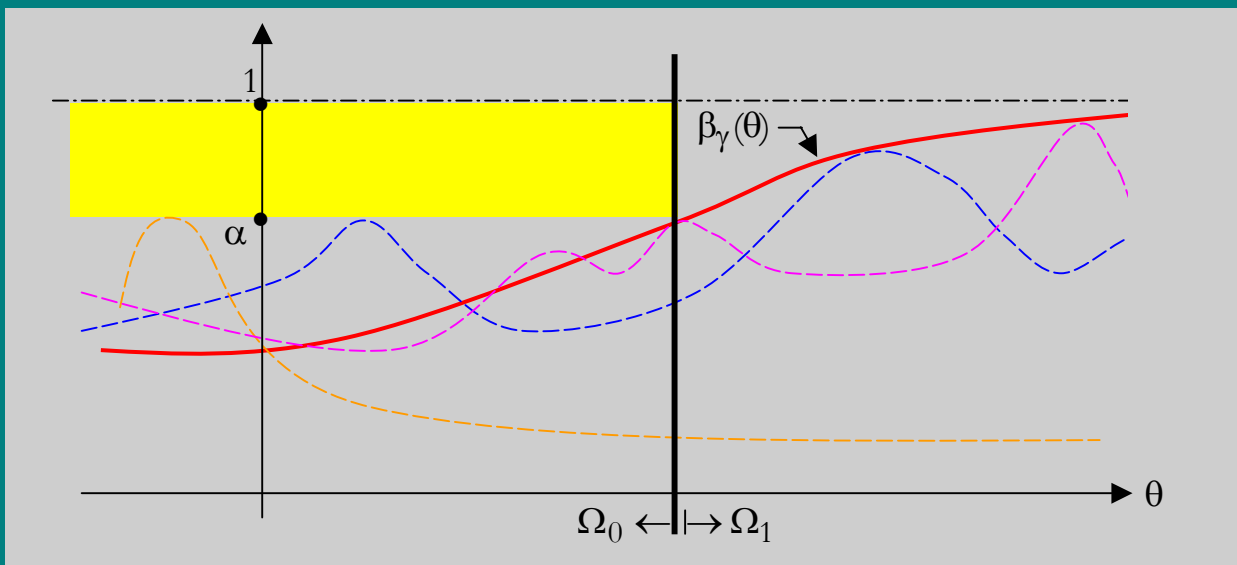


ESTADÍSTICA MATEMÁTICA

Volumen II Inferencia Estadística

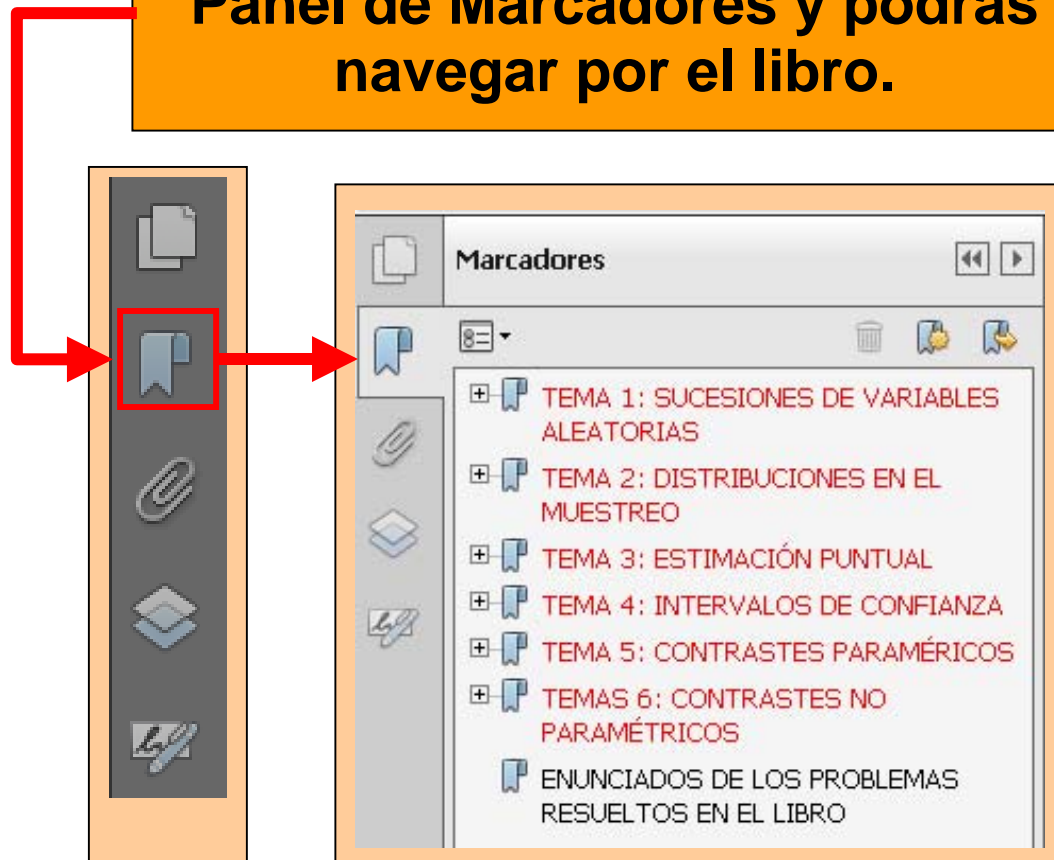


Rafael Cabrejas Hernansanz

→ fonemato.com

Aquí hay un videotutorial en el que
explicamos los contenidos de este libro.

Haciendo clic aquí se abrirá el Panel de Marcadores y podrás navegar por el libro.



INTRODUCCIÓN A LA ESTADÍSTICA MATEMÁTICA

Volumen II: Inferencia Estadística

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros métodos, sin permiso previo y por escrito de los titulares del Copyright.

© RAFAEL CABREJAS HERNANSANZ

Tema 2

Distribuciones en el muestreo

- 2.01 Muestreo aleatorio simple
- 2.02 Estadístico
- 2.03 Estadísticos más famosos
- 2.04 La media muestral
- 2.05 La varianza muestral
- 2.06 Muestreo de población normal

Hay "cosas" respecto de las que debes tener la misma certeza que tienes respecto a tu propio nombre ... y si el mismísimo Papa de Roma te lleva la contraria con una de tales cosas (por ejemplo, te dice que la varianza de una variable aleatoria es -7), debes contestar:

Su Santidad ha tenido un despiste o está mal informado

Y no debes asustarte si el Papa se empeña en llevarte la contraria e insiste en que la varianza es -7 , aunque esté vestido de pontifical, con la mayor educación y respeto, debes añadir:

Su Santidad no tiene ni puñetera idea de lo que dice



Seguridad en ti mism@, en la solvencia de tus conocimientos.

2.1 MUESTREO ALEATORIO SIMPLE

- Siendo "X" una variable aleatoria unidimensional (discreta o continua), se llama **población** de "X" al conjunto de individuos (ya sean personas, animales, plantaciones de patatas, etc.) en los que está definida la variable aleatoria "X".

Se llama **muestra de tamaño "n"** a todo conjunto de "n" observaciones de la variable "X". Se llama **muestreo** al procedimiento mediante el que obtenemos una muestra de la población. Asociada al muestreo de tamaño "n" de la variable "X" hay una variable aleatoria n-dimensional ($X_1; \dots; X_n$) que expresa los resultados que pueden presentarse al seleccionar una muestra de tamaño "n"; de ella diremos que es la **variable muestral**. De las variables unidimensionales X_1, \dots, X_n diremos que son las variables **componentes muestrales**. **Por ejemplo**, si "X" expresa el peso de los ciudadanos y tomamos una muestra tamaño 3, la variable muestral ($X_1; X_2; X_3$) es tridimensional, expresando X_i el aleatorio peso del i-ésimo ciudadano seleccionado ($i = 1, 2, 3$). Así, si la muestra seleccionada es la (60; 76; 90), resulta obvio que las variables X_1 , X_2 y X_3 se han concretado respectivamente en los números 60, 76 y 90.

- Como la selección de las muestras tiene como fin el estudio de las propiedades de la población, conviene que la muestra seleccionada sea representativa de ésta. El **muestreo aleatorio simple** se apoya en los siguientes supuestos:
 - 1) La muestra se obtiene al azar, de modo que todos los individuos de la población tienen la misma probabilidad de ser seleccionados, y las sucesivas elecciones de los "n" individuos que forman la muestra son independientes.
 - 2) La distribución de probabilidad "X" no se altera debido al muestreo.
 - 3) Las variables componentes muestrales X_1, \dots, X_n son independientes y tienen la misma distribución de probabilidad que la variable "X". Por tanto, la función de cuantía/densidad de la variable muestral es el producto de las respectivas funciones de cuantía/densidad de las variables X_1, \dots, X_n .



2.2 ESTADÍSTICO

Llamaremos **estadístico** a cualquier función real de la variable muestral $(X_1; \dots; X_n)$ que a su vez sea una variable aleatoria.

¡Nada de asustarse!

Un estadístico "T" es una variable aleatoria unidimensional corriente y moliente cuyo valor depende de los valores que tomen las variables X_1, \dots, X_n ; por tanto, para un estadístico "T" nos plantearemos las **historias de siempre**: ¿cuáles son las funciones de distribución, generatriz de momentos y de cuantía/densidad de "T"? ¿cuánto valen $E(T)$ y $V(T)$? ¿cuál es la probabilidad del suceso $T = t$? ¿cuál es la probabilidad del suceso $a < T \leq b$?

Llorar a chorros. Llorar la digestión. Llorar el sueño.
Llorar ante las puertas y los puertos.
Llorar de amabilidad y de amarillo.
Abrir las canillas, las compuertas del llanto.
Empaparnos el alma, la camiseta.
Inundar las veredas y los paseos, y salvarnos, a nado,
de nuestro llanto.
Asistir a los cursos de antropología, llorando.
Festejar los cumpleaños familiares, llorando.
Atravesar el África, llorando.
Llorar como un cacuy, como un cocodrilo...
si es verdad que los cacuyes y los cocodrilos
no dejan nunca de llorar.
Llorarlo todo, pero llorarlo bien.
Llorarlo con la nariz, con las rodillas.
Llorarlo por el ombligo, por la boca.
Llorar de amor, de hastío, de alegría.
Llorar de frac, de flato, de flacura.
Llorar improvisando, de memoria.
¡Llorar todo el insomnio
y todo el día!

Oliverio Gironde



Profe de Estadística después de corregir un examen

2.3 ESTADÍSTICOS FAMOSOS

Momentos muestrales respecto al origen

El estadístico llamado **momento muestral de orden "k" respecto al origen** se denota a_k , siendo:

$$a_k = \frac{1}{n} \cdot \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

El más famoso es el de primer orden, que llamaremos **media muestral** y denotaremos \bar{X} :

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

Momentos muestrales centrales

El estadístico llamado **momento muestral central de orden "k"** se denota m_k , siendo:

$$m_k = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 1, 2, \dots$$

El más famoso es el de segundo orden, que llamaremos **varianza muestral** y denotaremos S^2 :

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

Observa: el momento central de primer orden es el estadístico nulo:

$$\begin{aligned} m_1 &= \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^k \overset{\substack{\uparrow \\ \text{si } k=1}}{=} \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X}) = \\ &= \left(\frac{1}{n} \cdot \sum_{i=1}^n X_i \right) - \left(\frac{1}{n} \cdot n \cdot \bar{X} \right) = \bar{X} - \bar{X} = 0 \end{aligned}$$

El estadístico **cuasivarianza muestral** se denota S_c^2 , siendo:

$$S_c^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

Obviamente, es $S_c^2 = \frac{n}{n-1} \cdot S^2$ y $S^2 = \frac{n-1}{n} \cdot S_c^2$.

Recuerda que para toda variable "X" es $V(X) = E(X^2) - (E(X))^2$. Pues bien, esa misma relación hay entre la varianza muestral (S^2), el momento muestral de orden 2 respecto al origen (a_2) y la media muestral (\bar{X}); o sea: $S^2 = a_2 - \bar{X}^2$:

$$\begin{aligned}
 S^2 &= \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2 = S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i^2 - 2 \cdot X_i \cdot \bar{X} + \bar{X}^2) = \\
 &= \frac{1}{n} \cdot \left(\left(\sum_{i=1}^n X_i^2 \right) - \left(2 \cdot \bar{X} \cdot \sum_{i=1}^n X_i \right) + \left(n \cdot \bar{X}^2 \right) \right) = \\
 &\quad \boxed{\sum_{i=1}^n X_i = n \cdot \bar{X}} \quad \uparrow \\
 &= \frac{1}{n} \cdot \left(\left(\sum_{i=1}^n X_i^2 \right) - \left(2 \cdot n \cdot \bar{X}^2 \right) + \left(n \cdot \bar{X}^2 \right) \right) = \frac{1}{n} \cdot \left(\left(\sum_{i=1}^n X_i^2 \right) - \left(n \cdot \bar{X}^2 \right) \right) = \\
 &= \left(\frac{1}{n} \cdot \sum_{i=1}^n X_i^2 \right) - \bar{X}^2 = a_2 - \bar{X}^2
 \end{aligned}$$

Caso bidimensional

Si de una variable aleatoria bidimensional $(X; Y)$ se toma muestra de tamaño "n" (imagina que "X" expresa la aleatoria altura de los ciudadanos e "Y" su aleatorio peso), la variable muestral será $((X_1; Y_1); (X_2; Y_2); \dots; (X_n; Y_n))$; así:

- Las medias muestrales son:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i ; \bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i$$

- Los momentos muestrales de segundo orden son:

$$a_{2;0} = \frac{1}{n} \cdot \sum_{i=1}^n X_i^2 ; a_{0;2} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i^2$$

$$S_1^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2 ; S_2^2 = \frac{1}{n} \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2$$

- Los momentos muestrales orden p;q son:

$$a_{p;q} = \frac{1}{n} \cdot \sum_{i=1}^n X_i^p \cdot Y_i^q ; m_{p;q} = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^p \cdot (Y_i - \bar{Y})^q$$

- La **covarianza muestral**, que se denota S_{11} , es:

$$S_{11} = m_{1;1} = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y}) = \frac{1}{n} \cdot \left(\sum_{i=1}^n X_i \cdot Y_i \right) - \bar{X} \cdot \bar{Y}$$

- El **coeficiente de correlación muestral**, que se denota "r", es

$$r = \frac{S_{11}}{S_1 \cdot S_2}$$

FONEMATO 2.3.1

Se toma m.a.s de tamaño 2 de una población "X" cuya función de cuantía es

x	0	1	2	3
P(X = x)	1/8	3/8	3/8	1/8

Determinése la distribución de probabilidad de los siguientes estadísticos:

- 1) $T_1 = \bar{X}$; 2) $T_2 = X_1 + X_2$; 3) $T_3 = S^2$; 4) $T_4 = S_c^2$
 5) $T_5 = X_1 - X_2$; 6) $T_6 = 2 \cdot X_1 - X_2$; $T_7 = \max.(X_1; X_2)$

SOLUCIÓN

Las variables muestrales X_1 y X_2 son independientes y tienen igual distribución de probabilidad que la población "X":

x_1	0	1	2	3
P($X_1 = x_1$)	1/8	3/8	3/8	1/8

x_2	0	1	2	3
P($X_2 = x_2$)	1/8	3/8	3/8	1/8

La función de cuantía de la variable bidimensional $(X_1; X_2)$ es:

$$P(X_1 = x_1; X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2) \Rightarrow$$

pues X_1 y X_2 son independientes

$$\Rightarrow$$

X_1 / X_2	0	1	2	3
0	1/64	3/64	3/64	1/64
1	3/64	9/64	9/64	3/64
2	3/64	9/64	9/64	3/64
3	1/64	3/64	3/64	1/64

La siguiente tabla recoge toda la información necesaria para determinar la distribución de probabilidad de cada estadístico:

$(x_1; x_2)$	P($X_1 = x_1; X_2 = x_2$)	T_1	T_2	T_3	T_4	T_5	T_6	T_7
(0;0)	1/64	0	0	0	0	0	0	0
(0;1)	3/64	0'5	1	0'25	0'5	-1	-1	1
(0;2)	3/64	1	2	1	2	-2	-2	2
(0;3)	1/64	1'5	3	2'25	4'5	-3	-3	3
(1;0)	3/64	0'5	1	0'25	0'5	1	2	1
(1;1)	9/64	1	2	0	0	0	1	1
(1;2)	9/64	1'5	3	0'25	0'5	-1	0	2
(1;3)	3/64	2	4	1	2	2	-1	3
(2;0)	3/64	1	2	1	2	2	4	2
(2;1)	9/64	1'5	3	0'25	0'5	1	3	2
(2;2)	9/64	2	4	0	0	0	2	2
(2;3)	3/64	2'5	5	0'25	0'5	-1	1	3
(3;0)	1/64	1'5	3	2'25	4'5	3	6	3
(3;1)	3/64	2	4	1	2	2	5	3
(3;2)	3/64	2'5	5	0'25	0'5	1	4	3
(3;3)	1/64	3	6	0	0	0	3	3

Veamos cómo se ha calculado el valor que toma cada estadístico en una muestra concreta; por ejemplo, en la muestra (2;3):

$$1) \quad T_1 = \bar{X} = (X_1 + X_2)/2 \Rightarrow T_1(2;3) = (2 + 3)/2 = 2'5$$

$$2) \quad T_2 = X_1 + X_2 \Rightarrow T_2(2;3) = 2 + 3 = 5$$

$$3) \quad T_3 = S^2 = \frac{1}{2} \cdot \sum_{i=1}^2 (X_i - \bar{X})^2 = \frac{1}{2} \cdot ((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2) \Rightarrow \\ \Rightarrow T_3(2;3) = \frac{1}{2} \cdot ((2 - 2'5)^2 + (3 - 2'5)^2) = 0'25$$

$$4) \quad T_4 = S_c^2 = \frac{1}{2-1} \cdot \sum_{i=1}^2 (X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \Rightarrow \\ \Rightarrow T_4(2;3) = (2 - 2'5)^2 + (3 - 2'5)^2 = 0'5$$

$$5) \quad T_5 = X_1 - X_2 \Rightarrow T_5(2;3) = 2 - 3 = -1$$

$$6) \quad T_6 = 2 \cdot X_1 - X_2 \Rightarrow T_6(2;3) = 2 \cdot 2 - 3 = 1$$

$$7) \quad T_7 = \max.(X_1; X_2) \Rightarrow T_7(2;3) = \max.(2;3) = 3$$

Lo mismo que con el punto (2;3) se hace con los restantes 15 puntos que forman la distribución de masa de probabilidad asociada a la variable muestral $(X_1; X_2)$ que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 2 de la población "X".

1) La distribución de probabilidad del estadístico $T_1 = \bar{X}$ es:

t	0	0'5	1	1'5	2	2'5	3
$P(T_1 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_1 = 2) = 15/64$: la probabilidad del suceso $T_1 = 2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (1;3), (2;2) y (3;1) en que T_1 toma el valor 2:

$$P(T_1 = 2) = P(X_1 = 1; X_2 = 3) + P(X_1 = 2; X_2 = 2) + P(X_1 = 3; X_2 = 1) = \\ = \frac{3}{64} + \frac{9}{64} + \frac{3}{64} = \frac{15}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_1 .

