Práctico 10

jueves, 2 de junio de 2022

teorema central del límite

$X_1, \ldots, X_n$ MAS de X $\qquad E(X) = \mu \quad Var(X) = \sigma^2$

consideramos $\overline{X_n}$

$$E(\overline{X_n}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n} n\mu = \mu$$

$$Var(\overline{X_n}) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2$$

TCL: para n grande $\overline{X_n}$ tiene distribución normal

$$\overline{X_n} \sim_{aprox} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X_n} - \mu}{\sigma}\sqrt{n} \sim_{aprox} \mathcal{N}(0,1)$$

Ejercicio 1

Se realiza una encuesta para tratar de estimar el porcentaje de votos de determinado candidato. Al final de la misma se tendrán n respuestas, representadas en la muestra $X_1, X_2, \cdots, X_n$ de realizaciones de la variable de Bernoulli de parámetro p (la probabilidad de que una persona elegida al azar en la población vote por el candidato en cuestión).

1. A partir de la mencionada muestra ¿Qué valor usaría usted para estimar p?. $\overline{X_n}$
2. Usando la Desigualdad de Chevishoff (y antes de tomar la muestra) ¿Cuántos datos tomaría para que, con una probabilidad de al menos 0,95, el estimador no distara del valor de p más de un 0,03?. (Sugerencia: Demuestre y use que para cualquier valor de $p \in [0,1]$ se cumple $p(1-p) \leq 1/4$).
3. Usando el Teorema del Límite Central resolver la parte anterior y comparar con el resultado obtenido por la Desigualdad de Chevishoff. *(Las dos primeras partes de este ejercicio fueron resueltas en el práctico 8).*

$$P(|\overline{X_n} - p| \leq 0{,}03) \geq 0{,}95$$

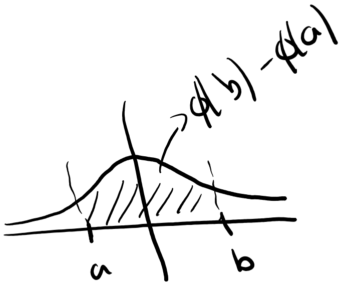
$X_1, \ldots, X_n$ MAS de X donde $X \sim Ber(p)$ $\qquad E(X) = p$

$$Var(X) = p(1-p)$$

$$\overline{X}_n \sim_{aprox} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

$$P(|\overline{X}_n - p| \leq 0{,}03) = P(-0{,}03 \leq \overline{X}_n - p \leq 0{,}03)$$

$$= P\left(\frac{-0{,}03}{\sqrt{p(1-p)}/\sqrt{n}} \leq \frac{\overline{X}_n - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq \frac{0{,}03}{\sqrt{p(1-p)}/\sqrt{n}}\right)$$



$$= P\left(\frac{-0{,}03\sqrt{n}}{\sqrt{p(1-p)}} \leq \underbrace{\frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n}}_{\sim \mathcal{N}(0,1)} \leq \frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

$$= \phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right) - \phi\left(-\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

$$= \phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right) - \left(1 - \phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right)\right)$$

$$= 2\phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1$$

buscamos n tal que $2\phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 0{,}95$

$$\Rightarrow \phi\left(\frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq \frac{1{,}95}{2} = 0{,}975$$

$$\Rightarrow \frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}} \geq \phi^{-1}(0{,}975) = 1{,}96$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850

$$\Rightarrow \frac{0{,}03\sqrt{n}}{\sqrt{p(1-p)}} \geq 1{,}96$$

$$\Rightarrow \sqrt{n} \geq \frac{1{,}96}{0{,}03}\sqrt{p(1-p)}$$

$$\Rightarrow n \geq \left(\frac{1{,}96}{0{,}03}\right)^2 p(1-p)$$

tenemos que $p(1-p) \leq \frac{1}{4}$ si $p \in [0,1]$

tomamos n tal que $n \geq \left(\frac{1{,}96}{0{,}03}\right)^2 \frac{1}{4} = 1067{,}11$

$$\Rightarrow \boxed{n \geq 1068}$$

Ejercicio 4

1. Si $X \sim \mathrm{BN}(k,p)$, con k muy grande, hallar una aproximación de su distribución. Sugerencia: Usar la descomposición de una Binomial Negativa BN como suma de geométricas.
2. Si $X \sim \mathrm{BN}(1600, 0{,}25)$, calcular $\mathbf{P}(X > 6200)$.

