Inferencia

Instituto Artek

Abril 17

Índice

1. Introducción

Ejercicio: En una universidad 1/3 de los estudiantes toma 9 horas de crédito, 1/3 toma 12 horas de crédito y 1/3 toma 15 horas de crédito. Si X representa las horas de crédito que toma un estudiante, la distribución de X es 1/3 para $x = 9, 12$ y 15. Encontrar la media y la varianza de X. ¿Qué tipo de distribución tiene X?

2. Repaso

Un **estimador** para un parámetro θ es una variable aleatoria,

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

un valor particular de $\hat{\Theta}$ denotado por $\hat{\theta}$ es una estimación de θ. El sesgo de un estimador $\hat{\Theta}$ para un parámetro θ por la ecuación,

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

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

3. Consistencia

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

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

cuando $n \to \infty$. 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.

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

Ejercicio:
$\mathbb{E}(\overline{X} - \mu) = \sigma^1/n$.

La relación de consistencia con la probabilidad se da mendiante la desigualdad de Chebyshev, para todo $\varepsilon > 0$,

$$\mathbb{P}(|\hat{\Theta}-\theta|>\varepsilon)\leq\frac{1}{\varepsilon^2}\mathbb{E}(\hat{\Theta}-\theta)^2,$$

esto también nos da consistencia en probabilidad, pues tenemos que si un estimador $\hat{\Theta}$ es consistente entonces

$$\mathbb{P}(|\hat{\Theta}-\theta|>\varepsilon)\to 0$$

cuando $n\to\infty$.

4. Intervalos de confianza

Los intervalos de confianza se interpretan como el nivel de confianza sobre la estimación de un parámetro. Por ejemplo, tomada la media muestral $\overline{X}$ y una estimación $\overline{x}=,86$ queremos encontrar un intervalo del tipo $,86\pm,4$, ahora nuestro interés se enfoca en calcular un error mediante los valores de la muestra. Digamos que nos interesa saber si dada una muestra, tenga media muestral entre $\mu-\varepsilon$ y $\mu+\varepsilon$ en un nivel de confianza del 95 %, en términos de probabilidades es

$$\mathbb{P}(\mu-\varepsilon<\overline{X}<\mu+\varepsilon)=,95$$

de manera equivalente

$$\mathbb{P}(\overline{X}-\varepsilon<\mu<\overline{X}+\varepsilon)=,95$$

cuando la muestra es muy grande, $\overline{X}$ tiene una distribución normal con media μ y varianza σ^2/n y así

$$\mathbb{P}(\frac{-\varepsilon}{\sigma/\sqrt{n}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<\frac{\varepsilon}{\sigma/\sqrt{n}})=,95$$

por el teorema limite central tenemos que la distribución limite $\overline{X}^*=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ es una $Norm(0,1)$.

Ejercicio: Halle el valor de $\frac{\varepsilon}{\sigma/\sqrt{n}}$ tal que

$$\mathbb{P}(0\leq\overline{X}^*\leq\frac{\varepsilon}{\sigma/\sqrt{n}})=,475$$

o

$$2 * \mathbb{P}(0 \leq \overline{X}^* \leq \frac{\varepsilon}{\sigma/\sqrt{n}}) = ,95$$

a los extremos $\overline{x} \pm \frac{\varepsilon}{\sigma/\sqrt{n}}$ se les llama limites de confianza.

Ejercicio: El departamento de tránsito de una cierta ciudad realizó una investigación para estimar el tiempo promedio μ, de reacción de un automovilista que conduce a una velocidad dada. Se hizo una estimación tomando una muestra aleatoria a 100 automovilistas donde se encontró que los tiempo de reacción son $\overline{x} = ,86$ y $s^2 = ,09$. Halle los limites de confianza del 95 %. Además del intervalo $\overline{x} - \varepsilon < \mu < \overline{x} - \varepsilon$.