(x|\mu)dx \\
&= \int_{\mathbb{R}} -\frac{1}{\sqrt{2\pi\sigma^2}} \frac{2(x-\mu)^2}{2\sigma^2} dx \\
&= -\frac{1}{\sigma} \underbrace{\int_{\mathbb{R}} \frac{u}{\sqrt{2\pi}} e^{-\frac{u}{2}} du}_{\mathbb{E}[N(0,1)]} = 0 \quad \text{usando el cambio de variable } \frac{x-\mu}{\sigma}
\end{aligned}$$

por lo que cumple el tercer supuesto.

Entonces

- $\lambda(x|\mu) = \frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$.
- $\lambda'(x|\mu) = \frac{1}{2\sigma^2} 2(x-\mu) = \frac{x-\mu}{\sigma^2}$.
- $\lambda''(x-\mu) = -\frac{1}{\sigma^2}$.

Por lo que

$$I(\mu) = -\mathbb{E}[\lambda''(x|\mu)] = \frac{1}{\mathrm{Var}(X)}$$

Definición. Suponga que $X = (X_1, \ldots, X_n)$ muestra de $f(x|\theta)$ donde f satisface las condiciones anteriores. Defina $\lambda_n = \ln f_n(x|\theta)$. La información de Fisher de X es

$$I_n(\theta) = \mathbb{E}[(\lambda'(x|\theta))^2] = -\mathbb{E}[\lambda_n''(x|\theta)].$$

Nota. Observe que

$$\lambda_n(x|\theta) = \ln f_n(x|\theta) = \sum_{i=1}^{n} \lambda(X_i|\theta)$$

lo que implica que

$$\lambda_n''(x|\theta) = \sum_{i=1}^{n} (X_i|\theta).$$

De esta forma,

$$I_n(\theta) = -\mathbb{E}[\lambda''(x|\theta)] = -\sum_{i=1}^{n} \mathbb{E}[\lambda''(X_i|\theta)] = nI(\theta).$$

Ejemplo. Clientes que entran a una tienda. Este se modela a partir de un proceso de Poisson. El tiempo de llegada entre cada cliente es independiente y se distribuye como $\mathrm{Exp}(\theta)$. Sea X el tiempo de arribo total de n clientes (n) fijo:

$$X \sim \sum_{i=1}^{n} \mathrm{Exp}(\theta) = \Gamma(n, \theta).$$

Así mismo, sea Y el número de clientes hasta el tiempo t:

$$Y \sim \mathrm{Poi}(\theta t)$$

¿Cuál variable contiene más información de θ?

Para Y,

- $f(y|\theta) = e^{-t\theta} \dfrac{(t\theta)^y}{y!}$.
- $\lambda(y|\theta) = t\theta + y \ln(t\theta) - \ln y!$.
- $\lambda'(y|\theta) = -t + \dfrac{ty}{t\theta}$.
- $\lambda''(y|\theta) = -\dfrac{y}{\theta^2}$.

Entonces,

$$I_Y(\theta) = -\mathbb{E}[\lambda''(y|\theta)] = \frac{\mathbb{E}[Y]}{\theta^2} = \frac{t}{\theta}.$$

Como ejercicio, verifique que $I_X(\theta) = \dfrac{n}{\theta^2}$.

Ambas variables tienen la misma información si

$$I_Y(\theta) = I_X(\theta) \implies \frac{t}{\theta} = \frac{n}{\theta^2} \implies n = t\theta.$$

9.4. Desigualdad de Cramer-Rao

Teorema. Si $X = (X_1, \ldots, X_n)$ muestra de $f(x|\theta)$. Todos los supuestos anteriores son válidos para f. Sea $T = r(X)$ un estadístico con varianza finita. Sea $m(\theta) = \mathbb{E}_\theta[T]$ y asuma que m es diferenciable. Entonces:

$$\mathrm{Var}_\theta(T) \geq \frac{[m'(\theta)]^2}{I_n(\theta)} = \frac{[m'(\theta)]^2}{nI(\theta)}.$$

La igualdad se da si y solo si existen funciones $u(\theta)$ y $v(\theta)$ que solo dependen de θ tales que

$$T = u(\theta)\lambda_n'(x|\theta) + v(\theta).$$

Prueba. Para el caso univariado:

$$\int_{\mathcal{X}} f'(x|\theta)dx = 0.$$

Para el caso multivariado:

$$\begin{aligned}\int_{\mathcal{X}^n} f_n'(x|\theta)dx_1 \cdots dx_n &= \int_{\mathcal{X}^n} [f(x1|\theta) \cdots f(x_n|\theta)]' dx_1 \cdots dx_n \\ &\frac{d}{d\theta}\int_{\mathcal{X}^n} f(x1|\theta) \cdots f(x_n|\theta) dx_1 \cdots dx_n = 0.\end{aligned}$$

Entonces

$$\mathbb{E}[\lambda_n'(X|\theta)] = \int_{\mathcal{X}^n} \frac{f_n'(x|\theta)}{f(x|\theta)} dx_1 \cdots dx_n = 0$$

Por lo tanto,

$$\begin{aligned}\mathrm{Cov}[T, \lambda_n'(X|\theta)] &= \mathbb{E}[T\lambda_n'(X|\theta)] - \mathbb{E}[T] \cdot 0 \\ &= \int_{\mathcal{X}^n} r(x)\frac{f_n'(x|\theta)}{f_n(x|\theta)} f_n(x|\theta) dx_1 \cdots dx_n \\ &= \frac{d}{d\theta}\int_{\mathcal{X}^n} r(x) f_n(x|\theta) dx_1 \cdots dx_n \\ &= \frac{d}{d\theta}\mathbb{E}_\theta[r(X)] = \frac{d}{d\theta}E_\theta[T] = m'(\theta)\end{aligned}$$

Considere el coeficiente de correlación

$$\rho = \frac{\mathrm{Cov}[T, \lambda_n'(X|\theta)]}{\sqrt{\mathrm{Var}(T)}\sqrt{\mathrm{Var}(\lambda_n'(X|\theta))}}.$$

Dado que $|p| \leq 1 \implies \rho^2 \leq 1$, se tiene que

$$\mathrm{Cov}[T, \lambda_n'(X|\theta)]^2 \leq \sqrt{\mathrm{Var}(T)}\sqrt{\mathrm{Var}(\lambda_n'(X|\theta))} \implies [m'(\theta)]^2 \leq \mathrm{Var}(T) I_n(\theta).$$

Entonces $\mathrm{Var}(T) \geq \dfrac{[m'(\theta)]^2}{I_n(\theta)}$.

Caso particular. Si T es un estimador insesgado de θ, entonces $\mathrm{Var}_\theta(T) \geq \dfrac{1}{I_n(\theta)}$.

Ejemplo. $X_1, \ldots, X_n \sim \mathrm{Exp}(\beta)$, $n > 2$.

- $f(x|\beta) = \beta e^{-\beta x}$, $x > 0$.
- $\lambda(x|\beta) = \ln f(x|\beta) = \ln \beta - \beta x$.
- $\lambda'(x|\beta) = \dfrac{1}{\beta} - x$.
- $\lambda'' = -\dfrac{1}{\beta^2}$.

Vea que

$$1 = \int_0^\infty \beta e^{-\beta x} dx = \lim_{u\to\infty} F(u) = \lim_{u\to\infty} [1 - e^{-\beta u}]$$

y el supuesto 3 se puede verificar por la diferenciabilidad de $1 - e^{-\beta u}$.

Así,

$$I(\beta) = -\mathbb{E}[\lambda''(x|\beta)] = \frac{1}{\beta^2}, \quad I_n(\beta) = \frac{n}{\beta^2}.$$

Considere el estadístico $T = \dfrac{n-1}{\sum_{i=1}^n X_i}$ es un estimador insesgado de β. La varianza de T es $\dfrac{\beta^2}{n-2}$.

La cota de Cramer Rao, si T es insesgado, es

$$\frac{1}{I_n(\beta)} = \frac{\beta^2}{n},$$

por lo que T no satisface la cota de Cramer Rao.

Ahora, estime $\theta = \dfrac{1}{\beta} = m(\beta)$. Un estimador insesgado de θ es $T = \bar{X}_n$:

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1] = \frac{1}{\beta} = \theta, \quad \operatorname{Var}(\bar{X}_n) = \frac{\operatorname{Var}(\bar{X}_1)}{n} = \frac{1}{n\beta^2}.$$

La cota de Cramer es

$$\frac{(m'(\beta))^2}{I_n(\beta)} = \frac{(-1/\beta^2)^2}{n/\beta^2} = \frac{\beta^2}{n\beta^4} = \frac{1}{n\beta^2}.$$

$\bar{X}_n$ satisface la cota de Cramer-Rao y además

$$\lambda'(X|\beta) = \frac{n}{\beta} - n\bar{X}_n = \frac{n}{\beta} - nT \implies T = \underbrace{-\frac{1}{n}}_{u(\beta)} \lambda_n'(X|\beta) + \underbrace{\frac{1}{\beta}}_{v(\beta)}.$$

9.5. Estimadores eficientes

Definición. T es un estimador eficiente de su esperanza $m(\theta)$ si su varianza es la cota de CR.

Ejemplo. $X_1, \ldots, X_n \sim \operatorname{Poi}(\theta)$. $\bar{X}_n$ es un estimador eficiente.

- Verosimilitud: $f_n(X|\theta) = e^{n\theta} \dfrac{\theta^{n\bar{X}_n}}{\prod X_i!}$.
- $\lambda_n(X|\theta) = -n\theta + n\bar{X}_n \ln\theta - \ln\prod X_i!$.
- $\lambda_n'(X|\theta) = -n + \dfrac{c\bar{X}_n}{\theta}$.
- $\lambda_n''(X) = -\dfrac{n\bar{X}_n}{\theta^2}$.

Entonces

$$\frac{n}{\theta^2}\mathbb{E}[\bar{X}_n] = \frac{n}{\theta}.$$

La cota de CR es $\dfrac{\theta}{n}$, pero

$$\mathrm{Var}(\bar{X}_n) = \frac{\mathrm{Var}(X_1)}{m} = \frac{\theta}{n}.$$

Por lo que $\bar{X}_n$ es eficiente.

Los otros candidatos para estimar θ

$$\sigma_1^2 = \frac{1}{n-1}s_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2,$$

y

$$\alpha\bar{X}_n + (1-\alpha)\hat{\sigma}_1^2$$

no son lineales con respecto a $\lambda'(X|\theta)$ por lo que tienen mayor varianza que $\bar{X}_n$.

9.6. Comportamiento asintótico del MLE

Teorema. Bajo las condiciones anteriores y si T es un estimador eficiente de $m'(\theta)$ y $m'(\theta) \neq 0$, entonces

$$\frac{1}{\sqrt{CR}}[T - m(\theta)] \xrightarrow{d} N(0,1)$$

Prueba. Recuerde que $\lambda_n'(X|\theta) = \sum_{i=1}^n \lambda'(X_i|\theta)$. Como X es una muestra, $\lambda'(X_i|\theta)$ son i.i.d, y

$$\mathbb{E}[\lambda'(X_i|\theta)] = 0, \quad \mathrm{Var}(\lambda'(X_i|\theta)) = I(\theta).$$

Como T es estimador eficiente de $m(\theta)$,

$$\mathbb{E}[T] = m(\theta), \quad \mathrm{Var}(T) = \frac{(m'(\theta))^2}{nI(\theta)}$$

y existen $u(\theta)$ y $v(\theta)$ tal que

$$T = v(\theta\lambda'(X|\theta)) + v(\theta).$$

- $\mathbb{E}[T] = u(\theta)\mathbb{E}[\lambda'(X|\theta)] + v(\theta) \implies v(\theta) = m(\theta)$.
- $\mathrm{Var}(T) = u^2(\theta)I_n(\theta) \implies v(\theta) = \dfrac{m'(\theta)}{nI(\theta)}$.

