EJERCICIOS RELACION 1

1º Sea X una variable uniforme en el intervalo (0,1). Dada una m.a.s. de tamaño 2, calcule la densidad de la media muestral

Solución:

La distribución uniforme en (a,b) tiene como función de densidad $X \rightsquigarrow U(a,b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en el resto} \end{cases}$$

Sea X_1, X_2 una m.a.s. Queremos hallar $f_{\bar{X}}(x)$ si el intervalo es (0,1) y $X \rightsquigarrow U(0,1)$ la función de densidad es:

$$f_X(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

La función de distribución $U(0,1)$ no es reproductiva. Sea X_1, X_2 una m.a.s. de tamaño 2. Damos los siguientes pasos:

$X_1 \rightsquigarrow U(0,1)$, $X_2 \rightsquigarrow U(0,1)$

1º Formamos la función de densidad de la variable bidimensional (X_1, X_2), que como son independientes X_1 y X_2 verifican:

$$f(x_1,x_2) = f(x_1) \cdot f(x_2) = \begin{cases} 1 & \text{si } 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{en el resto} \end{cases}$$

2º Hacemos el cambio de variable $w = \frac{x_1 + x_2}{2}$, $u = x_1$

Entonces la función de densidad de la variable (w,u) es:

$$f(w,u) = f(x_1,x_2) \cdot |J(x_1,x_2)|$$

Tenemos que poner x_1 y x_2 en función de w y u.

$x_1 = u$

$\frac{x_1+x_2}{2} = w \Rightarrow \quad x_2 = 2w - x_1 \Rightarrow \quad x_2 = 2w - u$

por tanto:

$$J(x_1,x_2) = \begin{vmatrix} \frac{\partial x_1}{\partial w} & \frac{\partial x_1}{\partial u} \\ \frac{\partial x_2}{\partial w} & \frac{\partial x_2}{\partial u} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 2 & -1 \end{vmatrix} = -2$$

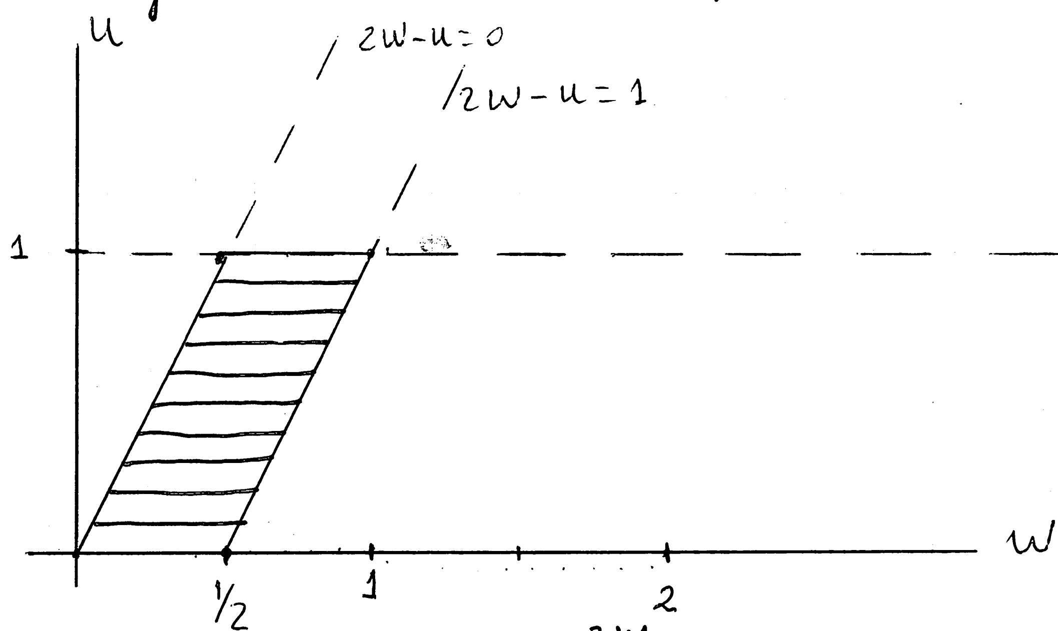
Por lo tanto:

$$f(w,u) = f(x_1,x_2) \cdot |-2| = +2 \cdot f(u, 2w-u) = \begin{cases} 2 \text{ si } 0 \le u \le 1, 0 < 2w-u < 1 \\ 0 \text{ en el resto} \end{cases}$$

$$f(w,u) = \begin{cases} 2 \text{ si } 0 \le u \le 1 \quad 0 \le 2w-u \le 1 \\ 0 \text{ en el resto} \end{cases}$$

Ahora lo que tenemos que hallar el la función de densidad marginal $f(w)$ que será la función de densidad de la media muestral.

Dibujamos el recinto $0 \le u \le 1 \quad 0 \le 2w-u \le 1$



$$f(w) = \int_{-\infty}^{\infty} f(w,u)\,du = \begin{cases} \int_0^{2w} 2\,du & \text{si } 0 \le w \le \frac{1}{2} \\ \int_{2w-1}^{1} 2\,du & \text{si } \frac{1}{2} \le w \le 1 \\ 0 \text{ en el resto} \end{cases}$$

La función de densidad es por tanto:

$$f(w) = \begin{cases} 4w & \text{si } 0 < w < \frac{1}{2} \\ 4(1-w) & \text{si } \frac{1}{2} < w < 1 \\ 0 \text{ en el resto} \end{cases}$$ esta es la función

de densidad de la media muestral $\bar{x} = w = \frac{x_1 + x_2}{2}$

2º ¿Cómo se distribuye la media muestral de una m.a.s. de tamaño n de una v.a. exponencial de parámetro λ? Compruebe mediante la correspondiente integral, cual es la esperanza de esa media muestral

a) Sea $X \rightsquigarrow Exp(\lambda)$. Tomamos una m.a.s $x_1, x_2, \dots x_n$. La distribución exponencial no es reproductiva, pero teniendo en cuenta que $Exp(\lambda) = G(1, \lambda)$ y la gamma si es reproductiva al ser $x_i \rightsquigarrow G(1, \lambda)$ $\forall i = 1 \dots n$ $\Rightarrow$

