Probabilidad

Desigualdades
12/05/2023

Lara Sánchez Peña[1]

UTDT 2023 - MEC

[1] Basado en las notas de Andrea Rotnitzky

Desigualdad de Cauchy-Schwarz

Sean X e Y dos variables aleatorias. Entonces

$$E[XY]^2 \leq E\left[X^2\right] E\left[Y^2\right]$$

Demostración:

Sea $0 \leq f(s) = E\left[(sX+Y)^2\right]$ para cualquier número real s. Entonces

$$\begin{aligned} f(s) &= E\left[(sX+Y)^2\right] \\ &= E\left[s^2X^2 + 2sXY + Y^2\right] \\ &= E\left[X^2\right]s^2 + 2E[XY]s + E\left[Y^2\right] \end{aligned}$$

Notemos que $f(s)$ es una función cuadrática en s, $f(s) = as^2 + bs + c$. Como $f(s) \geq 0$ si y sólo si $b^2 - 4ac \leq 0$, obtenemos que

$$(2E[XY])^2 - 4E\left[X^2\right]E\left[Y^2\right] \leq 0 \quad \Rightarrow \quad E[XY]^2 \leq E\left[X^2\right]E\left[Y^2\right]$$

Este resultado sirve para demostrar que el coeficiente de correlación $\rho \in [-1, 1]$.

El resultado C-S Aplicado a $X-E(X)$, $Y-E(Y)$

$$Cov(X,Y)^2 \leq Var(X)\,Var(Y)$$

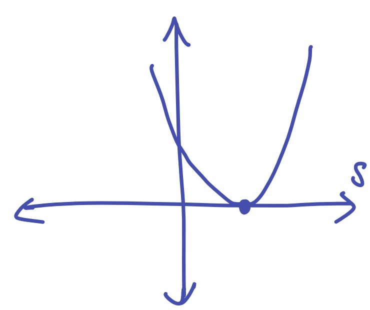
$$\rho^2 = \frac{E([X-E(X)][Y-E(Y)])^2}{E([X-E(X)]^2)\,E([Y-E(Y)]^2)} \leq 1$$

$$\rho^2 \leq 1 \iff -1 \leq \rho \leq 1.$$

Consideremos

$$0 \leq f(s) = E\left((sX+Y)^2\right) = E\left(s^2X^2 + 2sXY + Y^2\right)$$

para cualquier s



$$0 \leq s^2E(X^2) + 2sE(XY) + E(Y^2)$$

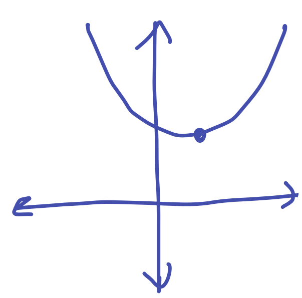
$$0 \leq as^2 + bs + c$$

Vamos a notar que, llamando $a = E(X^2) \geq 0$

$b = 2E(X\cdot Y)$

$c = E(Y^2) \geq 0$

o



las raíces salen de hacer

$$\frac{-b \pm \sqrt{b^2-4ac}}{2a} \implies$$ tiene que pasar que $b^2 - 4ac \leq 0$