Entonces $T = \dfrac{m'(\theta)}{nI(\theta)}\lambda'(X|\theta) + m(\theta)$. Por lo tanto,

$$\left[\frac{nI(\theta)}{m'(\theta)^2}\right]^{\frac{1}{2}}[T - m(\theta)] = \left[\frac{1}{nI(\theta)}\right]^{\frac{1}{2}}\lambda_n'(x|\theta) \xrightarrow[n\to\infty]{} N(0,1).$$

Teorema. Suponga que el MLE $\hat{\theta}_n$ se obtiene al resolver $\lambda'(x|\theta) = 0$. Además, $\lambda''(x|\theta)$ y $\lambda'''(x|\theta)$ existen y las condiciones anteriores son ciertas.

$$[nI(\theta)]^{1/2}(\hat{\theta} - \theta) \to N(0,1).$$

Ejemplo. $X_1, \ldots, X_n \sim N(0, \sigma^2)$, σ desconocida. $\hat{\sigma} = \left[\dfrac{1}{n}s_n^2\right]^{1/2}$ es MLE de σ y $I(\sigma) = \dfrac{2}{\sigma^2}$. Usando el teorema,

$$\sqrt{\frac{2n}{\sigma^2}} \underset{n\to\infty}{\sim} N\left(\sigma, \frac{\sigma^2}{2n}\right).$$

Verifique que

$$\hat{\sigma}_n \pm z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{2n}}$$

es un intervalo de confianza para σ.

Consecuencia en estimación bayesiana. La previa de θ es positiva y diferenciable con respecto a θ. Bajo todas las condiciones anteriores:

$$\theta|X \underset{n\to\infty}{\sim} N\left(\hat{\theta}_n, \frac{1}{nI(\hat{\theta}_n)}\right).$$

Nota: un IC para θ en este caso tiene un error estándar que depende del MLE.

Capítulo 10

Pruebas de hipótesis

10.1. Pruebas de hipótesis

Recordando el ejemplo de las nubes rociadas con químicos en donde log-lluvia $\sim N(\mu, \sigma^2)$, μ, σ desconocidos.

Hipótesis: $\mu > 4$ (nace a partir de una pregunta), es decir, si $\theta = (\mu, \sigma^2)$, ¿$\theta \in \{(\mu, \sigma^2) : \mu > 4\}$?

Para el caso bayesiano, ya calculamos $\mathbb{P}[\mu > 4|X]$. ¿Cómo resolverlo en el caso frecuentista?

Suponga que $\Omega = \Omega_0 \cup \Omega_1$ conjuntos disjuntos tales que

$$
\begin{aligned}
H_0 &: \text{hipótesis en donde } \theta \in \Omega_0. \\
H_1 &: \text{hipótesis en donde } \theta \in \Omega_1.
\end{aligned}
$$

Objetivo. Decidir si H_0 o H_1 es cierto, con los datos disponibles (problema de pruebas de hipótesis).

Definición. H_0 : hipótesis nula. H_1 : hipótesis alternativa. Una vez que se ha realizado una prueba de hipótesis si afirmamos $\theta \in \Omega_1$ decimos que *rechazamos* H_0. Si $\theta \in \Omega_0$, decimos que *no rechazamos* H_0.

Suponga que $X_1, \ldots, X_n \sim f(x|\theta)$, $\theta \in \Omega$, $\Omega = \Omega_0 \cup \Omega_1$ y queremos probar la hipótesis $H_0 : \theta \in \Omega_0$, $H_1 : \theta \in \Omega_1$.

Definición $(i = 0, 1)$

1) Si Ω_i tiene solamente un valor de θ, H_i es una **hipótesis simple**.

2) Si Ω_i tiene más de un valor de θ, H_i es una **hipótesis compuesta**.

3) **Hipótesis compuestas de una cola**. Si $\Omega_0 = (-\infty, \theta_0]$, $H_0 : \theta \geq \theta_0$, $H_1 : \theta > \theta_0$. Si $\Omega_0 = [\theta_0, +\infty)$, $H_0 : \theta \leq \theta_0$, $H_1 : \theta < \theta_0$.

4) Si $H_1 : \theta \neq \theta_0$ y $H_0 : \theta = \theta_0$ es una **hipótesis de 2 colas**.

10.2. Regiones críticas y estadísticas de prueba

Ejemplo. Si $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ desconocido, σ^2 conocido.

Queremos probar $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. La lógica es: rechazamos H_0 si μ está "muy alejado" de μ_0.

Seleccione un número c tal que se rechaza H_0 si $|\bar{X}_n - \mu_0| > c$. En general, suponga que queremos probar las hipótesis $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$.

En general, supónga que queremos probar las hipótesis $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$.

Cuando tenemos una muestra $X_1, \ldots, X_n \sim f(x|\theta)$. Sea $S_0 \subset \mathcal{X}$: conjunto en donde no se rechaza H_0 y $S_1 \subset \mathcal{X}$: conjunto en donde se rechaza H_0.

A S_1 se le llama **región crítica** de la prueba de hipótesis.

Nota. En la mayoría de los casos, la región crítica se define en términos de un estadístico $T = r(x)$.

Definición. Sea $ X$ una muestra aleatoria con distribución $f(x|\theta)$ y $T = r(X)$ un estadístico y $R \subset \mathbb{R}$. Suponga que se puede verificar las hipótesis al afirmar "rechazamos H_0 si $T \in R$", entonces T es un **estadístico** de prueba y R es la **región de rechazo** de la prueba.

Ejemplo. En el caso en que se rechaza H_0 si $|\bar{X}_n| > c$, $T = |\bar{X}_n - \mu_0|$ estadístico de prueba y (c, ∞) es la región de rechazo.

10.3. Función de potencia y tipos de error

Sea δ un procedimiento de prueba (basado en una región crítica o en un estadístico de prueba). Sea $\pi(\theta|\delta)$ (**función de potencia**) la probabilidad de que se rechace H_0 a través de δ para $\theta \in \Omega$.

Si S_1 es la región crítica de δ entonces $\pi(\theta|\delta) = \mathbb{P}(X \in S_1|\theta)$ para $\theta \in \Omega$.

Si δ se describe a través de un estadístico de prueba T con región de rechazo R, entonces $\pi(\theta|\delta) = \mathbb{P}(T \in R|\theta)$ para $\theta \in \Omega$.

Nota. Función de potencia ideal: $\pi(\theta|\delta) = 0$ si $\theta \in \Omega_0$, y $\pi(\theta|\delta) = 1$ si $\theta \in \Omega_1$.

Ejemplo.

- Estadístico de prueba: $T = |\bar{X}_n - \mu_0|$.
- Región de rechazo: $R = (c, \infty)$.

Como $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ desconocido, σ^2 conocido entonces $\bar{X}_n \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

- Función de potencia:

$$\begin{aligned}
\pi(\theta|\delta) = \mathbb{P}[T \in R|\mu] &= \mathbb{P}[|\bar{X}_n - \mu_0| > c|\mu] \\
&= \mathbb{P}[\bar{X}_n > \mu_0 + c|\mu] + \mathbb{P}[\bar{X}_n < \mu_0 - c|\mu] \\
&= \mathbb{P}\left[\sqrt{n}\frac{(\bar{X}_n - \mu)}{\sigma} > \frac{(\mu_0 + c - \mu)}{\sigma}\sqrt{n}\middle|\mu\right] + \mathbb{P}\left[\sqrt{n}\frac{(\bar{X}_n - \mu)}{\sigma} < \frac{(\mu_0 - c - \mu)}{\sigma}\sqrt{n}\middle|\mu\right] \\
&= 1 - \Phi\left(\sqrt{n}\frac{(\mu_0 + c - \mu)}{\sigma}\right) + \Phi\left(\sqrt{n}\frac{(\mu_0 - c - \mu)}{\sigma}\right)
\end{aligned}$$

Ejercicio: graficar la función de potencia $\pi(\mu|\delta)$ para distintos valores de μ.

Tipos de error:

- *Error Tipo I*: error de rechazar H_0 si $\theta \in \Omega_0$.
- *Error Tipo II*: error de no rechazar H_0 si $\theta \in \Omega_1$ en términos de la función de potencia.

Si $\theta \in \Omega_0$: $\pi(\theta|\delta)$ es el error tipo I. Si $\theta \in \Omega_1$: $1 - \pi(\theta|\delta)$ es el error tipo II.

Recuerde que el objetivo es hacer $\pi(\theta|\delta)$ pequeño cuando $\theta \in \Omega_0$. También se requiere que $\pi(\theta|\delta)$ sea grande cuando $\theta \in \Omega_1$. Una forma de alcanzar ese balance es seleccionar $\alpha_0 \in (0,1)$ tal que

$$\pi(\theta|\delta) \leq \alpha_0 \ \forall \theta \in \Omega_0 \quad (*)$$

y entre todas las pruebas que cumplan $(*)$ se selecciona aquella que maximice la potencia para $\theta \in \Omega_1$.

Otra forma es minimizar;

$$w_1 \cdot \text{Error I } + w_2 \cdot \text{Error II};$$

w_1, w_2 constantes.

Nota. Bajo la primera solución se produce una asimetría entre las hipótesis, ya que resulta difícil (o muy costoso) que ambas condiciones se cumplan. Por lo general, se le da más énfasis a $(*)$, por lo que se trata de controlar el error más serio (Error tipo I).

Definición. Una prueba que satisface $(*)$ se llama una **prueba de nivel** α_0 y decimos que la prueba está a un **nivel de significancia** α_0. Además el tamaño $\alpha(\delta)$ de una prueba δ se define como:

$$\alpha(\delta) = \sup_{\theta \in \Omega} \pi(\theta|\delta).$$

Corolario. Una prueba δ es una prueba de nivel α_0 si y solo si su tamaño es a lo sumo α_0 ($\alpha(\delta) \leq \alpha_0$).

Ejemplo. Suponga $X_1, \ldots, X_n \sim \text{Unif}(0,\theta)$, $\theta > 0$ desconocido. Se quiere probar las siguientes hipótesis:

$$H_0 : 3 \leq \theta \leq 4 \quad H_1 : \theta < 3 \text{ o } \theta > 4.$$

El MLE de θ es $Y_n = X_{(n)}$. Si n es grande, Y_n es muy cercano a θ.

