



Probabilidad y Estadística para IA

Examen

Alumna: Clara Bureu
Año: 2023

Ejercicio 1:

1. En un canal de comunicación se envían símbolos binarios en una proporción de tres *ceros* por cada cuatro *unos*. El ruido del canal genera errores en la transmisión, en particular se observó que con probabilidad $1/4$ un *cero* es recibido como un *uno*, mientras que la probabilidad de que un *uno* sea recibido como un *cero* es de $1/3$.
 - a) Si se recibe un *uno*, cuál es la probabilidad de que se haya enviado un *uno*?
 - b) Asumiendo independencia entre los símbolos, si se recibe el mensaje *cero-cero*, cuál es la distribución de los cuatro posibles mensajes que pudieron haber sido enviados?

En primer lugar, vamos a definir etiquetas para asociar a los cálculos:

EC = "Enviar un cero"

EU = "Enviar un uno"

RC = "Recibir un cero"

RU = "Recibir un uno"

Tal como se especifica en el enunciado, sabemos que:

$$P(RU|EC) = \frac{1}{4}$$

$$P(RC|EU) = \frac{1}{3}$$

Además, como dato tenemos que se envían 3 ceros por cada 4 unos:

1	1	1	1	0	0	0
---	---	---	---	---	---	---

Por lo tanto:

$$P(EC) = \frac{3}{7}$$

$$P(EU) = \frac{4}{7}$$

- a- Para conocer la probabilidad de haber enviado un uno dado que se recibió un uno podemos utilizar Bayes:

$$P(EU|RU) = \frac{P(RU|EU) * P(EU)}{P(RU)}$$

En primer lugar, vamos a calcular $P(RU|EU)$:

$$P(RU|EU) = 1 - P(RC|EU) = \frac{2}{3}$$

Además, necesitamos conocer $P(RU)$, podemos calcularla por Probabilidad Total:

$$P(RU) = P(RU|EU)P(EU) + P(RU|EC)P(EC) = \frac{2}{3} * \frac{4}{7} + \frac{1}{4} * \frac{3}{7} \cong 0.488$$

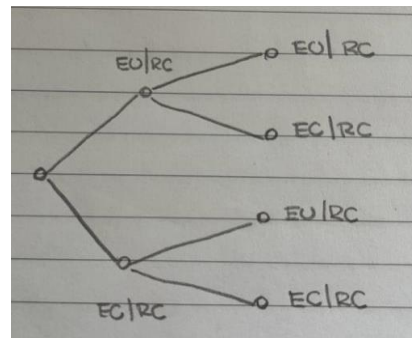
Por lo tanto, reemplazando tenemos que:

$$P(EU|RU) = \frac{P(RU|EU) * P(EU)}{P(RU)} = \frac{\frac{2}{3} * \frac{4}{7}}{0.488} \cong 0.78$$

Entonces, la probabilidad de haber enviado un uno dado que se recibió un uno es de aproximadamente 0.78.

b- Si se recibe un cero-cero se pudieron haber enviado 4 posibles combinaciones:

- Se envió un 11
- Se envió un 10
- Se envió un 01
- Se envió un 00



Por lo tanto, necesitamos conocer la probabilidad de haber enviado un uno dado que se recibió un cero:

Volvemos a usar el teorema de Bayes:

$$P(EU|RC) = \frac{P(RC|EU) * P(EU)}{P(RC)} = \frac{\frac{1}{3} * \frac{4}{7}}{1 - 0.488} \cong 0.36$$

$$P(EC|RC) = 1 - P(EU|RC) \cong 0.64$$

Ahora si podemos conocer las cuatro distribuciones dado que se recibió un cero-cero:

Probabilidad de haber enviado un 11 dado que se recibio 00

$$P(EUU|RCC) = 0.36 * 0.36 = 0.1296$$

Probabilidad de haber enviado un 10 dado que se recibio 00

$$P(EUC|RCC) = 0.36 * 0.64 = 0.2304$$

Probabilidad de haber enviado un 01 dado que se recibio 00

$$P(ECU|RCC) = 0.64 * 0.36 = 0.2304$$

Probabilidad de haber enviado un 00 dado que se recibio 00

$$P(ECC|RCC) = 0.64 * 0.64 = 0.4096$$

① 1 1 1 1 0 0 0

EC = "Enviar un cero"

