Probabilidad y Estadística

Agustín G. Bonifacio

UNSL

Teoremas Límites

Teorema (Desigualdad de Markov)

Sea X una v.a. que toma sólo valores mayores o iguales que cero. Entonces, para cualquier $a > 0$,

$$P([X \geq a]) \leq \frac{E(X)}{a}.$$

Teorema (Desigualdad de Chebyshev)

Sea X es una v.a. con $E(X) = \mu$ y $Var(X) = \sigma^2$. Entonces, para cualquier $k > 0$,

$$P([|X - \mu| \geq k]) \leq \frac{\sigma^2}{k^2}.$$

Ejemplo

Supongamos que sabemos que el número de ítems producidos en una fábrica en una semana es una v.a. con media igual a 50.

1. ¿Qué puede decirse acerca de la probabilidad de que la producción de esta semana sea de al menos 1.000 unidades?
2. Si la varianza de la producción semanal es igual a 100, ¿qué se puede decir sobre la probabilidad de que la producción de esta semana esté entre 400 y 600?

- La *ley débil de los grandes números* estudia la forma en que una sucesión de promedios de v.a. converge a un determinado valor.

Teorema (Ley Débil de los Grandes Números)

Sea $X_1, \ldots, X_n, \ldots$ una sucesión de v.a. independientes e idénticamente distribuidas. Supongamos que $E(X_i) = \mu$, $i = 1, \ldots, n, \ldots$ es finita. Entonces, para cualquier $\varepsilon > 0$,

$$P\left(\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right]\right) \longrightarrow 0 \text{ cuando } n \to \infty.$$

- La aplicación que daremos provee una justificación del uso de las frecuencias relativas como aproximaciones de la probabilidad de un evento. Anteriormente, la justificación era realizada a través de la presencia del fenómeno de *regularidad estadística*.

Sea $(\Omega, \mathcal{A}, P)$ un e.p. y $A \in \mathcal{A}$ un evento. Sea $p = P(A)$. Definamos la v.a. $X_A : \Omega \longrightarrow \mathbb{R}$ de la siguiente manera:

$$X_A(\omega) = \begin{cases} 1 & \text{si} \quad \omega \in A, \\ 0 & \text{o.c.} \end{cases}$$

Es fácil ver que $X_A \sim$ Bernoulli(p). Sea $X_1, \ldots, X_n, \ldots$ una muestra aleatoria de tamaño infinito para X_A. Como $E(X_i) = p$, tenemos

$$P\left(\left[\left|\frac{X_1 + \ldots + X_n}{n} - p\right| > \varepsilon\right]\right) \longrightarrow 0 \text{ cuando } n \to \infty.$$

Pero tener una muestra aleatoria infinita $X_1, \ldots, X_n, \ldots$ es equivalente a tener una muestra infinita $\omega_1, \ldots, \omega_n, \ldots$ y hacer $X_i = X_A(\omega_i)$. Por lo tanto,

$$\frac{X_1 + \ldots + X_n}{n} = \frac{N_A}{n}$$

de donde se deduce que

$$P\left(\left[\left|\frac{N_A}{n} - p\right| > \varepsilon\right]\right) \longrightarrow 0 \text{ cuando } n \to \infty.$$

- Dada una muestra aleatoria de n observaciones de una población con media μ y desviación estándar σ, queremos estudiar el comportamiento de la media muestral $\overline{X}_n$.
- Para la nueva variable aleatoria $\overline{X}_n$, se puede mostrar que:
  1. su media es igual a la media de la población muestreada, esto es, $E(\overline{X}_n) = \mu$,
  2. su desvío estándar es igual a $\sigma/\sqrt{n}$,
  3. si la población se distribuye de forma normal, entonces la media muestral también se distribuye de forma normal.
- Si la población no se distribuye de forma normal, ¿podemos decir algo más acerca de la distribución de $\overline{X}_n$? Sí.

La media muestral tenderá a distribuirse como una normal al aumentar el tamaño de la muestra, *independientemente de cuál sea la distribución de la población de la que estamos muestreando.*

Teorema Central de Límite (TCL)

Sea $X_1, \ldots, X_n, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuídas. Supongamos que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para cada $i = 1, \ldots, n, \ldots,$ y que tanto μ como σ^2 son finitas. Sea $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Entonces, la distribución de

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

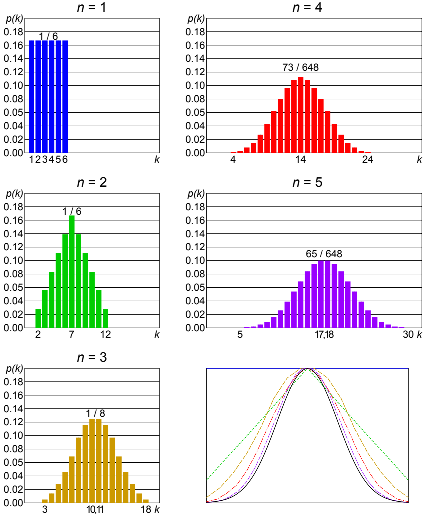
tiende a una normal estardarizada, esto es,

$$P\left(\left[a < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < b\right]\right) \longrightarrow \Phi(b) - \Phi(a) \;\; \text{cuando} \;\; n \to \infty,$$

donde $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$.

Para n grande,

- $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ tiende a distribuirse como $N(0,1)$,
- $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ tiende a distribuirse como $N(0,1)$.



Distribución de muestreo de suma de variables con distribución uniforme. Aquí vemos como al aumentar el tamaño de la muestra de $n = 1$ a $n = 5$, la distribución de las sumas converge (rápidamente) a la normal.

¿Qué tipo de preguntas nos permite responder (de manera aproximada) el TCL?

Problema: Supongamos que los sueldos en una gran empresa tienen media \$62.000 y desvío estándar de \$32.000.

1. Si se elige un empleado al azar, ¿cuál es la probabilidad de que su salario exceda los \$66.000?
2. Si se eligen 100 empleados al azar, ¿cuál es la probabilidad de que su salario medio exceda los \$66.000?

- La primera pregunta no se puede responder al menos que sepamos cómo se distribuyen los salarios.
- La segunda pregunta puede aproximarse utilizando el TCL.

¿Por qué es importante el TCL?

- El teorema fundamenta el uso extendido de la distribución normal en procedimientos de inferencia estadística aún en casos donde la población analizada podría tener una distribución que no es normal.

Aproximación Normal de la Distribución Binomial

Distribución Binomial con parámetros (n, p):

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{para} \quad x = 0, 1, \ldots, n.$$

$E(X) = np,$
$Var(X) = np(1-p).$

- Como toda $X \sim Bin(n,p)$ puede verse como una suma $X = \sum_{j=1}^{n} X_j$ con $X_j \sim Bern(p)$, el TCL nos asegura que para valores suficientemente grandes de n podemos aproximar la distribución binomial a través de la distribución normal.
- Importancia histórica: primer TCL (DeMoivre-Laplace).
- Importancia práctica: Cálculo de combinatorios difícil para n grande.
- Aproximación buena para valores de p cercanos a $\frac{1}{2}$ (sesgo).

Problema: Si se tiran 10 dados equilibrados, encontrar la probabilidad aproximada de que la suma de sus caras superiores esté entre 30 y 40. (Recuerde que si X_i es el valor de la cara superior del i-ésimo dado, entonces $E(X_i) = \frac{7}{2}$ y $Var(X_i) = \frac{35}{12}$).