Clase Práctica 5

Maite Angel

2022-05-10

Repaso

Desigualdad de Chebychev

Sea X una v.a. con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, entonces

$$\forall \varepsilon > 0 \qquad P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Teorema de LGN

Sea $X_1, ..., X_n$ v.a. iid (muestra) con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$ entonces:

$$\bar{X_n} \xrightarrow{p} \mu$$

Que converja en probabilidad ($\xrightarrow{p}$) es equivalente a decir $lim_{n\to\infty} P(|\bar{X_n} - \mu| > \varepsilon)) = 0 \qquad \forall \varepsilon > 0$

Teorema TCL

Sea $X_1, ..., X_n$ v.a. iid (muestra) con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$ entonces:

$$\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \xrightarrow{d} Z \qquad Z \sim N(0, 1)$$

Que converja en distribucion ($\xrightarrow{d}$) quiere decir que cuando n es suficientemente grande $\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \sim N(0, 1)$.

Obs: Mientras más simétrica y acampanada sea la distribución de las X_i, más rápidamente se obtiene una buena aproximación.

Ejercicio 1: Idea intuitiva de LGN y cuando no funciona.

Imaginemos un juego de apuestas, en el que el jugador, debe pagar un precio para participar.

Sean las v.a.:

- X= "resultado de un juego".
- Y= "dinero que se obtienen de un juego", $E(Y), V(Y) < \infty$, $Y = h(X)$

$Y_1, ...Y_n$ una muestra aleatoria de Y.

Mostrar que que promediando las Y_i podemos tener una noción de cuanto deberiamos pagar para jugar al juego si somos neutrales al riesgo. Calcular cual es ese valor si $h(x) : x^2$ y $X \sim U(0, 10)$. ¿Y si no tuviera la distribucion de la X?

Bonus: Ahora supongamos que sabemos un poco mas. El juego consiste en tirar una moneda sucesivamente de forma independiente hasta que sale cruz. Una vez ha salido cruz, el juego se detiene y se gana 2^k dólares si se realizaron k tiradas de la moneda. ¿Cuánto estaría dispuesto a pagar el jugador por este juego?

Solución

Vamos a justificar de alguna forma por que vale la idea de que si tenemos muchos experimentos de cuanta plata gana la gente en un juego al promediarla estoy cerca de saber cuanto podria ganar yo.

$$\frac{\sum_{i=1}^{n} Y_i}{n} \xrightarrow{LGNp} E(Y_1) = E(Y) = E(h(X)) = \int_{-\infty}^{+\infty} h(x) f_X(x) dx = \int_0^{10} x^2 \frac{1}{10} = \frac{1}{10}(\frac{10^3}{3} - \frac{0^3}{3}) \approx 33.33$$

Luego si somos neutrales al riesgo deberiamos pagar por participar como mucho lo que esperamos ganar, es decir 33.33 USD. Si no tuviera la distribución de la X simplemente me decido por lo que me de el promedio de mis observaciones Y_i.

Bonus

Tenemos las siguientes v.a.:

- X="la cantidad de tiradas", $R_X = \{1, 2, 3, ...\}$
- Y= "dinero en USD que se obtiene del juego.", $R_X = \{2^1, 2^2, 2^3, ...\}$

Y tenemos que

$$P(Y = 2^k) = P(X = k) = (\frac{1}{2})^k = \frac{1}{2^k}$$
$$E(Y) = \sum_{k=1}^{\infty} 2^k \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = \infty$$

Entonces si nos dejaramos guiar por el criterio anterior, a cualquier precio que nos propongan deberiamos aceptar jugar, lo cual no es para nada razonable.

En este escenario no seria correcto aplicar LGN, la esperanza no tiene porque ser el limite del promedio de observaciones.

Obs: Se la conoce como la Paradoja de San Petersburgo (La formulación la planteo Nicolaus Bernoulli a su primo y a Pierre de Montmort).

Ejercicio 2

Una empresa tiene acceso a todos los audios de whatsapp que alguna vez se grabaron en la plataforma. Escuchan 1000 audios con el fin de analizar su duración en minutos. Se sabe que la esperanza de la duración de un audio es 3 y la varianza 2. Las duraciones de los audios son independientes

Item 1: Probar que la probabilidad de que en promedio los audios duren entre 2 y 4 minutos es mayor a 0.99.

Item 2: Si sabemos que el 70% de las veces un audio dura mas de un minuto. Aproximar la probabilidad de que más de 68% de los audios duren mas de un minuto.