EU = "Enviar un uno"

RC = "Recibir un cero"

RU = "Recibir un uno"

Sabemos que
$$\begin{cases} P(EC) = 3/7 & P(EU) = 4/7 \\ P(RU|EU) = 1/4 & P(RC|EU) = 1/3 \end{cases}$$

a. Queremos conocer la $P(RU|EU)$: podemos utilizar Bayes.

$$P(EU|RU) = \frac{P(RU|EU)P(EU)}{P(RU)} \quad (*)$$

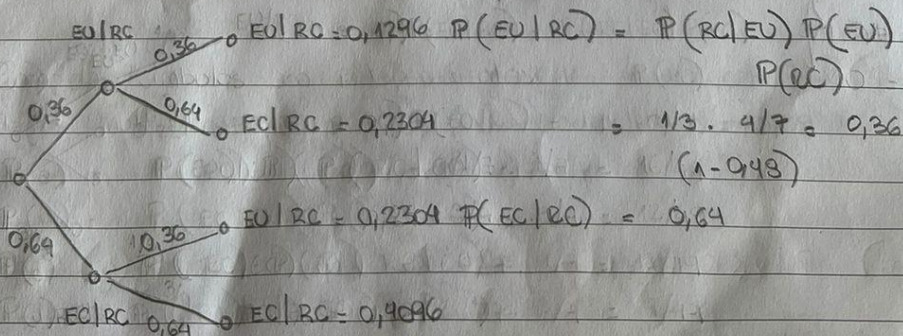
i) $P(RU|EU) = 1 - P(RC|EU) = 1 - 1/3 = 2/3$

ii)
$$P(RU) = P(RU|EU)P(EU) + P(RU|EC)P(EC)$$

$$2/3 \cdot 4/7 + 1/4 \cdot 3/7 \approx 0,4880$$

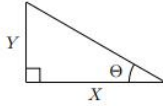
(*) $P(EU|RU) = \frac{2/3 \cdot 4/7}{0,4880} \approx 0,78$ Probabilidad de haber enviado un 1 dado que se recibió un 1.

b. Si se recibe un cero-cero se podría haber enviado



Ejercicio 2:

2. Para fabricar una prenda, Roberto necesita cortar retazos de tela triangulares. Cada recorte que hace es un triángulo rectángulo, cuyos catetos (X e Y) son variables aleatorias que dependen del ángulo de corte Θ , donde $\Theta \sim \text{Uniforme}(0, \pi/2)$. El molde sigue la siguiente figura:



- Encontrar la distribución condicional de $Y \mid X = x$.
- ¿Cuál es el soporte y a qué distribución corresponde $Y \mid X = x$?
- ¿Cuál es la probabilidad $\mathbf{P}[Y \leq 10 \mid X = 1]$?

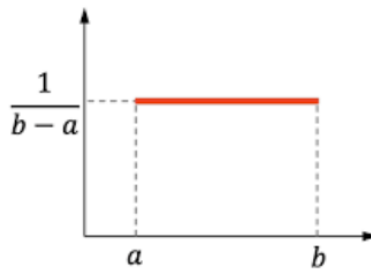
Sabemos que:

$$X = Y * \tan \theta$$

$$Y = X * \tan \theta$$

$$\theta = \tan^{-1} \frac{Y}{X}$$

Además, la distribución uniforme se representa como se muestra a continuación:



Con $a = 0$ y $b = \frac{\pi}{2}$

$$f_{\theta}(\theta) = \frac{2}{\pi}$$

- a- Dado que $Y = X * \tan \theta$ podemos usar la técnica de transformación de variables para encontrar la distribución de Y dado que $X=x$ (asumiendo que x es una constante)

$$f_Y(Y|X=x) = f_{\theta}(g^{-1}(y|X=x)) \left| \frac{dg^{-1}(y|X=x)}{dy} \right|$$