La prueba δ no rechaza H_0 si $2{,}9 < Y_n < 4$ y rechaza H_0 si $Y_n \geq 4$ o $Y_n \leq 2{,}9$. Entonces $R = (-\infty, 2{,}9] \cup [4, +\infty)$ y la función de potencia

$$\pi(\theta|\delta) = \mathbb{P}[Y_n \leq 2{,}9|\theta] + \mathbb{P}[Y_n \geq 4|\theta]$$

$\pi(\theta|\delta)$ se calcula en varios casos:

- Si $\theta \leq 2{,}9 \implies \mathbb{P}[Y_n \leq 2{,}9|\theta] = 1$ y $\mathbb{P}[Y_n \geq 4|\theta] = 0$.
- Si $2{,}9 < \theta \leq 4 \implies \mathbb{P}[Y_n \leq 2{,}9|\theta] = 1 = \prod_{i=1}^n \mathbb{P}[X_i \leq 2{,}9|\theta] = \left(\dfrac{2{,}9}{\theta}\right)^n$ y $\mathbb{P}[Y_n \geq 4|\theta] = 0$.
- Si $\theta > 4 \implies \mathbb{P}[Y_n \leq 2{,}9|\theta] = \left(\dfrac{2{,}9}{\theta}\right)^n$ y $\mathbb{P}[Y_n \geq 4|\theta] = 1 - \displaystyle\prod_{i=1}^n \mathbb{P}[X_i < 4|\theta] = 1 - \left(\dfrac{4}{\theta}\right)^n$.

Entonces

$$\pi(\theta|\delta) = \begin{cases} 1 & \text{si } \theta \leq 2{,}9 \\ \left(\dfrac{2{,}9}{\theta}\right)^n & \text{si } 2{,}9 < \theta \leq 4 \\ 1 + \left(\dfrac{2{,}9}{\theta}\right)^n - \left(\dfrac{4}{\theta}\right)^n & \text{si } \theta > 4 \end{cases}$$

Note, además, que el tamaño de prueba es

$$\alpha(\delta) = \sup_{3 \leq \theta \leq 4} \pi(\theta|\delta) = \sup_{3 \leq \theta \leq 4} \left(\frac{2{,}9}{\theta}\right)^n = \left(\frac{2{,}9}{3}\right)^n.$$

Si $n = 68 \implies \alpha(\delta) = \left(\dfrac{2{,}9}{3}\right)^{68} = 0{,}0997$.

Entonces δ es una prueba con nivel de significancia $\alpha_0 \geq 0{,}0997$.

¿Cómo diseñar una prueba para que tenga un cierto nivel de significancia?

Suponga que queremos probar $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$. Sea T un estadístico de prueba y suponga que si $T \geq c$, c constante, rechazamos H_0.

Si queremos que nuestra prueba tenga nivel de significancia α_0 entonces:

$$\pi(\theta|\delta) = \mathbb{P}(T \geq c|\theta) \text{ y } \sup_{\theta\in\Omega_0} \mathbb{P}[T \geq c|\theta] \leq \alpha_0 \quad (*)$$

Note que $\pi(\theta|\delta)$ es función no-creciente de c, entonces $(*)$ se cumple para valores grandes de c, si $\theta \in \Omega_0$. Si $\theta \in \Omega_1$, debemos escoger c pequeño para maximizar $\pi(\theta|\delta)$.

Ejemplo. En el caso normal, donde $H_0 : \mu = \mu_0$ y rechazamos H_0 si $|\bar{X}_n - \mu_0| \geq c$. Entonces:

$$\sup_{\theta\in\Omega_0} \mathbb{P}[T \geq c|\theta] = \mathbb{P}_{\mu_0}[|\bar{X}_n - \mu_0| \geq c] \geq \alpha_0.$$

Como bajo H_0: $Y = X_n - \mu_0 \sim N\left(0, \dfrac{\sigma^2}{n}\right)$, entonces podemos encontrar c tal que

$$\mathbb{P}[|\bar{X}_n - \mu_0| \geq c] = \alpha_0,$$

y cualquier c mayor va a cumplir $(*)$.

De esta manera el problema se convierte en encontrar c^* tal que $\mathbb{P}[|Z| > c^*] = \alpha_0$, donde $Z = \dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$.

Dado que $\mathbb{P}[|Z| \geq c^*] = 2[1 - \Phi(c^*)] = \alpha_0$, despejando se obtiene

$$\Phi(c^*) = 1 - \frac{\alpha_0}{2} \implies c^* = z_{1-\frac{\alpha_0}{2}}.$$

Procedimiento: rechazamos H_0 si

$$|Z| = \left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| \geq z_{1-\frac{\alpha_0}{2}}.$$

Ejemplo. $X_1, \ldots, X_n \sim \text{Ber}(p)$.

$$H_0 : p \leq p_0 \text{ vs } H_1 : p > p_0$$

Sea $Y = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$. Rechazo H_0 si $Y \leq c$.

El error tipo I es

$$\mathbb{P}[Y \geq x|p] = \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=0}^{n} \binom{n}{y} \underbrace{\left(\frac{p}{1-p}\right)^y (1-p)^n}_{g(p)}$$

$g(p)$ es monótona con respecto a p.Entonces

$$\sup_{p \leq p_0} \mathbb{P}[Y \geq c|p] = \mathbb{P}[Y \geq c|p_0] \leq \alpha_0.$$

Si $n = 10$, $p_0 = 0{,}3$, $\alpha_0 = 10\,\%$, entonces

c	0	1	2	3	4	5
$\mathbb{P}[Y \geq c|p_0]$	1	0.97	0.85	0.62	0.15	0.05

Para que el tamaño sea menor que $10\,\%$ seleccione $c > 5$. Si $c \in [5, 6]$ entonces el nivel de significancia es a lo sumo 0,15 y la prueba no cambia (ya que Y es una variable discreta).

Procedimiento: rechazamos $H_0 : p = 0{,}3$ si $Y \geq c$, $c \in [5, 6]$ con un nivel de significancia de $10\,\%$ a lo sumo.

10.4. Valor p

Restricción. El procedimiento de prueba depende de α_0.

Pregunta. ¿Será posible construir un estadístico que resuma el grado de evidencia en los datos en contra de H_0?

Respuesta. Cualquier procedimiento usa las siguientes dos fuentes:

1) El valor observado del estadístico de prueba.

2) Todos los valores de α_0 en donde rechazamos la nula.

Ejemplo (Normal). Si $Z = 2{,}78$, entonces se rechaza $H_0 : \mu = \mu_0$ si $|Z| =$

$2{,}78 > z_{1-\frac{\alpha_0}{2}}$, para cualquier α_0. Entonces,

$$\Phi(2{,}78) > 1 - d\frac{\alpha_0}{2} \implies \alpha_0 \geq 0{,}0054$$

que es el valor observado de significancia.

Definición. El **valor-**p es el nivel más pequeño de significancia en donde rechazaríamos H_0 bajo los datos observados.

Nota. El valor-p es un estadístico.

- Si valor-$p < \alpha_0$, rechazo H_0. (El valor-p es muy pequeño).
- Si valor-$p > \alpha_0$, no rechazo H_0. (El valor-p es muy grande).

Cálculo del valor-p

- Región de rechazo: $T \geq c$.
- Decisión de rechazo (*ejercicio*): para cada t, rechazamos H_0 si $T \geq t$ con $t \geq F^{-1}(1-\alpha_0)$, F distribución de T.

Entonces

$$F(t) \geq 1 - \alpha_0 \implies \alpha_0 \geq \mathbb{P}_\theta[T \geq t] \implies \alpha_0 \geq \sup_{\theta\in\Omega} P_\theta[T \geq t]$$

El tamaño de la prueba es $c = t$.

Ejemplo. Retomando el ejemplo con las variables aleatorias Bernouilli, rechazamos $H_0 : p \leq p_0$ si $Y \geq c$. Así,

$$\text{valor-}p = \sup_{p\in\Omega} P_p[Y \geq y] = P_p[Y \geq y]$$

Si $p_0 = 0{,}3, n = 10, y = 6$, el valor-$p$ es pbinom$(y = 6, p_0 = 3, n = 10) = 0{,}0473$.

10.5. Dualidad entre pruebas de hipótesis y regiones de confianza

Teorema. Sea $X = (X_1, \ldots, X_n)$ una muestra con distribución F_θ. Sea $g(\theta)$ una función tal que para cada valor g_0 de $g(\theta)$, existe una prueba con nivel

α_0 de las hipótesis:

$$H_{0,g_0} : g(\theta) = g_0 \text{ vs } H_{1,g_0} : g(\theta) \neq g_0.$$

Defina para cada $x \in X$

$$\omega(x) = \{g_0 : \delta_{g_0} \text{ no rechaza } H_{0,g_0} \text{ si } X = x\} \quad (*)$$

Sea $\gamma = 1 - \alpha_0$. Entonces

$$\mathbb{P}[g(\theta_0) \in \omega(x) | \theta = \theta_0] \geq \gamma, \ \forall \theta_0 \in \Omega.$$

Definición. Si $\omega(x)$ satisface $(*)$ $\forall \theta_0 \in \Omega$, entonces $\omega(x)$ es un **conjunto de confianza** con coeficiente γ donde $\gamma = 1 - \alpha_0$.

Teorema. Bajo las condiciones anteriores, si $\omega(x)$ es un conjunto de confianza para g_0, entonces construimos δ_{g_0}: no rechazo H_{0,g_0} si y solo si $g_0 \in \omega(X)$, entonces δ_{g_0} es una prueba con nivel $\alpha_0 = 1 - \gamma$ para H_{0,g_0}.

Ejemplo. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$ (desconocidos). En este caso $g(\theta) = \mu$. El intervalo de confianza con nivel *gamma* es

$$\bar{X}_n \pm t_{n-1, \frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}}.$$

La hipótesis de interés corresponde a

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0.$$

Por los teoremas anteriores, H_0 se rechaza si μ_0 no está en el IC, es decir, si y solo si

$$\mu_0 > \bar{X}_n + t_{n-1, \frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}} \text{ o } \mu_0 < \bar{X}_n - t_{n-1, \frac{1+\gamma}{2}} \frac{\sigma'}{\sqrt{n}},$$

que se puede resumir como

$$\left| \frac{\bar{X}_n - \mu_0}{\sigma'/\sqrt{n}} \right| > t_{n-1, 1-\frac{\alpha}{2}}.$$

Ejemplo. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, μ desconocido, σ^2 conocido. Construta un intervalo de confianza con nivel γ a partir de

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0.$$

Rechazamos H_0 si

$$\left| \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{1-\frac{\alpha_0}{2}}.$$

al nivel α_0. Usando los teoremas anteriores, una región de confianza con nivel $\gamma = 1 - \alpha_0$ satisface:

$$\mu \in \left\{ \left| \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right| < z_{1-\frac{\alpha_0}{2}} \right\} = \omega(x)$$

Por tanto,

$$\begin{aligned} \left| \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right| &\Leftrightarrow -\frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha_0}{2}} < \bar{X}_n - \mu < \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha_0}{2}} \\ &= \Leftrightarrow \bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha_0}{2}} < \mu < \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha_0}{2}} \end{aligned}$$

que es el IC con nivel γ para μ.

10.5.1. Dualidad en pruebas unilaterales

Si $X = (X_1, \ldots, X_n)$ es una muestra según F_θ y $g(\theta)$ es una función de variable real, suponga que para cada $g_0 \in I_m(g)$ existe una prueba δ_{g_0} con nivel α_0 de las hipótesis anteriores. Si

$$\omega(x) = \{g_0 : \delta_{g_0} \text{ no rechaza } H_{0,g_0} \text{ si } X = x\}$$

y si $\gamma = 1 - \alpha_0$, entonces $\omega(x)$ es una región de confianza para $g(\theta)$ con nivel γ.