$$[2E(X \cdot Y)]^2 - 4E(X^2)E(Y^2) \leq 0$$

$$\cancel{4}(E(XY))^2 \leq \cancel{4}E(X^2)E(Y^2)$$

Funciones cóncavas y convexas

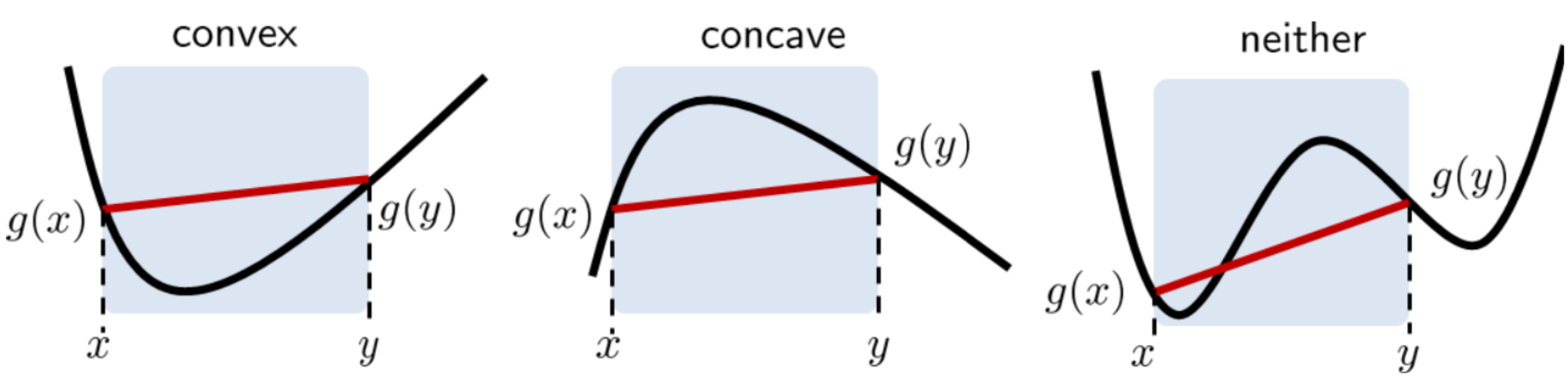
Una función $g(\cdot)$ es **convexa** si para cualquiera x e y, $\alpha \in [0,1]$ vale que

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y)$$

Si $g(x)$ es una función derivable dos veces, es convexa si $g''(x) > 0$.
Una función $g(x)$ es **cóncava** si $-g(x)$ es convexa.

Ejemplos de funciones cóncavas o convexas:

- $g(x) = \ln(x)$ es cóncava porque $g''(x) = -\frac{1}{x^2} < 0$
- $g(x) = x^2$ es convexa porque $g''(x) = 2 > 0$
- $g(x) = e^{-x}$ es convexa porque $g''(x) = e^{-x} > 0$
- $g(x) = \frac{1}{x}$ es convexa si $x > 0$ porque $g''(x) = \dfrac{2}{x^3} > 0$



Desigualdad de Jensen

ejemplo $g(x)=x^2$

$E(X^2) \geq [E(X)]^2 \iff Var(X) \geq 0.$

Sea X una v.a. y sea $g : \mathbb{R} \to \mathbb{R}$ una función **convexa**. Entonces

$$E(g(X)) \geq g(E(X))$$

Sea X una v.a. y sea $g : \mathbb{R} \to \mathbb{R}$ una función **cóncava**. Entonces

$$E(g(X)) \leq g(E(X))$$

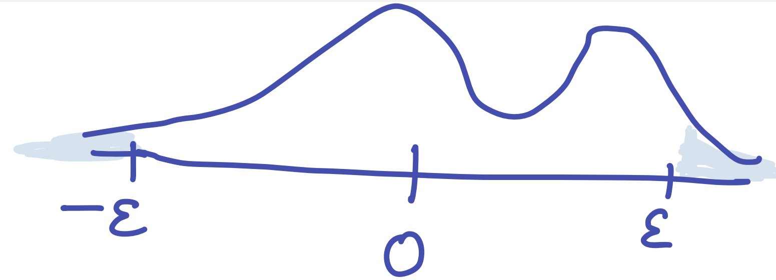
$g(x) = \ln x$

$E(\ln X) \leq \ln[E(X)].$

Entonces, usando Jensen, podemos afirmar que

- $E(\ln(X)) \leq \ln(E(X))$, tomando $g(x) = \ln(x)$
- $E(X^2) \geq (E(X))^2$, tomando $g(x) = x^2$
- $E(e^{-X}) \geq e^{-E(X)}$, tomando $g(x) = e^{-x}$
- $E\left(\dfrac{1}{X}\right) \geq \dfrac{1}{E(X)}$, tomando $g(x) = \dfrac{1}{x}$

Desigualdades de Markov y Chebyshev



Desigualdades de Markov y Chebyshev

1. **Markov:** Si $E(X) < \infty$, para todo $\varepsilon > 0$

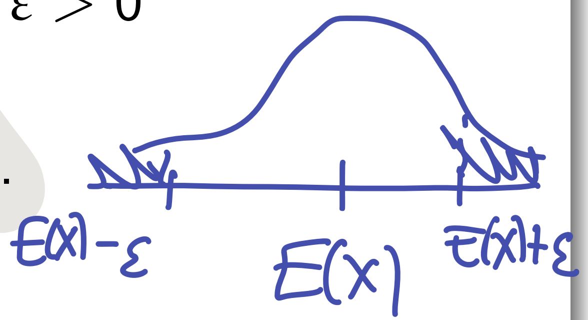
Si $E(X)<\infty$ entonces X está acotada en probabilidad.

$$P\left(|X| \geq \varepsilon\right) \leq \frac{E(|X|)}{\varepsilon}.$$

2. **Chebyshev:** Si $E(X) < \infty$ y $Var(X) < \infty$, para todo $\varepsilon > 0$

Si X tiene $E(X)$ y $Var(X)$ finitas $\Rightarrow$ "es poco probable" que X esté "lejos" de $E(X)$.

