4 - DESIGUALDAD DE CHEBYSHEV- LEY DE LOS GRANDES NUMEROS

La desigualdad de Chebyshev es una importante herramienta teórica. Entre otras aplicaciones constituirá un medio para comprender cómo la varianza mide la variabilidad de una dada variable aleatoria, con respecto a su esperanza matemática. También nos permitirá establecer con más precisión el hecho, reiteradamente señalando, de que la frecuencia relativa f_A de un suceso A asociado a un experimento aleatorio ε tiende, cuando el número de repeticiones de ε se hace infinitamente grande, a la probabilidad $P(A)$ (resultado conocido como la Ley de los grandes números). Pero además es de utilidad práctica pues, al constituir una cota de ciertas probabilidades, nos podrá servir como una estimación de esas mismas probabilidades.

4.1-Desigualdad de Chebyshev

Sea X una variable aleatoria cuya esperanza es $E(X)$, sea c un número real cualquiera y supongamos que $E\left[(X-c)^2\right]$ existe y es finito. Entonces

$$\boxed{\forall\varepsilon>0 \qquad P(|X-c|\geq\varepsilon)\leq\frac{1}{\varepsilon^2}E\left[(X-c)^2\right]}$$

Dem.) Consideraremos el caso en que X es una v.a. continua. El caso de una v.a. discreta se demuestra en forma similar cambiando integrales por sumas.

Sea, entonces, $f(x)$ la *fdp* de X. Tenemos:

$$P(|X-c|\geq\varepsilon)=\int_{x:|x-c|\geq\varepsilon} f(x)dx$$

Ahora bien, los valores de x que verifican $|x-c|\geq\varepsilon$ son los mismos que verifican $\frac{(x-c)^2}{\varepsilon^2}\geq 1$. Entonces, puesto que $f(x)\geq 0$ y que ambos miembros en la desigualdad anterior son también no negativos es:

$$P(|X-c|\geq\varepsilon)=\int_{x:|x-c|\geq\varepsilon} 1.f(x)dx\leq\int_{x:|x-c|\geq\varepsilon}\frac{(x-c)^2}{\varepsilon^2}.f(x)dx\leq\int_{-\infty}^{\infty}\frac{(x-c)^2}{\varepsilon^2}.f(x)dx$$

donde la última desigualdad proviene del hecho de que simplemente extiendo los límites de integración a todos los reales pero siendo siempre el integrando positivo. De manera que estoy agregando una contribución positiva sobre el valor de la integral $\int_{x:|x-c|\geq\varepsilon}\frac{(x-c)^2}{\varepsilon^2}.f(x)dx$.

Si tenemos presente la expresión para la esperanza de una función $H(X)$ de una variable aleatoria X:

$\int_{-\infty}^{\infty} H(x)f(x)dx = E[H(X)]$ y lo aplicamos a $H(x) = \frac{(x-c)^2}{\varepsilon^2}$, tenemos finalmente:

$$P(|X-c| \geq \varepsilon) \leq E\left[\frac{(X-c)^2}{\varepsilon^2}\right] = \frac{1}{\varepsilon^2}E\left[(X-c)^2\right]$$

Observación: La desigualdad lleva el nombre del matemático ruso que la descubrió. Su nombre aparece en una variedad de formas en la literatura: Chebyshev, Chebychev, Tchebyshev, etc.

Formas alternativas de la desigualdad de Chebyshev.

Podemos escribir la desigualdad de Chebyshev en una serie de formas alternativas:

a_1) Tenemos en primer lugar la forma que acabamos de demostrar:

$$P(|X-c| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}E\left[(X-c)^2\right]$$

a_2) Si consideramos el suceso complementario a $|X-c| \geq \varepsilon$, es decir $|X-c| < \varepsilon$, podemos escribir, recordando que $P(\overline{A}) = 1 - P(A)$:

$P(|X-c| < \varepsilon) = 1 - P(|X-c| \geq \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}E\left[(X-c)^2\right]$, esto es:

$$P(|X-c| < \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}E\left[(X-c)^2\right]$$

b_1) Si en a_1) elegimos $c = E(X)$ tenemos $P(|X-E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}E\left([X-E(X)]^2\right)$ y recordando la definición de la varianza: $V(X) = E\left([X-E(X)]^2\right)$ tenemos:

$$P(|X-E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

b_2) La correspondiente expresión para el suceso complementario es:

$$P(|X-E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2}$$

c_1) Si en a_1) elegimos $c = E(X)$ y además elegimos $\varepsilon = k\sigma_X = k\sqrt{V(X)}$, tenemos $P(|X-E(X)| \geq k\sigma_X) \leq \frac{V(X)}{(k\sigma_X)^2} = \frac{\sigma_X^2}{(k\sigma_X)^2}$, es decir:

$$P(|X - E(X)| \geq k\sigma_X) \leq \frac{1}{k^2}$$

c_2) Finalmente a correspondiente expresión para el suceso complementario en c_1) es:

$$P(|X - E(X)| < k\sigma_X) \geq 1 - \frac{1}{k^2}$$

A continuación daremos un ejemplo en el que podremos apreciar cómo la desigualdad de Chebyshev nos permite tener estimaciones de ciertas probabilidades que en algunos casos mejoran las "estimaciones" triviales dadas por el axioma i) de las probabilidades, esto es, que $0 \leq P(A) \leq 1$.

Ejemplo Deseamos estimar la probabilidad $P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X \right)$

a) Sin conocer la distribución
b) Sabiendo que $X \sim U\left[1 - \frac{1}{\sqrt{3}},\ 1 + \frac{1}{\sqrt{3}}\right]$.

a) Una estimación de la probabilidad en consideración cuando no se conoce la distribución, es decir que vale cualquiera sea la distribución puede tenerse usando la desigualdad de Chebyshev en la forma c_1) : $P(|X - E(X)| \geq k\sigma_X) \leq \frac{1}{k^2}$ con $k = \frac{3}{2}$.
Tenemos en consecuencia:

$P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X \right) \leq \frac{1}{(3/2)^2} = \frac{4}{9} = 0.44$. Vemos que, en este caso, la estimación mejora sustancialmente la cota superior trivial 1:

$$P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X \right) \leq 0.44 < 1$$

Notemos que esta estimación es aplicable cualquiera se la distribución de X y vale sea ésta discreta o continua.

b)Si sabemos que $X \sim U\left[1 - \frac{1}{\sqrt{3}},\ 1 + \frac{1}{\sqrt{3}}\right]$ tenemos toda la información y podemos calcular la probabilidad exactamente.

Tenemos:

$$E(X) = \frac{\left(1 + \frac{1}{\sqrt{3}}\right) + \left(1` - \frac{1}{\sqrt{3}}\right)}{2} = 1$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{\frac{\left[\left(1+\frac{1}{\sqrt{3}}\right)-\left(1`-\frac{1}{\sqrt{3}}\right)\right]^2}{12}} = \sqrt{\frac{\left(2/\sqrt{3}\right)^2}{12}} = \sqrt{\frac{4}{36}} = \sqrt{\frac{1}{9}} = \frac{1}{3}$$. Entonces:

$$P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X\right) = P\left(|X-1| \geq \frac{1}{2}\right) = 1 - P\left(|X-1| < \frac{1}{2}\right) = 1 - P\left(\frac{1}{2} \leq X \leq \frac{3}{2}\right)$$

$$P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X\right) = 1 - \frac{\frac{3}{2}-\frac{1}{2}}{\left(1+\frac{1}{\sqrt{3}}\right)-\left(1-\frac{1}{\sqrt{3}}\right)} = 1 - \frac{1}{\frac{2}{\sqrt{3}}} = 1 - \frac{\sqrt{3}}{2} = 0.134$$. Este valor exacto es menor que la cota superior dada por la desigualdad de Chebyshev:

$$P\left(|X - E(X)| \geq \frac{3}{2}\sigma_X\right) = 0.134 \leq 0.44 < 1.$$

La varianza como una medida de la concentración de la *fdp* de una v.a. alrededor de la esperanza.

Podemos usar las formas *b*) de la desigualdad de Chebyshev para interpretar a la varianza $V(X)$ como una medida de la variabilidad de la variable aleatoria X con respecto a su esperanza o en otras palabras de cómo la distribución de la v.a. X se concentra o dispersa con respecto a la esperanza $E(X)$.