Ejemplo (Bernoulli).

$$H_0 : p \leq p_0 \text{ vs } H_1 : p > p_0.$$

El criterio de rechazo al nivel α_0 es

$$Y = \sum_{i=1}^{n} X_i \geq c(p_0)$$

donde

$$\sup_{p \leq p_0} \mathbb{P}_p[Y \geq c] = \mathbb{P}_{p_0}[Y \geq c] \leq \alpha_0.$$

Entonces

$$\omega(x) = \{p_0 : Y < c(p_0)\} = \{p_0 : \text{valor-}p > \alpha_0\}.$$

Si $n = 10$, $Y = 6$, $\alpha_0 = 10\,\%$,

$$\omega(x) = \{p_0 : P_{p_0}[Y > 6] > 0{,}1\}.$$

Numéricamente, si $p_0 > 35{,}42\,\% \implies p_0 \in \omega(x)$, entonces $\omega(x) = (0{,}3542, 1]$ si $\alpha_0 = 10\,\%$ y es un IC para p_0 con nivel de $90\,\%$.

Ejemplo. $X = (X_1, \ldots, X_n) \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$ desconocido. Queremos probar

$$H_0 : \mu \leq \mu_0 \text{ vs } H_1 : \mu > \mu_0.$$

Por dualidad, basta con conocer un IC unilateral para μ:

$$\left(\bar{X}_n - t_{n-1,\gamma}\frac{\sigma'}{\sqrt{n}}, \infty\right).$$

Rechazamos H_0 si

$$\mu_0 \leq \bar{X}_n - t_{n-1,\gamma}\frac{\sigma'}{\sqrt{n}} \Leftrightarrow T = \frac{\bar{X}_n - \mu_0}{\sigma'/\sqrt{n}} \geq t_{n-1,\gamma}$$

(rechazando en la cola derecha de T).

10.5.2. Pruebas de cociente de verosimilitud (LRT)

Si $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_0^c = \Omega_1$. El **estadístico LRT** se define como

$$\Lambda(x) = \frac{\sup_{\theta \in \Omega_0} f_n(x|\theta)}{\sup_{\theta \in \Omega} f_n(x|\theta)}.$$

Una prueba de cociente de verosimilitud rechaza H_0 si $\Lambda(x) \leq k$, para una constante k.

Ejemplo. Supongamos que se observa Y el número de éxitos en el experimento Bernoulli(θ) con tamaño de muestra n.

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

- Verosimilitud: $f(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$.
- $\Omega_0 = \{\theta_0\}$, $\Omega_1 = [0,1] \setminus \{\theta_0\}$.
- Numerador: $f(y|\theta_0)$.
- Denominador: $f(y|\bar{y}) = \displaystyle\binom{n}{y} \bar{y}^y (1-\bar{y})^{n-y}$.

$$\Lambda(y) = \frac{f(y|\theta_0)}{f(y|\bar{y})} = \left(\frac{n\theta_0}{y}\right)^y \left(\frac{n(1-\theta_0)}{n-y}\right)^{n-y}, \quad y = 0, \ldots, n.$$

Si $n = 10$, $\theta_0 = 0{,}3$, $y = 6$, $\alpha_0 = 0{,}05$.

y	0	1	2	3	4	5	6	7	8	9	10
$\Lambda(y)$	0,028	0,312	0,773	1	0,797	0,418	0,147	0,034	0,005	3×10^{-4}	6×10^{-6}
$\mathbb{P}[Y = y|\theta = 0{,}3]$	0,028	0,121	0,233	0,267	0,200	0,103	0,037	0,009	0,001	1×10^{-4}	6×10^{-6}

Rechazamos H_0 con nivel $\alpha_0 = 0{,}05$ en $y \in \{10, 9, 8, 7, 0\}$ y $k \in [0{,}028, 0{,}147)$

si rechazo cuando $\Lambda(y) \leq k$. El tamaño de prueba es

$$\mathbb{P}_{0,3}[\text{Rechazo}] = \mathbb{P}_{0,3}[Y \in \{10, 9, 8, 7, 0\}] = 0{,}039.$$

Teorema. Sea Ω un abierto en $\mathbb{R}^p$ y suponga que H_0 especifica k coordenadas de θ, igualándolas a valores fijos. Asuma que H_0 es cierto y que todas las condiciones de regularidad de θ son ciertas.

$$-2 \ln \Lambda(x) \xrightarrow[H_0]{d} \chi^2_k.$$

Ejemplo. Del caso anterior, $k = 1$, $\alpha_0 = 5\,\%$. Rechazamos H_0:

$$-2 \ln \Lambda(y) > \chi^2_{1,1-0{,}05} = F^{-1}_{\chi^2_1}(0{,}95) = 3{,}841.$$

Rechazamos H_0 bajo la misma región del ejemplo anterior.

Capítulo 11

Pruebas con hipótesis simples

11.1. Hipótesis simples

Ejemplo. Sea X_i el tiempo de servicio del cliente #i en el sistema. El supuesto de independencia es poco válida.

Si no hay independencia y si $X_1, \ldots, X_n$ son observados

$$f_1(x) = \begin{cases} \dfrac{2(n!)}{(2 + \sum X_i)^{n+1}} & X_i > 0 \\ 0 & \text{si no} \end{cases}$$

Se asume que $X_i \sim \text{Exp}(1/2)$.

$$f_0(x) = \begin{cases} \dfrac{1}{2^n} e^{-\frac{1}{2}\sum X_i} & \text{si } X_i > 0 \\ 0 & \text{si no} \end{cases}$$

Si $H_0 : f = f_0$ vs $H_1 : f = f_1$, ¿Cuál hipótesis es cierta?

Podemos redefinir las hipótesis si $\Omega = \{\theta_0, \theta_1\}$ donde si $\theta = \theta_i$, seleccionamos $f = f_i$ y se prueba $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Asuma que $X_1, \ldots, X_n \sim f_i(X)$ donde se pueden tener dos posibilidades $(i = 0, 1)$. Sea $\Omega = \{\theta_0, \theta_1\}$ donde θ_1 es el parámetro que indica a cuál densidad se selecciona como hipótesis.

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1$$

Si δ es un procedimiento de prueba, se denota los errores tipo I y II:

- $\alpha(\delta) = \mathbb{P}[\text{Rechazo } H_0|\theta = \theta_0]$.
- $\beta(\delta) = \mathbb{P}[\text{No rechazo } H_0|\theta = \theta_1]$.

Del ejemplo anterior, si se asume (o se comprueba) que f_1 da probabilidades más altas que f_0 entonces un criterio de rechazo puede ser $X_1 > 4$ si se observa solo $n = 1$.

En este caso,

$$\alpha(\delta) = \mathbb{P}[X_1 > 4|\theta = \theta_0] = 1 - (1 - e^{-0{,}5\cdot 4}) = 0{,}135$$

$$\beta(\delta) = \mathbb{P}[X_1 < 4|\theta = \theta_1] = \int_0^4 \frac{2}{(2 + x_1)^2} dx_1 = 0{,}667.$$

Objetivo. Encontrar un procedimiento de prueba δ tal que $\alpha(\delta)$ y $\beta(\delta)$ se reduzcan simultáneamente o al menos si $a, b > 0$, que $a\alpha(\delta) + b\beta(\delta)$ sea mínimo.

Teorema. Sea δ^* un procedimiento de prueba tal que no se rechaza $H_0 : \theta = \theta_0$ si $af_0(x) > bf_1(x)$ y se rechaza H_0 si $af_0(x) < bf_1(x)$. Si $af_0(x) = bf_1(x)$ se puede rechazar o no H_0. Entonces para cualquier otro procedimiento de prueba δ

$$a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta).$$

Prueba. Caso discreto solamente.

Sea S_1 región crítica de δ (procedimiento arbitrario).

$$
\begin{aligned}
a\alpha(\delta) + b\beta(\delta) &= a \sum_{x \in S_1} f_0(x) + b \sum_{x \in S_1^c} f_1(x) \\
&= a \sum_{x \in S_1} f_0(x) + b\left[1 - \sum_{x \in S_1} f_1(x)\right] \\
&= b + \sum_{x \in S_1} (af_0 - bf_1(x))
\end{aligned}
$$

y lo anterior es mínimo si $af_0(x) - bf_1(x) < 0$ en toda la muestra y no hay punto en donde $af_0(x) - bf_1(x) > 0$.

Definición. **Cociente de verosimilitud**:

$$
\frac{f_1(x)}{f_0(x)}.
$$

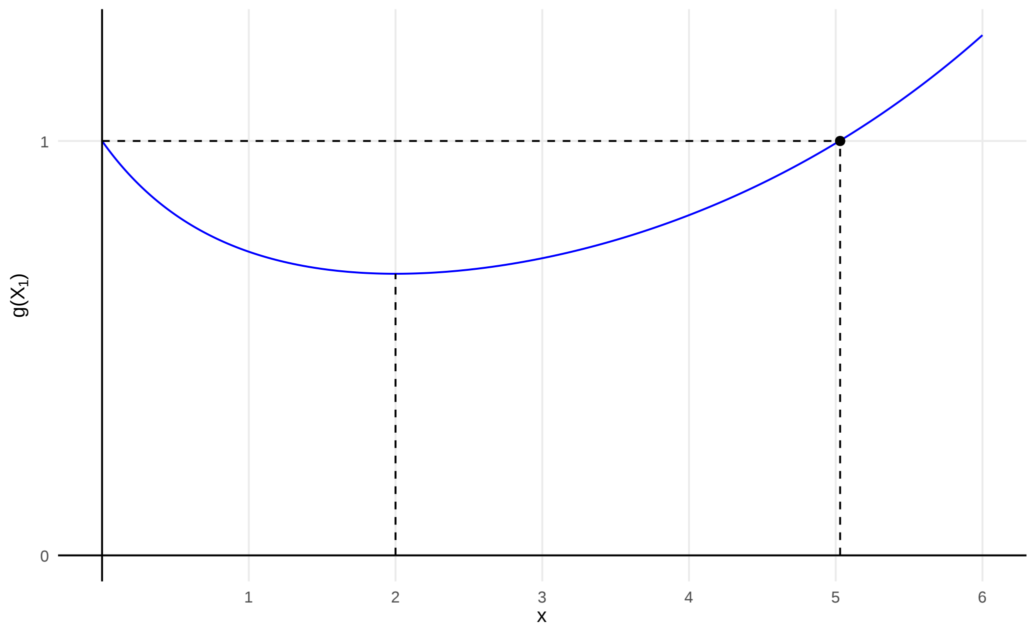
Note que el estadístico LR está relacionado con el anterior de la siguiente forma:

$$
\Lambda(x) = \frac{f_0(x)}{\text{máx}\{f_0(x), f_1(x)\}} = \frac{\sup_{\Omega_0} f(x|\theta)}{\sup_{\Omega} f(x|\theta)}.
$$

Corolario. Bajo las condiciones del teorema anterior, si $a, b > 0$ entonces la prueba δ para la cual $a\alpha(\delta) + b\beta(\delta)$ es un mínimo rechaza H_0 si el cociente de verosimilitud es mayor a $\dfrac{a}{b}$.

Del ejemplo de tiempo de servicio, en lugar de rechazar $H_0 : \theta = \theta_0$ si $X_1 > 4$ tome $a > b$ en el corolario anterior y rechace H_0 si

$$
\frac{f_1(x)}{f_0(x)} > 1 \Leftrightarrow \frac{4}{(2 + X_1)^2} \exp\left(\frac{X_1}{2}\right) > 1 \quad (*)
$$



Entonces $(*)$ es cierto si $X_1 > c$. Se puede comprobar numéricamente que $c \approx 5{,}03$.

Por lo tanto, rechazamos H_0 si $X_1 > 5{,}03$.