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{Var(X)}{\varepsilon^2}.$$



Equivalentemente, para todo $\varepsilon > 0$

$$P\left(|X - E(X)| \geq \sqrt{Var(X)}\varepsilon\right) \leq \frac{1}{\varepsilon^2}.$$

Nota: Es importante tener en cuenta que estas desigualdades valen si conocemos la esperanza y/o la varianza de la v.a. X.

Sea X el ingreso mensual de una familia elegida al azar en Arg.

$\varepsilon = 1/2\, E(X)$ $\qquad P\left(X \geq 1/2\, E(X) \right) \underset{\substack{\downarrow \\ \text{Markov}}}{\leq} \dfrac{E(X)}{1/2\, E(X)} = 2$

Como $X \geq 0$

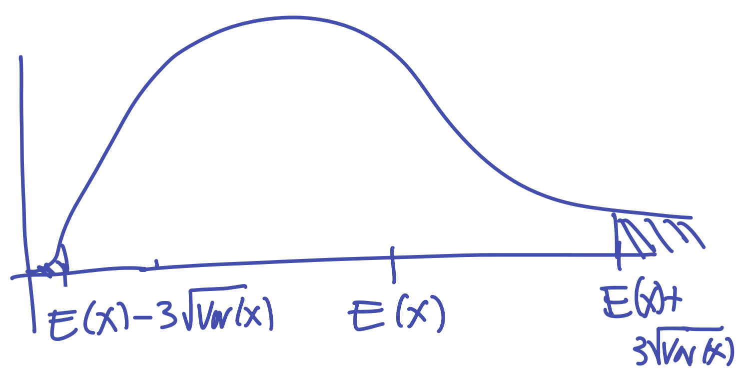
$|X| = X$.

$\varepsilon = 8\, E(X)$ $\qquad P\left(X \geq 8E(X) \right) \underset{\substack{\downarrow \\ \text{Markov}}}{\leq} \dfrac{E(X)}{8E(X)} = \dfrac{1}{8}$.

Usando Tchebyshev

$$P\left(|X - E(X)| \geq 3\sqrt{Var(X)} \right) \leq \frac{Var(X)}{3^2\, Var(X)} = \frac{1}{9}.$$

$\varepsilon = 3\sqrt{Var(X)}$



1° ll

- Presentás
  - Aprobar !!
  - no aprobás → vas al 2° llamado
    - aprobar 4 o C
    - desaprobás 2 o D
- no presentas
  - aprobás !!
  - no aprobás → D
  - no te presentás → U. Ausente.
    no afecta al promedio.

Desigualdad de Markov

- Markov: **La probabilidad de que X asuma valores muy por encima de $E(X)$ es relativamente pequeña.**

Por ejemplo: Sea X el ingreso mensual de un individuo elegido al azar de la población Argentina. Si tomamos $\varepsilon = 2E(X)$, la desigualdad de Markov[2] nos dice que

$$P\left(X \geq 2E(X)\right) \leq 1/2,$$

Es decir, no es posible que más de la mitad de la población tenga un ingreso de al menos el doble del ingreso promedio.

Resultado general: Si g es una función no negativa, $P(g(X) \geq 0) = 1$, donde $E(g(X)) < \infty$, luego

$$P\left(g(X) \geq \varepsilon\right) \leq \frac{E(g(X))}{\varepsilon}.$$

[2] Notar que, como $X \geq 0$, $|X| = X$.

Desigualdad de Chebyshev

- Chebyshev: **Intuitivamente, la desigualdad de Chebyshev dice que $Var(X)$ nos da una cota de la probabilidad de que X tome valores alejados de su esperanza. Si $Var(X)$ es pequeña, entonces es poco probable que X tome valores alejados de $E(X)$..**

Notar que la desigualdad de Chebyshev es un caso particular de la desigualdad de Markov. Usando la desigualdad de Markov y considerando $g(X) = |X - E(X)|$:

$$P\left(g(X) \geq \varepsilon\right) = P\left(|X - E(X)|^2 \geq \varepsilon^2\right)$$

$$\underbrace{\leq}_{\text{Markov}} \frac{E\left([X - E(X)]^2\right)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}. \quad \square$$

Para el ejemplo de los ingresos mensuales de la slide anterior: ¿cuál es como máximo la probabilidad de que el salario de una persona elegida al azar esté a más de 3 desvíos estándar de la media de salarios en la población?