X_1 = repeticiones hasta el primer exito

X_2 = repeticiones entre 1° y 2do exito

X_k = repeticiones entre el $(k-1)$-ésimo exito y el k-ésimo

1er exito   2do exito k-ésimo exito

$$X = X_1 + X_2 + \dots + X_k \quad \text{donde } X_i \sim \mathrm{Geo}(p)$$

$$X = k\overline{X_k}$$

buscamos aproximar la distribución de $k\overline{X_k}$

- distribución de $\overline{X_k}$

$X_1, \dots, X_k$ MAS de una $\mathrm{Geo}(p)$

$$E(X_1) = \frac{1}{p}$$

$$\mathrm{Var}(X_1) = \frac{1-p}{p^2}$$

$$\overline{X_k} \sim_{aprox} \mathcal{N}\left(\frac{1}{p}, \frac{(1-p)}{kp^2}\right)$$

- distribución de $k\overline{X_k}$

$$E(k\overline{X_k}) = k\,E(\overline{X_k}) = \frac{k}{p}$$

$$\mathrm{Var}(k\overline{X_k}) = k^2\,\mathrm{Var}(\overline{X_k}) = k^2\frac{(1-p)}{kp^2} = \frac{k(1-p)}{p^2}$$

$$X = k\overline{X_k} \sim_{aprox} \mathcal{N}\left(\frac{k}{p}, \frac{k(1-p)}{p^2}\right)$$

2) $X \sim \mathrm{BN}(1600, 0{,}25)$ $p_X(n) = C^{n-1}_{k-1}\, p^k (1-p)^{n-k}$

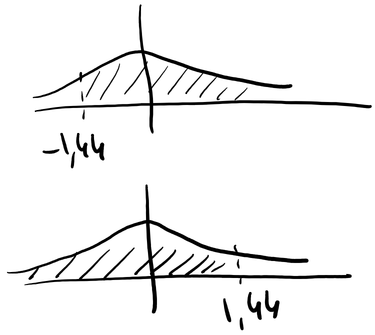
$P(X > 6200) = ?$

$$X \sim_{aprox} \mathcal{N}\left(\frac{1600}{0{,}25} , \frac{1600 \cdot 0{,}75}{(0{,}25)^2} \right)$$

$$X \sim_{aprox} \mathcal{N}(6400, 19200)$$

$$P(X > 6200) = P\Big(\underbrace{\frac{X-6400}{\sqrt{19200}}}_{\sim \mathcal{N}(0,1)} > \frac{6200-6400}{\sqrt{19200}} \Big)$$

$$= P\left(\frac{X-6400}{\sqrt{19200}} > -1{,}44 \right)$$



$$= 1 - \phi(-1{,}44)$$

$$= 1 - (1 - \phi(1{,}44))$$

$$= \phi(1{,}44) = 0{,}9251$$

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.519
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.559
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.598
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.636
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.673
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.708
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.742
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.773
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.802
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.828
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.853
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.874
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.894
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.911
1.4	0.9192	0.9207	0.9222	0.9236	$\boxed{0.9251}$	0.926
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.939

Intervalos de confianza

$X_1, \ldots, X_n$ MAS de X

queremos estimar un parámetro θ

tomamos dos estimadores $a(X_1, \ldots, X_n)$ y $b(X_1, \ldots, X_n)$

si $P(\theta \in [a(X_1,\dots,X_n), b(X_1,\dots,X_n)]) = 1-\alpha$ con $\alpha \in (0,1)$

decimos que $[a,b]$ es un intervalo de confianza $1-\alpha$

Vamos a estimar parámetros de $X \sim \mathcal{N}(\mu, \sigma^2)$

(1) asumiendo que conocemos σ^2 queremos estimar μ

busamos intervalos de confianza de la forma

$$[\overline{X}_n - k, \overline{X}_n + k]$$

queremos encontrar k tal que

$$P(\mu \in [\overline{X}_n - k, \overline{X}_n + k]) = 1-\alpha$$

$$P(\mu \in [\overline{X}_n - k, \overline{X}_n + k]) = P(\overline{X}_n - k \leq \mu \leq \overline{X}_n + k)$$