En primer lugar, vamos a buscar nuestra $g(y|X=x)$:

$$Y = g(y|X=x) = X * \tan \theta$$

Ahora buscamos la inversa de g :

$$\theta = g^{-1}(y|X=x) = \tan^{-1} \frac{y}{x}$$

Solo nos falta encontrar la derivada:

$$\left| \frac{dg^{-1}(y|X=x)}{dy} \right| = \frac{x}{x^2 + y^2}$$

Ahora reemplazamos y finalmente obtenemos la función $f_{y(Y|X=x)}$:

$$\begin{aligned} f_{y(Y|X=x)} &= f_{\theta}(g^{-1}(y|X=x)) \left| \frac{dg^{-1}(y|X=x)}{dy} \right| \\ &= \frac{2}{\pi} * \frac{x}{x^2 + y^2} \end{aligned}$$

Multiplicamos y dividimos por x:

$$\begin{aligned} f_{y(Y|X=x)} &= f_{\theta}(g^{-1}(y|X=x)) \left| \frac{dg^{-1}(y|X=x)}{dy} \right| \\ &= \frac{2}{\pi x} * \frac{x^2}{x^2 + y^2} \end{aligned}$$

Analizando esta función de distribución y comparándola con la tabla de distribuciones particulares, notamos que la distribución de Cauchy tiene la siguiente forma:

Cauchy	$\text{Cau}(x_0, \gamma)$	$\frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right]$	\mathbb{R}	$x_0 \in \mathbb{R}, \gamma > 0$
--------	---------------------------	--	--------------	----------------------------------

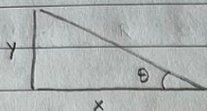
Entonces, la distribución condicional $Y | X = x$ es una Cauchy con $\gamma = x, x_0 = 0, x = y$.

- b- Dado que $Y = X * \tan \theta$, sabemos que θ se encuentra en el intervalo $[0, \frac{\pi}{2}]$ por lo tanto $\tan \theta$ tiene soporte en el intervalo $[0; \infty]$.
Por lo tanto, el soporte de $Y | X = x$ será $[0; \infty]$.
- c- Ahora integramos la función de distribución para hallar la probabilidad de que Y sea menor o igual a 10 dado que $x = 1$:

$$P[Y \leq 10 | x = 1] = \int_0^{10} \frac{2}{\pi x} * \frac{x^2}{x^2 + y^2} dy = \frac{2}{\pi * 1} \tan^{-1} 10 \cong 0.93$$

La probabilidad de que Y sea menor o igual a 10 dado que $x = 1$ es de 0.93

2)



$$\theta \sim U(0, \pi/2)$$

sabemos que: $x = y \cdot \tan(\theta)$

$$y = x \cdot \tan(\theta)$$

$$\theta = \arctan(y/x)$$

a.

$$\frac{1}{2}$$

x

y

 $\pi/2$

$$f_{\theta}(\theta) = \frac{1}{\pi/2 - 0} = \frac{2}{\pi} \quad \{0 \leq \theta \leq \pi/2\}$$

a) Dado que $y = x \cdot \tan(\theta)$, podemos usar la técnica de transformación de variables para encontrar la distribución de y dado $x = x$ (asumiendo que x es una constante).

$$f_Y(y|x) = f_{\theta}(g^{-1}(y/x)) \left| \frac{d\theta}{dy} \right|$$

$$\theta \sim U(0, \pi/2) \Rightarrow y = x \cdot \tan(\theta)$$

$$y = \underbrace{x \cdot \tan(\theta)}_{\text{despeja } \theta} = g(\theta) \Rightarrow \theta = \underbrace{\arctan\left(\frac{y}{x}\right)}_{g^{-1}(y)}$$

$$\left| \frac{d\theta}{dy} \right| = \frac{x}{x^2 + y^2}$$

$$f_Y(y) = \frac{2}{\pi} \cdot \frac{x^2}{x^2 + y^2}$$

$$x = 1$$

Distribución Cauchy

b) El soporte de $y|x=x$: Dado que $y = x \cdot \tan(\theta)$ y θ está en el intervalo $[0, \pi/2]$, $\tan(\theta)$ está entre $[0, \infty)$.
 \therefore Soporte de $y|x=x$ es $[0, \infty)$

$$c) P[y \leq 10 | x=1] = \int_0^{10} \frac{2}{\pi} \frac{x}{x^2 + y^2} dy = \frac{2}{\pi} \arctan(y) \approx 0.93$$

Ejercicio 3:

3. Un mayorista requiere caracterizar el stock diario X de un determinado producto. Se sabe que el stock $X | \Lambda = \lambda$ se distribuye como una $\text{Poisson}(\lambda)$, para un $\lambda > 0$ determinado. La logística se planifica en base a la distribución *a priori* $\Lambda \sim \text{Gamma}(10, 1)$. Se registraron a lo largo de 6 días los siguientes stocks i.i.d. $\underline{x} = [20, 5, 6, 30, 2, 5]$.