Demostrem la desigualdad de Tchebychev

$$P(|X-E(X)| \geq \varepsilon) = P(|X-E(X)|^2 \geq \varepsilon^2)$$

$$\downarrow$$

$$|X-E(X)| \geq \varepsilon \iff |X-E(X)|^2 \geq \varepsilon^2$$

$$P(|X-E(X)| \geq \varepsilon)$$

$$= P(|X-E(X)|^2 \geq \varepsilon^2) \underset{\text{Markov}}{\leq} \frac{E(|X-E(X)|^2)}{\varepsilon^2} = \frac{E((X-E(X))^2)}{\varepsilon^2}$$

Markov

$$P(|Y| \leq a) = \frac{E(|Y|)}{a}$$

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

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

Desigualdad de Chebyshev para $\overline{X}_n$ si $\underbrace{X_i \overset{iid}{\sim}}$

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}, \quad E(\overline{X}_n) = E(X), \quad Var(\overline{X}_n) = \frac{Var(X)}{n}.$$

- Para la variable aleatoria $\overline{X}_n$, que recordemos verifica $E(\overline{X}_n) = E(X)$ y $Var(\overline{X}_n) = Var(X)/n$, por Chebyshev se cumple que:

$$P\left(|\overline{X}_n - \overbrace{E(X)}^{E(\overline{X}_n)}| \geq \varepsilon\right) \leq \underbrace{\frac{Var(X)}{n}}_{Var(\overline{X}_n)} \frac{1}{\varepsilon^2}$$

- Interpretación: En otras palabras, cuando n es grande, la distribución de probabilidad de la variable aleatoria $\overline{X}_n$ está "concentrada" en torno a la constante $\mu = E(X)$.

- **¡Notar que este resultado no dice nada sobre ninguna realización de la media muestral en particular!**

$$E(\bar{X}_n) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \dots X_n)$$

$$= \frac{1}{n}\left[E(X_1) + \dots + E(X_n)\right] \underset{\downarrow}{=} \frac{1}{n}\left[E(X_1) + \dots + E(X_1)\right]$$

$X_i \sim iid$

$$= \frac{1}{n}\, n \cdot E(X_1)$$

$$= E(X_1).$$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n^2} Var(X_1 + \dots + X_n)$$

$$\underset{\downarrow}{=} \frac{1}{n^2}\left[Var(X_1) + \dots + Var(X_n)\right] \underset{\downarrow}{=} \frac{1}{n^2}\, n\, Var(X_1)$$

$X_i \sim indep$ $\qquad$ $X_i \sim iid$

$$= \frac{Var(X_1)}{n}.$$

Sean dato $X_i \sim iid\ Be(p)$

$$X_i = \begin{cases} 0 & \text{si la persona } i\text{-ésima no es hincha de Boca.} \\ 1 & \text{si la persona } i\text{-ésima es hincha de Boca} \end{cases}$$

Calcule el tamaño de muestra "n" de manera que

$$P(|\bar{X}_n - p| < 0.1) \geq 0.95$$

$$P(-0.1 < \bar{X}_n - p < 0.1) \geq 0.95.$$

Buscamos n que nos garantice que ocurra esto.

$$P(|\bar{X}_n - p| < 0.1) \geq 0.95$$

$\iff$

$$1 - P(|\bar{X}_n - p| \geq 0.1) \geq 0.95$$

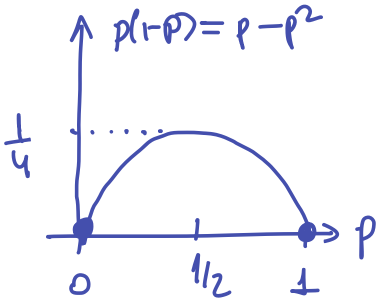
$$0.05 \geq P(|\bar{X}_n - p| \geq 0.1)$$

$$P(|\bar{X}_n - p| \geq 0.1) \leq 0.05.$$

Empezamos tomando Tchebychev para $\bar{X}_n$ y $\varepsilon = 0.1$

$$P(|\bar{X}_n - \underbrace{p}_{E(\bar{X}_n)=E(X)=p}| \geq 0.1) \leq \frac{Var(\bar{X}_n)}{0.1^2} = \frac{Var(X)}{n \cdot 0.1^2} = \frac{p(1-p)}{n \cdot 0.01}$$

$$P(|\bar{X}_n - p| \geq 0.1) \leq \frac{p(1-p)}{n \cdot 0.01} \leq \frac{1/4}{n \cdot 0.01}$$



no conozco p.