$\Rightarrow T = \sum_{i=1}^{n} x_i \rightsquigarrow G(n, \lambda)$. Por tanto la función de densidad de T es:

$$f_T(t) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} \cdot e^{-\lambda t} \cdot t^{n-1} & \text{si } t > 0 \quad p > 0 \\ 0 & \text{en el resto.} \end{cases}$$

Hay que tener en cuenta que si $n \in \mathbb{N} \Rightarrow \Gamma(n) = (n-1)!$

Por la propiedad del cambio de variable, como $\bar{x} = \frac{T}{n} \Rightarrow T = n\bar{x}$, por tanto

$$f_{\bar{x}}(x) = f_T(nx) \cdot \left| \frac{dT}{dx} \right| = n \cdot f_T(nx)$$

$$f_{\bar{x}}(x) = \begin{cases} \frac{n \cdot \lambda^n}{(n-1)!} \cdot e^{-\lambda n x} \cdot (xn)^{n-1} & \text{si } x > 0 \quad p > 0 \\ 0 \text{ en el resto} \end{cases}$$

Podemos poner entonces que:

$$f_{\bar{x}}(x) = \begin{cases} \dfrac{n\,\lambda^n}{(n-1)!} \cdot e^{-n\lambda x} \cdot x^{n-1} \cdot n^{n-1} & \text{si } x>0 \quad p>0 \\ 0 \text{ en el resto} \end{cases}$$

$$f_{\bar{x}}(x) = \begin{cases} \dfrac{(n\lambda)^n}{(n-1)!} \cdot e^{-(\lambda n)x} \cdot x^{n-1} & \text{si } x>0 \quad p>0 \\ 0 \quad \text{en el resto} \end{cases}$$

Luego $\bar{x}$ se distribuye según la siguiente gamma

$\bar{x} \rightsquigarrow G(n, n\lambda)$

b) $E[\bar{x}] = \int_{-\infty}^{\infty} x\, f_{\bar{x}}(x)\, dx = \int_0^{\infty} x \cdot \dfrac{(n\lambda)^n}{\Gamma(n)} \cdot e^{(-\lambda n)x} \cdot x^{n-1} dx$

$$E[\bar{x}] = \frac{(n\lambda)^n}{\Gamma(n)} \cdot \int_0^{\infty} e^{-(n\lambda)x} \cdot x^n\, dx$$

Teniendo en cuenta que:

$$\int_0^{\infty} e^{-(n\lambda)x} \cdot x^n\, dx = \frac{\Gamma(n+1)}{(\lambda n)^{n+1}}$$

$$E[\bar{x}] = \frac{(n\lambda)^n}{(n-1)!} \cdot \frac{n!}{(\lambda n)^{n+1}} = \frac{n}{\lambda n} = \frac{1}{\lambda}$$

3º Sean dos m.a.s. independientes entre sí, de tamaños n_1 y n_2, procedentes de variables con distribuciones no necesariamente idénticas, aunque sí con la misma varianza σ^2. Sean M_1 y M_2 las correspondientes medias muestrales y sea M la media muestral de la muestra conjunta formada por las n_1+n_2 variables. Demuestre que se cumple

$$Var(M-M_1) = n_2\sigma^2/(n_1(n_1+n_2))$$

Tomemos dos muestras aleatorias simples:

$$X_1, X_2, \dots X_{n_1} \Rightarrow M_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1}$$

$$Y_1, Y_2, \dots, Y_{n_2} \Rightarrow M_2 = \frac{\sum_{j=1}^{n_2} Y_j}{n_2}$$

Luego: $\sum_{i=1}^{n_1} X_i = n_1 \cdot M_1$

$$\sum_{j=1}^{n_2} Y_j = n_2 M_2$$

Como $M = \dfrac{\sum_{i=1}^{n_1} X_i + \sum_{j=1}^{n_2} Y_j}{n_1+n_2} \Rightarrow M = \dfrac{n_1M_1+n_2M_2}{n_1+n_2}$

Entonces:

$$M-M_1 = \frac{n_1 M_1 + n_2M_2}{n_1+n_2} - M_1 = \frac{\cancel{n_1M_1} + n_2M_2 - \cancel{M_1n_1} - M_1n_2}{n_1+n_2}$$

$M-M_1 = \dfrac{n_2(M_2-M_1)}{n_1+n_2}$. Por tanto:

$Var[M-M_1] = \dfrac{n_2^2}{(n_1+n_2)^2} \cdot Var[M_2-M_1]$. Si suponemos que las muestras son grandes, aplicando el teorema central del límite

$$X_i \rightsquigarrow N(\mu_X, \frac{\sigma_X}{\sqrt{n_1}}) \quad i=1\dots n_1$$

$$Y_j \rightsquigarrow N(\mu_Y, \frac{\sigma_Y}{\sqrt{n_2}}) \quad j=1\dots n_2$$

Por tanto:

$$\bar{X}-\bar{Y}=M_2-M_1 \rightarrow N\left(\mu_X-\mu_Y, \sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}\right)$$

