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 $\mathbf{P}(X > 6200)$.

1\. $X \sim BN(k, p)$

$$X = \underset{\substack{\uparrow \\ 1^{er} \text{ exito}}}{X_1} + \underset{\substack{\uparrow \\ 2^{do} \text{ exito}}}{X_2} + \dots + \underset{\substack{\uparrow \\ k\text{-ésimo exito}}}{X_k} \qquad \text{donde } X_1, \dots, X_k \text{ iid} \sim Geo(p)$$

$$E(X_1) = 1/p$$

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

$$E(X_1 + \dots + X_k) = \underbrace{E(X_1)}_{1/p} + \underbrace{E(X_2)}_{1/p} + \dots + \underbrace{E(X_k)}_{1/p} = k \cdot \frac{1}{p} = \frac{k}{p}$$

$$Var(X_1 + \dots + X_k) \underset{\substack{\uparrow \\ \text{las } X_i \text{ son independientes}}}{=} \underbrace{Var(X_1)}_{\frac{1-p}{p^2}} + \underbrace{Var(X_2)}_{\frac{1-p}{p^2}} + \dots + \underbrace{Var(X_k)}_{\frac{1-p}{p^2}} = \frac{k(1-p)}{p^2}$$

Por el teorema central del límite, la distribución aproximada de X es

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

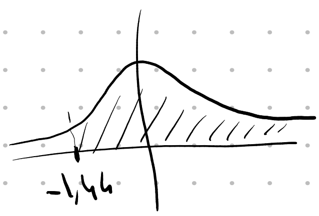
2\. $X \sim BN(\underset{\overset{\shortparallel}{k}}{1600}, \underset{p}{\underline{0{,}25}})$ calcular $P(X > 6200)$

$$X \sim_{aprox} \mathcal{N}\left(\frac{1600}{0{,}25}, \frac{1600(1-0{,}25)}{0{,}25^2}\right)$$

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

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

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



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

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	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265

intervalos de confianza

X variable aleatoria

queremos estimar un parámetro θ de la distribución de X

$X_1, \dots, X_n$ MAS de X

consideramos dos estimadores $a(X_1, \dots, X_n)$ y $b(X_1, \dots, X_n)$

nos preguntamos cuanto vale

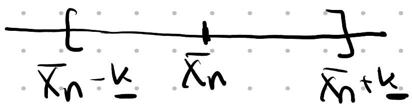
$$P(\theta \in [a(X_1, \dots, X_n), b(X_1, \dots, X_n)])$$

si $P(\theta \in [a,b]) = 1-\alpha$ $(\alpha \in (0,1))$ decimos que $[a,b]$ es un intervalo de confianza $1-\alpha$.

Vamos a estimar parámetros de $X \sim \mathcal{N}(\mu, \sigma^2)$ $X_1, \dots, X_n$ MAS de X

(1) conocemos σ^2 y queremos estimar μ

vamos a buscar un intervalo de confianza de esta forma



queremos k tal que

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

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

$$= P(\text{---} \leq \overline{X}_n \leq \text{---})$$

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

$$E(\overline{X}_n) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n}\left(\underbrace{E(X_1)}_{\mu} + \dots + \underbrace{E(X_n)}_{\mu}\right) = \frac{1}{n} n\mu = \mu$$

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

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

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

$$-\overline{X}_n \leq k - \mu$$

$$\overline{X}_n \geq -k + \mu$$

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

$$-\mu - k \leq -\overline{X}_n$$

$$\mu + k \geq \overline{X}_n$$

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

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

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

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

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

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

$$P(\mu \in [\overline{X}_n - k, \overline{X}_n + k]) = 2\phi\left(\frac{k}{\sigma}\sqrt{n}\right) - 1$$

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

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

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

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

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

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

el intervalo de confianza $1-\alpha$ es

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

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?

1. X = cantidad de líquido despachada

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

buscamos un intervalo de confianza 0,95 para μ

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

$1-\alpha = 0{,}95 \Rightarrow \boxed{\alpha = 0{,}05}$

$n = 36$

$\overline{X_n} = 2{,}25$

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

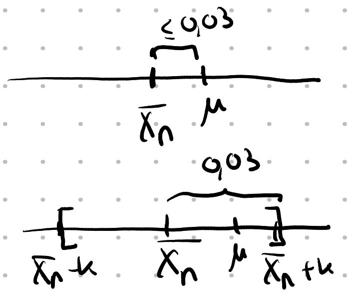
$$\frac{\sigma z_{\alpha/2}}{\sqrt{n}} = \frac{0{,}15\cdot 1{,}96}{\sqrt{36}} = 0{,}049$$

el intervalo de confianza es:

$$\left[2{,}25 - 0{,}049, 2{,}25 + 0{,}049\right] = \left[2{,}201, 2{,}299\right]$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.

