Probabilidad y Estadística

Clase 7

Gonzalo Blanco

2023

1. Teorema del Límite central

Dada una población, si tomamos muestras lo suficientemente grandes (mayores a 30 elementos) entonces la media muestral sigue una distribución gaussiana, no importa la distribución de probabilidad que siga la población. Además el valor medio de la distribución gaussiana de la media muestral tiende a la media de la población.

En términos de variables aleatorias lo puedo describir de la siguiente forma. Sean $X_1, X_2, ..., X_n$ variables aleatorias independientes con $E[X_i] = \mu$ y $V[X_i] = \sigma^2$ $\forall i$. No importa la distribución de estas variables aleatorias

$\Longrightarrow \bar{X} = \frac{\sum X_i}{n}$ sigue una distribución gaussiana con $E[\bar{X}] = \mu$ y varianza $Var(\bar{X}) = \frac{\sigma^2}{n}$.

Por supuesto que si $\bar{X} = \frac{\sum X_i}{n}$ sigue una distribución gaussiana entonces $S_n = \sum X_i$, tambíen seguirá una distribución gaussiana, en este caso con $E[S_n] = n\mu$ y $V[S_n] = n\sigma^2$.

Recordemos que estas variables aleatorias $\bar{X}$ y S_n, que siguen sus respectivas distribuciones gaussianas, las podemos transformar en variables aleatorias que sigan una distribución normal estándar N(0,1).

$\Longrightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ y $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, ambas siguen una N(0,1).

Ejemplo: Aplicado a una población.

Supongamos que tengo una población, tomo una muestra que llamo M_1 de 50 elementos seleccionados al azar, y a esta muestra le calculo su media muestral que llamo μ_1 y su varianza muestral σ_1^2. Repito el procedimiento, tomo una muestra M_2 de 50 elementos y calculo μ_2 y σ_2^2.

Supongamos entonces que repito esto 1000 veces, es decir, tome mil muestras y a cada una le estimé su media y su varianza muestral.

Entonces tengo el siguiente conjunto de datos

$M_1 \longrightarrow \mu_1$, σ_1^2

$M_2 \longrightarrow \mu_2$, σ_2^2

...

...

...
$M_{1000} \longrightarrow \mu_{1000}\ ,\ \sigma^2_{1000}$

Entonces si ahora hago un histograma de las **medias muestrales**, el mismo va a tener una forma aproximadamente gaussiana con valor medio $\mu = \frac{\sum \mu_i}{1000}$ y varianza $\sigma^2 = \frac{\sum \sigma_i^2}{1000}$.

Además, si tomo una muestra de forma aleatoria a cada elemento de mi muestra lo puedo considerar una variable aleatoria, ya que cuando se "saco" cada elemento i-ésimo para formar la muestra, este elemento podía ser cualquiera de mi población, por lo tanto a cada elemento x_i de una muestra lo puedo considerar como una varibale aleatoria independiente.

Ejemplo: de aplicación.
Supongamos que tengo 30 instrumentos electrónicos $D_1, D_2, D_3, ..., D_{30}$ que conforman un circuito de la siguiente manera: primero está en funcionamiento D_1, tan pronto como falla se activa D_2 y así sucesivamente hasta D_{30}, cuando falla el último instrumento entonces el circuito deja de operar. Supongamos que el tiempo de falla de cada D_i sigue una distribución exponencial con parámetro $\lambda = 0{,}1$ horas. Sea T : *el tiempo total de operación del circuito.* ¿ Cuál es la probabilidad de que el circuito opere más de 350 horas?

X_i :*tiempo de falla del instrumento i-ésimo.*
$X_i \sim Exp(0{,}1)$

$T = \sum\limits_{i=1}^{30} X_i$
$E[T] = E[\sum\limits_{i=1}^{30} X_i] = \sum\limits_{i=1}^{30} E[X_i] = 30\frac{1}{\lambda} = 300$
$V[T] = V[\sum X_i]$, como los X_i son independientes entre ellos.
$V[T] = V[\sum X_i] = V[T] = \sum V[X_i] = 30\frac{1}{\lambda^2} = 3000$

$\Longrightarrow$ por TLC la variable aletoria $Z = \frac{T-300}{3000} \sim N(0,1)$

$P(T > 350) = P(\frac{T-300}{\sqrt{3000}} > \frac{350-300}{\sqrt{3000}}) = 1 - P(\frac{T-300}{\sqrt{3000}} < \frac{350-300}{\sqrt{3000}}) = 1 - \Phi(\frac{350-300}{\sqrt{3000}}) = 1 - \Phi(0,9128) = 0,1814.$