Luego:

$$Var(M_2-M_1)=\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2} \quad \text{y como } \sigma_x=\sigma_y$$

$$Var(M_2-M_1)=\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}$$

Por tanto:

$$Var[M-M_1]=\frac{n_2^2}{(n_1+n_2)^2}\left(\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right)$$

$$Var[M-M_1]=\frac{n_2^{\cancel{2}}\cdot\sigma^2}{(n_1+n_2)^{\cancel{2}}}\cdot\frac{\cancel{n_1+n_2}}{n_1\cdot\cancel{n_2}}=\frac{n_2\,\sigma^2}{n_1(n_1+n_2)}$$

4º Una empresa de software está preparando el lanzamiento al mercado de su último sistema operativo a pesar de que es consciente de que el producto contiene errores. Sabe que en la última muestra de 1000 arranques fallaron 5. Para intentar reducir el nivel de errores del programa realiza algunas modificaciones. En una muestra de 395 arranques del programa ya modificado fallan 3. Si la tasa real del fallo del producto en ambos casos fuera del 1%, halle la probabilidad de cada uno de esos resultados muestrales. ¿Cuál de los dos era más probable?

Definimos la variable X = fallo en el arranque. Entonces X se distribuye según una distribución binomial de parámetro p = 0'01 (aquí el exito es que falle el arranque). La distribución de Bernouilli no es reproductiva pero la suma de n-variables de Bernouilli se distribuyen según una binomial de parámetros n y p, es decir

Si $X \rightsquigarrow b(0'01)$ y tomamos una m.a.s

$X_1, X_2, \dots X_n$ / $X_i \rightsquigarrow b(0'01)$ $i = 1 \dots n$

entonces:

$$T = \sum_{i=1}^{n} X_i \rightsquigarrow B(n, 0'01)$$

En el primer caso tomamos una m.a.s. de tamaño 1000: $X_1, X_2, \dots, X_{1000}$. Por tanto

$$T = \sum_{i=1}^{1000} X_i \rightsquigarrow B(1000, 0'01)$$

La probabilidad de que haya 5 fallos de arranque en esta muestra es:

$$P[T=5] = \binom{1000}{5} 0'01^5 \cdot 0'99^{995} = 0'03745.$$

En el segundo caso se toma una m.a.s. de tamaño 395 encontrandose 3 fallos. La probabilidad de encontrar 3 fallos de arranque en una muestra de 395 arranques es:

$$T = \sum_{i=1}^{395} X_i \rightsquigarrow B(395, 0'01)$$

$$P[T=3] = \binom{395}{3} \cdot 0'01^3 \cdot 0'99^{392} = 0'1983$$

Luego es más probable encontrar 3 fallos en el arranque de una muestra de 395 que encontrar 5 fallos en el arranque de una muestra de 1000.

Como las dos muestras son grandes ($1000 > 30$, $395 > 30$) podríamos haber aproximado la binomial a la normal. Aplicando el teorema central del límite:

• En el primer caso, como $T \rightsquigarrow B(1000, 0'01)$ entonces T se puede aproximar a una normal con media $\mu = n \cdot p = 1000 \cdot 0'01 = 10$ y desviación típica $\sigma = \sqrt{npq} = \sqrt{9'9}$. Entonces

$T \rightsquigarrow N(10, \sqrt{9'9})$. Como queríamos calcular $P[T=5]$ si pasamos a variable continua, cambiamos el concepto de probabilidad puntual por el de probabilidad en un intervalo, tenemos que hacer una corrección de continuidad

$P[T=5] \Rightarrow P[4'5 \leq T \leq 5'5]$ y si tipificamos
$T \rightsquigarrow B(1000, 5)$ $\quad T \rightsquigarrow N(10, \sqrt{9'9})$

$$P[4'5 \leq T \leq 5'5] = P\left[\frac{4'5-10}{\sqrt{9'9}} < T \leq \frac{5'5-10}{\sqrt{9'9}}\right]$$

$$= P[-1'74 \leq Z \leq -1'43] = P[Z \leq -1'43] - P[Z \leq -1'74]$$
$$= P[Z > 1'43] - P[Z > 1'74] = 1 - P(Z \leq 1'43) - (1 - P(Z \leq 1'74))$$
$$= P(Z \leq 1'74) - P(Z \leq 1'43) = 0'95907 - 0'92364 = 0'03543$$

De modo análogo se podría calcular $P[T=3]$ considerando que $T \rightsquigarrow N(3'95, \sqrt{3'91})$

5º Enuncie los resultados que son aplicables al comportamiento asintótico de la media muestral de una muestra aleatoria simple de una variable de media y varianza finitas. Comente las consecuencias que ello tiene para el comportamiento probabilístico de la media muestral.

Sea $X \rightsquigarrow F_X(\mu, \sigma)$, F_X no es conocida. Estudiar el comportamiento asintótico de la media y la varianza muestral consiste en considerar muestras de tamaño muy grande $(n \to \infty)$ y ver que pasa con la media y la varianza muestrales.

Si $n \to \infty$ entonces se trata de poblaciones grandes, según el teorema central del límite $X \rightsquigarrow N(\mu, \sigma)$ y por tanto si $n \to \infty$ se verifica que:

$$\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\frac{n s^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

Si $\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, podemos calcular la probabilidad de que la media muestral se encuentre en un intervalo, no podemos saber la probabilidad de que $\bar{x}$ tome un valor en concreto.