2. buscamos n para tener un intervalo de confianza 0,95 de μ tal que



buscamos $[\overline{X}_n - k, \overline{X}_n + k]$ con $k \leq 0{,}03$

$$k = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$$

$$\frac{z_{\alpha/2}\cdot\sigma}{\sqrt{n}} \leq 0{,}03 \Rightarrow \frac{\sqrt{n}}{z_{\alpha/2}\cdot\sigma} \geq \frac{1}{0{,}03}$$

$$\Rightarrow \sqrt{n} \geq \frac{z_{\alpha/2}\cdot\sigma}{0{,}03}$$

$$\Rightarrow n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{0{,}03}\right)^2 = \left(\frac{1{,}96\cdot 0{,}15}{0{,}03}\right)^2 = 96{,}04$$

$\Rightarrow$ tomamos $n = 97$

---

① conocemos σ^2 y queremos estimar μ

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

② no conocemos σ^2 y queremos estimar μ

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2 \leftarrow \text{es sesgado}$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2 \leftarrow \text{es insesgado}$$

buscamos un intervalo de confianza $1-\alpha$ para μ de la forma

$$\left[\overline{X_n} - ks_n, \overline{X_n} + ks_n\right]$$

$$P(\mu \in [\overline{X_n} - ks_n, \overline{X_n} + ks_n]) = P(\overline{X_n} - ks_n \leq \mu \leq \overline{X_n} + ks_n)$$

$$= P(-ks_n \leq \mu - \overline{X_n} \leq ks_n)$$

$$= P(-ks_n \leq \overline{X_n} - \mu \leq ks_n)$$

$$= P\left(-k \leq \frac{\overline{X_n} - \mu}{s_n} \leq k\right)$$

$$= P\left(-k\sqrt{n} \leq \frac{(\overline{X_n} - \mu)}{s_n}\sqrt{n} \leq k\sqrt{n}\right)$$

si $s_n = \sigma$ esto sería una $\mathcal{N}(0,1)$

<u>teorema</u>: $\frac{(\overline{X_n} - \mu)}{s_n}\sqrt{n}$ se distribuye como una T-student con $n-1$ grados de libertad.

$F_T^{(n-1)}$ = función de distribución de la T-student con $(n-1)$ grados de libertad

$$P(\mu \in [\overline{X_n} - ks_n, \overline{X_n} + ks_n]) = P\left(-k\sqrt{n} \leq \frac{(\overline{X_n} - \mu)}{s_n}\sqrt{n} \leq k\sqrt{n}\right)$$

$$= F_T^{(n-1)}(k\sqrt{n}) - F_T^{(n-1)}(-k\sqrt{n})$$

$$= 2F_T^{(n-1)}(k\sqrt{n}) - 1$$

queremos $P(\mu \in [\overline{X_n} - ks_n, \overline{X_n} + ks_n]) = 1-\alpha$

$$\Rightarrow 2F_T^{(n-1)}(k\sqrt{n}) - 1 = 1-\alpha$$

$$\Rightarrow 2F_T^{(n-1)}(k\sqrt{n}) = 2-\alpha$$

$$\Rightarrow F_T^{(n-1)}(k\sqrt{n}) = 1-\frac{\alpha}{2}$$

$$\Rightarrow k\sqrt{n} = \underbrace{\left(F_T^{(n-1)}\right)^{-1}\left(1-\frac{\alpha}{2}\right)}_{\boxed{t_{\alpha/2}(n-1)}} \quad \text{tabla}$$

$$\Rightarrow k = \frac{t_{\alpha/2}(n-1)}{\sqrt{n}}$$

entonces el intervalo de confianza $1-\alpha$ para μ si no conocemos σ^2 es

$$\left[\overline{X}_n - \frac{t_{\alpha/2}(n-1)\, S_n}{\sqrt{n}} \,,\, \overline{X}_n + \frac{t_{\alpha/2}(n-1)\, S_n}{\sqrt{n}}\right]$$

r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.997
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878