Es:

$$E(T_1) = E\left(\frac{1}{2} \cdot (X_1 + X_2)\right) = \frac{1}{2} \cdot (E(X_1) + E(X_2)) = \frac{1}{2} \cdot (1'5 + 1'5) = 1'5$$

$$E(X_i) = E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1'5, \quad \forall i = 1, 2$$

$$V(T_1) = V\left(\frac{1}{2} \cdot (X_1 + X_2)\right) = \frac{1}{4} \cdot (V(X_1) + V(X_2)) = \frac{1}{4} \cdot (0'75 + 0'75) = \frac{3}{8}$$

$$V(X_i) = V(X) = E(X^2) - (E(X))^2 = 3 - 1'5^2 = 0'75, \quad \forall i = 1, 2$$

$$E(X^2) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3$$

VENTANA

2) La distribución de probabilidad del estadístico $T_2 = X_1 + X_2$ es:

t	0	1	2	3	4	5	6
$P(T_2 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_2 = 5) = 6/64$: la probabilidad del suceso $T_2 = 5$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (2;3) y (3;2) en que T_2 toma el valor 5:

$$P(T_2 = 5) = P(X_1 = 2; X_2 = 3) + P(X_1 = 3; X_2 = 2) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_2 .

Es:

$$\begin{aligned} E(T_2) &= E(2 \cdot T_1) = 2 \cdot E(T_1) = 2 \cdot 1.5 = 3 \\ V(T_2) &= V(2 \cdot T_1) = 4 \cdot V(T_1) = 4 \cdot \frac{3}{8} = 1.5 \end{aligned}$$

$T_2 = 2 \cdot T_1$

3) La distribución de probabilidad del estadístico $T_3 = S^2$ es:

t	0	0.25	1	2.25
$P(T_3 = t)$	20/64	30/64	12/64	2/64

Veamos cómo hemos obtenido que $P(T_3 = 1) = 12/64$: la probabilidad del suceso $T_3 = 1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;2), (1;3), (2;0) y (3;1) en que T_3 toma el valor 1:

$$\begin{aligned} P(T_3 = 1) &= P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) + \\ &+ P(X_1 = 2; X_2 = 0) + P(X_1 = 3; X_2 = 1) = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{12}{64} \end{aligned}$$

De forma similar se hace para los restantes valores que puede tomar T_3 .

Es:

$$\begin{aligned} E(T_3) &= 0 \cdot \frac{20}{64} + 0.25 \cdot \frac{30}{64} + 1 \cdot \frac{12}{64} + 2.25 \cdot \frac{2}{64} = \frac{24}{64} \\ V(T_3) &= E(T_3^2) - (E(T_3))^2 = \frac{24}{64} - \left(\frac{24}{64}\right)^2 = \frac{15}{64} \end{aligned}$$

$$E(T_3^2) = 0^2 \cdot \frac{20}{64} + 0.25^2 \cdot \frac{30}{64} + 1^2 \cdot \frac{12}{64} + 2.25^2 \cdot \frac{2}{64} = \frac{24}{64}$$

En fonemato.com tienes el videotutorial en el que explicamos los contenidos de este libro.

4) La distribución de probabilidad del estadístico $T_4 = S_c^2$ es:

t	0	0'5	2	4'5
$P(T_4 = t)$	20/64	30/64	12/64	2/64

Veamos cómo hemos obtenido que $P(T_4 = 2) = 12/64$: la probabilidad del suceso $T_4 = 2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;2), (1;3), (2;0) y (3;1) en que T_4 toma el valor 2:

$$P(T_4 = 2) = P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) + P(X_1 = 2; X_2 = 0) + P(X_1 = 3; X_2 = 1) = \frac{3}{64} + \frac{3}{64} + \frac{3}{64} + \frac{3}{64} = \frac{12}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_4 .

Para una muestra aleatoria simple de tamaño "n" es $S_c^2 = n \cdot S^2 / (n - 1)$; así, si $n = 2$, es $T_4 = S_c^2 = 2 \cdot S^2 = 2 \cdot T_3$. En consecuencia:

$$E(T_4) = E(2 \cdot T_3) = 2 \cdot E(T_3) = 2 \cdot (24/64) = 48/64$$

$$V(T_4) = V(2 \cdot T_3) = 4 \cdot V(T_3) = 4 \cdot (15/64) = 60/64$$

5) La distribución de probabilidad del estadístico $T_5 = X_1 - X_2$ es:

t	-3	-2	-1	0	1	2	3
$P(T_5 = t)$	1/64	6/64	15/64	20/64	15/64	6/64	1/64

Veamos cómo hemos obtenido que $P(T_5 = -2) = 6/64$: la probabilidad del suceso $T_5 = -2$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;2) y (1;3) en que T_5 toma el valor -2:

$$P(T_5 = -2) = P(X_1 = 0; X_2 = 2) + P(X_1 = 1; X_2 = 3) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_5 .

Como $E(X_1) = E(X_2)$, es $E(T_5) = E(X_1 - X_2) = E(X_1) - E(X_2) = 0$.

Como $V(X_1) = V(X_2) = V(X) = 0'75$, es:

$$V(T_5) = V(X_1 - X_2) = V(X_1) + V(X_2) = 1'5$$

QUE QUEDE MUY, MUY CLARO

Siendo discreta la población "X", con muestra de tamaño 2, para calcular la función de cuantía del estadístico "Pepe" calculamos el valor que toma "Pepe" en cada punto de la distribución bidimensional discreta asociada a la variable muestral $(X_1; X_2)$ y después obramos en consecuencia.



6) La distribución de probabilidad del estadístico $T_6 = 2 \cdot X_1 - X_2$ es:

t	-3	-2	-1	0	1	2	3	4	5	6
$P(T_6 = t)$	1/64	3/64	6/64	10/64	12/64	12/64	10/64	6/64	3/64	1/64

Veamos cómo hemos obtenido que $P(T_6 = -1) = 6/64$: la probabilidad del suceso $T_6 = -1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;1) y (1;3) en que T_6 toma el valor -1:

$$P(T_6 = -1) = P(X_1 = 0; X_2 = 1) + P(X_1 = 1; X_2 = 3) = \frac{3}{64} + \frac{3}{64} = \frac{6}{64}$$

De forma similar se hace para los restantes valores que puede tomar T_6 .

Es:

$$E(T_6) = E(2 \cdot X_1 - X_2) = 2 \cdot E(X_1) - E(X_2) = 2 \cdot 1'5 - 1'5 = 1'5$$

$$E(X_1) = E(X_2) = 1'5$$

$$V(T_6) = V(2 \cdot X_1 - X_2) = 4 \cdot V(X_1) + V(X_2) = 3'75$$

$$V(X_i) = V(X) = 0'75, \forall i = 1, 2$$

7) La distribución de probabilidad del estadístico $T_7 = \max.(X_1; X_2)$ es:

t	0	1	2	3
$P(T_7 = t)$	1/64	15/64	33/64	15/64

Veamos cómo hemos obtenido que $P(T_7 = 1) = 15/64$: la probabilidad del suceso $T_7 = 1$ la obtenemos sumando las masas de probabilidad correspondientes a las muestras (0;1), (1;0) y (1;1) en que T_7 toma el valor 1:

$$P(T_7 = 1) = P(X_1 = 0; X_2 = 1) + P(X_1 = 1; X_2 = 0) + P(X_1 = 1; X_2 = 1) = \frac{3}{64} + \frac{3}{64} + \frac{9}{64} = \frac{15}{64}$$

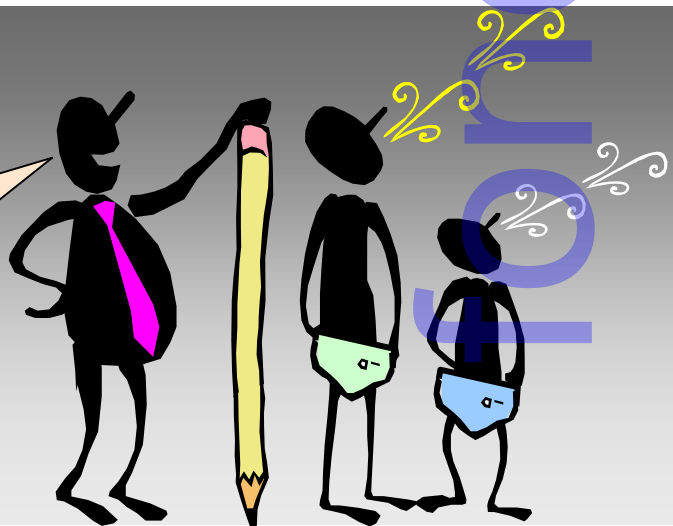
De forma similar se hace para los restantes valores que puede tomar T_7 . Es:

$$E(T_7) = 0 \cdot \frac{1}{64} + 1 \cdot \frac{15}{64} + 2 \cdot \frac{33}{64} + 3 \cdot \frac{15}{64} = \frac{126}{64}$$

$$V(T_7) = E(T_7^2) - (E(T_7))^2 = \frac{282}{64} - \left(\frac{126}{64}\right)^2$$

$$E(T_7^2) = 0^2 \cdot \frac{1}{64} + 1^2 \cdot \frac{15}{64} + 2^2 \cdot \frac{33}{64} + 3^2 \cdot \frac{15}{64} = \frac{282}{64}$$

¿Quién está todavía en pañales en lo que se refiere al uso de esas **mágicas ventanas** que me facilitan enormemente el desagradable trabajo de corregir exámenes?



FONEMATO 2.3.2

Sea "X" una población continua de la que se toma m.a.s de tamaño "n".

- 1) Determinése la distribución del mínimo valor muestral "Z".
- 2) Determinése la distribución del máximo valor muestral "Y".
- 3) Determinése la distribución de recorrido muestral $R = Y - Z$.
- 4) Aplíquese al caso $n = 3$ si la densidad de "X" es $f(x) = 2 \cdot x$, $x \in [0;1]$.
- 5) En las condiciones de 4), determinése las siguientes probabilidades

$$P(Y \leq 0'1) ; P(Y > 0'9) ; P(Z \leq 0'1) ; P(Z > 0'9) ; P(R \leq 0'5)$$

SOLUCIÓN

Siendo "f" y "F" las respectivas funciones de densidad y de distribución de "X", sea $(X_1; \dots; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño "n"; por tanto, X_1, \dots, X_n son independientes y tienen igual distribución de probabilidad que la población "X".

- 1) Siendo $Z = \min.(X_1; \dots; X_n)$, su función de distribución F_Z es:

$$F_Z(z) = P(Z \leq z) = P(\min.(X_1; \dots; X_n) \leq z) = 1 - P(\min.(X_1; \dots; X_n) > z) = 1 - P(X_1 > z; \dots; X_n > z)$$

pues X_1, \dots, X_n son independientes

$$= 1 - \prod_{i=1}^n P(X_i > z) = 1 - (1 - F(z))^n$$

$$P(X_i > z) = P(X > z) = 1 - F(z), \forall i = 1, 2, \dots, n$$

- Siendo $f_Z(z)$ la densidad de probabilidad de "Z", es:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = n \cdot (1 - F(z))^{n-1} \cdot \frac{dF(z)}{dz} = n \cdot (1 - F(z))^{n-1} \cdot f(z)$$

$$F_Z(z) = 1 - (1 - F(z))^n$$

$$dF(z)/dz = f(z)$$

- También podemos calcular la densidad $f_Z(z)$ de "Z" así:

$$P(z < Z < z + dz) = f_Z(z) \cdot dz = \underbrace{n}_{\text{a}} \cdot \underbrace{f(z)}_{\text{b}} \cdot \underbrace{dz}_{\text{c}} \cdot \left(\int_{z+dz}^{+\infty} f(x) \cdot dx \right)^{n-1} \Rightarrow$$

VENTANA

- a** El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales.
- b** Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z; z + dz)$.
- c** Probabilidad de las restantes $n - 1$ variables componentes muestrales tomen valores mayores que $z + dz$.

$$\Rightarrow f_Z(z) = n \cdot f(z) \cdot \left(\int_z^{+\infty} f(x) \cdot dx \right)^{n-1} = n \cdot (1 - F(z))^{n-1} \cdot f(z)$$

$$\int_z^{+\infty} f(x) \cdot dx = 1 - F(z)$$

dividimos los dos miembros por "dz", y consideramos que, por ser "dz" un infinitésimo, es $\int_{z+dz}^{+\infty} f(x) \cdot dx = \int_z^{+\infty} f(x) \cdot dx$

2) Siendo $Y = \max.(X_1; \dots; X_n)$, su función de distribución F_Y es:

$$F_Y(y) = P(Y \leq y) = P(\max.(X_1; \dots; X_n) \leq y) =$$

pues X_1, \dots, X_n son independientes

$$= P(X_1 \leq y; \dots; X_n \leq y) = \prod_{i=1}^n P(X_i \leq y) = (F(y))^n$$

$$P(X_i \leq y) = P(X \leq y) = F(y), \forall i = 1, 2, \dots, n$$

• Siendo $f_Y(y)$ la densidad de probabilidad de "Y", es:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = n \cdot (F(y))^{n-1} \cdot \frac{dF(y)}{dy} = n \cdot (F(y))^{n-1} \cdot f(y)$$

$$F_Y(y) = (F(y))^n$$

$$dF(y)/dy = f(y)$$

• También podemos calcular la densidad $f_Y(y)$ de "Y" así:

$$P(y < Y < y + dy) = f_Y(y) \cdot dy = n \cdot f(y) \cdot dy \cdot \left(\int_{-\infty}^y f(x) \cdot dx \right)^{n-1} \Rightarrow$$

a

b

c

VENTANA

- a El "papel" de "máximo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales.
- b Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(y; y + dy)$.
- c Probabilidad de las restantes $n - 1$ variables componentes muestrales tomen valores menores que "y".

$$\Rightarrow f_Y(y) = n \cdot f(y) \cdot \left(\int_{-\infty}^y f(x) \cdot dx \right)^{n-1} = n \cdot (F(y))^{n-1} \cdot f(y)$$

$$\text{es } \int_{-\infty}^y f(x) \cdot dx = F(y)$$

Dividimos los dos miembros por "dy"

- 3) Para determinar la distribución de probabilidad de $R = Y - Z$ debemos calcular previamente la densidad $g(z; y)$ de la variable bidimensional $(Z; Y)$; es:

$$P(z < Z < z + dz; y < Y < y + dy) = g(z; y) \cdot dz \cdot dy =$$

$$= \underbrace{n \cdot (n-1)}_a \cdot \underbrace{f(z) \cdot dz}_b \cdot \underbrace{f(y) \cdot dy}_c \cdot \underbrace{\left(\int_z^y f(x) \cdot dx \right)^{n-2}}_d \Rightarrow$$

VENTANA

- a** El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las "n" variables componentes muestrales, y el "papel" de "máximo valor muestral" puede hacerlo cualquiera de las restantes $n - 1$ variables componentes muestrales; por tanto, el número de pares distintos (mínimo; máximo) en que pueden concretarse las "n" observaciones muestrales es $n \cdot (n - 1)$.
- b** Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z; z + dz)$.
- c** Probabilidad de que otra de las variables componentes muestrales tome un valor en el intervalo $(y; y + dy)$.
- d** Probabilidad de las restantes $n - 2$ variables componentes muestrales tomen valores en el intervalo $(z + dz; y + dy)$. Como "dz" y "dy" son infinitésimos, despreciamos su influencia y consideramos el intervalo $(z; y)$.

$$\Rightarrow g(z; y) = n \cdot (n - 1) \cdot f(z) \cdot f(y) \cdot \left(\int_z^y f(x) \cdot dx \right)^{n-2}$$

dividimos los dos miembros por "dz.dy"

- Haciendo el cambio de variables

$$\begin{cases} S = Z \\ R = Y - Z \end{cases} \Rightarrow \begin{cases} Z = S \\ Y = S + R \end{cases}$$

el jacobiano es:

$$J = \begin{vmatrix} \partial z / \partial s & \partial z / \partial r \\ \partial y / \partial s & \partial y / \partial r \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

La densidad de probabilidad $h(s; r)$ de la variable $(S; R)$ es:

$$h(s; r) = g(s; s + r) \cdot |J| = n \cdot (n - 1) \cdot f(s) \cdot f(s + r) \cdot \left(\int_s^{s+r} f(x) \cdot dx \right)^{n-2}$$

La densidad de probabilidad $f_R(r)$ del recorrido muestral "R" es:

$$f_R(r) = \int_{s=-\infty}^{s=+\infty} h(s; r) \cdot ds$$

4) Si la densidad de "X" es $f(x) = 2 \cdot x$, $x \in [0;1]$, entonces:

$$F(x) = P(X \leq x) = \int_0^x 2 \cdot x \cdot dx = x^2$$

- Para $n = 3$, siendo $Z = \min.(X_1; X_2; X_3)$ e $Y = \max.(X_1; X_2; X_3)$, sus respectivas funciones de densidad de probabilidad son:

$$f_Z(z) = 3 \cdot (1 - F(z))^{3-1} \cdot f(z) = 3 \cdot (1 - z^2)^2 \cdot (2 \cdot z), z \in [0;1]$$

$$f_Y(y) = 3 \cdot (F(y))^{3-1} \cdot f(y) = 3 \cdot (y^2)^2 \cdot (2 \cdot y), y \in [0;1]$$

- La densidad de probabilidad $g(z; y)$ de la variable $(Z; Y)$ es:

$$g(z; y) = 3 \cdot (3 - 1) \cdot f(z) \cdot f(y) \cdot \left(\int_z^y f(x) \cdot dx \right)^{3-2} =$$

$$= 6 \cdot (2 \cdot z) \cdot (2 \cdot y) \cdot \int_z^y 2 \cdot x \cdot dx = 24 \cdot z \cdot y \cdot (y^2 - z^2), \text{ siendo } \begin{cases} 0 \leq z \leq 1 \\ z \leq y \leq 1 \end{cases}$$

- Haciendo el cambio de variables $\begin{cases} S = Z \\ R = Y - Z \end{cases} \Rightarrow \begin{cases} Z = S \\ Y = S + R \end{cases}$, cuyo jacobiano es 1, la densidad de probabilidad $h(s; r)$ de la variable $(S; R)$ es:

$$h(s, r) = g(s; s + r) \cdot |J| = 24 \cdot s \cdot (s + r) \cdot ((s + r)^2 - s^2) =$$

en $g(z; y)$ sustituimos "z" por "s" e "y" por "s+r"

$$= 24 \cdot s \cdot r \cdot (s + r) \cdot (2 \cdot s + r), \text{ siendo } \begin{cases} 0 \leq s \leq 1 \\ 0 \leq r \leq 1 - s \end{cases}$$

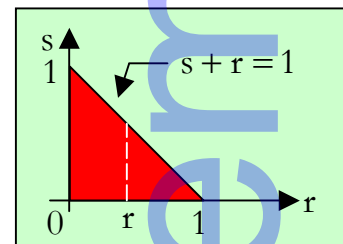
$$\text{si } \begin{cases} 0 \leq z \leq 1 \\ z \leq y \leq 1 \end{cases} \text{ y } \begin{cases} z = s \\ y = s + r \end{cases} \Rightarrow \begin{cases} 0 \leq s \leq 1 \\ s \leq s + r \leq 1 \end{cases} \Rightarrow \begin{cases} 0 \leq s \leq 1 \\ 0 \leq r \leq 1 - s \end{cases}$$

- Densidad de probabilidad $f_R(r)$ del recorrido "R":

$$f_R(r) = \int_{s=-\infty}^{s=+\infty} h(s; r) \cdot ds =$$

$$= \int_{s=0}^{s=1-r} 24 \cdot s \cdot r \cdot (s + r) \cdot (2 \cdot s + r) \cdot ds = \dots =$$

$$= 12 \cdot r \cdot (1 - r^2), r \in [0;1]$$



5) Siendo $F_Z(z) = 1 - (1 - F(z))^n$ y $F_Y(y) = (F(y))^n$, si $F(x) = x^2$ y $n = 3$, es:

$$F_Z(z) = 1 - (1 - z^2)^3 ; F_Y(y) = y^6$$

Por tanto:

$$* P(Y \leq 0'1) = F_Y(0'1) = 0'1^6$$

$$* P(Y > 0'9) = 1 - P(Y \leq 0'9) = 1 - F_Y(0'9) = 1 - 0'9^6$$

$$* P(Z \leq 0'1) = F_Z(0'1) = 1 - (1 - 0'1^2)^3$$

$$* P(Z > 0'9) = 1 - P(Z \leq 0'9) = 1 - F_Z(0'9) = 1 - (1 - (1 - 0'9^2)^3)$$

$$* P(R \leq 0'5) = \int_0^{0'5} f_R(r) \cdot dr = \int_0^{0'5} 12 \cdot r \cdot (1 - r^2) \cdot dr = \dots$$

FONEMATO 2.3.3

- 1) En el muestreo aleatorio simple de tamaño 5 de una población exponencial de parámetro "a", determínese la distribución del recorrido muestral.
- 2) Calcúlese la probabilidad de que el recorrido muestral sea mayor que 1.