Buscamos n /

$$\frac{1/4}{n \cdot 0.01} \leq 0.05$$

$$\frac{100 \cdot 20}{4} = \frac{1/4}{0.01 \cdot 0.05} \leq n$$

$500 \leq n$

$\boxed{n \geq 500}$

Ejemplo

Consideramos $\{X_1, \cdots, X_n\} \sim_{iid}$ y el la variable aleatoria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Queremos encontrar un tamaño de muestra n tal que

$$P\left(\left|\overline{X}_n - p\right| < 0{,}1\right) \geq 0{,}95,$$

equivalentemente

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq 0{,}05,$$

Por Chebyshev

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq \frac{Var(X)}{n}\frac{1}{0{,}1^2} = \frac{p(1-p)}{n}\frac{1}{0{,}1^2}.$$

Como $p(1-p) \leq 0{,}25$ para cualquier $p \in [0, 1]$, entonces:

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq \frac{0{,}25}{n}\frac{1}{0{,}1^2}.$$

Ejemplo

Finalmente, tenemos que encontrar n tal que:

$$P\left(\left|\overline{X}_n - p\right| \geq 0{,}1\right) \leq \frac{0{,}25}{n}\frac{1}{0{,}1^2}.$$

Por lo tanto, se tiene que cumplir:

$$\frac{0{,}25}{n}\frac{1}{0{,}1^2} \leq 0{,}05,$$

Despejando n obtenemos que:

$$\frac{0{,}25}{0{,}05}\frac{1}{0{,}1^2} \leq n,$$

o sea,

$$n \geq 500.$$

Desigualdades de Markov y Chebyshev

Las desigualdades de Markov y Cheby son cotas conservadoras

Consideremos $X \sim N(0,1)$. Por un lado

$$P(|X| \leq 1{,}96) = 2P(X \leq 1{,}96) - 1 = 0{,}95\,.$$

con lo cual

$$P(|X| \geq 1{,}96) = 0{,}05\,.$$

Se puede demostrar que $E(|X|) = \sqrt{2/\pi}$, entonces Markov nos dice

$$P(|X| \geq 1{,}96) \leq \frac{\sqrt{2/\pi}}{1{,}96} \approx 0{,}4\,.$$

La cota que da Markov es correcta, pero muy conservadora. ¿Por qué?

Comparando cotas

Hoeffding supone que X tiene soporte $X \subseteq [a,b]$

Comparemos las tres cotas para la probabilidad $P(|Z| > 1{,}96)$ si $Z \sim \mathcal{N}(0,1)$ usando las desigualdades de Markov, Chebyshev y Chernoff y calculando la probabilidad.

- **Markov:** $P(|Z| > 1{,}96) \leq 0{,}4$
- **Chebyshev:** $P(|Z| > 1{,}96) \leq \dfrac{1}{1{,}96^2} \approx 0{,}26$
- **Mill:** $P(|Z| > 1{,}96)\sqrt{\dfrac{2}{\pi}}\dfrac{e^{-\frac{1{,}96^2}{2}}}{1{,}96} \approx 0{,}0596$ $(Z \sim N(0,1))$.
- **Chernoff:** $P(|Z| > 1{,}96) \leq 2 \cdot e^{\frac{(t^*)^2}{2} - 1{,}96 t^*} \approx 0{,}293$
- **Cálculo exacto:** $P(|Z| > 1{,}96) = 0{,}05$

Donde en la desigualdad de Chernoff usamos que la normal es una distribución simétrica respecto del valor $z = 0$ y que el valor t^* que minimiza esa cota es $t^* = 1{,}96$.