De la expresión $P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$ vemos que, para un ε dado, si $V(X)$ es muy pequeño entonces la probabilidad de que X tome valores lejos de $E(X)$ es muy chica, es decir hay una gran probabilidad de que X tome valores próximos a $E(X)$. Inversamente si $V(X)$ es grande, la probabilidad de que X tome valores alejados de $E(X)$ puede ser también grande. Podemos precisar un poco más esto considerando el siguiente

Teorema. Si $V(X) = 0$ entonces $P[X = E(X)] = 1$. Decimos que $X = E(X)$ con probabilidad 1 (X es igual a su esperanza con probabilidad 1).

Dem.) Para cualquier $\varepsilon > 0$, si $V(X) = 0$ tenemos de b_2) :

$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2} \to P(|X - E(X)| < \varepsilon) = 1$, donde me quedo sólo con la igualdad porque la probabilidad no pude superar a 1.

Puesto que ε puede hacerse arbitrariamente pequeño, el teorema queda demostrado.

4.2 - La ley de los grandes números.

La ley de los grandes números establece en forma precisa el hecho que cuando el número de repeticiones de un experimento se hace muy grande, la frecuencia relativa f_A de un suceso *A* relacionado con el experimento converge en sentido probabilístico a la probabilidad $P(A)$. Daremos una versión de la Ley de los grandes números conocida como la ***forma de Bernoulli***.

Teorema (Forma de Bernoulli de la ley de los grandes números)
Sea un experimento probabilístico y sea *A* un suceso asociado con él. Consideremos *n* repeticiones independientes del experimento. Sea n_A el número de veces que ocurre *A* en las *n* repeticiones de forma tal que $f_A = \frac{n_A}{n}$ es la frecuencia relativa. Sea $P(A) = p$ (que se supone igual para todas las repeticiones). Entonces, para cualquier número $\varepsilon > 0$ se cumple
$$P\left(\left|f_A - p\right| < \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}.$$
Dem.)
De acuerdo con su significado n_A, el número de veces que ocurre el suceso *A* en las *n* repeticiones, es una variable aleatoria distribuida binomialmente con parámetros *n* y *p*: $n_A \sim B(n,p)$. Luego:

$$\begin{cases} E(n_A) = np \\ V(n_A) = np(1-p) \end{cases}$$

Por lo tanto
$$\begin{cases} E(f_A) = E\left(\frac{n_A}{n}\right) = \frac{1}{n}E(n_A) = p \\ V(f_A) = V\left(\frac{n_A}{n}\right) = \frac{1}{n^2}V(n_A) = \frac{p(1-p)}{n} \end{cases}$$
En consecuencia, aplicando a la v.a. f_A la desigualdad de Chebyshev en la forma b_2)
$P\left(\left|X - E(X)\right| < \varepsilon\right) \geq 1 - \frac{V(X)}{\varepsilon^2}$, es decir,
$P\left(\left|f_A - E(f_A)\right| < \varepsilon\right) \geq 1 - \frac{V(f_A)}{\varepsilon^2}$ llegamos a lo propuesto por el teorema:
$$P\left(\left|f_A - p\right| < \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}.$$

Es evidente que el *teorema* anterior implica que
$\lim_{n\to\infty} P\left(\left|f_A - p\right| < \varepsilon\right) = 1$ para todo $\varepsilon > 0$.

Entonces decimos que, en este sentido, la frecuencia relativa f_A converge a la probabilidad $P(A)$.

Observación: Esta convergencia, llamada ***convergencia en probabilidad*** difiere de la convergencia normalmente usada en Cálculo (límite aritmético).
Recordemos que una sucesión $\alpha_1, \alpha_2, ..., \alpha_n, ...$ tiene límite α o también que $\lim_{n \to \infty} \alpha_n = \alpha$ si $\forall \varepsilon > 0 \quad \exists \nu$ tal que $|\alpha_n - \alpha| < \varepsilon$ para todo $n > \nu$. Esto significa que, desde un n en adelante, α_n se aproxima permanentemente al valor límite α. En cambio cuando decimos que $f_A = n_A / n$ converge a $P(A)$, estamos significando que la probabilidad del suceso $\left\{ |f_A - P(A)| < \varepsilon \right\}$ puede hacerse arbitrariamente próximo a uno tomando un n suficientemente grande. Pero estamos hablando de probabilidad y no de certeza como en el caso del límite aritmético. Es decir, no significa que al tomar un *n* grande ocurra ciertamente que nos aproximemos más al valor de la probabilidad sino que existe una gran probabilidad de que eso ocurra.