SOLUCIÓN

- 1) Sea $X \approx \text{Exp.}(a)$ y $(X_1; \dots; X_5)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 5; por tanto, X_1, \dots, X_5 son independientes y tienen distribución $\text{Exp.}(a)$, como "X".

Si $Z = \min.(X_1; \dots; X_5)$ e $Y = \max.(X_1; \dots; X_5)$, el recorrido muestral "R" es $R = Y - Z$, y para determinar la distribución de probabilidad de "R" debemos calcular antes la densidad $g(z; y)$ de la variable bidimensional $(Z; Y)$.

Siendo "f" la función de densidad de probabilidad de la población "X", es:

$$\begin{aligned} P(z < Z < z + dz; y < Y < y + dy) &= g(z; y) \cdot dz \cdot dy = \\ &= \underbrace{5 \cdot (5 - 1)}_a \cdot \underbrace{f(z) \cdot dz}_b \cdot \underbrace{f(y) \cdot dy}_c \cdot \underbrace{\left(\int_z^y f(x) \cdot dx \right)^{5-2}}_d \Rightarrow \end{aligned}$$

a El "papel" de "mínimo valor muestral" puede hacerlo cualquiera de las 5 variables componentes muestrales, y el "papel" de "máximo valor muestral" puede hacerlo cualquiera de las restantes $5 - 1$ variables componentes muestrales; por tanto, el número de pares distintos (mínimo; máximo) en que pueden concretarse las "5" observaciones muestrales es $5 \cdot (5 - 1)$.

b Probabilidad de que una de las variables componentes muestrales tome un valor en el intervalo $(z; z + dz)$.

c Probabilidad de que otra de las variables componentes muestrales tome un valor en el intervalo $(y; y + dy)$.

d Probabilidad de que las restantes $5 - 2$ variables componentes muestrales tomen valores en el intervalo $(z + dz; y + dy)$. Como "dz" y "dy" son infinitésimos, despreciamos su influencia y consideramos el intervalo $(z; y)$.

VENTANA

dividimos los dos miembros por "dz.dy"

$$\Rightarrow g(z; y) = 20 \cdot f(z) \cdot f(y) \cdot \left(\int_z^y f(x) \cdot dx \right)^3 \Rightarrow$$

siendo $X \approx \text{Exp.}(a)$, es $f(x) = a \cdot e^{-a \cdot x}$, $x > 0$

$$\begin{aligned} \Rightarrow g(z; y) &= 20 \cdot (a \cdot e^{-a \cdot z}) \cdot (a \cdot e^{-a \cdot y}) \cdot \left(\int_z^y a \cdot e^{-a \cdot x} \cdot dx \right)^3 = \\ &= 20 \cdot (a \cdot e^{-a \cdot z}) \cdot (a \cdot e^{-a \cdot y}) \cdot \left(- (e^{-a \cdot x}) \right)_z^y^3 = \\ &= 20 \cdot a^2 \cdot e^{-a \cdot (z+y)} \cdot (e^{-a \cdot z} - e^{-a \cdot y})^3, \text{ siendo } 0 < z \leq y \end{aligned}$$

- Haciendo el cambio de variables $\begin{cases} S = Z \\ R = Y - Z \end{cases} \Rightarrow \begin{cases} Z = S \\ Y = S + R \end{cases}$, el jacobiano es:

$$J = \begin{vmatrix} \partial z / \partial s & \partial z / \partial r \\ \partial y / \partial s & \partial y / \partial r \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

La densidad de probabilidad $h(s;r)$ de la variable $(S;R)$ es:

$$h(s;r) = g(s; s+r) \cdot |J| = 20 \cdot a^2 \cdot e^{-a \cdot (s+s+r)} \cdot (e^{-a \cdot s} - e^{-a \cdot (s+r)})^3 =$$

en $g(z;y)$ sustituimos "z" por "s" e "y" por "s+r"

$$= 20 \cdot a^2 \cdot e^{-a \cdot r} \cdot e^{-5 \cdot a \cdot s} \cdot (1 - e^{-a \cdot r})^3, \text{ siendo } s > 0 \text{ y } r \geq 0$$

$$\text{si } \begin{cases} z > 0 \\ z \leq y \end{cases} \text{ y } \begin{cases} z = s \\ y = s + r \end{cases} \Rightarrow \begin{cases} s > 0 \\ s \leq s + r \end{cases} \Rightarrow \begin{cases} s > 0 \\ r \geq 0 \end{cases}$$

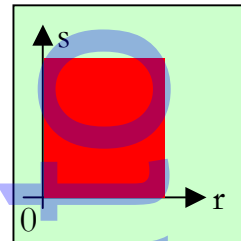
- La densidad de probabilidad $f_R(r)$ del recorrido muestral "R" es:

$$f_R(r) = \int_{s=-\infty}^{s=+\infty} h(s;r) \cdot ds = \int_{s=0}^{s=+\infty} h(s;r) \cdot ds =$$

$$= 20 \cdot a^2 \cdot e^{-a \cdot r} \cdot (1 - e^{-a \cdot r})^3 \cdot \int_{s=0}^{s=+\infty} e^{-5 \cdot a \cdot s} \cdot ds =$$

$$\int_{s=0}^{s=+\infty} e^{-5 \cdot a \cdot s} \cdot ds = \frac{1}{5 \cdot a} \cdot \int_{s=0}^{s=+\infty} e^{-5 \cdot a \cdot s} \cdot d(5 \cdot a \cdot s) = \frac{\Gamma(1)}{5 \cdot a} = \frac{1}{5 \cdot a}$$

$$= 4 \cdot a \cdot e^{-a \cdot r} \cdot (1 - e^{-a \cdot r})^3; r \geq 0$$



2) Es: $P(R > 1) = \int_1^{+\infty} f_R(r) \cdot dr = \int_1^{+\infty} 4 \cdot a \cdot e^{-a \cdot r} \cdot (1 - e^{-a \cdot r})^3 \cdot dr =$

AVISO $= \left((1 - e^{-a \cdot r})^4 \right)_1^{+\infty} = 1 - (1 - e^{-a})^4$

Si de una población continua "X" se toma muestra aleatoria simple de tamaño 9 (impar) y se obtiene la (13;16;9;7;18;6;8;5;3), **ordenando de menor a mayor los valores observados**, resulta $3 < 5 < 6 < 7 < 8 < 9 < 13 < 16 < 18$. Así, en esa muestra el **estadístico mediana muestral** se concreta en el número 8, que es la **observación inferior a la mitad de las restantes observaciones y superior a la otra mitad**.

¡EN EL SIGUIENTE EJERCICIO JUGAREMOS CON EL ESTADÍSTICO MEDIANA MUESTRAL!

Me temo que la mediana me va a doler mucho



¡Ánimo!...tras la **mediana muestral** como estrella, en el ejercicio 2.3.5 viene el **estadístico diferencia entre el sexto menor valor muestral y el tercer menor valor muestral**, que te cargará las pilas

FONEMATO 2.3.4

Sea "X" una población continua de la que se toma una muestra aleatoria simple de tamaño impar $2.n + 1$.

- 1) Determinése la distribución de la mediana muestral "M".
- 2) Particularícese en el caso $n = 2$ si la densidad de "X" es $f(x) = 2.x$, $x \in [0;1]$.
- 3) En las condiciones de 2), calcúlense la media y la varianza de "M".

SOLUCIÓN

- 1) Siendo f_M la función de densidad de probabilidad de "M", es:

$$f_M(m).dm = P(m < M < m + dm) = \\ = \frac{(2.n + 1)!}{n!.1!.n!} \cdot \left(\int_{-\infty}^m f(x).dx \right)^n \cdot (f(m).dm) \cdot \left(\int_{m+dm}^{+\infty} f(x).dx \right)^n \Rightarrow$$

En el experimento consistente en observar **una vez** el valor que toma la variable "X", por fuerza ha de ocurrir alguno de los siguientes sucesos:

- * $A_1 \equiv$ la variable "X" toma un valor no superior a "m"
- * $A_2 \equiv$ la variable "X" toma un valor en el intervalo $(m; m + dm)$
- * $A_3 \equiv$ la variable "X" toma un valor no inferior a "m + dm"

Siendo "f" la función de densidad de "X", es:

VENTANA

$$P(A_1) = \int_{-\infty}^m f(x).dx ; P(A_2) = f(m).dm ; P(A_3) = \int_{m+dm}^{+\infty} f(x).dx$$

Si el experimento se repite $2.n + 1$ veces, la aleatoria mediana muestral "M" toma un valor en el intervalo $(m; m + dm)$ sólo si el suceso A_1 ocurre "n" veces, el suceso A_2 ocurre una vez y el suceso A_3 ocurre "n" veces; y la probabilidad de tal contingencia es:

$$PR_{n;1;n}^{2n+1} \cdot (P(A_1))^n \cdot P(A_2) \cdot (P(A_3))^n$$

La probabilidad de que ocurra

$$\underbrace{A_1, A_1, \dots, A_1}_{\text{"n" veces}}, \underbrace{A_2}_{\text{una vez}}, \underbrace{A_3, A_3, \dots, A_3}_{\text{"n" veces}}$$

es $(P(A_1))^n \cdot P(A_2) \cdot (P(A_3))^n$, y debe multiplicarse por $PR_{n;1;n}^{2n+1}$, que es el número de **permutaciones con repetición** (alineaciones) que pueden formarse con A_1, A_2 y A_3 de modo que A_1 y A_3 estén presentes "n" veces y A_2 una vez, siendo:

$$PR_{n;1;n}^{2n+1} = \frac{(2.n + 1)!}{n!.1!.n!}$$

$$\Rightarrow f_M(m) = \frac{(2.n + 1)!}{n!.1!.n!} \cdot \left(\int_{-\infty}^m f(x).dx \right)^n \cdot f(m) \cdot \left(\int_m^{+\infty} f(x).dx \right)^n$$

dividimos por "dm" y pasamos del "dm" de la última integral

2) Si $n = 2$ y la densidad de "X" es $f(x) = 2.x$, $x \in [0;1]$, la densidad de "M" es:

$$f_M(m) = \frac{5!}{2!.1!.2!} \cdot \left(\int_{-\infty}^m f(x).dx \right)^2 \cdot f(m) \cdot \left(\int_m^{+\infty} f(x).dx \right)^2 =$$

$$= 30 \cdot \left(\int_0^m 2.x.dx \right)^2 \cdot (2.m) \cdot \left(\int_m^1 2.x.dx \right)^2 = \dots = 60.m^5.(1-m^2)^2, \quad m \in [0;1]$$

$f(x) = 2.x, \quad x \in [0;1]$

2) Es:

- $$E(M) = \int_{-\infty}^{+\infty} m.f_m(m).dm = \int_0^1 60.m^6.(1-m^2)^2.dm =$$

$m = \sqrt{u} \Rightarrow dm = du/(2.\sqrt{u})$
los límites de integración no cambian

$$= 60 \cdot \int_0^1 u^3.(1-u)^2 \cdot \frac{du}{2.\sqrt{u}} = 30 \cdot \int_0^1 u^{5/2}.(1-u)^2.du = 30.\beta(7/2;3) =$$

$\beta(r;s) = \int_0^1 u^{r-1}.(1-u)^{s-1}.du$
En nuestro caso es $\begin{cases} r-1=5/2 \\ s-1=3 \end{cases} \Rightarrow \begin{cases} r=7/2 \\ s=4 \end{cases}$

$$= 30 \cdot \frac{\Gamma(7/2).\Gamma(3)}{\Gamma(13/2)} = \frac{30.\Gamma(7/2).\Gamma(3)}{(11/2).(9/2).(7/2).\Gamma(7/2)} = \frac{30.\Gamma(3)}{(11/2).(9/2).(7/2)} = \frac{160}{231}$$

- $$V(M) = E(M^2) - (E(M))^2 = \frac{1}{2} - \left(\frac{160}{231}\right)^2$$

VENTANA

$$E(M^2) = \int_{-\infty}^{+\infty} m^2.f_m(m).dm = \int_0^1 60.m^7.(1-m^2)^2.dm =$$

$m = \sqrt{u} \Rightarrow dm = du/(2.\sqrt{u})$
los límites de integración no cambian

$$= 60 \cdot \int_0^1 u^{7/2}.(1-u)^2 \cdot \frac{du}{2.\sqrt{u}} = 30 \cdot \int_0^1 u^3.(1-u)^2.du =$$

$$= 30.\beta(4;3) = 30 \cdot \frac{\Gamma(4).\Gamma(3)}{\Gamma(7)} = 30 \cdot \frac{3!.2!}{6!} = \frac{1}{2}$$



FONEMATO 2.3.5

Si se toma m.a.s de tamaño 9 de una población "X" con distribución uniforme en el intervalo (0;1), determínese la distribución de probabilidad de la diferencia entre el sexto menor valor muestral y el tercer menor valor muestral.

SOLUCIÓN

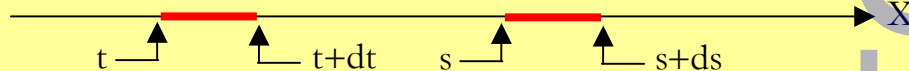
Siendo "S" el sexto menor valor muestral y "T" el tercer menor valor muestral, sea $Z = S - T$ su diferencia. Para hallar la distribución de probabilidad de $Z = S - T$ debemos calcular antes la función de densidad $g(t;s)$ de $(T;S)$; es:

$$P(t < T < t + dt; s < S < s + ds) = g(t;s) \cdot dt \cdot ds =$$

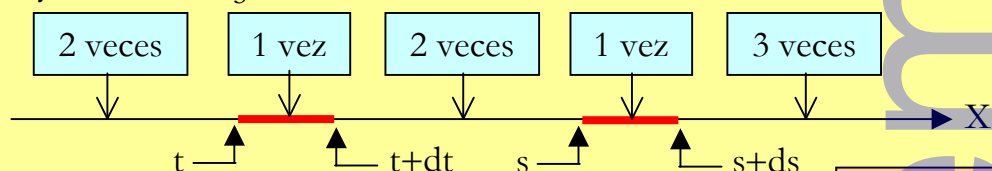
$$= \frac{9!}{2! \cdot 1! \cdot 2! \cdot 1! \cdot 3!} \cdot (P(A_1))^2 \cdot P(A_2) \cdot (P(A_3))^2 \cdot P(A_4) \cdot (P(A_5))^3 =$$

En el experimento consistente en observar una vez el valor que toma la variable "X", por fuerza ha de ocurrir alguno de los siguientes sucesos:

- * $A_1 \equiv$ la variable "X" toma un valor no superior a "t"
- * $A_2 \equiv$ la variable "X" toma un valor en el intervalo $(t; t + dt)$
- * $A_3 \equiv$ la variable "X" toma un valor en el intervalo $[t + dt; s]$
- * $A_4 \equiv$ la variable "X" toma un valor en el intervalo $(s; s + ds)$
- * $A_5 \equiv$ la variable "X" toma un valor no inferior a "s + ds"



Si el experimento se repite 9 veces, el tercer menor valor muestral "T" toma un valor en el intervalo $(t; t + dt)$ y el sexto menor valor muestral toma un valor en el intervalo $(s; s + ds)$ sólo si el suceso A_1 ocurre 2 veces, el suceso A_2 ocurre 1 vez, el suceso A_3 ocurre 2 veces, el suceso A_4 ocurre 1 vez y el suceso A_5 ocurre 3 veces.



La probabilidad de tal contingencia es:

$$PR_{2;1;2;1;3}^9 \cdot (P(A_1))^2 \cdot P(A_2) \cdot (P(A_3))^2 \cdot P(A_4) \cdot (P(A_5))^3$$

VENTANA

La probab. de que ocurra $\underbrace{A_1, A_1}_{2 \text{ veces}}, \underbrace{A_2}_{\text{una vez}}, \underbrace{A_3, A_3}_{2 \text{ veces}}, \underbrace{A_4}_{\text{una vez}}, \underbrace{A_5, A_5, A_5}_{3 \text{ veces}}$ es

$(P(A_1))^2 \cdot P(A_2) \cdot (P(A_3))^2 \cdot P(A_4) \cdot (P(A_5))^3$ y debe multiplicarse por $PR_{2;1;2;1;3}^9$, que es el número de **permutaciones con repetición** que pueden formarse con A_1, A_2, A_3, A_4 y A_5 de modo que A_1 y A_3 estén presentes 2 veces, A_2 y A_4 estén presentes 1 vez y A_5 esté presente 3 veces, siendo $PR_{2;1;2;1;3}^9 = 9! / (2! \cdot 1! \cdot 2! \cdot 1! \cdot 3!)$.

$$= \frac{9!}{2!.2!.3!} \cdot (t)^2 \cdot (dt) \cdot (s-t)^2 \cdot (ds) \cdot (1-t)^3$$

Si $X \approx U(0;1)$, su función de densidad es $f(x)=1$, $x \in (0;1)$; por tanto:

VENTANA

$$P(A_1) = P(X \leq t) = \int_0^t dx = t$$

$$P(A_2) = P(t < X < t + dt) = f(t) \cdot dt = dt$$

$$P(A_3) = P(t + dt \leq X \leq s) = P(t \leq X \leq s) = \int_t^s dx = (s - t)$$

$$P(A_4) = P(s < X \leq s + ds) = f(s) \cdot ds = ds$$

$$P(A_5) = P(X \geq s + ds) = P(X \geq s) = \int_s^1 dx = (1 - s)$$

Por tanto:

$$g(t;s) = \frac{9!}{2!.2!.3!} \cdot (t)^2 \cdot (s-t)^2 \cdot (1-t)^3$$

siendo

$$0 < t < 1 ; 0 < s < 1 ; t \leq s$$

Haciendo el cambio de variables $\begin{cases} Z = S - T \\ W = S \end{cases} \Rightarrow \begin{cases} T = W - Z \\ S = W \end{cases}$, el jacobiano es:

$$J = \begin{vmatrix} \partial t / \partial w & \partial t / \partial z \\ \partial s / \partial w & \partial s / \partial z \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} = 1$$

La densidad de probabilidad $h(w;z)$ de la variable $(W;Z)$ es:

$$h(w;z) = g(w-z;w) \cdot |J| =$$

en $g(t;s)$ sustituimos "t" por "w-z" y "s" por "w"

$$\begin{aligned} &= \frac{9!}{2!.2!.3!} \cdot (w-z)^2 \cdot (w-(w-z))^2 \cdot (1-w)^3 = \\ &= \frac{9!}{2!.2!.3!} \cdot (w-z)^2 \cdot z^2 \cdot (1-w)^3 \end{aligned}$$

siendo:

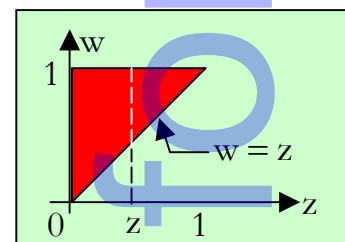
$$0 < w-z < 1 ; 0 < w < 1 ; z \geq 0$$

$$\text{si } \begin{cases} 0 < t < 1 \\ 0 < s < 1 \\ t \leq s \end{cases} \text{ y } \begin{cases} t = w - z \\ s = w \end{cases} \Rightarrow \begin{cases} 0 < w - z < 1 \\ 0 < w < 1 \\ w - z \leq w \end{cases} \Rightarrow \begin{cases} 0 < w - z < 1 \\ 0 < w < 1 \\ z \geq 0 \end{cases}$$

La densidad $f_Z(z)$ de "Z" ($Z = S - T$) es:

$$\begin{aligned} f_Z(z) &= \int_{w=-\infty}^{w=+\infty} h(w;z) \cdot dw = \\ &= \int_{w=z}^{w=1} \frac{9!}{2!.2!.3!} \cdot (w-z)^2 \cdot z^2 \cdot (1-w)^3 \cdot dw = \dots \end{aligned}$$

siendo $z \in [0;1)$.