Item 3: ¿Cuantos audios deberiamos analizar para que la probabilidad aproximada de encontrar 500 audios que duren mas de un minuto sea mayor 0.95?

Solucion

Item 1

Sean $X_1, ..., X_{1000}$ las duraciones de los 1000 audios que se escuchan, si $X =$"duracion de un audio", $E(X) = 3, V(X) = 2$.

$$P(2 \leq \bar{X} \leq 4) = P(2-3 \leq \bar{X}-3 \leq 4-3) = P(|\bar{X}-3| \leq 1) = 1 - P(|\bar{X}-3| \geq 1)$$

Además

- $E(\bar{X}) = \frac{\sum E(X_i)}{1000} = \frac{1000*3}{1000} = 3$
- $V(\bar{X}) = \frac{\sum V(X_i)}{1000^2} = \frac{1000*2}{1000^2} = \frac{2}{1000}$

Luego por Chebychev:

$$1 - P(|\bar{X}-3| \geq 1) = 1 - P(|\bar{X} - E(\bar{X})| \geq 1) \geq 1 - \frac{2}{1000 * 1^2} \geq 0.998 \geq 0.99$$

Item 2

Sea $Y =$"cantidad de audios que duran mas de un minuto", $Y \sim Bi(1000, 0.7)$, nos piden aproximar $P(Y > 680)$.

Si nos pidieran calcularla de forma exacta y estuvieramos en un parcial escrito sin R deberiamos resolver la siguiente suma de 250 sumandos:

$$P(Y > 680) = \sum_{k=751}^{1000} \binom{1000}{k} 0.7^k (1-0.7)^{1000-k}$$

Suerte con eso.. pero por suerte solo queremos aproximar y tenemos el TCL que sale a nuestro rescate. Pensemos a $Y = \sum_{i=1}^{1000} A_i$ donde cada A_i="si el audio numero i duro mas de un minuto", $A_i \sim Be(0.7)$. Modelamos a la binomial como una suma de Bernoullis.

$$P(Y > 680) = 1 - P(Y \leq 680) = 1 - P(\sum_{i=1}^{1000} A_i \leq 680) = 1 - P(\frac{\sqrt{1000}(\bar{A}-0.7)}{\sqrt{0.7*0.3}} \leq \frac{\sqrt{1000}(\frac{680}{1000}-0.7)}{\sqrt{0.7*0.3}})$$

Por TCL aproximadamente $\frac{\sqrt{1000}(\bar{A}-0.7)}{\sqrt{0.7*0.3}} \sim N(0,1)$, luego

$$= 1-P(\frac{\sqrt{1000}(\bar{A}-0.7)}{\sqrt{0.7*0.3}} \leq \frac{\sqrt{1000}(\frac{680}{1000}-0.7)}{\sqrt{0.7*0.3}}) \approx 1-\Phi(-1.380131) = 1-(1-\Phi(1.380131)) = \Phi(1.380131) = 0.9162269$$

Item 3

Planteamos la misma cuenta que antes pero ahora estamos buscando $P(\sum_{i=1}^{n} A_i > 500) \geq 0.95$, entonces

$$P(\sum_{i=1}^{n} A_i > 500) = 1 - P(\frac{\sqrt{n}(\bar{A}-0.7)}{\sqrt{0.7*0.3}} \leq \frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}}) \approx 1 - \Phi(\frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}})$$

Luego queremos que

$$1 - \Phi(\frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}}) \geq 0.95$$

$$\Phi(\frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}}) \leq 0.05$$

$$\frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}} \leq \Phi^{-1}(0.05)$$

$$\frac{\sqrt{n}(\frac{500}{n}-0.7)}{\sqrt{0.7*0.3}} \leq -1.644854$$

$$\sqrt{n}(\frac{500-0.7n}{n}) \leq -1.644854\sqrt{0.7*0.3}$$

$$500 - 0.7n \leq -1.644854\sqrt{0.7*0.3}\sqrt{n}$$

$$0 \leq -0.7537668\sqrt{n} - 500 + 0.7n$$

Usemos $m = \sqrt{n}$

$$0 \leq -0.7537668m - 500 + 0.7m^2 = (m - 27.27)(m + 26.1931)$$

Para que se cumpla esto debe ocurrir que $m \geq 27.27$ ó $m \leq -26.1931$, como estamos pensando que $m = \sqrt{n}$, los negativos no son un resultado posibles.

Luego necesitamos que $m \geq 27.27$ osea $n \geq 744$, chequeamos $1-pnorm((\sqrt{(}742)*((500/744)-0.7))/(\sqrt{(}0.7*0.3))) = 0.9519491$.