6º Sea una población $N(\mu = 5, \sigma^2 = 0'01)$ y una m.a.s. de tamaño 16 extraída de ella. Halle:

a) $P(5 < \bar{x} < 5'2)$

b) $P(S^2 < 0'019125)$

a) Sabemos que $\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(5, \frac{0'1}{\sqrt{16}}\right) \Rightarrow$

$\Rightarrow \bar{x} \rightsquigarrow N(5, 0'025)$

Por tanto:

$$P(5 < \bar{x} < 5'2) = P\left(\frac{5-5}{0'025} < \frac{\bar{x}-5}{0'025} < \frac{5'2-5}{0'025}\right)$$

$$= P(0 < Z < 8) = P(Z<8) - P(Z<0) = 1 - \frac{1}{2} = \frac{1}{2}$$

b) Sabemos que $\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$, es decir

$\frac{16 S^2}{0'01} \rightsquigarrow \chi^2_{15} \Rightarrow 1600 S^2 \rightsquigarrow \chi^2_{15}$. Por tanto

$$P(S^2 < 0'019125) = P(1600 S^2 < 1600 \cdot 0'019125) =$$

$$= P(Z < 30'6) = P(\chi^2_{15} < 30'6) = 0'99$$

7º De una población normal de media μ y varianza 4'5 se extraen 2 m.a.s. independientes entre sí, de tamaño n. ¿Qué valor ha de tener n para que podamos estar seguros, con una confianza de al menos del 95%, de que las medias muestrales difieren entre sí dos unidades como máximo?

Sabemos que si $X \rightsquigarrow N(\mu_x, \sigma_x)$, $Y \rightsquigarrow N(\mu_y, \sigma_y)$ y tomamos 2 m.a.s. en cada v.a:

$x_1, x_2, \dots x_n$

$y_1, y_2, \dots y_m$

Entonces la diferencia de las medias muestrales $\bar{X} - \bar{Y}$ se distribuye como:

$$\bar{X} - \bar{Y} \rightsquigarrow N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

En este caso $\mu_x = \mu_y = \mu$, $\sigma_x^2 = \sigma_y^2 = 4'5$ por lo que:

$\bar{X} - \bar{Y} \rightsquigarrow N\left(0, \frac{3}{\sqrt{n}}\right)$. Ha de verificarse que:

$$P[|\bar{X}-\bar{Y}| \leq 2] \geq 0'95 \iff P[-2 \leq \bar{X}-\bar{Y} \leq 2] \geq 0'95$$

$$\iff P\left[\frac{-2\sqrt{n}}{3} \leq Z \leq \frac{2\sqrt{n}}{3}\right] \geq 0'95, \quad Z = \frac{(\bar{X}-\bar{Y})\sqrt{n}}{3}$$

$$P\left(\bar{Z} \leq \frac{2\sqrt{n}}{3}\right) - P\left[Z \leq -\frac{2\sqrt{n}}{3}\right] \geq 0'95$$

$$P\left(\bar{Z} \leq \frac{2\sqrt{n}}{3}\right) - P\left[Z \geq \frac{2\sqrt{n}}{3}\right] \geq 0'95$$

$$2P\left(\bar{Z} \leq \frac{2\sqrt{n}}{3}\right) - 1 \geq 0'95 \Rightarrow P\left(\bar{Z} \leq \frac{2\sqrt{n}}{3}\right) \geq \frac{1'95}{2}$$

$P\left(\bar{Z} \leq \frac{2\sqrt{n}}{3}\right) \geq 0'975$. Si buscamos en la tabla de la normal:

$$\frac{2\sqrt{n}}{3} \geq 1'96 \Rightarrow n = \left(\frac{3 \cdot 1'96}{2}\right)^2 \geq 8'6436 \simeq 9.$$

8º Consideremos dos muestras de tamaño n y m respectivamente, de una población $X \rightsquigarrow N(\mu_x, \sigma_x^2)$ la primera e $Y \rightsquigarrow N(\mu_y, \sigma_y^2)$ la segunda. Supongamos que las muestras son independientes entre sí sean $\bar{X}$ e $\bar{Y}$ las respectivas medias muestrales. y S_{cx}^2, S_{cy}^2 las cuasi varianzas. Determine la distribución

a) $$Z = \frac{n-1}{\sigma_x^2} S_{cx}^2 + \frac{m-1}{\sigma_y^2} S_{cy}^2$$

b) $$\frac{\left[(\bar{X}-\bar{Y}) - (\mu_x - \mu_y)\right]\sqrt{n+m-2}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\ \sqrt{Z}}$$

a) Según el lema de Fisher:

$$\frac{n-1}{\sigma_x^2} \cdot S_{cx}^2 \rightsquigarrow \chi_{n-1}^2$$

$$\frac{m-1}{\sigma_y^2} \cdot S_{cy}^2 \rightsquigarrow \chi_{m-1}^2$$

Como la distribución JI-CUADRADO es reproductiva:

$$z = \frac{n-1}{\sigma_x^2} S_{cx}^2 + \frac{m-1}{\sigma_y^2} S_{cy}^2 \rightsquigarrow \chi^2_{m+n-1}$$

b) Según el lema de Fisher:

$$\bar{x} \rightsquigarrow N\left(\mu_x, \frac{\sigma_x^2}{n}\right) \quad \bar{y} \rightsquigarrow N\left(\mu_y, \frac{\sigma_y^2}{m}\right)$$

También sabemos que:

$\bar{x} - \bar{y} \rightsquigarrow N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)$. Si tipificamos:

$$\frac{(\bar{x}-\bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \rightsquigarrow N(0,1)$$

Según la definición de la t de Studen:

$$\frac{\dfrac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}}{\sqrt{\dfrac{z}{m+n-2}}} = \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{m+n-2}}{m+n-2}}} \rightsquigarrow t_{m+n-2}$$

es decir que:

$$\frac{[(\bar{x}-\bar{y}) - (\mu_x-\mu_y)]\sqrt{m+n-2}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \cdot \sqrt{z}} \rightsquigarrow t_{m+n-2}$$

9) Se toma una muestra aleatoria de tamaño 5 de una población normal con media 2'5 y varianza 36.