Criterio de Neyman-Pearson. Encontrar un procedimiento δ tal que

1) $\alpha(\delta) \leq \alpha_0$ (α_0: nivel de significancia).

2) $\beta(\delta)$ es mínimo.

Lema de Neyman-Pearson. Suponga que δ' es un procedimiento de prueba que no rechaza H_0 si $f_1(x) < kf_0(x)$ rechaza H_0. Si $f_1(x) > kf_0(x)$ y decide cualquiera de los dos si $f_1(x) = kf_0(x)$ para $k > 0$. Si δ es otro procedimiento de prueba tal que $\alpha(\delta) \leq \alpha(\delta')$, entonces $\beta(\delta) \geq \beta(\delta')$. Si $\alpha(\delta) < \alpha(\delta')$, $\beta(\delta) > \beta(\delta')$.

Prueba. Tome $a = k$ y $b = 1$ en el corolario y teoremas anteriores. Como

$$k\alpha(\delta') + \beta(\delta') \leq k\alpha(\delta') + \beta(\delta'),$$

entonces

$$\alpha(\delta) \leq \alpha(\delta') \implies \beta(\delta') \geq \beta(\delta').$$

Consecuencia. Si queremos encontrar una prueba δ' que satisfaga el criterio de Neyman-Pearson, debemos encontrar k tal que $\alpha(\delta') = \alpha_0$, y se rechace H_0 si $f_1(x) > kf_0(x) \Leftrightarrow \dfrac{f_0(x)}{f_1(x)} < k^{-1}$.

Ejemplo. Suponga que $X_1, \ldots, X_n \sim N(\theta, 1)$ y se quiere probar $H_0 : \theta = 0$ vs $H_1 : \theta = 1$ usando una prueba según el criterio de Neyman-Pearson con $\alpha = 0{,}05$.

Note que

- $f_0(x) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum X_i^2\right]$.
- $f_1(x) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum (X_i - 1)^2\right]$.

Entonces

$$\begin{aligned}\frac{f_1(x)}{f_0(x)} &= \exp\left[-\frac{1}{2}\sum (X_i^2 - 2X_i + 1 - X_1^2)\right] \\ &= \exp\left[n\bar{X}_n - \frac{n}{2}\right] = \exp\left[n\left(\bar{X}_n - \frac{1}{2}\right)\right]\end{aligned}$$

Rechazamos H_0 si

$$\frac{f_1(x)}{f_0(x)} = \exp\left[n\left(\bar{X}_n - \frac{1}{2}\right)\right] > k \Leftrightarrow \bar{X}_n > \underbrace{\frac{1}{2} + \frac{\ln k}{n}}_{k'}.$$

Entonces buscamos k' tal que

$$\mathbb{P}[\bar{X}_n > k' | \theta = 0] = 0{,}05 \Leftrightarrow \mathbb{P}\left[\frac{\bar{X}_n}{1/\sqrt{n}} > \frac{k'}{1/\sqrt{n}}\middle| \theta = 0\right] = 0{,}05$$

Despejando,

$$k'\sqrt{n} = z_{0{,}95} \implies k' = \frac{z_{0{,}95}}{\sqrt{n}}.$$

Entonces, entre todas las pruebas en donde $\alpha(\delta) \leq 0{,}05$, la que tiene el error tipo II más pequeño es la que rechaza H_0 si

$$\bar{X}_n > \frac{z_{0,95}}{\sqrt{n}} = \frac{1{,}645}{\sqrt{n}}.$$

El error tipo II de esta prueba sería

$$\begin{aligned}\beta(\delta') &= \mathbb{P}[\bar{X}_n < 1{,}645 n^{-1/2} | \theta = 1] \\ &= \mathbb{P}\left[Z < \frac{1{,}645 n^{-1/2} - 1}{n^{-1/2}}\right] = \Phi(1{,}645 - n^{1/2})\end{aligned}$$

Si $n = 9$, por ejemplo, $\beta(\delta') = \Phi(1{,}645 - 3) = 0{,}0877$.

Ejemplo. $X_1, \ldots, X_n \sim \mathrm{Ber}(p)$ y considere las hipótesis

$$H_0 : p = 0{,}2 \text{ vs } H_1 : p = 0{,}4.$$

Queremos encontrar un procedimiento de prueba en donde $\alpha(\delta) = 0{,}05$ y $\beta(\delta)$ es mínimo. Sea $y = \sum X_i$.

$$\begin{aligned}f_0(x) &= 0{,}2^y 0{,}8^{n-y} \\ f_1(x) &= 0{,}4^y 0{,}6^{n-y}\end{aligned}$$

Entonces el cociente de verosimilitud es

$$\frac{f_1(x)}{f_0(x)} = \left(\frac{3}{4}\right)^n \left(\frac{8}{3}\right)^y$$

y se rechaza H_0 si

$$\begin{aligned}\frac{f_1(x)}{f_0(x)} > k &\Leftrightarrow -n \ln\left(\frac{4}{3}\right) + y \ln\left(\frac{8}{3}\right) > \ln k \\ &\Leftrightarrow y > \frac{\ln k + n\ln(4/3)}{\ln(8/3)} = k'.\end{aligned}$$

Entonces basta con encontrar k' tal que

$$\mathbb{P}(Y > k'|p = 0{,}2) = 0{,}05,$$

pero como Y es una variable discreta (Binomial), no es posible encontrar ese k'. Note que

$$\mathbb{P}(Y > 4|p = 0{,}2) = 0{,}0328$$

$$\mathbb{P}(Y > 3|p = 0{,}2) = 0{,}1209$$

Por lo tanto, se puede especificar una prueba con nivel 0.05, $\alpha(\delta) = 0{,}0328$ y potencia mínima si $Y > 4$ como región de rechazo.

11.2. Prueba t

Suponga que $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, con (μ, σ^2) desconocidos, y considere las siguientes hipótesis:

$$H_0 : \mu \leq \mu_0 \text{ vs } H_1 : \mu > \mu_0.$$

Recuerde que si $U = \dfrac{\bar{X}_n - \mu_0}{\sigma'/\sqrt{n}}$, entonces la prueba rechaza H_0 si $U \geq c$. Si $\mu = \mu_0$ entonces $U \sim t_{n-1}$.

Si $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$, entonces se rechaza H_0 si $U \leq c$.

Definición. Considere las hipótesis $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$. Decimos que una prueba de hipótesis δ es **insesgada** si $\forall\theta \in \Omega_0$ y $\forall\theta \in \Omega_1$:

$$\pi(\theta|\delta) \leq \pi(\theta'|\delta).$$

11.2.1. Propiedades de las pruebas t

Teorema. Sea $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Sea U definido anteriormente, $c = t_{n-1,1-\alpha_0}$. Sea δ la prueba que rechaza H_0 si $U \geq c$. Entonces

I) $\pi(\mu, \sigma^2|\delta) = \alpha_0$ si $\mu = \mu_0$.

II) $\pi(\mu, \sigma^2|\delta) < \alpha_0$ si $\mu > \mu_0$.

III) $\pi(\mu, \sigma^2|\delta) > \alpha_0$ si $\mu > \mu_0$.

IV) $\pi(\mu, \sigma^2|\delta) \to 0$ si $\mu \to -\infty$.

V) $\pi(\mu, \sigma^2|\delta) \to 1$ si $\mu \to +\infty$.

Entonces, la prueba tiene tamaño α_0 y es insesgada.

Prueba. Ver en el libro.

En el caso en donde $H_0 : \mu \geq \mu_0$ las desigualdades se intercambian y la prueba también tiene tamaño α_0 y es insesgada.

Teorema. Bajo cualquiera de los dos casos anteriores, sea U el valor observado de U. Entonces, el valor-p de la prueba δ que rechaza $H_0 : \mu \leq \mu_0$ es $1-T_{n-1}(u)$ donde T_{n-1} es c.d.f de t_{n-1} y si se rechaza $H_0\mu \geq \mu_0$, el valor-p es $T_{n-1}(u)$.

Prueba. El caso $H_0 : \mu \leq \mu_0$ es análogo al cálculo del valor-p que se hizo en el capítulo anterior. El caso $H_0 : \mu \geq \mu_0$ se rechaza si

$$U \leq T_{n-1}^{-1}(\alpha_0) \Leftrightarrow T_{n-1}(u) \leq \alpha_0.$$

Es decir, el nivel más pequeño de significancia observada es $T_{n-1}(u)$

Considere el caso $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu > \mu_0$.

- Región de rechazo: $U \geq c$ con $U = \dfrac{\bar{X}_n - \mu_0}{\sigma'/\sqrt{n}}$.

- *Ejercicio*: es una prueba insesgada con nivel α_0 si $c = t_{n-1,1-\alpha_0}$.

- Valor-p: si observamos $U = u$, se rechaza H_0 si $u \geq t_{n-1,1-\alpha_0}$,

$$T_{n-1}(u) \geq T_{n-1}(t_{n-1,1-\alpha_0}) = 1 - \alpha_0 \implies 1 - T_{n-1(u)} = T_{n-1}(u).$$

- Función de potencia:

$$\begin{aligned}
\mathbb{P}[\text{Rechazo}|\mu] &= \mathbb{P}\left[\frac{\bar{X}_n-\mu_0}{\sigma'/\sqrt{n}} \geq t_{n-1,1-\alpha_0}\middle|\mu\right] \\
&== \mathbb{P}\left[\frac{\bar{X}_n+\mu-\mu-\mu_0}{\sigma'/\sqrt{n}} \geq t_{n-1,1-\alpha_0}\middle|\mu\right] \\
&= \mathbb{P}\left[\underbrace{\frac{\bar{X}_n-\mu}{\sigma'/\sqrt{n}}}_{\Delta} + \frac{\mu-\mu_0}{\sigma'/\sqrt{n}} \geq t_{n-1,1-\alpha_0}\middle|\mu\right]
\end{aligned}$$

Observe que

$$\Delta = \frac{\bar{X}_n-\mu}{\sigma'/\sqrt{n}}\cdot\frac{\sigma}{\sigma} = \frac{\dfrac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma}\sim N(0,1)}{\dfrac{\sigma'}{\sigma}=\sqrt{\dfrac{\chi^2}{n-1}}} \sim t_{n-1}.$$

De igual forma, vea que

$$U = \frac{\dfrac{\sqrt{n}(\bar{X}_n-\mu_0)}{\sigma}}{\dfrac{\sigma'}{\sigma}} = \frac{\dfrac{\sqrt{n}}{\sigma}(\bar{X}_n-\mu)+\overbrace{\dfrac{\sqrt{n}}{\sigma}(\mu-\mu_0)}^{\psi}\sim N(\psi,1)}{\dfrac{\sigma'}{\sigma}}.$$

Definición. Si Y, W son independientes con $W\sim N(\psi,1)$ y $Y\sim\chi^2_m$, entonces X se distribuye como una **t-Student no centrada** con parámetro ψ si

$$X=\frac{W}{\sqrt{\dfrac{Y}{m}}}.$$

Si $T_m(t|\psi)$ es c.d.f de X, entonces

$$\pi(\mu|\delta)=T_{n-1}(t_{n-1,1-\alpha_0}).$$

En el caso que la prueba sea $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu < \mu_0$.

$$\pi(\mu|\delta) = \mathbb{P}[U \leq t_{n-1,1-\alpha_0}] = T_{n-1}(t_{n-1,\alpha_0}).$$

Conclusión: a partir del error tipo II se puede determinar un tamaño de muestra dado, siempre y cuando existan restricciones sobre μ y σ^2.