- Encontrar la distribución *a posteriori* de Λ .
- Dada la muestra, estime la probabilidad de que una nueva observación X sea mayor a 30.

Sabemos que:

- El stock diario X de un producto determinado distribuye Poisson con parámetro λ .

Poisson	Poi(μ)	$(\mu^x e^{-\mu})/x!$
---------	--------------	-----------------------

- La logística se planifica con una *a priori* que distribuye Gamma con parámetros 10,1.

Gamma	$\Gamma(\nu, \lambda)$	$\frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$
-------	------------------------	--

- Debemos calcular *a posteriori*. Es posible realizar esto usando Bayes.

$$f_{\Theta | X=x}(\theta) \propto f_{X | \Theta=\theta}(x) * \pi_{\theta}$$

- Podemos reemplazar las funciones de distribución correspondientes:

$$f_{\Lambda | X=x}(\lambda) \propto \frac{\prod_{i=1}^n e^{-\lambda} \lambda^{x_i}}{x_i!} * \frac{1}{\Gamma(10)} \lambda^{10-1} e^{-\lambda} \mathbf{I}\{\lambda \geq 0\}$$

Reordenando los términos, tenemos que:

$$f_{\Lambda | X=x}(\lambda) \propto \lambda^{\sum_{i=1}^n x_i} * \lambda^{10-1} * e^{-\lambda n} * \frac{1}{\Gamma(10)x_i!} \mathbf{I}\{\lambda \geq 0\}$$

$$f_{\Lambda | X=x}(\lambda) \propto \lambda^{\sum_{i=1}^n x_i + 9} * e^{-(n+1)\lambda} * \frac{1}{\Gamma(10)x_i!} \mathbf{I}\{\lambda \geq 0\}$$

Notamos que la *A posteriori* $f_{\Lambda | X=x}(\lambda)$ distribuye $\text{Gamma} \sim \Gamma(78, 7)$

- Cálculo de probabilidades:

$$\begin{aligned} P(x > 30) &= \\ &= 1 - P(x < 30) = \\ &= 1 - \sum_{i=1}^{30} \int_0^{30} \lambda^{x_i} * e^{-\lambda} * \lambda^{78} * e^{-7} * \frac{7^{78}}{\Gamma(78)x_i!} d\lambda \\ &= \\ &= 1 - \int_0^{30} \lambda^{\sum_{i=1}^n x_i} * e^{-n\lambda} * \lambda^{78} * e^{-7} d\lambda = \\ &= 1 - \frac{7^{78}}{\Gamma(78)x_i!} * \int_0^{30} \lambda^{\sum_{i=1}^{30} x_i + 78} * e^{-(n+7)\lambda} d\lambda = \end{aligned}$$

Como $\int_0^{30} \lambda^{\sum_{i=1}^{30} xi + 78} * e^{-(n+7)\lambda} d\lambda =$ es la integral de una distribución del tipo Gamma sabemos que esa integral aproximadamente da 1.

$$= 1 - \frac{7^{78}}{\Gamma(78)x_i!} \cong 1$$

b)

$$\begin{aligned}
 P(x > 30) &= 1 - \sum_{x=0}^{30} \int_0^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{7^{78}}{\Gamma(78)} \lambda^{78} e^{-7\lambda} d\lambda \\
 &= 1 - \sum_{x=0}^{30} \frac{7^{78}}{\Gamma(78)} \frac{1}{x!} \int_0^{\infty} \lambda^{78+x} e^{-8\lambda} d\lambda \\
 &= 1 - \sum_{x=0}^{30} \frac{7^{78}}{\Gamma(78)} \frac{1}{x!} \cdot 8^{-(x+79)} \Gamma(x+79) \approx 0.003
 \end{aligned}$$

Integral definida

$$\int_0^{\infty} \lambda^{78+x} \exp(-8\lambda) d\lambda = 8^{-x-79} \Gamma(x+79) \text{ para } \operatorname{Re}(x) > -79$$