a) Halle una cota superior de la varianza muestral con probabilidad de al menos 0'9

b) ¿Cual es la probabilidad de que la media muestral esté entre 1'3 y 3'5 si la varianza muestral se sabe que está entre 30 y 40?

a) $X \rightsquigarrow N(2'5, 36=\sigma^2)$

Sabemos que $\dfrac{5 S^2}{36} \rightsquigarrow \chi^2_4$. Si K es la cota superior:

$$P[S^2 \leq K] = 0'9 \Rightarrow P\left[\frac{5S^2}{36} \leq \frac{5K}{36}\right] \geq 0'9$$

$$P\left[\chi^2_4 \leq \frac{5K}{36}\right] \geq 0'9$$

$$\frac{5K}{36} \geq 1'06 \Rightarrow K \geq \frac{36 \cdot 1'06}{5} = 7'632 \simeq 8$$

b) Sabemos, según el lema de Fisher, que:

$$\bar{X} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(2'5, \frac{6}{\sqrt{5}}\right) = N(2'5, 2'68)$$

Nos piden que calculemos la expresión:

$$P[1'3 \leq \bar{X} \leq 3'5 / 30 \leq S^2 \leq 40] = P[1'3 \leq \bar{X} \leq 3'5]$$

ya que según el lema de Fisher la media muestral $\bar{X}$ y la varianza muestral son independientes.

$$P[1'3 \leq \bar{X} \leq 3'5] = P\left[\frac{1'3-2'5}{2'68} < \bar{Z} < \frac{3'5-2'5}{2'68}\right]$$

siendo $\bar{Z} = \dfrac{\bar{X}-2'5}{2'68}$.

Lo que tenemos que calcular es:

$$P[-0'45 \leq \bar{Z} \leq 0'37] = P(\bar{Z} \leq 0'37) - P(Z \leq -0'45)$$

$$= P(\bar{Z} \leq 0'35) - 1 + P(Z \leq 0'45) = 0'6368 - 1 + 0'92647$$

Luego:

$$P[1'3 \leq \bar{X} \leq 3'5] = 0'56327.$$

10º Sean $S_{c_1}^2$ y $S_{c_2}^2$ las cuasivarianzas muestrales de dos muestras normales e independientes de tamaños 5 y 4 respectivamente. Supongamos que las varianzas poblacionales son iguales. Encuentre el punto crítico c que verifica:

$$P\left[\frac{S^2_{c_1}}{S^2_{c_2}} < c\right] = 95\%$$

Según el lema de Fisher se verifica que

$$\frac{(n-1)S_c^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}, \text{ en este caso}$$

$$\frac{4S_{c_1}^2}{\sigma^2} \rightsquigarrow \chi^2_4$$

$$\frac{3S_{c_2}^2}{\sigma^2} \rightsquigarrow \chi^2_3 \text{ . Entonces:}$$

$$\frac{\dfrac{4S_{c_1}^2}{4\sigma^2}}{\dfrac{3S_{c_2}^2}{3\sigma^2}} = \frac{\dfrac{\chi^2_4}{4}}{\dfrac{\chi^2_3}{3}} = \frac{S_{c_1}^2}{S_{c_2}^2} \rightsquigarrow F_{4\,3}$$

Nos piden $P\left[\frac{S^2_{c_1}}{S^2_{c_2}} < c\right] \geq 0'95 \Rightarrow$

$$P[F_{3\,4} < c] = 0'95 \Rightarrow c \geq 6'59 \Rightarrow c \geq 7$$

11.- Dados dos poblaciones normales $X \rightsquigarrow N(m_1, \sigma^2)$ e $Y \rightsquigarrow N(m_2, \sigma^2)$ independientes entre sí; de cada una se extrae una m.a.s. de tamaño n. Obtenga la distribución que sigue $\frac{S_{c_1}^2}{S_{c_2}^2}$ y calcule $P[S_1^2 < 9 S_2^2]$ si $n=3$ (S^2 indica la varianza muestral y S_c^2 la cuasivarianza muestral)

Según el lema de Fisher

$$\frac{(n-1)\cdot S_{c_1}^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

$$\frac{(n-1)\cdot S_{c_2}^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$. Por tanto

$$\frac{\dfrac{\cancel{(n-1)}\cdot S_{c_1}^2}{\cancel{(n-1)}\,\sigma^2}}{\dfrac{\cancel{(n-1)}\cdot S_{c_2}^2}{\cancel{(n-1)}\,\sigma^2}} = \frac{\dfrac{\chi^2_{n-1}}{n-1}}{\dfrac{\chi^2_{n-1}}{n-1}} \rightsquigarrow F_{n-1,n-1}$$

Por tanto $\frac{S_{c_1}^2}{S_{c_2}^2} \rightsquigarrow F_{n-1,n-1}$.

Si $n=3$ se verifica que $\frac{S_{c_1}^2}{S_{c_2}^2} \rightsquigarrow F_{2\,2}$

Como $S^2 = \frac{(n-1)\,S_{c_1}^2}{n}$ entonces:

$$\frac{S_1^2}{S_2^2} = \frac{\dfrac{(n-1)\,S_{c_1}^2}{n}}{\dfrac{(n-1)\,S_{c_2}^2}{n}} = \frac{S_{c_1}^2}{S_{c_2}^2} \rightsquigarrow F_{2\,2}$$

$$P[S_1^2 < 9S_2^2] = P\left[\frac{S_1^2}{S_2^2} < 9\right] = P[F_{2\,2} < 9] = 0'9$$

12º

Sean dos muestras aleatorias simples de tamaño 11 de una misma variable aleatoria normal. Se desea que las muestras sean tales que la suma de sus medias muestrales no se aleje del doble de la media poblacional en más de la quinta parte de lo que vale la desviación típica poblacional, y que la suma de las varianzas muestrales sea menor que el 99% de lo que vale la propia varianza poblacional. ¿Cuál es la probabilidad de que las muestras cumplan al mismo tiempo ambas condiciones? Si una vez escogidas las muestras se comprueba que la suma de medias muestrales de hecho se aleja del doble de la media poblacional en más de la mitad de la desviación típica poblacional, ¿cuál es la probabilidad de que las varianzas muestrales cumplan la condición expresada para ellas anteriormente?

