Tarea 3

Rodrigo Carvajal Pizarro (rcarvaja@astro.puc.cl)
`https://github.com/racarvajal/Tarea3_Astro_Stats.git`

Resolución

- **Distribución de variables aleatorias**
  - χ^2_ν
$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi^2_\nu$$
  - t_ν
$$T = Z\sqrt{\frac{\nu}{V}} \sim t_\nu$$
  - F_{ν_1,ν_2}
$$\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \sim F_{\nu_1,\nu_2}$$

- **Sesgo y varianza de estimador.**

  El sesgo da cuenta de lo alejado que está el promedio de los datos obtenidos del valor esperado según la distribución.

  La varianza indica la dispersión de los datos con respecto al promedio de estos.

  Esta diferencia puede verse en la Fig. 1. Se distingue la diferencia entre la esperanza del estimador y la media del parámetro. También, puede notarse que tanto el estimador como el parámetro tienen varianzas mayores a cero y diferentes entre sí.

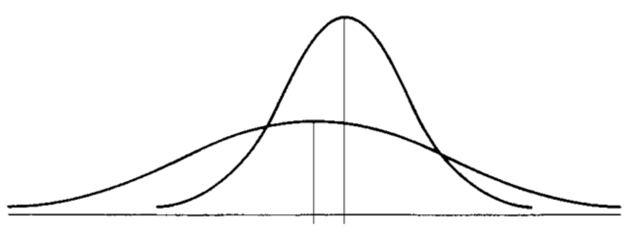

Figura 1: Diferencia entre bias y varianza para un estimador. Imagen tomada de `http://what-when-how.com/statistics/b632-method-to-binomial-distribution-statistics/`

  Se dice que un estimador es no sesgado cuando el valor de la esperanza del estimador es igual al valor del mismo parámetro. Dicho de otro modo, cuando el valor esperado de un estimador es el mismo parámetro.

- **Error cuadrático medio (MSE)**

$$MSE = \mathbb{E}[(\theta - \hat{\theta})^2] = \mathbb{E}[(\theta - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \hat{\theta})^2]$$
$$MSE = \mathbb{E}[(\theta - \mathbb{E}^2[\hat{\theta}])] + 2\mathbb{E}[(\theta - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \hat{\theta})] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2]$$

  En el paradigma frecuentista, θ es un valor constante.

$$MSE = \mathbb{E}[(\theta - \mathbb{E}^2[\hat{\theta}])] + 2(\theta - \mathbb{E}[\hat{\theta}])\mathbb{E}[\mathbb{E}[\hat{\theta}] - \hat{\theta}] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2]$$

Para el segundo término, se tiene que:

$$2(\theta - \mathbb{E}[\hat{\theta}])\mathbb{E}[\mathbb{E}[\hat{\theta}] - \hat{\theta}] = 2(\theta - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \mathbb{E}[\hat{\theta}]) = 2(\theta - \mathbb{E}[\hat{\theta}]) \times 0 = 0$$

Por lo tanto,

$$MSE = \mathbb{E}[(\theta - \mathbb{E}^2[\hat{\theta}])] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2] = [\mathbb{E}[\theta] - \mathbb{E}[\hat{\theta}]]^2 + \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$$

$$MSE = (\theta - \mathbb{E}[\hat{\theta}])^2 + \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$$

$$MSE = \text{sesgo}^2 + \mathbb{V}[\hat{\theta}]$$

- **Mejor estimador**

  Una buena manera de responder a la pregunta sobre cuál estimador es mejor sobre otro es utilizando el Error Cuadrático Medio (MSE) para compararlos. Esta medida (MSE) da cuenta de qué tan alejados están los valores que se han estimado (parámetros) de aquellos intrínsecos a la distribución de los datos. Esta lejanía se mide como una distancia cuadrática (euclidiana). Como se ve en la pregunta anterior, el MSE, además, entrega información sobre la separación entre el valor esperado del estimador del parámetro y el valor real de éste y sobre la varianza del estimador.

  De este modo, para ambos estimadores (S^2 y $\hat{\sigma}^2$) se calculará el error cuadrático medio y sus valores se compararán. Aquel estimador que presente un MSE menor, será considerado como mejor que el otro.