$\Gamma(x)$ es la función gamma

$\operatorname{Re}(z)$ es la parte real d

Integral indefinida

$$\int \lambda^{78+x} \exp(-8\lambda) d\lambda = -8^{-x-79} \Gamma(x+79, 8\lambda) + \text{constante}$$

$\Gamma(\alpha, x)$ es la función gamma incompleta

Entrada

$$\left(\sum_{x=0}^{30} 7^{78} \times \frac{1}{\Gamma(78) \times 30!} \right) ((-8)^{-30-79} \Gamma(30+79))$$

$\Gamma(x)$ es la función gamma

$n!$ es la función factorial

Resultado

$$\begin{aligned}
 &-(27566471617048239433957743232323609657645790645940522405241 \cdot \\
 &145765822987748882217849939311189559 / \\
 &8543948143683640329580086824678208458410818089426611079 \cdot \\
 &788166431288878903122562200091848347746304) \approx \\
 &-0.00322643
 \end{aligned}$$

3)

a. Stoca $x|\lambda = \lambda \Rightarrow$ Distribuye Poisson con parámetro λ .

A Priori $\lambda \sim \text{Gamma}(10, 1)$

Podemos calcular A Posteriori usando Bayes:

$$p(\lambda|x) \propto \underbrace{p(x|\lambda)}_{\text{La verosimilitud}} \cdot \underbrace{p(\lambda)}_{\text{a priori}}$$

$$p(x|\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \quad \text{donde } n=6 \text{ y } x=[20, 5, 6, 30, 2, 5]$$

$$p(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \quad \text{donde } \Gamma \text{ es la función Gamma}$$

$$a=10, b=1$$

$$\therefore \text{a posteriori} \rightarrow p(\lambda|x) \propto p(x|\lambda) \cdot p(\lambda)$$

$$p(\lambda|x) = \left(\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \left(\frac{1}{\Gamma(10)} \lambda^{10-1} e^{-\lambda} \right)$$

$$p(\lambda|x) \propto \left(\lambda^{\sum x_i} e^{-\lambda n} \lambda^{10-1} e^{-\lambda} \right) \left(\frac{1}{x_i! \Gamma(10)} \right) \mathbb{I}\{\lambda \geq 0\}$$

$$p(\lambda|x) \propto \left(\lambda^{\sum x_i + 9} e^{-(n+1)\lambda} \right) \frac{1}{x_i! \Gamma(10)} \mathbb{I}\{\lambda \geq 0\}$$

$$\text{A posteriori } p(\lambda|x) \text{ distribuye Gamma } \left(\left(\sum x_i + 9 + 1 \right), (n+1) \right)$$

$$\Gamma(78, 7)$$

b. Cálculo de probabilidades:

$$P(x > 30) = 1 - P(x \leq 30) = 1 - \left[\sum_{i=1}^{30} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \frac{78}{\Gamma(78)} \lambda^{78} e^{-7\lambda} d\lambda \right]$$

$x|\lambda \sim \Gamma(78, 8)$
 $x \sim \text{Poisson}(\lambda)$
 $\mathbb{I}\{\lambda \geq 0\}$

$$P(x > 30) = 1 - \int_0^\infty \frac{e^{-\lambda n} \lambda^{\sum x_i}}{x_i! \Gamma(78)} \frac{78}{\Gamma(78)} \lambda^{78} e^{-7\lambda} d\lambda = 1 - \int_0^\infty \frac{e^{-(n+7)\lambda} \lambda^{\sum x_i + 78}}{\Gamma(78) x_i!} d\lambda$$

integral 1.

$$= 1 - \frac{8^{78}}{\Gamma(78) x_i!} \approx 1$$

Asamblea

Ejercicio 4:

4. La producción anual (en kilos) de fruta de un manzano es una variable aleatoria con distribución normal con desvío estándar (conocido) de 40 kilos. Para estudiar la peso medio producido se plantan n árboles, suponiendo que el peso producido por cada uno es independiente de los demás.
- a) Encontrar el tamaño muestral n_0 más chico posible de manera de garantizar que un intervalo de confianza para μ de nivel 0.95 tenga una longitud no mayor a 10. *Sugerencia: buscar el intervalo de confianza para un n genérico y analiza como afecta a la longitud del intervalo.*
- b) Se observa una muestra aleatoria de tamaño n_0 encontrado en a) con promedio \bar{x} y se considera el test de hipótesis $\begin{cases} H_0 : \mu = 100 \\ H_1 : \mu \neq 100 \end{cases}$, ¿Cuál es el nivel de significancia del test $1\{|\bar{x} - 100| > 5\}$?