$X \rightsquigarrow N(\mu, \sigma)$ Tomamos una m.a.s $x_1, x_2, \dots x_{11}$

$Y \rightsquigarrow N(\mu, \sigma)$ Tomamos una m.a.s. $x_1, x_2, \dots x_{11}$

Según el lema de Fisher

i) $\bar{x}, \bar{Y}, S_x^2, S_y^2$ son independientes. Además

$\bar{x} \rightsquigarrow N(\mu, \frac{\sigma}{\sqrt{11}})$

$\bar{Y} \rightsquigarrow N(\mu, \frac{\sigma}{\sqrt{11}})$ y como la distribución normal es reproductiva podemos afirmar que:

$$\bar{x}+\bar{Y} \rightsquigarrow N(2\mu, \frac{2\sigma}{\sqrt{11}}) \Rightarrow \frac{\bar{x}+\bar{Y}-2\mu}{\frac{2\sigma}{\sqrt{11}}} \rightsquigarrow N(0,1)$$

$$\Rightarrow \frac{\sqrt{11}(\bar{x}+\bar{y}-2\mu)}{2\sigma} \rightsquigarrow N(0,1).$$

ii) $\frac{11 S_x^2}{\sigma^2} \rightsquigarrow \chi^2_{10}$

$\frac{11 \cdot S_y^2}{\sigma^2} \rightsquigarrow \chi^2_{10}$ Como la distribución Ji-Cuadrado es reproductiva:

$$\frac{11\,S^2_x}{\sigma^2} + \frac{11\,S^2_y}{\sigma^2} = \frac{11}{\sigma^2}\left(S^2_x + S^2_y\right) \rightsquigarrow \chi^2_{20}$$

Nos piden que hallemos la siguiente probabilidad:

$$P\left[\,|\bar{x}+\bar{y}-2\mu| < \frac{1}{5}\sigma \;\cap\; S^2_x + S^2_y \leq 0'99\sigma^2\right]$$

Como $\bar{x}$, $\bar{y}$, S^2_x y S^2_y son independientes esta probabilidad se puede expresar también como:

$P\left[\,|\bar{x}+\bar{y}-2\mu| < \frac{1}{5}\sigma\right] \cdot P\left[S^2_x + S^2_y \leq 0'99\sigma^2\right]$, vamos a calcular cada una de estas probabilidades por separado:

$$P\left[|\bar{x}+\bar{y}-2\mu| < \frac{1}{5}\sigma\right] = P\left[-\frac{1}{5}\sigma < \bar{x}+\bar{y}-2\mu < \frac{1}{5}\sigma\right] =$$

$$= P\left[\frac{-\frac{1}{5}\sigma}{\frac{2\sigma}{\sqrt{11}}} < \frac{\bar{x}+\bar{y}-2\mu}{\frac{2\sigma}{\sqrt{11}}} < \frac{\frac{1}{5}\sigma}{\frac{2\sigma}{\sqrt{11}}}\right] =$$

$$= P\left[-\frac{\sqrt{11}}{10} < Z < \frac{\sqrt{11}}{10}\right] = P\left[Z < \frac{\sqrt{11}}{10}\right] - P\left[Z < -\frac{\sqrt{11}}{10}\right]$$

$$= 2P\left[Z < \frac{\sqrt{11}}{10}\right] - 1 = \underset{\downarrow \text{Tabla de la normal}}{2 \cdot 0'6293} - 1 = 0'2586$$

$$P\left[S^2_x + S^2_y \leq 0'99\sigma^2\right] = P\left[\frac{S^2_x + S^2_y}{\sigma^2} \leq 0'99\right] =$$

$$= P\left[\frac{11\left(S^2_x + S^2_y\right)}{\sigma^2} < 11 \cdot 0'99\right] = P\left[\chi^2_{20} < 10'89\right] =$$

$$= 0'90$$

Luego:

$$P\left[|\bar{x}+\bar{y}-2\mu| < \frac{1}{5}\sigma \cap S^2_x + S^2_y < 0'99\sigma^2\right] = 0'2586 \cdot 0'90 = 0'23274$$

13º La duración (en horas) de una cierta marca de bombillas sigue una distribución $N(\mu = 1000, \sigma^2 = 100^2)$. Se desea enviar una muestra de bombillas de modo que la duración media muestral no difiera de la media poblacional en más de 50h. con una probabilidad de al menos 0'95. Halle el tamaño que debe tener la muestra. Repita el ejercicio si la distribución concreta de la duración de las bombillas es desconocida (aunque la media y la desviación típica se mantienen)

Como $X \rightsquigarrow N(1000, 100) \Rightarrow \bar{X} \rightsquigarrow N\left(1000, \frac{100}{\sqrt{n}}\right)$