FONEMATO 2.3.6

De una población $X \approx N(0; \sigma)$ se toma m.a.s de tamaño 2.

Demuéstrese que los estadísticos "Z" y "T" son incorrelados, siendo:

$$Z = (X_1 + X_2)/2 ; T = X_1^2 + X_2^2$$

SOLUCIÓN

Sea $(X_1; X_2)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2; por tanto, X_1 y X_2 son independientes y tienen distribución $N(0; \sigma)$, como la población "X".

Para demostrar que los estadísticos "Z" y "T" son incorrelados debemos demostrar que su covarianza es nula.

$$\text{COV.}(Z; T) = E(Z.T) - E(Z).E(T) \stackrel{?}{=} 0$$

VENTANA

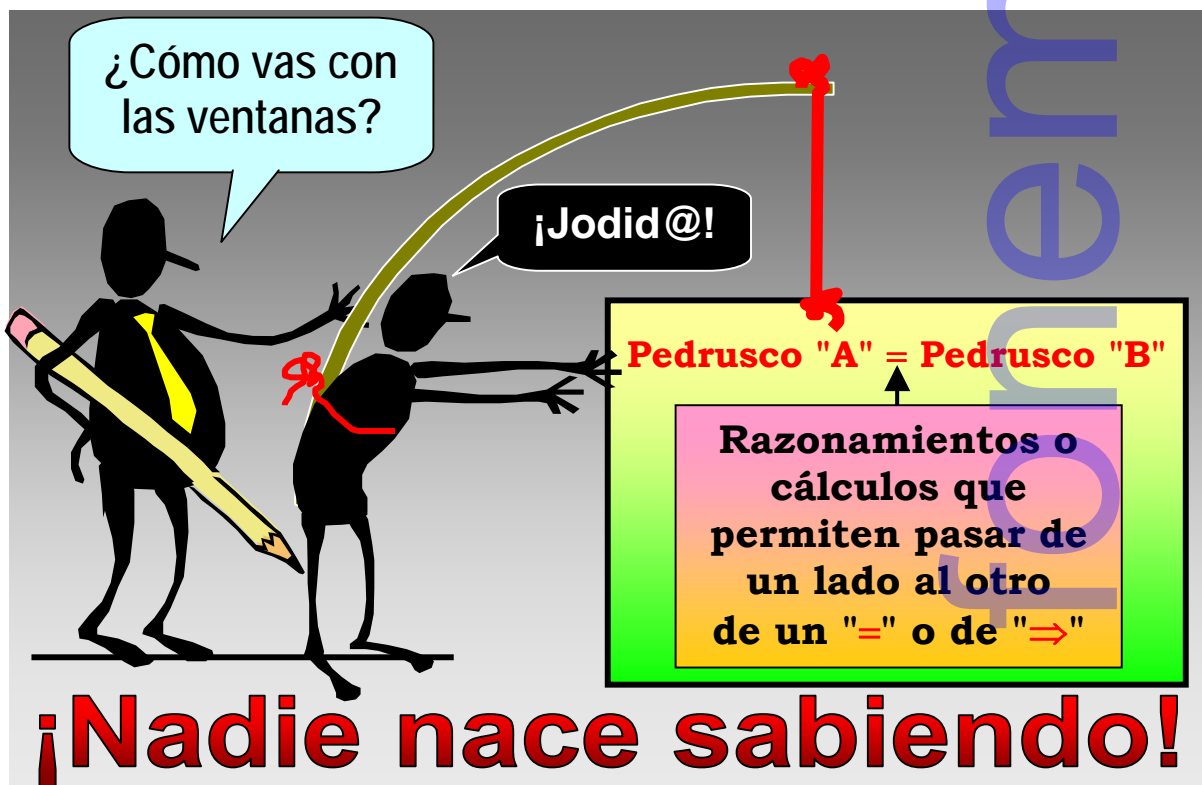
$$* E(Z) = E\left(\frac{1}{2} \cdot (X_1 + X_2)\right) = \frac{1}{2} \cdot (E(X_1) + E(X_2)) = \frac{1}{2} \cdot (0 + 0) = 0$$

$$\begin{aligned} * E(Z.T) &= E\left(\frac{1}{2} \cdot (X_1 + X_2) \cdot (X_1^2 + X_2^2)\right) = \\ &= \frac{1}{2} \cdot E(X_1^3 + X_1 \cdot X_2^2 + X_1^2 \cdot X_2 + X_2^3) = \\ &= \frac{1}{2} \cdot (E(X_1^3) + E(X_1 \cdot X_2^2) + E(X_1^2 \cdot X_2) + E(X_2^3)) \end{aligned}$$

pues las variables aleatorias X_1 y X_2 son independientes

$$= \frac{1}{2} \cdot (E(X_1^3) + E(X_1) \cdot E(X_2^2) + E(X_1^2) \cdot E(X_2) + E(X_2^3)) = 0$$

como en la distribución $N(0; \sigma)$ son nulos todos los momentos de orden impar respecto al origen, es $E(X_1^3) = E(X_1) = E(X_2) = E(X_2^3) = 0$



2.4 LA MEDIA MUESTRAL

Siendo \bar{X} el estadístico media muestral en el muestreo aleatorio simple de tamaño "n" de una población "X" de media μ y varianza σ^2 , es:

$$E(\bar{X}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

$E(X_i) = \mu, \forall i = 1, 2, \dots, n$
 $V(X_i) = \sigma^2, \forall i = 1, 2, \dots, n$

$$V(\bar{X}) = V\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

pues X_1, X_2, \dots, X_n son independientes

O sea, el centro de gravedad de la distribución de masa de probabilidad asociada al estadístico \bar{X} coincide con el centro de gravedad de la distribución de masa de probabilidad asociada a "X", pero la variabilidad de \bar{X} , expresada por $V(\bar{X})$, es la n-ésima parte de la variabilidad de "X", expresada por $V(X)$. **Por ejemplo**, si $n = 100$, la varianza de \bar{X} es la centésima parte de la varianza de "X".

En el Tema 3, para expresar que el valor esperado del estadístico media muestral \bar{X} siempre coincide con el valor esperado μ de la población "X", diremos que \bar{X} es un **estimador insesgado** de la media poblacional μ .

Función generatriz de momentos del estadístico media muestral

Si $\Phi_X(t)$ es la f.g.m. de la variable "X", la f.g.m. de \bar{X} es $\Phi(t) = (\Phi_X(t/n))^n$:

$$\Phi(t) = E(e^{t \cdot \bar{X}}) = E(e^{t \cdot (X_1 + X_2 + \dots + X_n) / n}) = E\left(\prod_{i=1}^n e^{t \cdot X_i / n}\right) =$$

pues X_1, X_2, \dots, X_n son independientes

como X_1, X_2, \dots, X_n tienen la misma distribución de probabilidad que "X", es $E(e^{t \cdot X_i / n}) = E(e^{t \cdot X / n}), \forall i = 1, 2, \dots, n$

$$= \prod_{i=1}^n E(e^{t \cdot X_i / n}) = (E(e^{t \cdot X / n}))^n = (\Phi_X(t/n))^n$$

pues $E(e^{\text{Pepe} \cdot X}) = \Phi_X(\text{Pepe})$

Por ejemplo, siendo $X \approx G(p; a)$ es $\Phi_X(t) = (1 - (t/a))^{-p}$, y la f.g.m. de \bar{X} es:

$$\Phi(t) = (\Phi_X(t/n))^{-n} = \left(1 - \frac{t}{a \cdot n}\right)^{-p \cdot n} = \left(1 - \frac{t}{a \cdot n}\right)^{-p \cdot n}$$

que corresponde a una variable con distribución $G(p \cdot n; a \cdot n)$.

Función de cuantía/densidad del estadístico media muestral

La determinación de la función de cuantía/densidad de la media muestral \bar{X} puede ser cosa fácil o menos fácil.

- Si el tamaño muestral "n" es grande (≥ 30), el teorema Central del Límite nos permitirá aproximar la distribución de \bar{X} mediante la normal con media y varianza las de \bar{X} ; es decir, mediante la $N(\mu; \sigma/\sqrt{n})$.
- **Para muchas variables famosas, con independencia del tamaño muestral, podremos determinar la función de cuantía/densidad del estadístico media muestral a través del estadístico suma total.**

Ejemplo 1: si la variable "X" tiene distribución de Bernoulli de parámetro "p", siendo $T = X_1 + \dots + X_n \approx B(n; p)$ y $\bar{X} = (X_1 + \dots + X_n)/n = T/n$, es:

$$P(\bar{X} = x) = P(T/n = x) = P(T = n \cdot x) \uparrow$$

$$T \approx B(n; p) \Rightarrow P(T = t) = \binom{n}{t} \cdot p^t \cdot (1-p)^{n-t}, \quad t = 0, 1, 2, \dots, n$$

$$= \binom{n}{n \cdot x} \cdot p^{n \cdot x} \cdot (1-p)^{n-n \cdot x}, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1$$

Ejemplo 2: si la variable "X" tiene distribución $B(k; p)$, siendo

$$T = X_1 + \dots + X_n \approx B(n \cdot k; p); \quad \bar{X} = (X_1 + \dots + X_n)/n = T/n$$

es:

$$P(\bar{X} = x) = P(T/n = x) = P(T = n \cdot x) \uparrow$$

$$T \approx B(n \cdot k; p) \Rightarrow P(T = t) = \binom{n \cdot k}{t} \cdot p^t \cdot (1-p)^{n \cdot k - t}, \quad t = 0, 1, 2, \dots, n \cdot k$$

$$= \binom{n \cdot k}{n \cdot x} \cdot p^{n \cdot x} \cdot (1-p)^{n \cdot k - n \cdot x}, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots, k$$

Ejemplo 3: si la variable "X" tiene distribución $P(\theta)$, siendo

$$T = X_1 + \dots + X_n \approx P(n \cdot \theta); \quad \bar{X} = (X_1 + \dots + X_n)/n = T/n$$

es:

$$P(\bar{X} = x) = P(T/n = x) = P(T = n \cdot x) \uparrow$$

$$T \approx P(n \cdot \theta) \Rightarrow P(T = t) = \frac{e^{-n \cdot \theta} \cdot (n \cdot \theta)^t}{t!}, \quad t = 0, 1, 2, \dots$$

$$= \frac{e^{-n \cdot \theta} \cdot (n \cdot \theta)^{n \cdot x}}{(n \cdot x)!}, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots$$

Ejemplo 4: si la variable "X" tiene distribución $G(p)$, siendo

$$T = X_1 + \dots + X_n \approx BN(n; p); \quad \bar{X} = (X_1 + \dots + X_n)/n = T/n$$

es: $P(\bar{X} = x) = P(T/n = x) = P(T = n \cdot x) =$

$$T \approx \text{BN}(n; p) \Rightarrow P(T = t) = \binom{t + n - 1}{t} \cdot (1 - p)^t \cdot p^n, \quad t = 0, 1, 2, \dots$$

$$= \binom{n \cdot x + n - 1}{n \cdot x} \cdot (1 - p)^{n \cdot x} \cdot p^n, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots$$

Ejemplo 5: si la variable "X" tiene distribución $\text{BN}(k; p)$, siendo

$$T = X_1 + \dots + X_n \approx \text{BN}(n \cdot k; p); \quad \bar{X} = (X_1 + \dots + X_n)/n = T/n$$

es:

$$P(\bar{X} = x) = P(T/n = x) = P(T = n \cdot x) =$$

$$T \approx \text{BN}(n \cdot k; p) \Rightarrow P(T = t) = \binom{t + n \cdot k - 1}{t} \cdot (1 - p)^t \cdot p^{n \cdot k}, \quad t = 0, 1, 2, \dots$$

$$= \binom{n \cdot x + n \cdot k - 1}{n \cdot x} \cdot (1 - p)^{n \cdot x} \cdot p^{n \cdot k}, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots$$

Ejemplo 6: si la variable "X" tiene distribución $\text{Exp.}(a) \equiv G(1; a)$, siendo

$$T = X_1 + \dots + X_n \approx G(n; a)$$

la función de densidad de probabilidad de "T" es:

$$f_T(t) = \frac{a^n}{\Gamma(n)} \cdot t^{n-1} \cdot e^{-a \cdot t}, \quad t > 0$$

Así, siendo $\bar{X} = T/n (\Rightarrow T = n \cdot \bar{X})$ la aleatoria media muestral y denotando $g(x)$ a la función densidad de probabilidad de \bar{X} , es:

$$\begin{aligned} g(x) &= f_T(n \cdot x) \cdot \left| \frac{d(n \cdot x)}{dx} \right| = \frac{a^n}{\Gamma(n)} \cdot (n \cdot x)^{n-1} \cdot e^{-a \cdot n \cdot x} \cdot n = \\ &= \frac{(a \cdot n)^n}{\Gamma(n)} \cdot x^{n-1} \cdot e^{-(a \cdot n) \cdot x}, \quad x > 0 \end{aligned}$$

que es la densidad de una variable con distribución $G(n; a \cdot n)$.

Ejemplo 7: si la variable "X" tiene distribución $G(p; a)$, siendo

$$T = X_1 + \dots + X_n \approx G(p \cdot n; a)$$

la función de densidad de probabilidad de "T" es:

$$f_T(t) = \frac{a^{p \cdot n}}{\Gamma(p \cdot n)} \cdot t^{p \cdot n - 1} \cdot e^{-a \cdot t}, \quad t > 0$$

Por tanto, siendo $\bar{X} = T/n (\Rightarrow T = n \cdot \bar{X})$ la aleatoria media muestral y denotando $g(x)$ a la función densidad de probabilidad de \bar{X} , es:

$$\begin{aligned} g(x) &= f_T(n \cdot x) \cdot \left| \frac{d(n \cdot x)}{dx} \right| = \frac{a^{p \cdot n}}{\Gamma(p \cdot n)} \cdot (n \cdot x)^{p \cdot n - 1} \cdot e^{-a \cdot n \cdot x} \cdot n = \\ &= \frac{(a \cdot n)^{p \cdot n}}{\Gamma(p \cdot n)} \cdot x^{p \cdot n - 1} \cdot e^{-a \cdot n \cdot x}, \quad x > 0 \end{aligned}$$

que es la densidad de una variable con distribución $G(p \cdot n; a \cdot n)$.

¡NO ESTAMOS DESCUBRIENDO LA PÓLVORA!

Los 7 ejemplos precedentes son auténticas bobadas que aprendimos a lidiar cuando estudiamos los cambios de variable unidimensionales: se conoce la función de cuantía/densidad de una variable "T" (la suma total) y nos planteamos el cálculo de la función de cuantía/densidad de otra variable \bar{X} (la media muestral) que está ligada con "T" por la relación monótona $T = n \cdot \bar{X}$.

Ejemplo 8: si se toma una muestra de tamaño 2 de una variable "X" con distribución $U(0;3)$, el problema de determinar la distribución de probabilidad del estadístico media muestral es distinto a los anteriores, pues **al contrario que en los ejemplos precedentes, ahora no tenemos ni puta idea de la distribución de probabilidad del estadístico suma total.**

Siendo "X" continua y muestra de tamaño 2, para calcular la densidad de la media muestral "W" (la denotamos "W" en vez de \bar{X} , por comodidad), trabajaremos como aprendimos en su momento: partiendo de la densidad $f(x_1; x_2)$ de la variable muestral $(X_1; X_2)$, calcularemos la densidad $g(w; u)$ de la variable bidimensional $(W; U)$, siendo $U = X_1$ ó $U = X_2$, y después calcularemos la densidad de "W" como marginal de la bidimensional $(W; U)$.

Vamos al tajo: como las variables X_1 y X_2 son independientes, la función de densidad $f(x_1; x_2)$ de la variable bidimensional $(X_1; X_2)$ es el producto de las respectivas funciones de densidad de X_1 y X_2 ; por tanto, como X_1 y X_2 tienen distribución $U(0;3)$, es:

$$f(x_1; x_2) = f_1(x_1) \cdot f_2(x_2) \stackrel{\uparrow}{=} 1/9, \quad 0 < x_1 < 3, \quad 0 < x_2 < 3$$

$$X_i \approx U(0;3) \Rightarrow f_i(x_i) = 1/3, \quad 0 < x_i < 3, \quad i = 1, 2$$

Recuerda: si $f(x_1; x_2)$ es la densidad de la variable bidimensional continua $(X_1; X_2)$ definida en un dominio $D \subseteq \mathbb{R}^2$, siendo $X_1 = h_1(W; U)$ y $X_2 = h_2(W; U)$, la densidad $g(w; u)$ de la variable bidimensional $(W; U)$, es:

$$g(w; u) = f(h_1(w; u); h_2(w; u)) \cdot |J| \quad ; \quad (w; u) \in D^*$$

donde D^* es el dominio en que se transforma "D" como consecuencia del cambio de variables, y "J" es el jacobiano asociado a dicho cambio de variables.