En primer lugar, como sabemos que el desvío estándar es conocido vamos a buscar qué pivote se adapta mejor a nuestro problema:

Dada \underline{X}_n una m.a. de una distribución $\mathcal{N}(\mu, \sigma^2)$ definimos algunos pivotes:

- Para la media con varianza conocida: $U(\underline{X}, \mu) = \frac{(\bar{X} - \mu)}{\sigma} \sqrt{n} \sim \mathcal{N}(0, 1)$

Por lo tanto, nuestro pivote distribuirá como una Normal (0, 1).

a- Tenemos que:

$$P\left(a < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < b\right) = 0.95$$

$$b = 1.96$$

$$a = -b = -1.96$$

$$-1.96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

$$-1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\mu \in \bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Queremos que $2 * 1.96 * \frac{\sigma}{\sqrt{n}} < 10$, entonces $n = 246$.

El tamaño muestral mínimo debe ser $n = 246$.

- b- Se llama nivel de significancia del test a la máxima probabilidad de cometer un error de tipo I, siendo el error de tipo I el error que se comete al rechazar una hipótesis que era verdadera. La idea es que tipo de error sea muy bajo.

$$\alpha = P(|\bar{X} - 100| < Z_{\frac{\alpha}{2}} * \frac{40}{\sqrt{246}})$$

$$\alpha = P(100 - Z_{\frac{\alpha}{2}} * \frac{40}{\sqrt{246}} < \bar{X} < 100 + Z_{\frac{\alpha}{2}} * \frac{40}{\sqrt{246}})$$

$$95 = 100 - Z_{\frac{\alpha}{2}} * \frac{40}{\sqrt{246}} \therefore Z_{\frac{\alpha}{2}} = -1.96$$

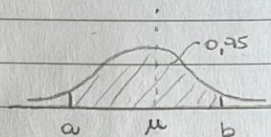
$$105 = 100 - Z_{\frac{\alpha}{2}} * \frac{40}{\sqrt{246}} \therefore Z_{\frac{\alpha}{2}} = 1.96$$

Por lo que vimos en el ítem a) el $\alpha = 0.05$

4) $\begin{cases} \sigma = 40 \text{ kilos} \\ n \text{ árboles?} \\ \text{peso de árboles iid.} \end{cases}$ me: $X \sim N(\mu, \sigma^2)$

Busco el pivote $O(\bar{x}, \mu) = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

$P\left(a < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < b\right) = 0.95$



$b = 1.96$
 $a = -b = -1.96$

$$-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96$$

$$-1.96 \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < \frac{\sigma}{\sqrt{n}} 1.96$$

$$\mu \in \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Quiero que $2 \cdot 1.96 \frac{\sigma}{\sqrt{n}} < 10 \therefore 2 \cdot 1.96 \frac{40}{10} < \sqrt{n}$

$$245.86 < n$$

\therefore El tamaño muestral mínimo debe ser $n = 246$

b) $H_0 \rightarrow \mu = 100$
 $H_1 \rightarrow \mu \neq 100$

$$\delta(x) = \mathbb{I} \left\{ \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| > z_{\alpha/2} \right\}$$

$$\alpha = P_{\mu=100} \left(\left| \bar{x} - 100 \right| > z_{\alpha/2} \frac{40}{\sqrt{246}} \right)$$

$$\alpha = P \left(100 - z_{\alpha/2} \frac{40}{\sqrt{246}} < \bar{x} < 100 + z_{\alpha/2} \frac{40}{\sqrt{246}} \right)$$

$$95 = 100 - z_{\alpha/2} \frac{40}{\sqrt{246}} \rightarrow z = 1.96$$

$$105 = 100 + z_{\alpha/2} \frac{40}{\sqrt{246}} \rightarrow z = 1.96$$

Asamblea