11.2.2. Prueba t pareada

Ejemplo. Considere una muestra de n pacientes que fuman X cantidad al día. Sean t_1 el momento de la observación y t_2 el tratamiento. El consumo de cigarrillos en el individuo $\#i$ es

$$D_i = X_i^{t_2} - X_i^{t_1}, \quad i = 1, \ldots, n.$$

Otro ejemplo es tomar $Y_i^{t_1,t_2}$ el log-daño en los muñecos de prueba, donde t_1 corresponde al conductor y t_2 al acompañante, entonces

$$X_i = Y_i^{t_1} - Y_i^{t_2} = \ln\left(\frac{\text{daño}^{t_1}}{\text{daño}^{t_2}}\right) \implies \text{daño}^{t_2} \cdot e^{X_i} = \text{daño}^{t_1}$$

Evaluemos la prueba $H_0 : \mu \leq 0$ vs $H_1 : \mu > 0$ al $1\,\%$. Si $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ ambos parámetros desconocidos, y $n = 164$, $\bar{X}_n = 0{,}2199$, $\sigma' = 0{,}5342$, rechazamos H_0 si

$$U = \frac{0{,}2199 - 0}{\dfrac{0{,}5342}{\sqrt{164}}} = 5{,}271 > t_{163,1-0{,}01} = 2{,}35.$$

El valor-p de la prueba es

$$1 - \mathbb{P}[t_{163} < 5{,}271] = 1 \times 10^{-6} < 1\,\%.$$

Entonces rechazo H_0 con nivel de significancia de 1.

Suponga que la diferencia media entre conductor y pasajero es $\dfrac{\sigma}{4}$. ¿Cuál es el error tipo II?

$$\mu = \frac{\sigma}{4} \implies \psi = \frac{\mu - \mu_0}{\sigma/\sqrt{n}} = \frac{\sigma/4 - 0}{\sigma/\sqrt{164}} = \frac{\sqrt{164}}{3} = 3{,}2.$$

El error tipo II es $T_{163}(2{,}35|\psi = 3{,}2) = 1 - 0{,}802 = 0{,}198$.

11.2.3. Pruebas t de dos colas

- Región de rechazo: $|U| \geq t_{n-1,1-\frac{\alpha_0}{2}}$.
- Función de potencia:

$$\pi(\mu|\delta) = \mathbb{P}[U \geq t_{n-1,1-\frac{\alpha_0}{2}}|\mu] + \mathbb{P}[U \leq t_{n-1,1-\frac{\alpha_0}{2}}|\mu] = T_{n-1}(-c|\psi) + 1 - T_{n-1}(c|\psi).$$

- Valor-p: si observamos $U = u$, rechazamos H_0 si

$$|u| \geq t_{n-1,1-\frac{\alpha_0}{2}} \Leftrightarrow T_{n-1}(|U|) \geq 1 - \frac{\alpha_0}{2} \Leftrightarrow \alpha_0 \geq \underbrace{2[1 - T_{n-1}(|u|)]}_{\text{valor-}p}.$$

Propiedad. La prueba-t unilateral es un LRT.

Sea $f_n(x|\mu)$ la función de verosimilitud de una muestra de distribuciones normales y considere

$$\Lambda(x) = \frac{\sup_{\mu \leq \mu_0} f_n(x|\mu)}{\sup_\mu f_n(x|\mu)}.$$

El MLE en Ω es $(\bar{X}_n, \hat{\sigma}^2)$, entonces

$$\sup_\mu f_n(x|\mu) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-n/2}.$$

El MLE en Ω_0, si $\bar{X}_n < \mu_0$ es $\bar{X}_n$, por lo que $\Lambda(x) = 1$.

Si $\bar{X}_n > \mu_0$, se puede probar que $f_n(x|\mu)$ se maximiza si μ está lo más cerca posible de $\bar{X}_n$, que, en el subconjunto Ω_0 sería μ_0. Entonces $\mu = \mu_0$, $\hat{\sigma}_0 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$ y

$$\sup_\mu f_n(x|\mu) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} e^{-n/2}.$$

Por lo tanto $\Lambda(x) = \begin{cases} \left(\dfrac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} & \text{si } \bar{X}_n > \mu_0 \\ 1 & \text{si no} \end{cases}$

Ejercicio: si u es el valor observado del estadístico U, verifique que $\Lambda(x)$ es monótono decreciente con respecto a u.

Por lo tanto, para $k < 1$ existe c tal que

$$\Lambda(x) \leq k \Leftrightarrow u \geq c.$$

Ejercicio: encuentre c.

Se concluye que LRT es una prueba t.

Capítulo 12

Prueba de comparación de medias en 2 poblaciones

12.1. Comparación de medias normales

Asuma que $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu_1, \sigma^2)$ y $Y_1, \ldots, Y_n \overset{i.i.d}{\sim} N(\mu_2, \sigma^2)$. Los parámetros desconocidos son μ_1, μ_2, σ^2. Asuma que (X_i, Y_i) son independientes y la varianza es la misma (homocedasticidad).

Hipótesis: $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$.

Notación: $\bar{X}_m, \bar{Y}_n$, $S_X^2 = \sum_{i=1}^m (X_i - \bar{X}_m)^2$, $S_Y^2 = \sum_{i=1}^m (Y_i - \bar{Y}_n)^2$.

Teorema. Considere

$$U = \frac{(m+n-2)^{1/2}(\bar{X}_m - \bar{Y}_n)}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2} (S_X^2 + S_Y^2)^{1/2}}.$$

Si $\mu_1 = \mu_2 \implies U \sim t_{m+n-2}$.

Prueba. Vea que, bajo el supuesto que $\mu_1 = \mu_2$, $\bar{X}_n - \bar{Y}_n$ se distribuye como una normal con parámetros:

- $\mathbb{E}[X_n - \bar{Y}_n] = \mu_1 - \mu_2 = 0$.
- $\text{Var}(\bar{X}_m - \bar{Y}_n) = \text{Var}(\bar{X}_m) + \text{Var}(\bar{Y}_n) = \dfrac{\sigma^2}{m} + \dfrac{\sigma^2}{n} = \left(\dfrac{1}{m} + \dfrac{1}{n}\right)\sigma^2$.

Entonces

$$Z = \frac{\bar{X}_m - \bar{Y}_n}{\sigma\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}} \underset{\mu_1=\mu_2}{\sim} N(0,1).$$

Así mismo, se sabe que $\frac{S_X^2}{\sigma^2} \sim \chi^2_{m-1}$ y $\frac{S_Y^2}{\sigma^2} \sim \chi^2_{n-1}$.

Nota: no depende de H_0.

Como (X, Y) son independientes, $\frac{S_X^2}{\sigma^2}$ y $\frac{S_Y^2}{\sigma^2}$ son independientes. Así,

$$W = \frac{S_X^2 + S_Y^2}{\sigma^2} \sim \chi^2_{m+n-2}.$$

Entonces

$$U = \frac{Z}{\sqrt{\frac{W}{m+n-2}}} = \frac{\frac{\bar{X}_m - \bar{Y}_n}{\sigma\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}}}{\sqrt{\frac{1}{m+n-2}\left(\frac{S_X^2 + S_Y^2}{\sigma^2}\right)}} \sim t_{m+n-1}.$$

12.2. Prueba t de dos muestras

Dada una región de rechazo $U \geq c$,

$$\begin{aligned}\sup_{\mu_1 \leq \mu_2} \mathbb{P}[U \geq c|\mu_1, \mu_2, \sigma^2] \leq \alpha_0 &\implies \mathbb{P}[U \geq c|\mu_1 = \mu_2, \sigma^2] = 1 - T_{n+m-2}(c) \leq \alpha_0 \\ &\implies c = T^{-1}_{n+m-2}(1-\alpha_0)\end{aligned}$$

Rechazo H_0 si $U > T^{-1}_{n+m-2}(1-\alpha_0) : \delta$.

Teorema. La función de potencia $\pi(\mu_1, \mu_2, \sigma^2|\delta)$ tiene las siguientes propiedades:

I. $\pi(\mu_1, \mu_2, \sigma^2|\delta) = \alpha_0$ si $\mu_1 = \mu_2$.

II. $\pi(\mu_1, \mu_2, \sigma^2|\delta) < \alpha_0$ si $\mu_1 < \mu_2$.

III. $\pi(\mu_1, \mu_2, \sigma^2|\delta) > \alpha_0$ si $\mu_1 > \mu_2$.

Conclusión. δ es una prueba insesgada con tamaño α_0.

IV. Los límites cuando $\mu_1 - \mu_2 \to -\infty(+\infty)$ son los mismos del caso de una muestra.

Observe que para el caso II: $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$.

$$\delta : \text{Rechazo } H_0 \text{ si } U < T^{-1}_{n+m-2}(\alpha_0) = -T^{-1}_{n+m-2}(1 - \alpha_0).$$

Los p-valores son:

- Caso I: $1 - T_{n+m-2}(u)$ si observamos $U = u$.
- Caso II: $T_{n+m-2}(u)$.

Ejemplo. Considere la log-precipitación de 26 observaciones de nubes con químicos, $X_1, \ldots, X_{26}$ y 26 sin químicos $Y_1, \ldots, Y_{26}$.

Supuestos: $X_i \sim N(\mu_1, \sigma^2)$, $Y_j \sim N(\mu_2, \sigma^2)$.

Hipótesis: $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$.

Con los siguientes datos: $\bar{X}_m = 5{,}13$, $\bar{Y}_n = 3{,}99$, $S^2_X = 63{,}96$, $S^2_Y = 67{,}39$, se tiene que

$$U = \frac{50^{1/2}(5{,}13 - 3{,}99)}{\left(\dfrac{1}{26} + \dfrac{1}{26}\right)^{1/2} (63{,}96 + 67{,}39)^{1/2}} = 2{,}544.$$

A un nivel de confianza del 99 % ,

$$T_{n+m-2}(1 - \alpha_0) = T^{-1}_{50}(99\,\%) = 2{,}403 \implies U > T^{-1}_{50}(99\,\%)$$

y el valor-p: $1 - T_{50}(2{,}544) = 0{,}007$.

Interpretación: rechazamos al nivel 1 % de significancia la hipótesis de que las nubes irradiadas tienen una log-precipitación media menor a la de las nubes no irradiadas.

12.2.1. Prueba de 2 colas

Hipótesis. $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ (Prueba ANOVA).

- Prueba. δ : Rechazo H_0 si $|U| \geq T^{-1}_{m+n-2}\left(1 - \frac{\alpha_0}{2}\right)$.
- Valor-p: $2 - T_{m+n-2}(|u|)$ donde $U = u$.

Ejemplo. Menas de cobre. Sean $X_1, \ldots, X_8$ la cantidad de cobre (g) en 8 menas en la localización 1, y $X_1, \ldots, X_{10}$ en 10 menas en la localización 2. Los datos son $\bar{X}_8 = 2{,}6, \bar{Y}_{10} = 2{,}3, S^2_X = 0{,}32$ y $S^2_Y = 0{,}22$ ¿Las dos localizaciones generan el mismo nivel de cobre?

$H_0 : \mu_1 = \mu_2$, $X_i \sim N(\mu_1, \sigma^2)$, $Y_j \sim N(\mu_2, \sigma^2)$.

Se tiene que $U = 3{,}442$.Si $\alpha_0 = 1\,\%$, $T^{-1}_{16}\left(1 - \frac{0{,}01}{2}\right) = T^{-1}_{16}(0{,}995) = 2{,}921$.