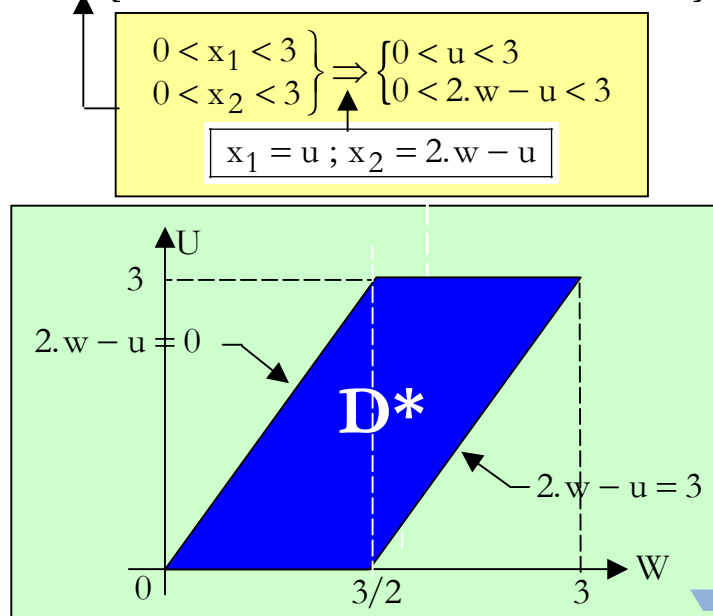
* Siendo $W = (X_1 + X_2)/2$ y $U = X_1$, al despejar X_1 y X_2 , resulta:

$$X_1 = h_1(W; U) = U ; X_2 = h_2(W; U) = 2.W - U$$

* El jacobiano es $J = \begin{vmatrix} \partial x_1 / \partial w & \partial x_1 / \partial u \\ \partial x_2 / \partial w & \partial x_2 / \partial u \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 2 & -1 \end{vmatrix} = -2 \Rightarrow |J| = 2$

* El dominio $D = \{(x_1; x_2) \in \mathbb{R}^2 / 0 < x_1 < 3, 0 < x_2 < 3\}$ se transforma en:

$$D^* = \{(w; u) \in \mathbb{R}^2 / 0 < u < 3, 0 < 2.w - u < 3\}$$

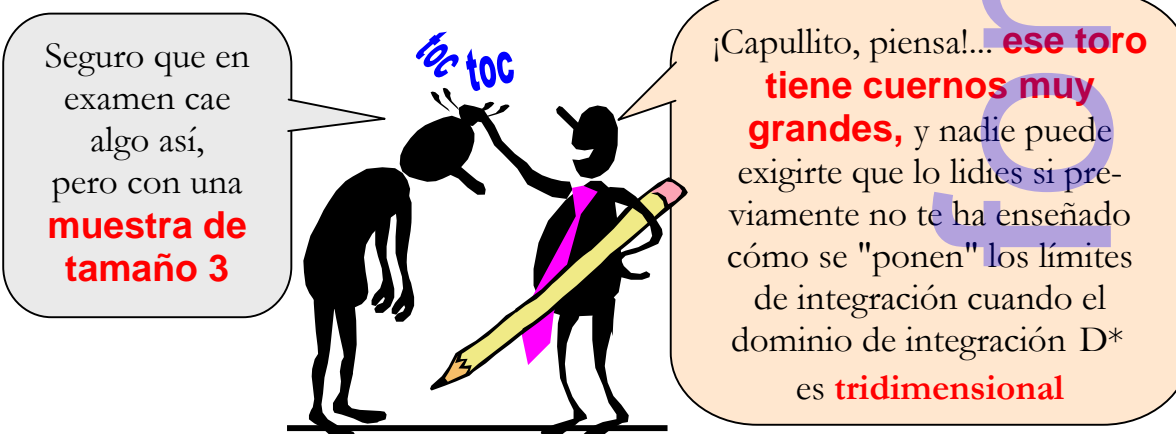


* Función de densidad $g(w; u)$ de la variable $(W; U)$:

$$g(w; u) = f(h_1(w; u); h_2(w; u)) \cdot |J| = 2/9 ; (w; u) \in D^*$$

* Como queda dicho, calculamos la densidad $g_1(w)$ del estadístico media muestral "W" como marginal de la bidimensional $(W; U)$:

$$\begin{aligned} * g_1(w) &= \int_{u=-\infty}^{u=+\infty} g(w; u) \cdot du = \\ &= \begin{cases} \text{si } 0 < w < 3/2 & \text{---} = \int_{u=0}^{u=2.w} 2 \cdot du / 9 = 4 \cdot w / 9 \\ \text{si } 3/2 \leq w < 3 & \text{---} = \int_{u=2.w-3}^{u=3} 2 \cdot du / 9 = (12 - 4 \cdot w) / 9 \\ \text{resto} & \text{---} = 0 \end{cases} \end{aligned}$$



Naturalmente, **una vez conocida la distribución de probabilidad del estadístico media muestral, puede marearse la perdiz con cualquiera de las puñetas típicas de una variable aleatoria unidimensional**; es decir, pueden pedirte la media o la varianza de "W", la función generatriz de momentos de "W", la función de distribución de "W", los coeficientes de asimetría y de curtosis de "W", probabilidades de sucesos relacionados con "W", etc. ... y nada nuevo bajo el sol:

$$* E(W) = \int_0^3 w \cdot g_1(w) \cdot dw = \int_0^{3/2} \frac{4 \cdot w^2}{9} \cdot dw + \int_{3/2}^3 \frac{12 \cdot w - 4 \cdot w^2}{9} \cdot dw = \dots$$

$$* V(W) = E(W^2) - (E(W))^2 = \dots$$

$$E(W^2) = \int_0^3 w^2 \cdot g_1(w) \cdot dw = \int_0^{3/2} \frac{4 \cdot w^3}{9} \cdot dw + \int_{3/2}^3 \frac{12 \cdot w^2 - 4 \cdot w^3}{9} \cdot dw$$

$$* \Phi(t) = E(e^{t \cdot W}) = \int_0^3 e^{t \cdot w} \cdot g_1(w) \cdot dw = \int_0^{3/2} \frac{4 \cdot w}{9} \cdot e^{t \cdot w} \cdot dw + \int_{3/2}^3 \frac{12 - 4 \cdot w}{9} \cdot e^{t \cdot w} \cdot dw = \dots$$

$$* P(0'3 < W < 1'7) = \int_{0'3}^{1'7} g_1(w) \cdot dw = \int_{0'3}^{3/2} \frac{4 \cdot w}{9} \cdot dw + \int_{3/2}^{1'7} \frac{12 - 4 \cdot w}{9} \cdot dw = \dots$$

Observa: una vez conocida la función de densidad de probabilidad del estadístico media muestral "W", el cálculo de la función de densidad de probabilidad $r(t)$ del estadístico suma total $T = 2 \cdot W$ ($\Rightarrow W = T/2$) se reduce a un cambio de variable unidimensional; es:

$$r(t) = g_1(t/2) \cdot \left| \frac{d(t/2)}{dt} \right| = \frac{1}{2} \cdot g_1(t/2) =$$

$$= \begin{cases} \text{si } 0 < t < 3 & \text{---} = \frac{4}{9} \cdot \frac{t}{2} \cdot \frac{1}{2} \\ \text{si } 3 \leq t < 6 & \text{---} = \frac{1}{9} \cdot \left(12 - 4 \cdot \frac{t}{2} \right) \cdot \frac{1}{2} \\ \text{resto} & \text{---} = 0 \end{cases}$$

Que quede claro: el método seguido en este ejemplo para calcular la distribución de probabilidad de la media muestral (siendo continua la población "X", con muestra de tamaño 2), puedes emplearlo para cualquier estadístico.

Por ejemplo, para calcular la función de densidad del estadístico $Z = X_1 \cdot X_2$, partiendo de la densidad $f(x_1; x_2)$ de la variable muestral $(X_1; X_2)$, calcularemos la densidad $s(z; v)$ de la variable bidimensional $(Z; V)$, siendo $V = X_1$ ó $V = X_2$, y después calcularemos la densidad de "Z" como marginal de la bidimensional $(Z; V)$. Vamos al tajo:

* Siendo $Z = X_1 \cdot X_2$ y $V = X_1$, al despejar X_1 y X_2 , resulta:

$$X_1 = \gamma_1(Z; V) = V ; X_2 = \gamma_2(Z; V) = Z/V$$

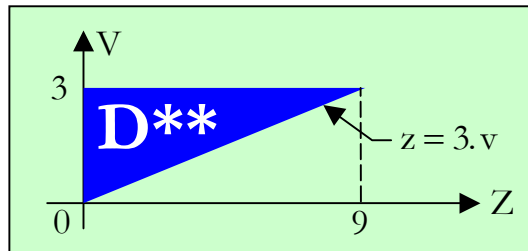
* El jacobiano es $J = \begin{vmatrix} \partial x_1 / \partial z & \partial x_1 / \partial v \\ \partial x_2 / \partial z & \partial x_2 / \partial v \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1/v & -z/v^2 \end{vmatrix} = \frac{1}{v}$

* El dominio $D = \{(x_1; x_2) \in \mathbb{R}^2 / 0 < x_1 < 3, 0 < x_2 < 3\}$ se transforma en:

$$D^{**} = \{(z; v) \in \mathbb{R}^2 / 0 < v < 3, 0 < z < 3 \cdot v\}$$

$$\begin{matrix} 0 < x_1 < 3 \\ 0 < x_2 < 3 \end{matrix} \Rightarrow \begin{cases} 0 < v < 3 \\ 0 < z/v < 3 \Rightarrow 0 < z < 3 \cdot v \end{cases}$$

$$\boxed{x_1 = v; x_2 = z/v}$$



* Función de densidad $s(z; v)$ de la variable $(Z; V)$:

$$s(z; v) = f(\gamma_1(z; v); \gamma_2(z; v)) \cdot |J| = \frac{1}{9 \cdot |v|} ; (z; v) \in D^{**}$$

* Como queda dicho, calculamos la densidad $s_1(z)$ de "Z" como marginal de la bidimensional $(Z; V)$:

$$s_1(z) = \int_{v=-\infty}^{v=+\infty} s(z; v) \cdot dv = \int_{v=z/3}^{v=3} \frac{1}{9 \cdot v} \cdot dv = \frac{1}{9} \cdot \ln \frac{9}{z}, 0 < z < 9$$

$$\text{en } D^{**} \text{ es } v > 0; \text{ por tanto: } \frac{1}{9 \cdot |v|} = \frac{1}{9 \cdot v}$$

Ejemplo 9: considera que se toma una muestra de tamaño 2 de una variable discreta "X" cuya función de cuantía es

x	0	1	2
P(X = x)	0'5	0'2	0'3

El problema de determinar la función de cuantía del estadístico media muestral es distinto de los anteriores, pues ahora, además de desconocer la distribución de probabilidad del estadístico suma total, hay una **putadita adicional: la función de cuantía de la variable discreta "X" no tiene una expresión genérica para todo valor que puede tomar "X".**

Siendo "X" discreta y muestra de tamaño 2, la determinación de la función de cuantía de la media muestral (variable aleatoria discreta) es una gilipollez que aprendimos en su momento: calcularemos el valor que toma la media muestral en cada posible muestra (o sea, en cada punto de la distribución bidimensional discreta asociada a la variable muestral $(X_1; X_2)$), y después obraremos en consecuencia.

Vamos al tajo: determinamos la función de probabilidad o cuantía de la variable muestral $(X_1; X_2)$: como X_1 y X_2 tienen la misma distribución que " X ", es:

$$\frac{x_1}{P(X_1 = x_1)} \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline & 0'5 & 0'2 & 0'3 \end{array} ; \frac{x_2}{P(X_2 = x_2)} \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline & 0'5 & 0'2 & 0'3 \end{array}$$

Como las variables X_1 y X_2 son independientes, la función de cuantía de la variable bidimensional $(X_1; X_2)$ es el producto de las respectivas funciones de cuantía de X_1 y X_2 ; por tanto:

$$P(X_1 = x_1; X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2) \Rightarrow$$

$$\Rightarrow \begin{array}{c|ccc} X_1 / X_2 & 0 & 1 & 2 \\ \hline 0 & 0'25 & 0'10 & 0'15 \\ 1 & 0'10 & 0'04 & 0'06 \\ 2 & 0'15 & 0'06 & 0'09 \end{array}$$

Ahora calculamos el valor que toma $\bar{X} = (X_1 + X_2)/2$ en cada punto de la distribución de probabilidad de la variable bidimensional $(X_1; X_2)$:

$(x_1; x_2)$	$P(X_1 = x_1; X_2 = x_2)$	$\bar{X} = (X_1 + X_2)/2$
(0;0)	0'25	0
(0;1)	0'10	0'5
(0;2)	0'15	1
(1;0)	0'10	0'5
(1;1)	0'04	1
(1;2)	0'06	1'5
(2;0)	0'15	1
(2;1)	0'06	1'5
(2;2)	0'09	2

A la vista del valor que toma $\bar{X} = (X_1 + X_2)/2$ en cada posible muestra y de la probabilidad que corresponde a esa muestra, con la gorra determinamos la función de cuantía del estadístico \bar{X} ; es:

$$\left\{ \begin{array}{l} P(\bar{X} = 0) = P(0;0) = 0'25 \\ P(\bar{X} = 0'5) = P(0;1) + P(1;0) = 0'20 \\ P(\bar{X} = 1) = P(0;2) + P(1;1) + P(2;0) = 0'34 \\ P(\bar{X} = 1'5) = P(1;2) + P(2;1) = 0'12 \\ P(\bar{X} = 2) = P(2;2) = 0'09 \end{array} \right\} \Rightarrow$$

$$\Rightarrow \frac{x}{P(\bar{X} = x)} \begin{array}{c|ccccc} & 0 & 0'5 & 1 & 1'5 & 2 \\ \hline & 0'25 & 0'20 & 0'34 & 0'12 & 0'09 \end{array}$$

Como queda dicho en el ejemplo anterior, una vez conocida la distribución de probabilidad del estadístico media muestral \bar{X} , puede marcarse la perdiz no cualquiera de las puñetas típicas de una variable aleatoria unidimensional:

$$* E(\bar{X}) = 0.0'25 + 0'5.0'20 + 1.0'34 + 1'5.0'12 + 2.0'09 = \dots$$

$$* V(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}))^2 = \dots$$

$$E(\bar{X}^2) = 0^2 \cdot 0'25 + 0'5^2 \cdot 0'20 + 1^2 \cdot 0'34 + 1'5^2 \cdot 0'12 + 2^2 \cdot 0'09$$

$$* \Phi(t) = E(e^{t \cdot \bar{X}}) = e^{0 \cdot t} \cdot 0'25 + e^{0'5 \cdot t} \cdot 0'20 + e^{1 \cdot t} \cdot 0'34 + e^{1'5 \cdot t} \cdot 0'12 + e^{2 \cdot t} \cdot 0'09 = \dots$$

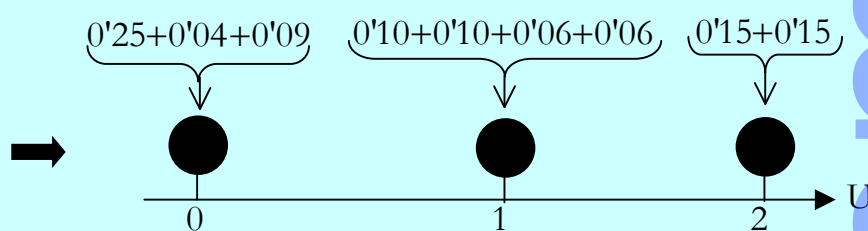
$$* P\left(\frac{1}{\bar{X} - 0'7} \leq 2\right) = P\left(\bar{X} - 0'7 \geq \frac{1}{2}\right) = P(\bar{X} \geq 1'2) = 0'12 + 0'09$$

¡Que quede claro!

Siendo "X" discreta y muestra de tamaño 2, para calcular la función de cuantía de un estadístico "Pepe", haremos lo que acabamos de hacer con el estadístico media muestral: calcularemos el valor que toma "Pepe" en cada punto de la distribución bidimensional discreta asociada a la variable muestral $(X_1; X_2)$, y después obraremos en consecuencia.

Por ejemplo, para el estadístico "U" tal que $U = |X_1 - X_2|$, es:

$(x_1; x_2)$	(0;0)	(0;1)	(0;2)	(1;0)	(1;1)	(1;2)	(2;0)	(2;1)	(2;2)
$P(x_1; x_2)$	0'25	0'10	0'15	0'10	0'04	0'06	0'15	0'06	0'09
$U = X_1 - X_2 $	0	1	2	1	0	1	2	1	0



Una vez conocida la distribución de probabilidad del estadístico "U", puede marearse la perdiz con cualquier puñeta típica de una variable unidimensional:

$$* E(U) = 0.0'38 + 1.0'32 + 2.0'30 = \dots$$

$$* V(U) = E(U^2) - (E(U))^2 = \dots$$

$$E(U^2) = 0^2 \cdot 0'38 + 1^2 \cdot 0'32 + 2^2 \cdot 0'30$$

$$* \Phi(t) = E(e^{t \cdot U}) = e^{0 \cdot t} \cdot 0'38 + e^{1 \cdot t} \cdot 0'32 + e^{2 \cdot t} \cdot 0'30 = \dots$$

$$* P(1/(U - 0'7) \leq 2) = P(U - 0'7 \geq 1/2) = P(U \geq 1'2) = 0'30$$

FONEMATO 2.4.1

- 1) Si el 70 % de los estudiantes se chupan el dedo y seleccionamos 100 estudiantes al azar, determínese la probabilidad de que la proporción de estudiantes que se chupan el dedo esté entre 0'65 y 0'8.
- 2) Determínese un intervalo en el que con probabilidad 0'95 puede esperarse que se encuentre la proporción muestral de chupadedos.

SOLUCIÓN

- 1) Si chuparse el dedo es un "éxito", la población "X" objeto de estudio tiene distribución de Bernoulli de parámetro 0'7, pues dicen que el 70 % de los estudiantes se lo chupan. Así, siendo $(X_1; \dots; X_{100})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 100, el estadístico media muestral $\bar{X} = (X_1 + \dots + X_{100})/100$ expresa la proporción de estudiantes que se chupan el dedo en la muestra; y **como el tamaño muestral es grande, podemos aproximar la distribución de probabilidad de \bar{X} mediante la normal con media y varianza las de \bar{X}** , que son:

$$\begin{array}{c} E(\bar{X}) = E(X) = 0'7 \\ \uparrow \\ \boxed{X \approx B(p) \Rightarrow E(X) = p \text{ y } V(X) = p \cdot (1 - p)} \\ \downarrow \\ V(\bar{X}) = \frac{V(X)}{100} = \frac{0'7 \cdot 0'3}{100} = 0'0021 \end{array}$$

Por tanto:

$$\begin{aligned} P(0'65 < \bar{X} < 0'8) &\cong P(0'65 < N(0'7; \sqrt{0'0021}) < 0'8) = \\ &= P\left(N(0;1) < \frac{0'8 - 0'7}{\sqrt{0'0021}}\right) - P\left(N(0;1) < \frac{0'65 - 0'7}{\sqrt{0'0021}}\right) = \dots \end{aligned}$$

- 2) En la distribución de probabilidad de la variable $N(0'7; \sqrt{0'0021})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la $N(0'7; \sqrt{0'0021})$ es simétrica respecto al punto 0'7, **entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto** (puede demostrarse que éste es el de menor amplitud). Por tanto, debemos determinar $c > 0$ de modo que $P(|\bar{X} - 0'7| < c) = 0'95$:

$$\begin{aligned} P(|\bar{X} - 0'7| < c) &= 0'95 \Rightarrow P\left(|N(0;1)| < \frac{c}{\sqrt{0'0021}}\right) = 0'95 \Rightarrow \\ \Rightarrow 2 \cdot P\left(N(0;1) < \frac{c}{\sqrt{0'0021}}\right) - 1 &= 0'95 \Rightarrow P\left(N(0;1) < \frac{c}{\sqrt{0'0021}}\right) = 0'975 \Rightarrow \\ \Rightarrow \frac{c}{\sqrt{0'0021}} &= 1'96 \Rightarrow c = 1'96 \cdot \sqrt{0'0021} = 0'0898 \end{aligned}$$

O sea, si tomamos sucesivas muestras de tamaño 100 de $X \approx B(0'7)$, el 95 % de las veces sucederá que el estadístico media muestral $\bar{X} \approx N(0'7; \sqrt{0'0021})$ toma un valor en el intervalo $(0'7 - 0'0898; 0'7 + 0'0898)$.

Tema 2: Distribuciones en el muestreo

FONEMATO 2.4.2

La producción diaria de tomate de una empresa oscila entre 6000 y 10000 kg. Determinése la probabilidad de que la producción media durante 320 días supere los 8000 kg.

SOLUCIÓN

Supuesto que la variable "X" que expresa la producción diaria tiene distribución $U(6000;10000)$, y siendo $(X_1; \dots; X_{320})$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño 320, el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_{320})/320$. Como el tamaño muestral es grande, podemos aproximar la distribución de probabilidad de \bar{X} mediante la normal con media y varianza las de \bar{X} , que son:

$$E(\bar{X}) = E(X) = \frac{6000 + 10000}{2} = 8000$$

$$X \approx U(a;b) \Rightarrow E(X) = \frac{a+b}{2} \text{ y } V(X) = \frac{(b-a)^2}{12}$$

$$V(\bar{X}) = \frac{V(X)}{320} = \frac{4000^2}{12}$$

Por tanto:

$$P(\bar{X} > 8000) \cong P(N(8000; 4000/\sqrt{12}) > 8000) = P(N(0;1) > 0) = 0.5$$



FONEMATO 2.4.3

De una población "X" se toma muestra aleatoria simple de tamaño 81. En los siguientes casos, determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre el estadístico media muestral.

- 1) Población de Poisson de parámetro 5.
- 2) Población $B(4;0'3)$.
- 3) Población $\text{Exp.}(0'02)$.
- 4) Población $U(-1;1)$.
- 5) Población con densidad $f(x) = 2 \cdot x$, $0 \leq x \leq 1$.
- 6) Población geométrica de parámetro 0'1.
- 7) Población $G(20;2)$.

