

Inferencia

Instituto Artek

14 de abril de 2023

Índice

1. Recordatorio	2
2. Estimadores	2
3. Estimadores insesgados	2
4. Eficiencia	3
5. Consistencia	4

1. Recordatorio

Dada una muestra x_1, \dots, x_n . Estos valores se consideran como n observaciones de una variable X . X es nuestro modelo de interés, pues es el que creemos describe a una población. Sin embargo en contexto de la estadística, es mejor considerar a la muestra x_1, \dots, x_n como valores de X_1, \dots, X_n variables aleatorias con misma distribución que X pero que son independientes ya que la muestra es aleatoria. Por tanto si X tiene valor esperado μ entonces al considerar a

$$S_n = X_1 + \dots + X_n$$

se tiene que

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = n\mu = n\mathbb{E}(X),$$

por la independencia de la variables también se cumple que,

$$\mathbb{V}ar(S_n) = \mathbb{V}ar(X_1) + \dots + \mathbb{V}ar(X_n) = n\sigma^2 = n\mathbb{V}ar(X).$$

a S_n se le llama **suma muestral**, note que la suma muestral es una variable aleatoria mientras que $x_1 + \dots + x_n$ es la **suma de la muestra** y es un número.

Notemos además, para $\bar{X} = S_n/n$ se tiene que $\mathbb{E}(\bar{X}) = \mu$.

2. Estimadores

Supongamos el caso en que tenemos una distribución con parámetro desconocido θ . La idea es aproximar este parámetro mediante $\hat{\theta}$ mediante los datos de la muestra x_1, \dots, x_n . Por ejemplo supongamos una distribución normal con σ^2 conocida y μ desconocida entonces podríamos pensar una aproximación $\hat{\mu} = \frac{1}{n}(x_1 + \dots + x_n)$, es importante recalcar que $\hat{\mu}$ depende de la muestra.

En general, se define a un **estimador** para un parámetro θ a una variable aleatoria

$$\hat{\Theta} = f(X_1, \dots, X_n),$$

un valor particular de $\hat{\Theta}$ denotado por $\hat{\theta}$ es una estimación de θ .

3. Estimadores insesgados

Definimos el sesgo de un estimador $\hat{\Theta}$ para un parámetro θ por la ecuación

$$Sesg(\Theta) = \mathbb{E}(\Theta) - \theta$$

Si $Sesg(\Theta) = 0$ decimos que el estimador Θ es insesgado, en caso contrario decimos que es sesgado.

Ejercicio: Consideremos una muestra de 100 personas y sea X el número de personas a favor de un candidato político.

1. ¿Cuál es la distribución de X ?
2. Considera a $\hat{P} = X/100$. ¿Es un estimador? ¿Qué tipo de estimador es?

Regresando al ejemplo anterior. Supongamos que tenemos una observación $x = 45$ entonces una estimación es $\hat{p} = 45/100 = .45$. Además supongamos una nueva muestra de 20 personas y considera ahora Y a la cantidad de personas que están a favor de cierto político. Considera ahora los siguientes estimadores para p ,

$$\hat{P}_1 = 1/2 \left[\frac{X + 5Y}{100} \right], \quad (1)$$

$$\hat{P}_2 = 1/2 \left[\frac{X + Y}{120} \right] \quad (2)$$

Supón ahora que tienes observaciones $x = 40$ e $y = 15$ calcula,

1. \hat{p}_1 y \hat{p}_2
2. ¿Cuál de las dos estimaciones escoger? ¿Hay una estimación mejor que otra?
3. Calcula las varianzas de \hat{P}_1 y \hat{P}_2

4. Eficiencia

Dados Θ_1 y Θ_2 dos estimadores insesgados dedimos que el estimador Θ_1 es mas eficiente que $\hat{\Theta}_2$ si

$$\mathbb{V}ar(\hat{\Theta}_1) < \mathbb{V}ar(\hat{\Theta}_2).$$

y la eficiencia relativa es

$$\mathbb{V}ar(\hat{\Theta}_2)/\mathbb{V}ar(\hat{\Theta}_1).$$

Si un estimador insesgado es mas eficiente que cualquier otro, decimos que es absolutamente eficiente. Para estimadores sesgados tenemos la definición

$$\mathbb{E}(\hat{\Theta}_2 - \theta)^2 / \mathbb{E}(\hat{\Theta}_1 - \theta)^2.$$

¿Qué podemos decir cuando $\text{Var}(\hat{\Theta}_2)/\text{Var}(\hat{\Theta}_1) > 1$, $\text{Var}(\hat{\Theta}_2)/\text{Var}(\hat{\Theta}_1) = 1$ o $\text{Var}(\hat{\Theta}_2)/\text{Var}(\hat{\Theta}_1) < 1$?

5. Consistencia

Un estimador $\hat{\Theta}$ es **consistente** de un parámetro θ

$$\mathbb{E}(\hat{\Theta} - \theta)^2 \rightarrow 0$$

cuando $n \rightarrow \infty$. A la ecuación anterior se le llama usualmente como **error cuadrático medio** y es una medida de concentración de $\hat{\Theta}$ alrededor de θ .

En general se relaciona con la varianza de la siguiente manera.

$$\mathbb{E}(\hat{\Theta} - \theta)^2 = (\mathbb{E}(\hat{\Theta}) - \theta)^2 + \text{Var}(\hat{\Theta})$$