Rechazamos H_0 pero el valor-p es $2[1 - T_{16}(|3{,}442|)] = 0{,}003$.

Interpretación: rechazamos al 1 % de significancia la hipótesis de una diferencia no significativa entre las cantidades medias de cobre en cada localización.

Ejercicio. La prueba t de 2 muestras es un LRT.

12.3. Prueba F

Definición Si Y y W son variables aleatorias independientes, $Y \sim \chi^2_m$ y $W \sim \chi^2_n$, $m, n \in \mathbb{Z}^+$. Defina

$$X = \frac{Y/m}{W/n} \sim F_{m,n}$$

X tiene una distribución F con m y n grados de libertad.

Propiedades:

1. Si $X \sim F_{m,n} \implies 1/X \sim F_{n,m}$.
2. Si $Y \sim t_n \implies Y^2 \sim F_{1,n}$.

Sean $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu_1, \sigma_1^2)$ y $Y_1, \ldots, Y_n \overset{i.i.d}{\sim} N(\mu_2, \sigma_2^2)$.

Considere el esquema

$$\begin{array}{rcl} U \sim t_{n-1} & & U^2 \sim F_{1,n-1} \\ H_0 : \mu = \mu_0 & \Leftrightarrow & H_0 : \mu = \mu_0 \\ |U| \geq |c| & & U^2 \geq c^* \end{array}$$

Bajo el esquema anterior y si (X, Y) son independientes, considere:

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \text{ vs } H_1 : \sigma_1^2 > \sigma_2^2$$

y tome $\alpha_0 \in (0, 1)$.

La lógica de esta prueba es, como $\dfrac{S_X^2}{\sigma_1^2} \sim \chi^2_{m-1}$ y $\dfrac{S_X^2}{\sigma_1^2} \sim \chi^2_{n-1}$, calculamos

$V^* = \dfrac{\dfrac{S_X^2/\sigma_1^2}{m-1}}{\dfrac{S_Y^2/\sigma_1^2}{n-1}} \sim F_{m-1,n-1}$. Bajo el supuesto de homocedasticidad,

$V = \dfrac{\dfrac{S_X^2}{m-1}}{\dfrac{S_Y^2}{n-1}} \sim F_{m-1,n-1}$.

δ : Rechazo H_0 si $V \geq c$.

Teorema. La distribución de $V^* \sim F_{m-1,n-1}$ y si $\sigma_1 = \sigma_2$, $V \sim F_{m-1,n-1}$.

Usando el δ anterior

$$\sup_{\sigma_1^2 \leq \sigma_2^2} \mathbb{P}[V \geq c | \mu_1 \mu_2, \sigma_1^2, \sigma_2^2] \leq \alpha_0,$$

resuelve

$$\mathbb{P}[V \geq c | \mu_1, \mu_2, \sigma_1^2, \sigma_2^2] = \alpha_0 \implies c = F^{-1}_{m-1,n-1}(1 - \alpha_0) =: G^{-1}_{m-1,n-1}(1 - \alpha_0).$$

Teorema. si δ se define según lo anterior,

I.
$$\begin{aligned}\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|\delta) &= \mathbb{P}[V \geq G^{-1}_{m-1,n-1}(1-\alpha_0)]\\ &= \mathbb{P}\left[V^* \geq \frac{\sigma_2^2}{\sigma_1^2}c\right]\\ &= 1 - G_{m-1,n-1}\left(\frac{\sigma_2^2}{\sigma_1^2}c\right)\end{aligned}$$

II. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,|\delta) = \alpha_0$ si $\sigma_1^2 = \sigma_2^2$.

III. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|\delta) < \alpha_0$ si $\sigma_1^2 < \sigma_2^2$.

IV. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|\delta) > \alpha_0$ si $\sigma_1^2 > \sigma_2^2$.

V. $\dfrac{\sigma_1^2}{\sigma_2^2} \to 0 \implies \pi \to 0.$

VI. $\dfrac{\sigma_1^2}{\sigma_2^2} \to \infty \implies \pi \to 1.$

Por (i)-(iv) δ es insesgada con tamaño α_0.

El valor-p es $1 - G_{m-1,n-1}(v)$, $V = v$.

Ejemplo. $X_1,\dots,X_6 \sim N(\mu_1,\sigma_1^2)$, $S_X^2 = 30$, $Y_1,\dots,Y_{21} \sim N(\mu_2,\sigma_2^2)$, $S_Y^2 =$ 30.

La hipótesis nula es $H_0 : \sigma_1^2 \leq \sigma_2^2$.

Se calcula $V = \dfrac{30/5}{40/20} = 3$ y $F^{-1}_{5,20}(1-0{,}05) = 2{,}71$.

El valor-p corresponde a $1 - G_{5,20}(3) = 0{,}035$.

Si $\alpha_0 = 1\,\%$, no rechazo. Si $\alpha_0 = 5\,\%$ rechazo.

12.3.1. Prueba de 2 colas (prueba de homocedasticidad)

Bajo las hipótesis $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, se rechaza si $V \geq c_2$ o $V \leq c_1$ con c_1, c_2 tales que

$$\mathbb{P}[V \leq c_1] = \frac{\alpha_0}{2} \text{ y } \mathbb{P}[V \geq c_2] = \frac{\alpha_0}{2} \implies c_1 = G^{-1}_{m-1,n-1}\left(\frac{\alpha_0}{2}\right) c_1 = G^{-1}_{m-1,n-1}\left(\frac{\alpha_0}{2}\right)$$

Ejemplo. Mismo ejemplo de las nubes. Queremos probar $H_0 : \sigma_1^2 = \sigma_2^2$. Calculamos

$$V = \frac{\frac{63{,}96}{25}}{\frac{67{,}39}{25}} = 0{,}9491$$

Se tiene que $c_1 = G^{-1}_{25,25}(0{,}0025) = 0{,}4484$ y $c_2 = G^{-1}_{25,25}(0{,}975) = 2{,}23$.

Si observamos $V = v$, podemos rechazar si

$$v \leq G^{-1}_{m-1,n-1}\left(\frac{\alpha_0}{2}\right) \implies 2G_{m-1,n-1}(v) \leq \alpha_0$$

o tambien si

$$v \geq G^{-1}_{m-1,n-1}\left(1 - \frac{\alpha_0}{2}\right) \implies G_{m-1,n-1}(v) \geq 1 - \frac{\alpha_0}{2} \implies \alpha_0 \geq 2G_{m-1,n-1}(v)$$

Por lo tanto, el p-valor es

$$\text{valor-}p = 2\,\text{mín}[1 - G_{m-1,n-1}(v), G_{m-1,n-1}(v)]$$

Ejercicio. Verifique que en este caso da 0.9.

Al ser mayor al 5 %, no rechaza la hipótesis de homocedasticidad.

Propiedad. La prueba F es un LRT.

Capítulo 13

Pruebas de hipótesis bayesianas

13.1. Pruebas de hipótesis bayesianas

Suponga que $\Omega = \{\theta_0, \theta_1\}$. Si $\theta = \theta_i$ $(i = 0, 1)$, $X_1, \ldots, X_n \sim f_i(x)$. Considere las hipótesis

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1.$$

Hay dos decisiones, d_0 : no rechazo H_0 y d_1 : rechazo H_0.

Asuma que si selecciono d_1 cuando H_0 es cierto, la pérdida es w_0. Por el contrario, si selecciono d_1 cuando H_0 es cierto, la pérdida es w_1.

	d_0	d_1
θ_0	0	w_0
θ_1	w_1	0

Sean $\pi_0 = \mathbb{P}[H_0 \text{ es cierta}] = \mathbb{P}[H_0]$ y $\pi_1 = \mathbb{P}[H_1 \text{ es cierta}] = 1 - \mathbb{P}[H_0]$.

Considere δ el procedimiento de prueba. El valor esperado de la pérdida corresponde a

$$r(\delta) = \mathbb{E}[\text{Pérdida}|H_0]\mathbb{P}[H_0] + \mathbb{E}[\text{Pérdida}|H_1]\mathbb{P}[H_1].$$

Dado H_0,

$$\begin{aligned}\mathbb{E}[\text{Pérdida}|H_0] &= w_0 \cdot \mathbb{P}[\text{Seleccione } d_1|\theta_0] + 0 \cdot \mathbb{P}[\text{Seleccione } d_0|\theta_0] \\ &= w_0 \cdot \text{Error I} = w_0\alpha(\delta).\end{aligned}$$

Por el otro lado,

$$\begin{aligned}\mathbb{E}[\text{Pérdida}|H_1] &= 0 \cdot \mathbb{P}[\text{Seleccione} d_1|\theta_0] + w_1 \cdot \mathbb{P}[\text{Seleccione} d_0|\theta_0] \\ &= w_1 \cdot \text{Error II} = w_1\beta(\delta).\end{aligned}$$

Entonces $r(\delta) = w_0\alpha(\delta) + w_1\beta(\delta)$.

El procedimiento δ que minimiza $r(\delta)$ se llama procedimiento de prueba bayesiana. Por el teorema de Neyman-Pearson y tomando $a = \pi_0 w_0$ y $b = \pi_1 w_1$, el δ que soluciona el problema es:

$$\delta : \text{Rechazo } H_0 \text{ si } \pi_0 w_0 f_0(x) < \pi_1 w_1 f_1(x).$$

o si $\pi_0 w_0 f_0(x) = \pi_1 w_1 f_1(x)$, $\dfrac{f_1(x)}{f_0(x)} > \dfrac{a}{b} = k$.

Nota. La decisión δ es invariante a multiplicacione por escalar en el costo.

Ejemplo.

a. Calculemos

$$\pi(\theta_0|x) = \frac{\pi_0 f_0(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}$$
$$\pi(\theta_1|x) = \frac{\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}$$

b. Esperanza de la pérdida

$$\mathbb{E}[\text{Pérdida}|x] = \mathbb{E}[\text{Pérdida}|\theta_0, x] + \mathbb{E}[\text{Pérdida}|\theta_1, x].$$

Si $\delta = d_0$,

$$\mathbb{E}_\delta[\text{Pérdida}|X] = w_1\pi(\theta_1|x) = \frac{w_1\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}.$$

Si $\delta = d_1$,

$$\mathbb{E}_\delta[\text{Pérdida}|X] = \frac{w_0\pi_0 f_0(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}.$$

Minimizar $\mathbb{E}[\text{Pérdida}|x]$ con respecto a δ es equivalente a rechazar H_0 bajo el criterio anterior.

Conclusión: es equivalente construir la decisión en cualquiera de los dos criterios (previa o probabilidad posterior).

c. Rechazo H_0 si $\mathbb{P}[H_0 \text{ es cierto}|X] \leq \dfrac{w_1}{w_0 + w_1}$.

Rechazamos H_0 si

$$\frac{w_0\pi_0 f_0(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)} \leq \frac{w_1\pi_1 f_1(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}.$$

Entonces

$$w_0\mathbb{P}[H_0|x] \leq w_1\mathbb{P}[H_1|x] = w_1[1 - \mathbb{P}[H_0|x]] \implies \mathbb{P}[H_0|x] \leq \frac{w_1}{w_0 + w_1}.$$

Caso general: $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$.

13.2. Hipótesis de una cola

Asuma la misma función de pérdida $L(\theta, d_i)$:

	d_0	d_1
H_0	0	w_0
H_1	w_1	0

y considere la hipótesis $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.