SOLUCIÓN

Como es sabido, en el muestreo aleatorio simple de tamaño $n \geq 30$ de una población "X" con media μ y varianza σ^2 , Lindeberg-Levy permiten aproximar la distribución de probabilidad del estadístico media muestral \bar{X} mediante la normal con media y varianza las de \bar{X} ; o sea, mediante la $N(\mu; \sigma/\sqrt{n})$.

En la distribución de probabilidad de la variable $N(\mu; \sigma/\sqrt{n})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, como la $N(\mu; \sigma/\sqrt{n})$ es simétrica respecto al punto μ , entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto (puede demostrarse que éste es el de menor amplitud).

Por tanto, debemos determinar $c > 0$ de modo que $P(|\bar{X} - \mu| < c) = 0'95$:

$$\begin{aligned} P(|\bar{X} - \mu| < c) = 0'95 &\Rightarrow P\left(|N(0;1)| < \frac{c}{\sigma/\sqrt{n}}\right) = 0'95 \Rightarrow \\ \Rightarrow 2P\left(N(0;1) < \frac{c}{\sigma/\sqrt{n}}\right) - 1 = 0'95 &\Rightarrow P\left(N(0;1) < \frac{c}{\sigma/\sqrt{n}}\right) = 0'975 \Rightarrow \\ \Rightarrow \frac{c}{\sigma/\sqrt{n}} = 1'96 &\Rightarrow c = 0'217 \cdot \sigma \end{aligned}$$

si $n = 81$

- 1) Si $X \approx P(5)$ es $E(X) = \mu = 5$ y $V(X) = \sigma^2 = 5$.

Así, es $P(|\bar{X} - 5| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{5} = 0'48$; o sea, si tomamos sucesivas muestras de tamaño 81 de $X \approx P(5)$, el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(5; \sqrt{5}/9)$ toma un valor en el intervalo $(5 - 0'48; 5 + 0'48)$.

- 2) Si $X \approx B(4;0'3)$ es $E(X) = \mu = 4 \cdot 0'3 = 1'2$ y $V(X) = \sigma^2 = 4 \cdot 0'3 \cdot 0'7 = 0'84$.

Así, es $P(|\bar{X} - 1'2| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{0'84} = 0'19$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx B(4;0'3)$, el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(1'2; \sqrt{0'84}/9)$ se concreta en algún punto del intervalo $(1'2 - 0'19; 1'2 + 0'19)$.

3) Si $X \approx \text{Exp.}(0'02)$ es $E(X) = \mu = 1/0'02 = 50$ y $V(X) = \sigma^2 = 1/0'02^2 = 2500$.

Así, es $P(|\bar{X} - 50| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{2500} = 10'95$; o sea, si tomamos sucesivas muestras de tamaño 81 de $X \approx \text{Exp.}(0'02)$, el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(50; \sqrt{2500}/9)$ toma un valor en el intervalo $(50 - 10'95; 50 + 10'95)$.

4) Si $X \approx U(-1;1)$ es $E(X) = \mu = (-1+1)/2 = 0$ y $V(X) = \sigma^2 = (1-(-1))^2/12 = 1/3$.

Así, es $P(|\bar{X} - 0| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{1/3} = 0'12$; o sea, si tomamos sucesivas muestras de tamaño 81 de $X \approx U(-1;1)$, el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(0; \sqrt{1/243})$ toma un valor en $(-0'12; 0'12)$.

5) Si la densidad de la población "X" es $f(x) = 2 \cdot x$, $0 \leq x \leq 1$, entonces:

$$E(X) = \int_0^1 x \cdot f(x) \cdot dx = \int_0^1 2 \cdot x^2 \cdot dx = 2/3$$

$$V(X) = \sigma^2 = E(X^2) - (E(X))^2 = 1/18$$

$$E(X^2) = \int_0^1 x^2 \cdot f(x) \cdot dx = \int_0^1 2 \cdot x^3 \cdot dx = 1/2$$

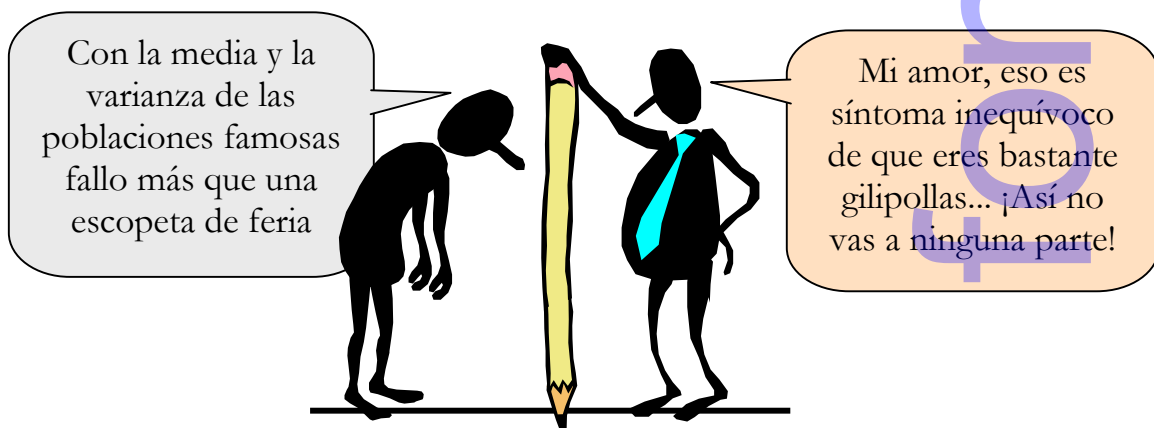
Así, es $P(|\bar{X} - 2/3| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{1/18} = 0'05$; o sea, si tomamos sucesivas muestras de tamaño 81 de "X", el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(2/3; \sqrt{1/1458})$ se concreta en algún punto del intervalo $((2/3) - 0'05; (2/3) + 0'05)$.

6) Si $X \approx G(0'1)$ es $E(X) = \mu = (1 - 0'1)/0'1 = 9$ y $V(X) = \sigma^2 = (1 - 0'1)/0'1^2 = 90$.

Así, es $P(|\bar{X} - 9| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{90} = 2'05$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx G(0'1)$, el 95 % de las veces sucederá que la media muestral $\bar{X} \approx N(9; \sqrt{90}/9)$ toma un valor en $(9 - 2'05; 9 + 2'05)$.

7) Si $X \approx G(20;2)$ es $E(X) = \mu = 20/2 = 10$ y $V(X) = \sigma^2 = 20/2^2 = 5$.

Así, es $P(|\bar{X} - 10| < c) = 0'95$ si $c = 0'217 \cdot \sqrt{5} = 0'48$; o sea, tomando sucesivas muestras de tamaño 81 de $X \approx G(20;2)$, el 95 % de las veces sucederá que $\bar{X} \approx N(10; \sqrt{5}/9)$ toma un valor en el intervalo $(10 - 0'48; 10 + 0'48)$.



2.5 LA VARIANZA MUESTRAL

Siendo S^2 el estadístico varianza muestral en el muestreo aleatorio simple de tamaño "n" de una población "X" de media μ y varianza σ^2 , es:

$$E(S^2) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n} \cdot E\left(\left(\sum_{i=1}^n (X_i - \mu)^2\right) - n \cdot (\bar{X} - \mu)^2\right) =$$

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i + (\mu - \mu) - \bar{X})^2 = \sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2 = \\ &= \sum_{i=1}^n ((X_i - \mu)^2 - 2 \cdot (X_i - \mu) \cdot (\bar{X} - \mu) + (\bar{X} - \mu)^2) = \\ &= \left(\sum_{i=1}^n (X_i - \mu)^2\right) - 2 \cdot (\bar{X} - \mu) \cdot \left(\sum_{i=1}^n (X_i - \mu)\right) + n \cdot (\bar{X} - \mu)^2 = \end{aligned}$$

$$\sum_{i=1}^n (X_i - \mu) = \left(\sum_{i=1}^n X_i\right) - n \cdot \mu = n \cdot \bar{X} - n \cdot \mu$$

$$\sum_{i=1}^n X_i = n \cdot \bar{X}$$

$$\begin{aligned} &= \left(\sum_{i=1}^n (X_i - \mu)^2\right) - 2 \cdot (\bar{X} - \mu) \cdot (n \cdot \bar{X} - n \cdot \mu) + n \cdot (\bar{X} - \mu)^2 = \\ &= \left(\sum_{i=1}^n (X_i - \mu)^2\right) - n \cdot (\bar{X} - \mu)^2 \end{aligned}$$

$$= \frac{1}{n} \cdot E\left(\sum_{i=1}^n (X_i - \mu)^2\right) - E\left((\bar{X} - \mu)^2\right) = \frac{1}{n} \cdot (n \cdot \sigma^2) - \frac{\sigma^2}{n} = \frac{n-1}{n} \cdot \sigma^2$$

$$* E\left(\sum_{i=1}^n (X_i - \mu)^2\right) = \sum_{i=1}^n E((X_i - \mu)^2) = \sum_{i=1}^n V(X_i) = n \cdot \sigma^2$$

$$E(X_i) = \mu \Rightarrow E((X_i - \mu)^2) = V(X_i), \forall i = 1, 2, \dots, n$$

$$* E\left((\bar{X} - \mu)^2\right) = E\left((\bar{X} - E(\bar{X}))^2\right) = V(\bar{X}) = \sigma^2/n$$

$$\mu = E(\bar{X})$$

Observa: siendo $S_c^2 = n \cdot S^2 / (n-1)$ el **estadístico cuasivarianza muestral**, es:

$$E(S_c^2) = E\left(\frac{n}{n-1} \cdot S^2\right) = \frac{n}{n-1} \cdot E(S^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2$$

En el Tema 3, para expresar que el $E(S_c^2) = \sigma^2$, diremos que S_c^2 es un **estimador insesgado** de σ^2 . Como $E(S^2) = (n-1) \cdot \sigma^2 / n \neq \sigma^2$, el estadístico varianza muestral S^2 no es un estimador insesgado de σ^2 ; diremos que S^2 es un estimador **asintóticamente insesgado** de σ^2 para expresar de modo rápido y eficiente que $\lim_{n \rightarrow \infty} E(S^2) = \lim_{n \rightarrow \infty} (n-1) \cdot \sigma^2 / n = \sigma^2$.

Tema 2: Distribuciones en el muestreo

© Rafael Cabrejas Hernansanz

FONEMATO 2.5.1

Determinése la distribución de probabilidad de la varianza muestral en el muestreo aleatorio simple de tamaño 2 de una población $Z \approx U(0;1)$.

SOLUCIÓN

Siendo $Z \approx U(0;1)$, su densidad es $f(z) = 1, z \in (0;1)$. La densidad "g" de la variable muestral $(X;Y)$ que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2 de la población $Z \approx U(0;1)$ es el producto de las densidades de "X" e "Y", pues éstas son independientes:

$$g(x;y) = f(x).f(y) = 1, x \in (0;1), y \in (0;1)$$

pues $X \approx U(0;1)$ e $Y \approx U(0;1)$

Siendo $U = (X + Y)/2$ la media muestral, la varianza muestral "T" es:

$$\begin{aligned} T &= \frac{1}{2} \cdot ((X - U)^2 + (Y - U)^2) = \frac{1}{2} \cdot \left(\left(X - \frac{X+Y}{2} \right)^2 + \left(Y - \frac{X+Y}{2} \right)^2 \right) = \\ &= \frac{1}{2} \cdot \left(\left(\frac{X-Y}{2} \right)^2 + \left(\frac{Y-X}{2} \right)^2 \right) = \frac{1}{4} \cdot (X - Y)^2 \end{aligned}$$

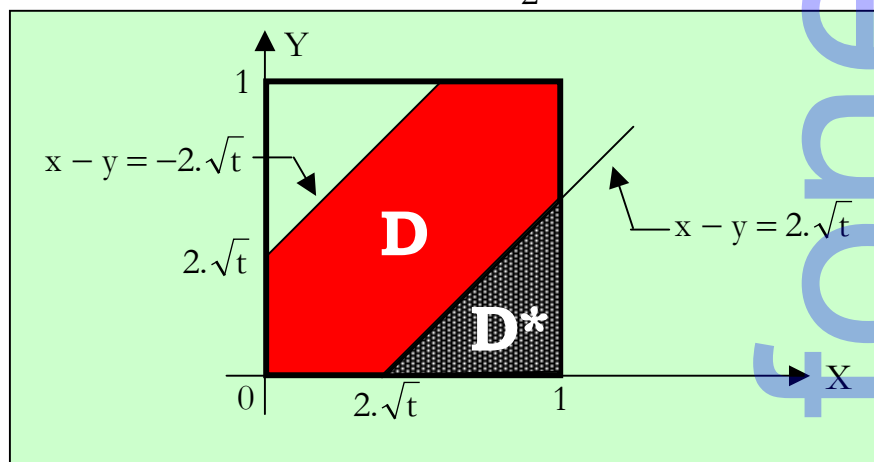
- Función de distribución de "T":

$$\begin{aligned} F_T(t) &= P(T \leq t) = P\left(\frac{1}{4} \cdot (X - Y)^2 \leq t\right) = P((X - Y)^2 \leq 4 \cdot t) = \\ &= P(-2 \cdot \sqrt{t} \leq X - Y \leq 2 \cdot \sqrt{t}) = \iint_D g(x;y) \cdot dx \cdot dy = \end{aligned}$$

siendo "g" la función de densidad de la variable muestral $(X;Y)$ y
 $D = \{(x;y) \in \mathbb{R}^2 / -2 \cdot \sqrt{t} \leq x - y \leq 2 \cdot \sqrt{t}, 0 < x < 1, 0 < y < 1\}$

siempre hay algún capullito de alhelí que no se da cuenta de que al ser constante la densidad $g(x;y)$ es innecesario lidiar la integral doble, pues el asunto se reduce a calcular áreas

$$= 1 - 2 \cdot (\text{Área de } D^*) = 1 - 2 \cdot \frac{(1 - 2 \cdot \sqrt{t})^2}{2} = 4 \cdot \sqrt{t} - 4 \cdot t$$



- La función de densidad de "T" es $f_T(t) = \frac{dF_T(t)}{dt} = \frac{2}{\sqrt{t}} - 4, 0 < t < 1/4$.

FONEMATO 2.5.2

Determinése la distribución de probabilidad de la varianza muestral en el muestreo aleatorio simple de tamaño 2 de una población $Z \approx \text{Exp.}(1)$.

SOLUCIÓN

Siendo $Z \approx \text{Exp.}(1)$, su densidad es $f(z) = e^{-z}$, $z > 0$. La densidad "g" de la variable muestral $(X; Y)$ que expresa los resultados que pueden presentarse al tomar m.a.s de tamaño 2 de la población $Z \approx \text{Exp.}(1)$ es el producto de las densidades de "X" e "Y", pues éstas son independientes:

$$g(x; y) = f(x) \cdot f(y) = e^{-(x+y)}, \quad x > 0, \quad y > 0$$

pues $X \approx \text{Exp.}(1)$ e $Y \approx \text{Exp.}(1)$

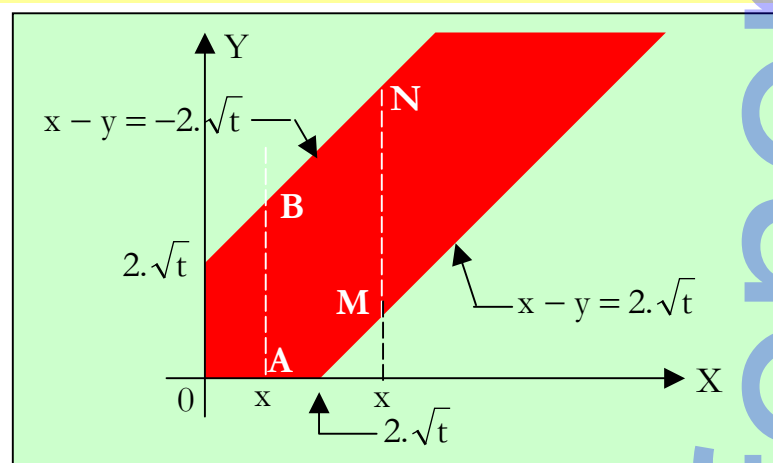
Siendo $U = (X + Y)/2$ el estadístico media muestral, el estadístico varianza muestral "T" es:

$$\begin{aligned} T &= \frac{1}{2} \cdot ((X - U)^2 + (Y - U)^2) = \\ &= \frac{1}{2} \cdot \left(\left(X - \frac{X+Y}{2} \right)^2 + \left(Y - \frac{X+Y}{2} \right)^2 \right) = \\ &= \frac{1}{2} \cdot \left(\left(\frac{X-Y}{2} \right)^2 + \left(\frac{Y-X}{2} \right)^2 \right) = \frac{1}{4} \cdot (X - Y)^2 \end{aligned}$$

- Función de distribución de "T":

$$\begin{aligned} F_T(t) &= P(T \leq t) = P\left(\frac{1}{4} \cdot (X - Y)^2 \leq t\right) = P((X - Y)^2 \leq 4t) = \\ &= P(-2\sqrt{t} \leq X - Y \leq 2\sqrt{t}) = \iint_D g(x; y) \cdot dx \cdot dy = \end{aligned}$$

siendo "g" la función de densidad de la variable muestral $(X; Y)$ y
 $D = \{(x; y) \in \mathbb{R}^2 / -2\sqrt{t} \leq x - y \leq 2\sqrt{t}, \quad x > 0, \quad y > 0\}$