Apéndice

Las slides a partir de aquí son optativas

Funciones convexas y combinaciones convexas

Sea $f : \mathbb{R} \to \mathbb{R}$ una función. Diremos que f es convexa, si dados $x, y \in \mathbb{R}$ y $\alpha \in [0, 1]$, se verifica que:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

- **Observación 1:** Si f es de clase C^2, entonces f es convexa, si y sólo si $f''(x) \geq 0$
- **Observación 2:** Una función convexa en $\mathbb{R}$ es necesariamente continua. Además es posible probar que su derivada $f'(x)$ existe salvo quizás para un conjunto a lo sumo numerable de valores de x, y que f' es creciente.
- **Ejercicio:** Una combinación convexa de los x_i es una combinación lineal

  $$\sum_{i=1}^{n} \alpha_i x_i \text{ en la que } 0 \leq \alpha_i \text{ y } \sum_{i=1}^{n} \alpha_i = 1$$

  Pruebe que si $f : \mathbb{R} \to \mathbb{R}$ es una función convexa y $\sum_{i=1}^{n} \alpha_i x_i$ es una combinación convexa, entonces:

  $$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

Demostración de Jensen (para v.a. discretas)

Desigualdad de Jensen Si $g : \mathbb{R} \to \mathbb{R}$ es una función convexa, entonces:

$$g(E[X]) \leq E[g(X)])$$

Hagamos la demostración primero, en el caso que X toma sólo finitos valores. Sea $p_i = P(X = x_i)$. Entonces

$$E(X) = \sum_{i=1}^{n} p_i x_i$$

es una combinación convexa de los valores de X. Como X es una función convexa,

$$g(E[X]) = g\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i g(x_i) = E[g(X)]$$

Este resultado sirve para demostrar la relación entre $E(g(X))$ y $g(E(X))$. Por ejemplo, se usa en economía para cuando hay incertidumbre y en estadística se usa, por ejemplo, para el teorema de Rao-Blackwell.

Demostración de Jensen (para v.a. discretas)

Si X toma un número numerable de valores, x_i con probabilidades p_i, entonces hacemos lo siguiente: para cada $n \in \mathbb{N}$ definamos,

$$s_n = \sum_{i=1}^{n} p_i$$

y notamos que

$$\sum_{i=1}^{n} \frac{p_i}{s_n} x_i$$

es una combinación convexa. Entonces, como g es convexa:

$$g\left(\sum_{i=1}^{n} \frac{p_i}{s_n} x_i\right) \leq \sum_{i=1}^{n} \frac{p_i}{s_n} g\left(x_i\right)$$

Cuando $n \to +\infty$, tenemos que $s_n \to 1$. Entonces, utilizando la continuidad de g, obtenemos que:

$$g(E[X]) = g\left(\sum_{i=1}^{\infty} p_i x_i\right) \leq \sum_{i=1}^{\infty} p_i g\left(x_i\right) = E[g(X)]$$

Demostración de la desigualdad de Markov

Demostración para variables discretas.

Llamemos
$$A = \{x : |x| < \varepsilon\}.$$

Llamemos $I_A(x)$ a la función indicadora del conjunto A, es decir que $I_A(x) = 1$ si $x \in A$ e $I_A(x) = 0$ en otro caso, entonces:

$$\begin{aligned}
E(|X|) = \sum_x |x| p_X(x) &= \sum_x |x| p_X(x) I_A(x) + \sum_x |x| p_X(x) I_{A^c}(x) \\
&\geq \sum_x |x| p_X(x) I_{A^c}(x) \\
&\geq \varepsilon \sum_x p_X(x) I_{A^c}(x) \\
&\geq \varepsilon P(X \in A^c) = \varepsilon P(|X| \geq \varepsilon)
\end{aligned}$$

□

Para demostrar el caso general considerá $A \equiv \{x : g(x) < \varepsilon\}$.

Demostración de la desigualdad de Markov

Demostración para variables continuas.

Llamemos

$$A = \{x : |x| < \varepsilon\}.$$

Llamemos $I_A(x)$ a la función indicadora del conjunto A, es decir que $I_A(x) = 1$ si $x \in A$ e $I_A(x) = 0$ en otro caso, entonces:

$$\begin{aligned}
E(|X|) = \int\limits_x |x| f_X(x) dx &= \int\limits_x |x| f_X(x) I_A(x) dx + \int\limits_x |x| f_X(x) I_{A^c}(x) dx \\
&\geq \int\limits_x |x| f_X(x) I_{A^c}(x) dx \\
&\geq \varepsilon \int\limits_x f_X(x) I_{A^c}(x) dx \\
&\geq \varepsilon P\left(X \in A^c\right) = \varepsilon P\left(|X| \geq \varepsilon\right)
\end{aligned}$$

□

Para demostrar el caso general considerá $A \equiv \{x : g(x) < \varepsilon\}$.