Del enunciado se desprende que si tomamos una muestra aleatoria simple de tamaño n entonces

$$P[\,|\bar{X}-\mu| < 50\,] \geq 0'95 \iff P[-50 < \bar{X}-\mu < 50] \geq 0'95$$

$$\Rightarrow P\left[\frac{-50}{\frac{100}{\sqrt{n}}} < \frac{\bar{X}-\mu}{\frac{100}{\sqrt{n}}} < \frac{50}{\frac{100}{\sqrt{n}}}\right] \geq 0'95$$

$$P\left[-\frac{\sqrt{n}}{2} < Z < \frac{\sqrt{n}}{2}\right] \geq 0'95$$

$$2P\left[Z < \frac{\sqrt{n}}{2}\right] - 1 \geq 0'95 \Rightarrow P\left[Z < \frac{\sqrt{n}}{2}\right] \geq \frac{1'95}{2}$$

$$P\left[Z < \frac{\sqrt{n}}{2}\right] \geq 0'975 \Rightarrow \frac{\sqrt{n}}{2} \geq 1'96 \Rightarrow n \geq 3'92^2$$

$n \geq 15'36$

Recordemos el teorema de Chebyshev

Sea X una v.a. distribuida con media μ y desviación típica σ. Entonces $\forall K > 0$ se verifica que:

$$P[\,|X-\mu| < K\sigma\,] \geq 1 - \frac{1}{K^2}$$

. Si queremos

Si la distribución no es conocida sabemos que:

$E[\bar{X}] = \mu$

$E[S^2] = \dfrac{\sigma^2}{n}$

La media es ahora μ y la varianza $\frac{\sigma^2}{n}$

Aplicando el teorema de Chebyshev:

$$P\left[|x-\mu| < k \cdot \frac{\sigma}{\sqrt{n}} \right] \geq 1 - \frac{1}{k^2}$$

$$k \cdot \frac{\sigma}{\sqrt{n}} = 50 \Rightarrow k = \frac{50 n}{\sigma} = \frac{50 n}{100} = \frac{n}{2}$$

Entonces:

$$P\left[|x-\mu| < 50 \right] \geq 1 - \frac{1}{\frac{n^2}{4}} = 1 - \frac{4}{n^2} \geq 0'95$$

$$\frac{4}{n^2} \leq 0'05 \Rightarrow \frac{4}{0'05} < n^2 \Rightarrow n > \sqrt{80} = 8'94$$

Luego la muestra ha de ser $n > 9$

14 a) Sea una población normal. Enuncie los resultados muestrales que se emplean para realizar inferencia sobre la varianza poblacional tanto si la media es conocida como si no. Comente en qué consiste y a qué es debido el problema de la pérdida de grados de libertad en los estadísticos muestrales. ¿Qué consecuencias cree que tiene para la inferencia que se realiza con ellos?

b) Sea una m.a.s. $x_1, x_2, \dots x_n$ de una población $N(0, \sigma^2)$. Razone como se distribuiría el estadístico

$$T = \frac{\sum_1^n x_i^2}{(\sum x_i)^2/n} - 1$$

a) Inferencia sobre σ^2 con μ conocida.

Si μ es conocida, como $X \rightsquigarrow N(\mu, \sigma)$, si tomamos una m.a.s $x_1, x_2, \dots x_n$ entonces se verifica que $x_i \rightsquigarrow N(\mu, \sigma) \Rightarrow \frac{x_i - \mu}{\sigma} \rightsquigarrow N(0,1)$. Por tanto $\frac{(x_i-\mu)^2}{\sigma^2} \rightsquigarrow N(0,1)^2 \Rightarrow \sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2} \rightsquigarrow \chi^2_n$

Si μ es conocida entonces:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \rightsquigarrow \chi^2_n$$

Inferencia sobre σ^2 siendo μ desconocida.
Según el lema de Fisher tenemos que

$$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1} \quad , \quad S^2 = \frac{\sum(x_i-\bar{x})^2}{n}$$

Podemos concluir diciendo que con la media desconocida perdemos grados de libertad y con esta perdida de grados de libertad se pierde precisión.

b) $$T = \frac{\sum x_i^2}{(\sum x_i)^2/n} - 1$$

Numerador: como $x_i \rightsquigarrow N(0,\sigma) \Rightarrow \frac{x_i}{\sigma} \rightsquigarrow N(0,1)$

$$\Rightarrow \frac{x_i^2}{\sigma^2} \rightsquigarrow \chi^2_1 \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \rightsquigarrow \chi^2_n$$

Denominador: $x_i \rightsquigarrow N(0,\sigma^2) \Rightarrow \sum_{i=1}^{n} x_i \rightsquigarrow N(0, n\sigma^2)$

ya que la población normal es reproductiva $\Rightarrow$

$$\Rightarrow \frac{\sum_{i=1}^{n} x_i}{\sqrt{n}\sigma} \rightsquigarrow N(0,1) \Rightarrow \frac{(\sum_{i=1}^{n} x_i)^2}{n\sigma^2} \rightsquigarrow \chi^2_1$$

Luego:

$$Y = \frac{\frac{1}{\cancel{n\sigma^2}} \cdot \sum_{i=1}^{n} x_i^2}{\frac{(\sum x_i)^2}{\cancel{n\sigma^2}}} = \frac{\frac{\chi^2_n}{n}}{\frac{\chi^2_1}{1}} \rightsquigarrow F_{n,1}$$

Es decir que:

$$Y = \frac{\sum_{i=1}^{n} x_i^2}{(\sum_{i=1}^{n} x_i)^2} \rightsquigarrow F_{n,1}$$

La distribución del estadístico que nos preguntan es:

$T = n \cdot Y - 1 \Rightarrow Y = \frac{T+1}{n}$. Mediante la fórmula del cambio de variable

$$f_T(t) = f_Y\left(\frac{t+1}{n}\right) \cdot \left|\frac{dY}{dT}\right| = f_Y\left(\frac{T+1}{n}\right) \cdot \frac{1}{n}$$

donde $Y \rightsquigarrow F_{n,1}$

Como la función de densidad de Y es:

$$f_Y(y) = \begin{cases} \dfrac{\left(\frac{n}{1}\right)^{\frac{n}{2}}}{\beta\left(\frac{n}{2}, \frac{1}{2}\right)} \cdot y^{\frac{n}{2}} \cdot \left(\frac{ny}{1} + 1\right)^{\frac{m+n}{2}} & \text{si } y > 0 \\ 0 & \text{en el resto} \end{cases}$$