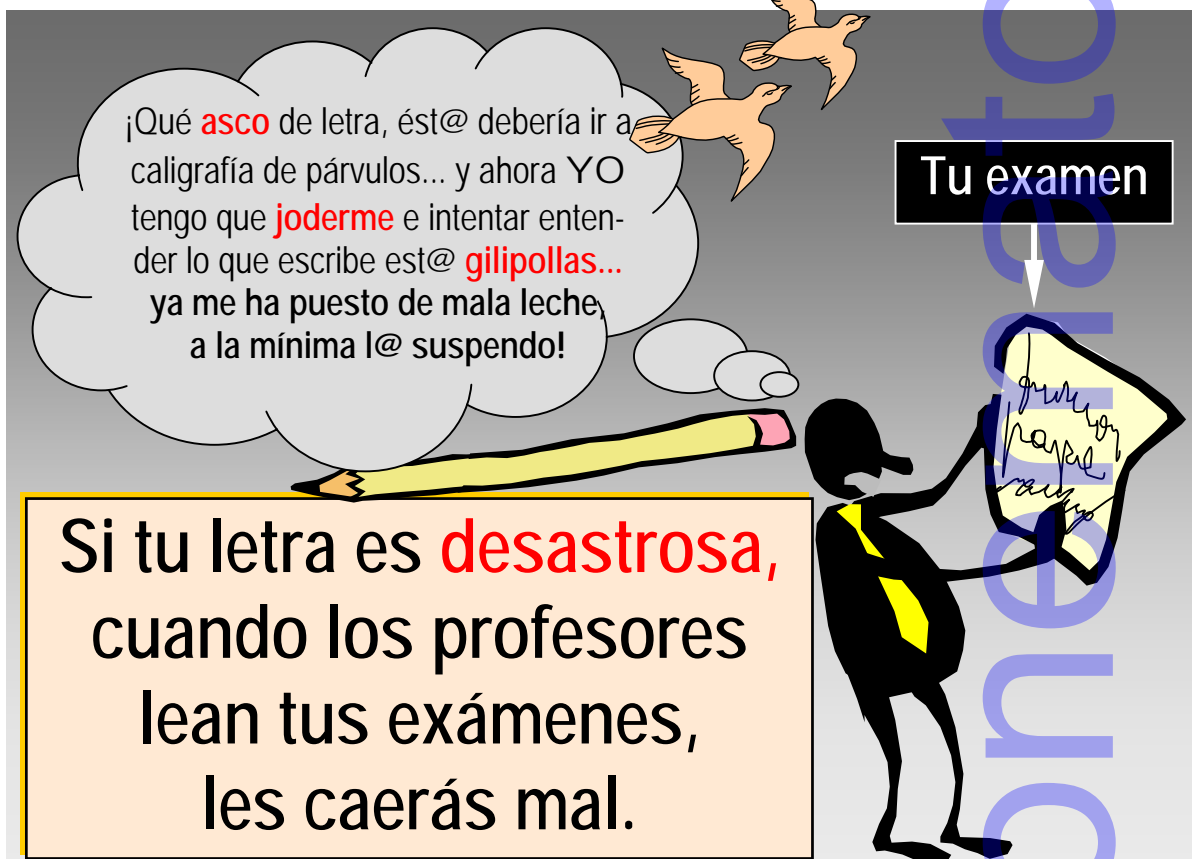


$$\begin{aligned} &= \int_{x=0}^{x=2\sqrt{t}} \left(\int_{y_A=0}^{y_B=x+2\sqrt{t}} e^{-(x+y)} \cdot dy \right) \cdot dx + \\ &+ \int_{x=2\sqrt{t}}^{x=+\infty} \left(\int_{y_M=x-2\sqrt{t}}^{y_N=x+2\sqrt{t}} e^{-(x+y)} \cdot dy \right) \cdot dx = \end{aligned}$$

$$\begin{aligned}
&= \int_{x=0}^{x=2\sqrt{t}} e^{-x} \cdot (-e^{-y})_{y=0}^{y=x+2\sqrt{t}} \cdot dx + \int_{x=2\sqrt{t}}^{x=+\infty} e^{-x} \cdot (-e^{-y})_{y=x-2\sqrt{t}}^{y=x+2\sqrt{t}} \cdot dx = \\
&= \int_{x=0}^{x=2\sqrt{t}} e^{-x} \cdot (1 - e^{-x-2\sqrt{t}}) \cdot dx + \int_{x=2\sqrt{t}}^{x=+\infty} e^{-x} \cdot (e^{-x+2\sqrt{t}} - e^{-x-2\sqrt{t}}) \cdot dx = \\
&= \int_{x=0}^{x=2\sqrt{t}} e^{-x} \cdot dx - \int_{x=0}^{x=2\sqrt{t}} e^{-2x-2\sqrt{t}} \cdot dx \\
&\quad + \int_{x=2\sqrt{t}}^{x=+\infty} e^{-2x+2\sqrt{t}} \cdot dx - \int_{x=2\sqrt{t}}^{x=+\infty} e^{-2x-2\sqrt{t}} \cdot dx = \\
&= (-e^{-x})_0^{2\sqrt{t}} + \frac{1}{2} \cdot (e^{-2x-2\sqrt{t}})_0^{2\sqrt{t}} - \\
&\quad - \frac{1}{2} \cdot (e^{-2x+2\sqrt{t}})_{2\sqrt{t}}^{+\infty} + \frac{1}{2} \cdot (e^{-2x-2\sqrt{t}})_{2\sqrt{t}}^{+\infty} = \\
&= (1 - e^{-2\sqrt{t}}) + \frac{e^{-6\sqrt{t}} - e^{-2\sqrt{t}}}{2} - \frac{e^{-\infty} - e^{-2\sqrt{t}}}{2} + \frac{e^{-\infty} - e^{-6\sqrt{t}}}{2} = \\
&= 1 - e^{-2\sqrt{t}}, \quad t > 0
\end{aligned}$$

- Función de densidad de "T":

$$f_T(t) = dF_T(t)/dt = \frac{1}{\sqrt{t}} \cdot e^{-2\sqrt{t}}, \quad t > 0$$



En fonemato.com tienes los vídeos en los que explicamos los contenidos de este libro.

2.6 MUESTREO DE POBLACIÓN NORMAL

Siendo $X \approx N(\mu; \sigma)$, sea $(X_1, \dots; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar m.a.s. de tamaño "n".

• Distribución de la media muestral si se conoce σ

Siendo $\bar{X} = (X_1 + \dots + X_n)/n$ una combinación lineal de variables normales independientes, la distribución de \bar{X} es normal con media y varianza las de \bar{X} ; es decir, \bar{X} tiene distribución $N(\mu; \sigma/\sqrt{n})$. Por tanto, la f.g.m. de \bar{X} es

$$\Phi(t) = e^{\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2 \cdot n}}$$

• Teorema de Fisher

En el muestreo aleatorio simple de una variable $X \approx N(\mu; \sigma)$, los estadísticos (variables aleatorias unidimensionales) media muestral \bar{X} y varianza muestral S^2 son independientes. En efecto, la f.g.m. de la variable (n+1)-dimensional $(\bar{X}; X_1 - \bar{X}; \dots; X_n - \bar{X})$ es:

$$\begin{aligned} \Phi(t; t_1; \dots; t_n) &= E\left(\exp\left(t \cdot \bar{X} + t_1 \cdot (X_1 - \bar{X}) + \dots + t_n \cdot (X_n - \bar{X})\right)\right) = \\ &= E\left(\exp\left(t \cdot \bar{X} + \sum_{i=1}^n t_i \cdot (X_i - \bar{X})\right)\right) = \\ &= E\left(\exp\left(\frac{1}{n} \cdot \sum_{i=1}^n t \cdot X_i + \sum_{i=1}^n t_i \cdot (X_i - \bar{X})\right)\right) = \\ &= E\left(\exp\left(\sum_{i=1}^n X_i \cdot \left(\frac{t}{n} + t_i - \frac{1}{n} \cdot \sum_{i=1}^n t_i\right)\right)\right) = \\ &= E\left(\exp\left(\sum_{i=1}^n X_i \cdot \left(t_i - \frac{t_1 + t_2 + \dots + t_n - t}{n}\right)\right)\right) \end{aligned}$$

por comodidad en la notación, hacemos $\delta = \frac{1}{n} \cdot \sum_{i=1}^n t_i$

$$= E\left(\exp\left(\sum_{i=1}^n X_i \cdot \left(t_i - \frac{n \cdot \delta - t}{n}\right)\right)\right) = E\left(\prod_{i=1}^n \exp\left(X_i \cdot \left(t_i - \frac{n \cdot \delta - t}{n}\right)\right)\right) =$$

pues $E(\exp(\text{Pepe} \cdot X_i)) = \Phi_i(\text{Pepe})$, siendo Φ_i la f.g.m. de X_i

$$= \prod_{i=1}^n E\left(\exp\left(X_i \cdot \left(t_i - \frac{n \cdot \delta - t}{n}\right)\right)\right) = \prod_{i=1}^n \Phi_i\left(t_i - \frac{n \cdot \delta - t}{n}\right) =$$

la f.g.m de la variable $X_i \approx N(\mu; \sigma)$ es $\Phi_i(\lambda) = \exp\left(\mu \cdot \lambda + \frac{1}{2} \cdot \lambda^2 \cdot \sigma^2\right)$

$$\begin{aligned}
&= \prod_{i=1}^n \exp \left(\left(t_i - \frac{n \cdot \delta - t}{n} \right) \cdot \mu + \frac{1}{2} \cdot \left(t_i - \frac{n \cdot \delta - t}{n} \right)^2 \cdot \sigma^2 \right) = \\
&= \exp \left(\left(t_i - \frac{n \cdot \delta - t}{n} \right) \cdot \mu + \frac{1}{2} \cdot \sigma^2 \cdot \sum_{i=1}^n \left(t_i - \frac{n \cdot \delta - t}{n} \right)^2 \right) = \\
&= \exp \left(\left(\left(\sum_{i=1}^n t_i \right) - n \cdot \delta + t \right) \cdot \mu + \frac{1}{2} \cdot \sigma^2 \cdot \sum_{i=1}^n \left(t_i - \frac{n \cdot \delta - t}{n} \right)^2 \right) = \\
&\quad \boxed{\delta = \frac{1}{n} \cdot \sum_{i=1}^n t_i \Rightarrow \sum_{i=1}^n t_i = \delta \cdot n} \quad \uparrow \\
&= \exp \left(\mu \cdot t + \frac{\sigma^2}{2 \cdot n^2} \cdot \sum_{i=1}^n (t + n \cdot (t_i - \delta))^2 \right) = \\
&= \exp \left(\mu \cdot t + \frac{\sigma^2}{2 \cdot n^2} \cdot \sum_{i=1}^n (t^2 + n^2 \cdot (t_i - \delta)^2 + 2 \cdot t \cdot n \cdot (t_i - \delta)) \right) = \\
&= \exp \left(\mu \cdot t + \frac{\sigma^2}{2 \cdot n^2} \cdot \left(n \cdot t^2 + n^2 \cdot \left(\sum_{i=1}^n (t_i - \delta)^2 \right) + 2 \cdot t \cdot n \cdot \left(\sum_{i=1}^n (t_i - \delta) \right) \right) \right) = \\
&= \exp \left(\mu \cdot t + \frac{\sigma^2}{2 \cdot n^2} \cdot \left(n \cdot t^2 + n^2 \cdot \left(\sum_{i=1}^n (t_i - \delta)^2 \right) + 2 \cdot t \cdot n \cdot \left(\left(\sum_{i=1}^n t_i \right) - n \cdot \delta \right) \right) \right) = \\
&\quad \boxed{\delta = \frac{1}{n} \cdot \sum_{i=1}^n t_i \Rightarrow \left(\sum_{i=1}^n t_i \right) - n \cdot \delta = 0} \quad \uparrow \\
&= \exp \left(\mu \cdot t + \frac{\sigma^2}{2 \cdot n^2} \cdot \left(n \cdot t^2 + n^2 \cdot \left(\sum_{i=1}^n (t_i - \delta)^2 \right) \right) \right) = \\
&= \exp \left(\mu \cdot t + \frac{1}{2} \cdot \frac{\sigma^2}{n} \cdot t^2 \right) \cdot \exp \left(\frac{\sigma^2}{2} \cdot \left(\sum_{i=1}^n (t_i - \delta)^2 \right) \right) \quad \uparrow = \Phi_1(t) \cdot \Phi_2(t_1; \dots; t_n) \\
&\quad \boxed{\Phi_1(t) = \exp \left(\mu \cdot t + \frac{1}{2} \cdot \frac{\sigma^2}{n} \cdot t^2 \right) ; \Phi_2(t_1; \dots; t_n) = \exp \left(\frac{\sigma^2}{2} \cdot \left(\sum_{i=1}^n (t_i - \delta)^2 \right) \right)}
\end{aligned}$$

Como $\Phi_1(t)$ es la f.g.m de $\bar{X} \approx N(\mu; \sigma/\sqrt{n})$, entonces $\Phi_2(t_1; \dots; t_n)$ es necesariamente la f.g.m de la variable n-dimensional $(X_1 - \bar{X}; \dots; X_n - \bar{X})$. Como la f.g.m. de la variable $(\bar{X}; X_1 - \bar{X}; \dots; X_n - \bar{X})$ es el producto de las respectivas f.g.m de las variables \bar{X} y $(X_1 - \bar{X}; \dots; X_n - \bar{X})$, éstas son independientes; por tanto, también son independientes \bar{X} y S^2 , siendo:

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

• Distribución de $n.S^2/\sigma^2$

En el muestreo aleatorio simple de tamaño "n" de una variable $X \approx N(\mu; \sigma)$, el estadístico $n.S^2/\sigma^2$ tiene distribución χ^2_{n-1} .

En efecto, al calcular la media del estadístico S^2 vimos que:

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \left(\sum_{i=1}^n (X_i - \mu)^2 \right) - n.(\bar{X} - \mu)^2$$

Por tanto, es $\sum_{i=1}^n (X_i - \mu)^2 = \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) + n.(\bar{X} - \mu)^2$; y así:

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \left(\frac{1}{\sigma^2} \cdot \sum_{i=1}^n (X_i - \bar{X})^2 \right) + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \Rightarrow$$

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2 \Rightarrow \sum_{i=1}^n (X_i - \bar{X})^2 = n.S^2$$

$$\Rightarrow \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{n.S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \Rightarrow \frac{n.S^2}{\sigma^2} \approx \chi^2_{n-1}$$

$$* X_i \approx N(\mu; \sigma) \Rightarrow \frac{X_i - \mu}{\sigma} \approx N(0;1) \Rightarrow \left(\frac{X_i - \mu}{\sigma} \right)^2 \approx \chi^2_1 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \approx \chi^2_n$$

pues X_1, \dots, X_n son independientes

$$* \bar{X} \approx N(\mu; \sigma/\sqrt{n}) \Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0;1) \Rightarrow \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \approx \chi^2_1$$

* Como $\frac{n.S^2}{\sigma^2}$ y $\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \approx \chi^2_1$ son independientes (por serlo

S^2 y \bar{X}), si su suma tiene distribución χ^2_n , ha de ser $\frac{n.S^2}{\sigma^2} \approx \chi^2_{n-1}$



• Distribución de media muestral si no se conoce σ

Siendo independientes $\bar{X} \approx N(\mu; \sigma/\sqrt{n})$ y S^2 , también son independientes las variables $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0;1)$ y $\frac{n.S^2}{\sigma^2} \approx \chi_{n-1}^2$; por tanto, recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable $Pepa \approx N(0;1)$ y la raíz cuadrada de una variable $Juana \approx \chi_k^2$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), resulta que:

$$Z = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{n.S^2/\sigma^2}{n-1}}} = \frac{\bar{X} - \mu}{S/\sqrt{n-1}} \approx t_{n-1}$$

Siendo $S_c^2 = n.S^2/(n-1)$, es $S/\sqrt{n-1} = S_c/\sqrt{n}$; por tanto, también es:

$$Z = \frac{\bar{X} - \mu}{S_c/\sqrt{n}} \approx t_{n-1}$$

• Distribución de la diferencia entre las medias muestrales de sendas poblaciones normales

Considera que de las variables independientes $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$ se toman muestras de tamaños respectivos "n" y "m", siendo \bar{X}_1 y \bar{X}_2 las respectivas medias muestrales y S_1^2 y S_2^2 las varianzas muestrales

◆ Varianzas poblacionales conocidas

Siendo $\bar{X}_1 \approx N(\mu_1; \sigma_1/\sqrt{n})$ y $\bar{X}_2 \approx N(\mu_2; \sigma_2/\sqrt{m})$, es:

$$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2; \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

◆ Varianzas poblacionales no conocidas e iguales

Siendo $\bar{X}_1 \approx N(\mu_1; \sigma/\sqrt{n})$ y $\bar{X}_2 \approx N(\mu_2; \sigma/\sqrt{m})$, es:

$$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2; \sigma \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

Además $n.S_1^2/\sigma^2 \approx \chi_{n-1}^2$ y $m.S_2^2/\sigma^2 \approx \chi_{m-1}^2$, y siendo éstas independientes, es:

$$\frac{n.S_1^2}{\sigma^2} + \frac{m.S_2^2}{\sigma^2} = \frac{n.S_1^2 + m.S_2^2}{\sigma^2} \approx \chi_{n+m-2}^2$$

Recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable $Pepa \approx N(0;1)$ y la raíz cuadrada de una variable $Juana \approx \chi_k^2$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si consideramos que

$$Pepa = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \approx N(0;1)$$

$$Juana = \frac{n \cdot S_1^2 + m \cdot S_2^2}{\sigma^2} \approx \chi_{n+m-2}^2$$

es:

$$Z = \frac{\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{(n \cdot S_1^2 + m \cdot S_2^2) / \sigma^2}{n + m - 2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n \cdot S_1^2 + m \cdot S_2^2}{n + m - 2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \approx t_{n+m-2}$$

♦ Varianzas poblacionales desconocidas y distintas, con muestras grandes

Si las muestras son grandes (tamaño ≥ 30), las varianzas muestrales son muy buenas estimaciones de las varianzas poblacionales; por tanto, considerando que $\sigma_1^2 = S_1^2$ y que $\sigma_2^2 = S_2^2$, sucede que:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \approx N(0;1)$$

♦ Varianzas poblacionales desconocidas y distintas, con muestras pequeñas

Para muestra pequeñas, es:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \approx t_{n^*}$$

donde n^* es el entero más cercano a la solución de la ecuación

$$\frac{1}{n^*} = \frac{c^2}{n-1} + \frac{(1-c)^2}{m-1}$$

siendo:

$$c = \frac{S_1^2 / (n-1)}{\frac{S_1^2}{n-1} + \frac{S_2^2}{m-1}}$$

• Distribución del cociente entre las varianzas muestrales de sendas poblaciones normales

Sean las variables independientes $X \approx N(\mu_1; \sigma_1)$ e $Y \approx N(\mu_2; \sigma_2)$ de las que se toman muestras de tamaños respectivos "n" y "m", siendo S_1^2 y S_2^2 las respectivas varianzas muestrales.

Recordando que la variable F-Snedecor con r;s grados de libertad es el cociente entre una variable $Pepa \approx \chi_r^2$ y una variable $Juana \approx \chi_s^2$ dividida cada una por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si $Pepa = n.S_1^2/\sigma_1^2 \approx \chi_{n-1}^2$ y $Juana = m.S_2^2/\sigma_2^2 \approx \chi_{m-1}^2$, entonces:

$$Z = \frac{\frac{n.S_1^2/\sigma_1^2}{n-1}}{\frac{m.S_2^2/\sigma_2^2}{m-1}} = \frac{n.(m-1).S_1^2}{m.(n-1).S_2^2} \approx F_{n-1;m-1} \Rightarrow$$

$$S_{c1}^2 = \frac{n.S_1^2}{n-1}; S_{c2}^2 = \frac{m.S_2^2}{m-1}$$

$$\Rightarrow Z = \frac{\sigma_2^2.S_{c1}^2}{\sigma_1^2.S_{c2}^2} \approx F_{n-1;m-1}$$

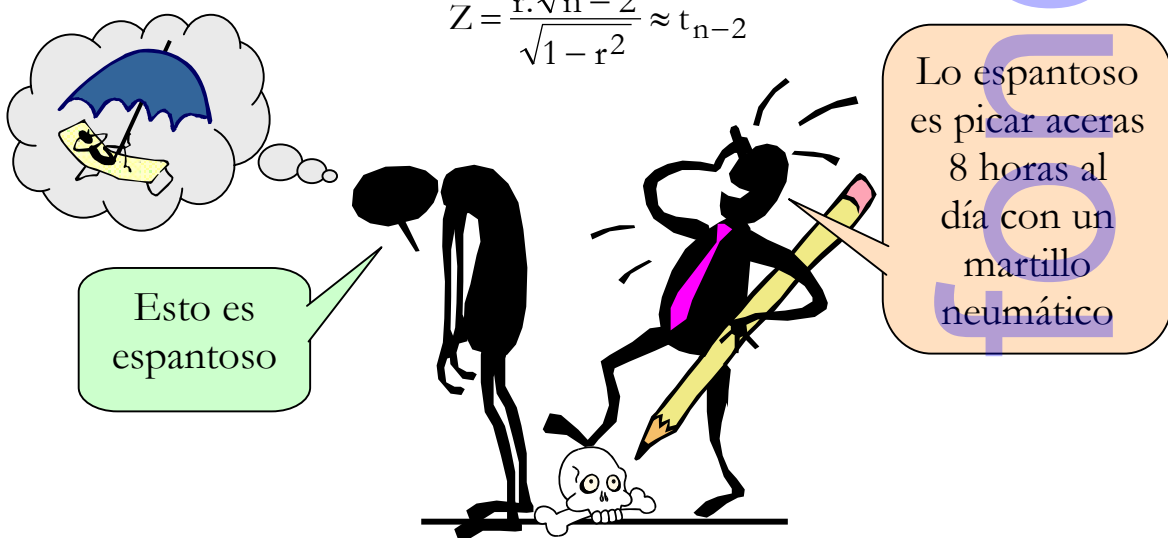
• Distribución del coeficiente de correlación muestral de sendas poblaciones normales

Siendo "r" el coeficiente de correlación muestral correspondiente a una m.a.s de tamaño "n" de la población normal bidimensional (X;Y) con coeficiente de correlación $\rho = \text{COV.}(X;Y)/\sqrt{V(X).V(Y)}$, según Fisher, es:

$$W = \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \approx N\left(\frac{1}{2} \cdot \text{Ln} \frac{1+\rho}{1-\rho}; \frac{1}{\sqrt{n-3}}\right)$$

Si $\rho = 0$ (\Leftrightarrow "X" e "Y" son independientes; único caso de \Leftrightarrow), entonces:

$$Z = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \approx t_{n-2}$$



FONEMATO 2.6.1

Una máquina de refrescos está regulada de modo que la cantidad de bebida que suministra tiene distribución normal con media 22'5 dl. y desviación típica 1'5 dl.