$$= P(-\mu - k \leq -\overline{X}_n \leq -\mu + k)$$

$$= P(\mu - k \leq \overline{X}_n \leq \mu + k)$$

$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

$$= P\left(\frac{\mu - k - \mu}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu + k - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(-\frac{k\sqrt{n}}{\sigma} \leq \underbrace{\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}}_{\sim \mathcal{N}(0,1)} \leq \frac{k\sqrt{n}}{\sigma}\right)$$

$$= \phi\left(\frac{k\sqrt{n}}{\sigma}\right) - \phi\left(-\frac{k\sqrt{n}}{\sigma}\right)$$

$$= 2\phi\left(\frac{k\sqrt{n}}{\sigma}\right) - 1$$

buscamos k tal que $2\phi\left(\frac{k\sqrt{n}}{\sigma}\right) - 1 = 1-\alpha$

$$\Rightarrow \phi\left(\frac{k\sqrt{n}}{\sigma}\right) = 1 - \frac{\alpha}{2}$$

$$\Rightarrow \frac{k\sqrt{n}}{\sigma} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right) = z_{1-\frac{\alpha}{2}}$$

$$\Rightarrow \boxed{k = \frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\sqrt{n}}}$$

intervalo de confianza $1-\alpha$ para μ conociendo σ^2 es:

$$\left[\overline{X}_n - \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} \,,\, \overline{X}_n + \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}\right]$$

Ejercicio 5

1. Una máquina de refrescos está ajustada de tal manera que la cantidad de líquido despachada se distribuye aproximadamente en forma normal con una desviación estándar igual a 0,15 decilitros. Encontrar un intervalo de confianza 0,95 para la media de todos los refrescos que sirve esta máquina si una muestra aleatoria de 36 refrescos tiene un contenido promedio de 2,25 decilitros.
2. ¿Qué tan grande tiene que ser la muestra si se desea tener una confianza del 95 % de que la media muestral no difiera en más de 0,03 decilitros de la media real μ?

X = cantidad de líquido despachado

$X \sim \mathcal{N}(\mu, (0{,}15)^2)$

buscamos un intervalo de confianza $\overbrace{0{,}95}^{1-\alpha}$ para μ

$$\left[\overline{X}_n - \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} \,,\, \overline{X}_n + \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}\right]$$

donde $\alpha = 0{,}05$

$\sigma = 0{,}15$

$n = 36$

$\overline{X_{36}} = 2{,}25$

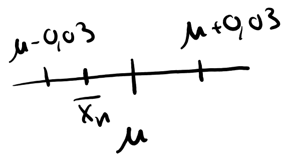
$$z_{1-\alpha/2} = \phi^{-1}\left(1-\frac{\alpha}{2}\right) = \phi^{-1}\left(1-\frac{0{,}05}{2}\right) = \phi^{-1}(0{,}975) = 1{,}96$$

$$\frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 0{,}15}{6} = 0{,}049$$

el intervalo de confianza $0{,}95$ es

$$\left[2{,}25 - 0{,}049 \,;\, 2{,}25 + 0{,}049\right]$$

2. ¿Qué tan grande tiene que ser la muestra si se desea tener una confianza del 95 % de que la media muestral no difiera en más de 0,03 decilitros de la media real μ?



$$\left[\overline{X_n} - k \,,\, \overline{X_n} + k\right] \quad \text{con } k = \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} \quad \text{y } \alpha = 0{,}05$$

buscamos n de forma que $\dfrac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} \leq 0{,}03$

$$\Rightarrow \frac{1{,}96 \cdot 0{,}15}{\sqrt{n}} \leq 0{,}03$$

$$\Rightarrow \frac{1}{\sqrt{n}} \leq \frac{0{,}03}{1{,}96 \cdot 0{,}15}$$

$$\Rightarrow n \geq \frac{1{,}96^2 \cdot 0{,}15^2}{0{,}03^2} = 96{,}04$$

$\boxed{n \geq 97}$