Teorema. Suponga que $f_n(x|\theta)$ tiene un cociente de verosimilitud monótono con respecto al estadístico $T = r(x)$. Es decir, $R(x) = \dfrac{f_n(x|\theta_1)}{f_n(x|\theta_0)} = g(r(x))$ con $g(r(x))$ monótono con respecto a $r(x)$.

Asuma la función de pérdida anterior. Entonces el procedimiento bayesiano de prueba rechaza H_0 cuando $T \geq c$.

Definición. Si $f_n(x|\theta)$ es una verosimilitud y $T = r(x)$ un estadístico, decimos que $f_n(x|\theta)$ tiene un MLR con respecto a T si para $\theta_1, \theta_2 \in \Omega$ tales que $\theta_1 < \theta_2$, $\dfrac{f_n(x|\theta_2)}{f_n(x|\theta_1)}$ depende de x a través de $r(x)$ y es una función monótona de $r(x)$.

Ejemplo. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, σ^2 conocido. Si $\mu_1 < \mu_2$:

$$\begin{aligned}\frac{f_n(x|\mu_2)}{f_n(x|\mu_1)} &= \frac{(2\pi\sigma^2)^{-n/2} \exp\left[-\dfrac{1}{2\sigma^2}\sum(X_i-\mu_2)^2\right]}{(2\pi\sigma^2)^{-n/2} \exp\left[-\dfrac{1}{2\sigma^2}\sum(X_i-\mu_1)^2\right]} \\ &= \exp\left[\frac{n(\mu_2-\mu_1)}{\sigma^2}\bar{X}_n - \frac{1}{2}(\mu_2+\mu_1))\right] = g(T)\end{aligned}$$

Entonces $g(T)$ es monótono creciente con respecto a $T = \bar{X}_n$ y por lo tanto tiene un MLE con respecto a $\bar{X}_n$.

Prueba. Recuerde que

$$\pi(\theta|x) = \frac{f_n(x|\theta)\pi(\theta)}{\int f_n(x|\psi)\pi(\psi)d\psi}$$

$$\mathcal{L}(x) = \frac{\mathbb{E}[\text{Pérdida}|x,d_0]}{\mathbb{E}[\text{Pérdida}|x,d_1]} = \frac{\dfrac{1}{\text{cte}}\displaystyle\int_{\theta_0}^{\infty} w_1 f_n(x|\theta)\pi(\theta)d\theta}{\dfrac{1}{\text{cte}}\displaystyle\int_{-\infty}^{\theta_0} w_0 f_n(x|\theta)\pi(\theta)d\theta}$$
$$= \frac{w_1}{w_0}\frac{\displaystyle\int_{\theta_0}^{\infty} f_n(x|\theta)\pi(\theta)d\theta}{\displaystyle\int_{-\infty}^{\theta_0} f_n(x|\theta)\pi(\theta)d\theta} \geq 1$$

Buscamos rechazar si $l(x) \geq 1$ (prueba bayesiana) y si existe una función monótona tal que $l(x) \geq 1 \Leftrightarrow T \geq c$, entonces ambas pruebas son iguales. Basta con probar que $l(x)$ es una función monótona creciente de T.

Sea $X_1, X_2 \in \mathcal{X}$ tal que $r(X_1) \leq r(X_2)$. Entonces

$$l(X_1) - l(X_2) = \frac{w_1 \displaystyle\int_{\theta_0}^{\infty} f_n(x_1|\theta)\pi(\theta)d\theta}{w_0 \displaystyle\int_{-\infty}^{\theta_0} f_n(x_1|\theta)\pi(\theta)d\theta} - \frac{w_1 \displaystyle\int_{\theta_0}^{\infty} f_n(x_2|\theta)\pi(\theta)d\theta}{w_0 \displaystyle\int_{-\infty}^{\theta_0} f_n(x_2|\theta)\pi(\theta)d\theta} \leq 0.$$

Si simplificamos la expresión en una sola fracción, el numerador es de la forma

$$\int_{\theta_0}^{\infty}\int_{-\infty}^{\theta_0} \pi(\theta)\pi(\psi)[f_n(x_1|\theta)f_n(x_2|\psi) - f_n(x_2|\theta)f_n(x_1|\psi)]\, d\psi d\theta$$

y el denominador es siempre positivo, por lo que basta con que el numerador sea negativo.

Como $f_n(x|\theta)$ tiene un MLR, si $r(x_1) \leq r(x_2)$ y $-\infty < \psi \leq \theta_0 \leq \theta < +\infty$,

$$\frac{f_n(x_1|\theta)}{f_n(x_1|\psi)} \leq \frac{f_n(x_2|\theta)}{f_n(x_2|\psi)} \Leftrightarrow f_n(x_1|\theta)f_n(x_2|\psi) \leq f_n(x_2|\theta)f_n(x_1|\psi)$$

Entonces $l(x_1) \leq l(x_2)$ y por tanto ambas pruebas son equivalentes.

Ejemplo. Diferencias porcentuales entre calorías observadas y calorías en publicidad para 20 productos preparados.

$X_1, \dots, X_{20} \sim N(\theta, 100)$, $\theta \sim N(0, 60)$

La media posterior es

$$\frac{100 \cdot 0 + 20 \cdot 60\bar{X}_{20}}{100 + 20 \cdot 60} = 0{,}923\bar{X}_{20}.$$

y $\sigma_1^2 = 4{,}62$.

La hipótesis de interés es $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$.

δ: Rechazo H_0 si $\mathbb{P}[H_0|\bar{X}_{20}] \leq \dfrac{w_1}{w_0 + w_1}$, donde

$$\mathbb{P}[\theta \leq 0|\bar{X}_{20}] = \mathbb{P}\left[Z\middle|\frac{-0{,}923\bar{X}_{20}}{4{,}62}\right] = \Phi(-0{,}429\bar{X}_{20}).$$

Bajo δ:

$$\begin{aligned}\Phi(-0{,}429\bar{X}_{20}) \leq \frac{w_1}{w_0 + w_1} = \beta &\implies -0{,}429\bar{X}_{20} \leq \Phi^{-1}(\beta)\\ &\implies \bar{X}_{20} \geq \frac{-\Phi^{-1}(\beta)}{0{,}429}\end{aligned}$$

Si $w_0 = w_1 \implies \beta = 1/2$ y $\Phi(1/2) = 0$. Por lo tanto $\bar{X}_{20} \geq 0$.

Interpretación. Si $\bar{X}_{20} \geq c$ entonces aceptamos la hipótesis de que $\theta > 0$ (en términos de la aplicación).

13.3. Hipótesis de 2 colas

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

La significancia práctica indica que "ser igual a θ_0 significa estar cerca".

Replanteamos H_0, tomando $d > 0$:

$$H_0 : |\theta - \theta_0| \leq d \text{ vs } H_1 : |\theta - \theta_0| > d.$$

En el ejemplo anterior, $(\theta_0 = 0)$

$$\mathbb{P}[H_0|\bar{X}_{20}] = \mathbb{P}[|\theta| \leq d|\bar{X}_{20}] = \Phi\left(\frac{d - 0{,}1154}{4{,}62^{1/2}}\right) - \Phi\left(\frac{-d - 0{,}1154}{4{,}62^{1/2}}\right) = g(d)$$

En el caso normal, si $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ ambos parámetros desconocidos y $\tau = \dfrac{1}{\sigma^2}$, recuerde que

$$[\mu, \tau] \sim \text{Normal-Gamma}(\mu_0, \lambda_0, \alpha_0, \beta_0) \implies \left(\frac{\lambda_0 \alpha_0}{\beta_0}\right)^{\frac{1}{2}} (\mu - \mu_0) \sim t_{2\alpha_0}$$

y la marginal de μ se usa en el cálculo $\mathbb{P}[H_0|x]$.

Ejemplo. Residuos de un pesticida en apio. $X_1, \ldots, X_{77} \sim N(\mu, \sigma^2)$. Usamos una previa impropia de (μ, σ^2),

$$\pi(\mu, \tau) \propto \tau^{-1}$$

.

Recuerde, además, que

$$U = \left(\frac{n(n-1)}{s_n^2}\right)^{\frac{1}{2}} (\mu - \bar{X}_n) \sim t_{n-1}$$

en el nivel posterior.

Nos interesa probar $H_0 : \mu \geq 55$ vs $H_1 : \mu < 55$.

Los datos son $\bar{X}_{77} = 50{,}23$, $s_{77}^2 = 34106$.

$$\begin{aligned}
\mathbb{P}[H_0|X] &= \mathbb{P}[\mu \geq 55|X] \\
&= \mathbb{P}\left[\frac{\mu - \bar{X}_{77}}{\left(\frac{\sigma'_2}{77}\right)^{1/2}}\right] \leq \mathbb{P}\left[\frac{55 - 50{,}23}{\left(\frac{\sigma'_2}{77}\right)^{1/2}}\right] \\
&= 1 - T_{76}[1{,}974] = 0{,}026
\end{aligned}$$

Note que $\dfrac{-\bar{X}_{77} - \overbrace{55}^{\mu_0}}{\left(\frac{\sigma_2'}{77}\right)^{1/2}} = -U$

donde U es el estadístico de prueba en el caso frecuentista y

$$\mathbb{P}[H_0|X] = \mathbb{P}[\underbrace{t_{n-1} \geq -\overbrace{u}^{\text{Observado}}}_{\text{Región de rechazoen la prueba frecuentista}}|X] \leq \alpha_0$$

Interpretación: aceptamos la hipótesis de que el valor medio del pesticida es menor o igual a 55 ante una función de pérdida en donde $w_0 = w_1$.

Teorema. Sean $X_1, \ldots, X_m \sim N(\mu_1, \tau)$ y $Y_1, \ldots, Y_n \sim N(\mu_2, \tau)$ dos muestras y $\pi(\mu_1, \mu_2, \tau) \propto \tau^{-1}$, $\tau > 0$. Entonces

$$(m+n-2)^{1/2} \frac{\mu_1 - \mu_2 - (\bar{X}_m - \bar{Y}_n)}{\left(\dfrac{1}{m} + \dfrac{1}{n}\right)^{1/2} (s_X^2 + s_Y^2)^{1/2}} \sim t_{m+n-2}$$

condicionado en (X, Y).

Prueba: Ejercicio.

Consecuencia. Si queremos probar $H_0 : \mu_1 - \mu_2 \leq 0$ vs $H_1 : \mu_1 - \mu_2 > 0$,

$$\mathbb{P}[\mu_1 - \mu_2 \leq 0|x, y] = \mathbb{P}\left[t_{m+n-2} \leq \frac{-(\bar{X}_m - \bar{Y}_n)}{\left(\dfrac{1}{m} + \dfrac{1}{n}\right)^{1/2} (s_X^2 + s_Y^2)^{1/2}} (m+n-2)^{1/2}\right] = T_{m+n-2})(-u).$$

donde u es el valor observado de la prueba de 2 muestras en el caso frecuentista.

Rechazamos H_0 si

$$T_{m+n-2}(-u) \leq \frac{w_1}{w_0 + w_1} = \alpha_0 \Leftrightarrow -u \leq T^{-1}_{m+n-2}(\alpha_0) \implies u \geq -T^{-1}_{m+n-2}(\alpha_0) = T^{-1}_{m+n2-2}(1 - \alpha_0)$$

Es la misma prueba con $\alpha_0 = \dfrac{w_1}{w_0 + w_1}$ con distinta interpretación.

Otro caso particular es la prueba de varianzas en el caso normal con previas impropias.