  — $\mathbf{S^2}$

  Primero, se obtendrá el sesgo de este estimador. Para ello, se construye una variable aleatoria que cumpla las condiciones indicadas en el enunciado (que distribuya como χ^2_{n-1}). Luego, y utilizando las propiedades del cálculo de esperanzas, se volverá a una expresión para el estimador original.

$$\mathbb{E}\left[\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\right] = n - 1$$

$$\Rightarrow \mathbb{E}\left[S^2 = \frac{\sigma^2}{n-1}\left(\frac{n-1}{\sigma^2}S^2\right)\right] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

$$\Rightarrow \text{sesgo} = \mathbb{E}[S^2] - \sigma^2 = \sigma^2 - \sigma^2 = 0$$

  Por lo tanto, S^2 es un estimador insesgado.

  Para el cálculo de la varianza, se seguirá el mismo procedimiento que en el caso del sesgo. Se construirá una variable aleatoria que se comporta como χ^2_{n-1} y se regresará, luego, al estimador inicial.

$$\mathbb{V}\left[\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\right] = 2(n-1)$$

$$\Rightarrow \mathbb{V}\left[S^2 = \frac{\sigma^2}{n-1}\left(\frac{n-1}{\sigma^2}S^2\right)\right] = \left(\frac{\sigma^2}{n-1}\right)^2 \times 2(n-1) = \frac{2\sigma^4}{n-1}$$

  De este modo, se puede obtener el error cuadrático medio para el estimador.

$$MSE(S^2) = \text{sesgo}^2 + \mathbb{V}[\hat{S^2}] = 0 + \frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n-1}$$

— $\hat{\sigma}^2$

Para este estimador, el procedimiento será el mismo que para el anterior.

$$\mathbb{E}\left[\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\right] = n-1$$

$$\Rightarrow \mathbb{E}\left[\hat{\sigma}^2 = \frac{n-1}{n}S^2 = \left(\frac{n-1}{n}\right)\frac{\sigma^2}{n-1}\left(\frac{n-1}{\sigma^2}S^2\right)\right] = \frac{\sigma^2}{n}(n-1) = \sigma^2\frac{n-1}{n}$$

$$\Rightarrow \text{sesgo} = \sigma^2\frac{n-1}{n} - \sigma^2 = -\frac{\sigma^2}{n}$$

Para la varianza,

$$\mathbb{V}\left[\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\right] = 2(n-1)$$

$$\Rightarrow \mathbb{V}\left[\hat{\sigma}^2 = \frac{n-1}{n}S^2 = \left(\frac{n-1}{n}\right)\frac{\sigma^2}{n-1}\left(\frac{n-1}{\sigma^2}S^2\right)\right] = \left(\frac{\sigma^2}{n}\right)^2(n-1) = \sigma^4\frac{n-1}{n^2}$$

Así, el error cuadrático medio será:

$$MSE(\hat{\sigma}^2) = \text{sesgo}^2 + \mathbb{V}[\hat{\sigma}^2] = 0 + \frac{2\sigma^4}{n-1} = \left(-\frac{\sigma^2}{n}\right)^2 + \sigma^4\frac{n-1}{n^2} = \frac{\sigma^4}{n^2}n = \frac{\sigma^4}{n}$$

Ahora, deben compararse los errores cuadráticos medios de ambos estimadores para conocer cuál es mejor. En general, esta comparación dependerá del valor de n. Se analizará gráficamente el comportamiento de $MSE(S^2) - MSE(\hat{\sigma}^2)$ para saber cuál es mayor y en qué rango. Esta comparación puede verse en la Fig. 2.

$$MSE(S^2) - MSE(\hat{\sigma}^2) = \frac{2\sigma^4}{n-1} - \frac{\sigma^4}{n} = \frac{n+1}{n(n-1)}\sigma^4$$

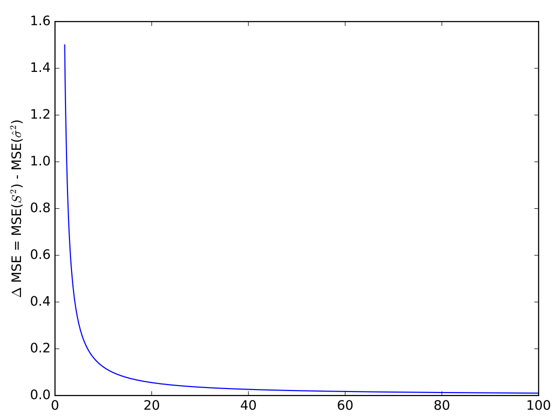

Figura 2: Comparación entre $MSE(S^2)$ y $MSE(\hat{\sigma}^2)$

De la figura puede verse que esta diferencia es siempre positiva. Esto significa que $MSE(S^2)$ es siempre mayor que $MSE(\hat{\sigma}^2)$ bajo la condición de que $n \geq 2$.