La función de densidad de T es:

$$f_T(t) = \begin{cases} \cancel{\frac{1}{n}} \cdot \dfrac{\cancel{n}}{\beta\left(\frac{n}{2}, \frac{1}{2}\right)} \cdot \left(\frac{t+1}{n}\right)^{\frac{n}{2}} \cdot (t+2)^{\frac{m+1}{2}} & \text{si } \frac{T+1}{n} > 0 \\ 0 & \text{en el resto} \end{cases}$$

15. Considerese una m.a.s. $X_1, X_2, \dots X_n$ extraída de una población $N(1, \sigma^2)$. Sean los estadísticos $T_1 = S_X^2$, $T_2 = S_{cX}^2$, $T_3 = \left(\sum (X_i - 1)^2 / n\right)$. Halle sus medias y sus varianzas. Comente las propiedades estadísticas que haya utilizado.

a) Media y varianza de S_X^2

En general, para cualquier población, se verifica que la varianza muestral S^2 y la varianza poblacional σ^2 están relacionadas por:

$S^2 = \frac{n-1}{n} \sigma^2$, esto nos indica que la varianza muestral es siempre menor que la varianza poblacional (cosa que siempre es de esperar). La media poblacional o esperanza de $S^2 = \frac{n-1}{n}\sigma^2$ es:

$E[S^2] = E[\frac{n-1}{n}\sigma^2]$ Al ser constante $\frac{n-1}{n}\sigma^2$

$\Rightarrow E[S^2] = \frac{n-1}{n}\sigma^2$

Como estamos en una población normal $N(\mu,\sigma^2)$ podemos aplicar el lema de Fisher:

$\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$. Por tanto:

$$E[\frac{nS^2}{\sigma^2}] = \frac{n^2}{\sigma^2} \cdot E[S^2] = E[\chi^2_{n-1}] = n-1$$

$$\frac{n^2}{\sigma^2} \cdot E[S^2] = n-1 \Rightarrow E[S^2] = \frac{n-1}{n}\sigma^2$$

Como $\frac{nS^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1} \Rightarrow var[\frac{nS^2}{\sigma^2}] = Var[\chi^2_{n-1}]$

$$\Rightarrow \frac{n^2}{\sigma^4} \cdot Var[S^2] = 2(n-1) \Rightarrow Var[S^2] = \frac{2(n-1)\cdot\sigma^4}{n^2}$$

Media y varianza de $T_2 = S^2_{cx}$

Como $nS^2 = (n-1)S^2_c$ entonces $\frac{nS^2}{\sigma^2} = \frac{(n-1)S^2_c}{\sigma^2}$ y este estadístico se distribuye según χ^2_{n-1}, es decir

$\frac{n-1}{\sigma^2}S^2_c \rightsquigarrow \chi^2_{n-1}$. Por tanto:

$$E[\frac{n-1}{\sigma^2} \cdot S^2_c] = E[\chi^2_{n-1}] = n-1$$

$$\frac{n-1}{\sigma^2} \cdot E[S^2_c] = n-1 \Rightarrow E[S^2_c] = \sigma^2$$

$$Var[\frac{n-1}{\sigma^2} S^2_c] = Var[\chi^2_{n-1}] = 2(n-1) \Rightarrow$$

$$\Rightarrow \frac{(n-1)^2}{\sigma^4} \cdot Var[S^2_c] = 2(n-1) \Rightarrow Var[S^2_c] = \frac{2\sigma^4}{n-1}$$

Medio y varianza de $T_3 = \frac{\sum (x_i - 1)^2}{n}$

$x_1, x_2, \dots, x_n$ m.a.s de $X \rightsquigarrow N(1, \sigma)$. Por tanto:

$x_i \rightsquigarrow N(1, \sigma) \quad \forall i = 1 \dots n. \Rightarrow \frac{x_i - 1}{\sigma} \rightsquigarrow N(0,1) \Rightarrow$

$$\Rightarrow \frac{(x_i - 1)^2}{\sigma^2} \rightsquigarrow \chi_1^2 \Rightarrow \sum_{i=1}^{n} \frac{(x_i - 1)^2}{\sigma^2} \rightsquigarrow \chi_n^2$$

$$\frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} (x_i - 1)^2 \rightsquigarrow \chi_n^2$$

El estadístico T_3 se puede expresar como:

$T_3 = \frac{\sigma^2}{n} \cdot \frac{\sum (x_i - 1)^2}{\sigma^2}$. Por tanto:

$$E[T_3] = E\left[\frac{\sigma^2}{n} \cdot \frac{\sum (x_i - 1)^2}{\sigma^2}\right] = E\left[\frac{\sigma^2}{n} \cdot \chi_n^2\right] =$$

$$= \frac{\sigma^2}{n} \cdot E\left[\chi_n^2\right] = \frac{\sigma^2}{n} \cdot n = \sigma^2$$

Para la varianza tenemos que:

$$Var[T_3] = Var\left[\frac{\sigma^2}{n} \cdot \frac{\sum (x_i - 1)^2}{\sigma^2}\right] = Var\left[\frac{\sigma^2}{n} \cdot \chi_n^2\right] =$$

$$= \frac{\sigma^4}{n^2} \cdot Var\left[\chi_n^2\right] = \frac{\sigma^4}{n^2} \cdot 2n = \frac{2\sigma^4}{n}$$