Esto implica, y bajo el criterio que se ha utilizado (MSE), S^2 es un mejor estimador que $\hat{\sigma}^2$ de σ^2.

- **Estimadores de Máxima Verosimilitud, MLE**

  Un estimador de máxima verosimilitud se obtiene al optimizar la verosimilitud de un modelo. O, para efectos de mejor computación, al optimizar el logaritmo de la verosimilitud. Es decir,

$$MLE = \text{máx}\,\mathcal{L} = \text{máx}\ln\left(p(x_i|\theta)\right)$$

- **Distribución a posteriori**

  La distribución a posteriori da cuenta de la probabilidad de obtener ciertos valores de los parámetros dados los datos. Esto se expresa como:

$$p(\theta|x_i) \propto p(x_i|\theta) \times p(\theta)$$

- **Estimadores maximum-a-posteriori, MAP**

  El estimador maximum-a-posterior proviene de optimizar el valor de la probabilidad a posteriori. También, puede optimizarse el logaritmo de esta probabilidad.

$$MAP = \text{máx}\,p(\theta|x_i) = \text{máx}[p(x_i|\theta) \times p(\theta)]$$

- **MAP, MLE y suma de residuos al cuadrado**

  Entre el MLE y el MAP, la relación tiene que ver con la inclusión de un factor, el *prior*, que indica la probabilidad que tienen los parámetros. De este modo, se pueden relacionar dos maneras de estimar el valor para un estimador; desde el paradigma frecuentista (MLE) y el paradigma bayesiano (MAP). A través de una fórmula, se ve que:

$$p(\theta|x_i) \propto p(x_i|\theta) \times p(\theta)$$

$$\Rightarrow MAP \propto MLE$$

  Si ahora se toma el Maximum Likelihood Estimator, se puede llegar a una relación con los estimadores obtenidos al minimizar la suma de los residuales al cuadrado, RSS ($\sum_{i=1}^{n}(y_i - f(x_i|\theta))^2$).

  Primero, se debe maximizar el logaritmo de la verosimilitud:

$$\mathcal{L} = \ln\left(\prod_{i=1}^{n} p(y_i|x_i) \cdot p(x_i)\right)$$

$$\mathcal{L} = \ln\left(\prod_{i=1}^{n} p(y_i|x_i)\right) + \ln\left(\prod_{i=1}^{n} p(x_i)\right)$$

  Si se asume "gausianidad" en los errores de x, es decir:

$$p(y|x) \sim \mathcal{N}(f(x|\theta), \sigma^2)$$

$$\Rightarrow \mathcal{L} = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i|\theta))^2}{2\sigma^2}\right)\right) + \ln\left(\prod_{i=1}^{n} p(x_i)\right)$$

  Como el segundo término en el lado derecho de la igualdad es independiente de los parámetros, este valor se toma como una constante a la hora de optimizar $\mathcal{L}$. De este modo, se tendrá:

$$\mathcal{L} = \sum_{i=1}^{n} \left(\ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_i - f(x_i|\theta))^2}{2\sigma^2} \right) \right) \right) + \ln \left(\prod_{i=1}^{n} p(x_i) \right)$$

$$\mathcal{L} = \sum_{i=1}^{n} \left(\ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \left(\frac{y_i - f(x_i|\theta)}{\sigma} \right) \right) + \ln \left(\prod_{i=1}^{n} p(x_i) \right)$$

Y, pensando en términos de θ, se puede escribir

$$\mathcal{L} = \text{constante} - \frac{1}{\sigma} \sum_{i=1}^{n} (y_i - f(x_i|\theta))$$

Por lo tanto, los tres procedimientos (MLE, MAP y RSS) es equivalente en el caso que los errores asociados sean gaussianos.