Determinéase la probabilidad de que al obtener una muestra de 36 bebidas, el contenido medio supere los 24 dl.

SOLUCIÓN

Si $X \approx N(22'5; 1'5)$ expresa la cantidad de bebida que suministra la máquina, siendo $(X_1; \dots; X_{36})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 36, el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_{36})/36$, que tiene distribución normal, por ser **combinación lineal de normales independientes**, siendo:

$$E(\bar{X}) = E(X) = 22'5 ; V(\bar{X}) = V(X)/36 = 1'5^2/36$$

Por tanto:

$$P(\bar{X} > 24) = P(N(22'5; 1'6/6) > 24) = P\left(N(0;1) > \frac{24 - 22'5}{1'6/6}\right) = \dots$$

EL PROFESOR QUE CORRIGE EL EXAMEN

Todos los profesores son de la misma opinión: corregir exámenes no gusta a nadie, no es trabajo agradable enfrentarse por n-ésima vez a la tarea de leer y puntuar un montón de folios escritos por principiantes que en la mayoría de los casos no tienen ni idea y sólo escriben barbaridades y estupideces sobre el asunto de sota, caballo y rey que por j-ésima vez cae en examen.

Por eso, **cuando un profe se sienta a corregir exámenes no suele estar de buen humor.**

Así las cosas, lo que escribamos o dibujemos en

examen debe **diferenciarnos positivamente** de los demás, y para conseguir tal diferenciación y que **al profe se le caigan los pantalones**, basta **escribir o dibujar pensando que no se lo sabe y por tanto hay que llevarle de la mano**, explicándole todos los aspectos más relevantes de las **conexiones neuronales** que establezcamos en cada caso.



FONEMATO 2.6.2

De una población normal de varianza 1000 se toma m.a.s. de tamaño 100. Détermínese la probabilidad de que la media muestral supere a la media poblacional en no menos de 0'2.

SOLUCIÓN

Siendo $X \approx N(\mu; 10)$ y $(X_1; \dots; X_{100})$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño 100, el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_n)/n$.

La distribución de probabilidad de \bar{X} es normal, pues \bar{X} es combinación lineal de normales independientes, siendo:

$$E(\bar{X}) = E(X) = \mu ; V(\bar{X}) = V(X)/100 = 1.$$

Así:

$$P(\bar{X} - \mu \geq 0'2) = P\left(N(0;1) \geq \frac{0'2}{1}\right) = \dots$$



FONEMATO 2.6.3

Siendo $X \approx N(100;12)$, determínese el tamaño muestral mínimo que debe tomarse para garantizar, con probabilidad no inferior a 0'95, que la media muestral no difiera de la media poblacional en más de 2 unidades.

SOLUCIÓN

Siendo $X \approx N(100;12)$ y $(X_1; \dots; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_n)/n$.

La distribución de probabilidad de \bar{X} es normal, por ser combinación lineal de normales independientes, siendo

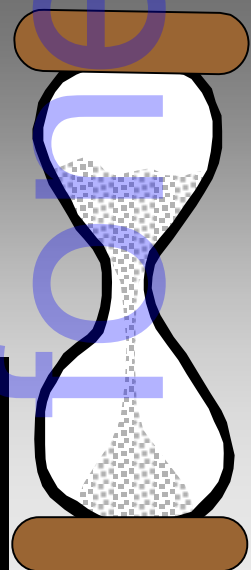
$$E(\bar{X}) = E(X) = 100 ; V(\bar{X}) = V(X)/n = 144/n.$$

Debemos determinar "n" de modo que $P(|\bar{X} - 100| \leq 2) \geq 0'95$:

$$\begin{aligned} P(|\bar{X} - 100| \leq 2) &\geq 0'95 \Rightarrow \\ \Rightarrow P\left(|N(0;1)| \leq \frac{2}{12/\sqrt{n}}\right) &\geq 0'95 \Rightarrow \\ \Rightarrow 2 \cdot P\left(N(0;1) \leq \frac{2}{12/\sqrt{n}}\right) - 1 &\geq 0'95 \Rightarrow \\ \Rightarrow P\left(N(0;1) \leq \frac{2}{12/\sqrt{n}}\right) &\geq 0'975 \Rightarrow \\ \Rightarrow \frac{2}{12/\sqrt{n}} &\geq 1'96 \Rightarrow \\ \Rightarrow n &\geq 138'29 \Rightarrow \\ \Rightarrow n &\geq 139 \end{aligned}$$

En el año 1967, el Sistema Internacional de Unidades estableció que un segundo equivale a 9.192.631.770 periodos de radiación de la transición entre los dos niveles hiperfinos del isótopo 133 del átomo de cesio (^{133}Cs), medidos a cero grados kelvin.

No es verdad que para ti tu tiempo sea oro, es mucho más: el oro lo venden en todas partes, pero tu tiempo no puedes comprarlo ni con todo el oro del mundo.



FONEMATO 2.6.4

Siendo $X \approx N(\mu;1)$, determínese el tamaño muestral mínimo que debe tomarse para garantizar, con probabilidad no inferior a 0'95, que la media muestral no diferirá de la media poblacional en más de 0'5 unidades.

SOLUCIÓN

Siendo $X \approx N(\mu;1)$ y $(X_1; \dots; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_n)/n$.

La distribución de probabilidad de \bar{X} es normal, pues \bar{X} es combinación lineal de normales independientes, siendo:

$$E(\bar{X}) = E(X) = \mu ; V(\bar{X}) = V(X)/n = 1/n.$$

Debemos determinar "n" de modo que $P(|\bar{X} - \mu| \leq 0'5) \geq 0'95$:

$$\begin{aligned} P(|\bar{X} - \mu| \leq 0'5) &\geq 0'95 \Rightarrow \\ \Rightarrow P\left(|N(0;1)| \leq \frac{0'5}{1/\sqrt{n}}\right) &\geq 0'95 \Rightarrow \\ \Rightarrow 2 \cdot P\left(N(0;1) \leq \frac{0'5}{1/\sqrt{n}}\right) - 1 &\geq 0'95 \Rightarrow \\ \Rightarrow P\left(N(0;1) \leq \frac{0'5}{1/\sqrt{n}}\right) &\geq 0'975 \Rightarrow \frac{0'5}{1/\sqrt{n}} \geq 1'96 \Rightarrow \\ \Rightarrow n &\geq 15'21 \Rightarrow n \geq 16 \end{aligned}$$



FONEMATO 2.6.5

En horas, la duración de un tipo de bombillas tiene distribución $N(1000;100)$. Se desea enviar una muestra de modo que la duración media muestral no difiera de la poblacional en más de 50 horas con probabilidad no inferior a 0'95.

- 1) Determinése el tamaño de la muestra.
- 2) Determinése el tamaño de la muestra si de "X" sólo se sabe que tiene media 1000 y desviación típica 100.

SOLUCIÓN

Siendo $X \approx N(1000;100)$ y $(X_1; \dots; X_n)$ la variable muestral que expresa los resultados que pueden presentarse al tomar muestra aleatoria simple de tamaño "n", el estadístico media muestral es $\bar{X} = (X_1 + \dots + X_n)/n$.

La distribución de probabilidad de \bar{X} es normal, pues \bar{X} es combinación lineal de normales independientes, siendo:

$$E(\bar{X}) = E(X) = 1000 ; V(\bar{X}) = V(X)/n = 100^2/n.$$

Hay que determinar "n" de modo que $P(|\bar{X} - 1000| \leq 50) \geq 0'95$:

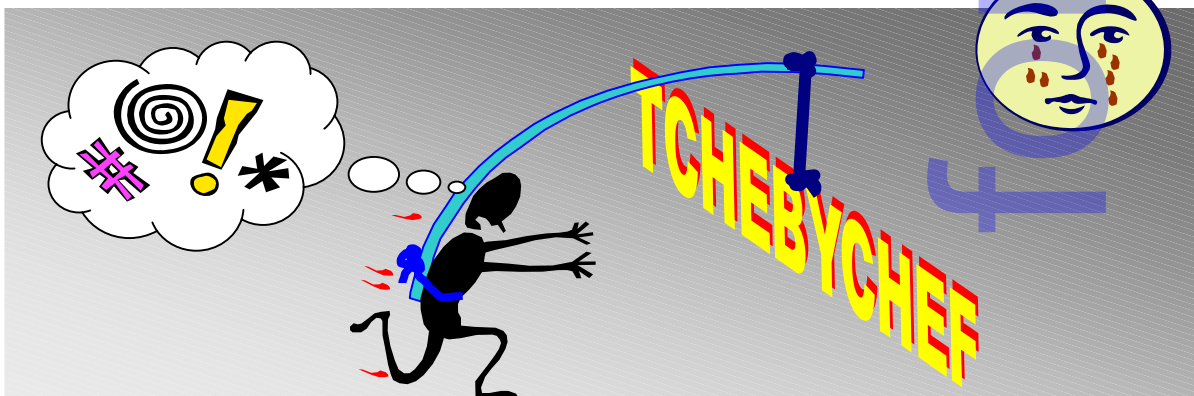
$$P(|\bar{X} - 1000| \leq 50) \geq 0'95 \Rightarrow P\left(|N(0;1)| \leq \frac{50}{100/\sqrt{n}}\right) \geq 0'95 \Rightarrow$$

$$P(|N(0;1)| \leq a) = 2.P(N(0;1) \leq a) - 1$$

$$\begin{aligned} \Rightarrow 2.P\left(N(0;1) \leq \frac{50}{100/\sqrt{n}}\right) - 1 &\geq 0'95 \Rightarrow P\left(N(0;1) \leq \frac{50}{100/\sqrt{n}}\right) \geq 0'975 \Rightarrow \\ \Rightarrow \frac{50}{100/\sqrt{n}} &\geq 1'96 \Rightarrow n \geq 15'36 \Rightarrow n \geq 16 \end{aligned}$$

- 2) Si $E(\bar{X}) = E(X) = 1000$ y $V(\bar{X}) = V(X)/n = 100^2/n$, como desconocemos la distribución de probabilidad de "X", lidiamos mediante **Tchebychef**:

$$\begin{aligned} P(|\bar{X} - 1000| \leq 50) &= P(|\bar{X} - E(\bar{X})| \leq 50) \geq 1 - \frac{V(\bar{X})}{50^2} = \\ &= 1 - \frac{100^2/n}{50^2} \geq 0'95 \Rightarrow n \geq 80 \end{aligned}$$



FONEMATO 2.6.6

De una población $X \approx N(5; 0'1)$ se toma m.a.s de tamaño 16.

Determinése la probabilidad del suceso $5 < \bar{X} < 5'2$, la del suceso $S^2 \leq 0'019125$, la del suceso $S > 0'1$ y la del suceso $5 < \bar{X} < 5'2 / S^2 \leq 0'019125$.

SOLUCIÓN

En el muestreo aleatorio simple de tamaño "n" de una población $N(\mu; \sigma)$, siendo

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i ; S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$$

es:

$$\bar{X} \approx N(\mu; \sigma/\sqrt{n}) ; n.S^2/\sigma^2 \approx \chi_{n-1}^2$$

Así, si $n = 16$, $\mu = 5$ y $\sigma = 0'1$, es:

$$\bar{X} \approx N(5; 0'1/4) ; 16.S^2/0'1^2 \approx \chi_{15}^2$$

- $P(5 < \bar{X} < 5'2) = P(5 < N(5; 0'1/4) < 5'2) = \dots$
- $P(S^2 \leq 0'019125) = P(16.S^2/0'1^2 \leq 16.0'019125/0'1^2) = P(\chi_{15}^2 \leq 30'6) = 0'99$
- $P(S > 0'1) = P(S^2 > 0'1^2) = P(16.S^2/0'1^2 > 16.0'1^2/0'1^2) = P(\chi_{15}^2 > 16) = \dots$
- $P(5 < \bar{X} < 5'2 / S^2 \leq 0'019125) = P(5 < \bar{X} < 5'2)$

en el muestreo de una población normal, los estadísticos media muestral y varianza muestral son independientes



FONEMATO 2.6.7

De una población $X \approx N(13;6)$ se toma m.a.s de tamaño 9.

- 1) Determinése un intervalo en el que puede esperarse, con probabilidad 0'99, que se encuentre la media muestral.
- 2) Repita el apartado 1) si $n = 100$ y si $n = 10000$.
- 3) ¿Qué diría si en una m.a.s. de tamaño 10000 la media muestral es 15?

SOLUCIÓN

- 1) Es sabido que en el muestreo aleatorio simple de tamaño "n" de una población $X \approx N(\mu; \sigma)$, el estadístico media muestral \bar{X} tiene distribución $N(\mu; \sigma/\sqrt{n})$.

Por tanto, si $n = 9$ y $X \approx N(13;6)$, es $\bar{X} \approx N(13; 6/\sqrt{9}) \equiv N(13;2)$.

En la distribución de probabilidad de $\bar{X} \approx N(13;2)$ hay infinitud de intervalos que "encierran" masa 0'99, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, **como la distribución $N(13;2)$ es simétrica respecto al punto 13, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto** (puede demostrarse que éste es el de menor amplitud). Por tanto,

debemos determinar $c > 0$ de modo que $P(|\bar{X} - 13| < c) = 0'99$

$$\begin{aligned} P(|\bar{X} - 13| < c) &= 0'99 \Rightarrow P(|N(0;1)| < \frac{c}{2}) = 0'99 \Rightarrow \\ \Rightarrow 2 \cdot P(N(0;1) < \frac{c}{2}) - 1 &= 0'99 \Rightarrow P(N(0;1) < \frac{c}{2}) = 0'995 \Rightarrow \\ \Rightarrow \frac{c}{2} &= 2'57 \Rightarrow c = 5'14 \end{aligned}$$

Así, tomando muestras de tamaño 9 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico media muestral \bar{X} se concretará en un punto del intervalo $(13 - 5'14; 13 + 5'14) \equiv (7'86; 18'14)$.

- 2) Si $n = 100$ entonces $\bar{X} \approx N(13; 6/\sqrt{100}) \equiv N(13; 0'6)$; por tanto:

$$\begin{aligned} P(|\bar{X} - 13| < c) &= 0'99 \Rightarrow P(|N(0;1)| < \frac{c}{0'6}) = 0'99 \Rightarrow \\ \Rightarrow \frac{c}{0'6} &= 2'57 \Rightarrow c = 1'54 \end{aligned}$$

Así, tomando muestras de tamaño 100 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico muestral \bar{X} se concretará en un punto del intervalo $(13 - 1'54; 13 + 1'54) \equiv (11'46; 14'54)$.

- Si $n = 10000$ entonces $\bar{X} \approx N(13; 6/\sqrt{10000}) \equiv N(13; 0'06)$; por tanto:

$$P(|\bar{X} - 13| < c) = 0'99 \Rightarrow P(|N(0;1)| < \frac{c}{0'06}) = 0'99 \Rightarrow \frac{c}{0'06} = 2'57 \Rightarrow c = 0'154$$

Así, tomando muestras de tamaño 10000 de la población $X \approx N(13;6)$, el 99 % de las veces ocurrirá que el estadístico muestral \bar{X} se concretará en un punto del intervalo $(13 - 0'154; 13 + 0'154) \equiv (12'846; 13'154)$.

3) **Pregunta:** ¿qué dirías si en una m.a.s. de tamaño 10000 se obtiene $\bar{X} = 15$?

Respuesta: depende

Pregunta: ¿de qué depende?

Respuesta: entre el blanco y el negro que describimos a continuación hay grises que sin duda el lector puede imaginar.

El blanco: tranquilidad total

Si estamos podridos de dinero y podemos tomar muestras de tamaño 10000 constantemente (por ejemplo, una muestra cada cinco minutos), estaremos felices mientras sea mayor o igual que 0'99 la proporción de muestras en que la media muestral \bar{X} se concreta en algún punto del intervalo (12'846;13'154). Mientras estemos felices, el que en una muestra concreta (la de las 13'35) suceda que $\bar{X} = 15 \notin (12'846;13'154)$ no nos preocupará en absoluto: lo achacaremos a causas aleatorias por las que no merece la pena obsesionarse, pues es lógico y normal que, si tomamos muchas muestras, algunas sean "raras" (en este ejemplo, si $n = 10000$, las muestras "raras" son aquéllas en que la media muestral no toma un valor en el intervalo (12'846;13'154)) pero estaremos tranquilos mientras la proporción de muestras "raras" no supere el 1 %.

El negro: mosqueo total

Si somos unos pobretones y para pagar la m.a.s. de tamaño 10000 hemos tenido que romper nuestra hucha cerdito de barro y empeñar el oso de peluche, entonces, si ocurre que $\bar{X} = 15 \notin (12'846;13'154)$ nos pegamos un tiro, pues el que haya sucedido tal cosa indica que algo falla estrepitosamente o que somos cenizos.

- **Falla la tracción:** los que han tomado la muestra estaban borrachos, o son tan inútiles que no saben tomar una m.a.s, o están comprados por la "competencia" y deliberadamente nos suministran datos falsos, o
- **Falla la información:** el responsable del desastre es el sinvergüenza que nos ha dicho que $X \approx N(13;6)$ no es cierto que $X \approx N(13;6)$.
- **Somos gafes:** la tracción hace bien su trabajo y es cierto que $X \approx N(13;6)$, pero hemos tenido la desgracia de seleccionar una muestra "rara".

Debes saber que este tipo de triángulos carecen de solución: si nos obsesionamos en investigar qué ha pasado, los encargados de la "tracción" dirán que "ellos" no han sido, el que nos ha dicho que $X \approx N(13;6)$ nos jurará por Snoopy que "su" información es buena y como la muestra no puede hablar, no habrá más remedio que pedir un crédito al banco y seleccionar otra muestra, o suicidarnos.

Observa: si en la muestra se obtuviera $\bar{X} = 12'83 \notin (12'846;13'154)$, diríamos lo mismo que si $\bar{X} = 15 \notin (12'846;13'154)$, pues $\bar{X} = 12'83$ también corresponde a una muestra "rara". No obstante, la media muestral 12'83 deja de ser "rara" si sólo consideramos "raras" las muestras en que $\bar{X} \notin (11'46;14'54)$, que es el intervalo que corresponde a $n = 100$; lo que no sucede con una media muestral 15, que sigue siendo "rara" para $n = 100$.

Tema 2: Distribuciones en el muestreo

© Rafael Cabrejas Hernansanz

FONEMATO 2.6.8

Si de una población $X \approx N(17;4)$ se toma m.a.s de tamaño 10, determínese la cota que superará la desviación típica muestral con probabilidad 0'99.

SOLUCIÓN

En el m.a.s de tamaño "n" de $X \approx N(\mu;\sigma)$, siendo S^2 el estadístico varianza muestral, es $n.S^2/\sigma^2 \approx \chi_{n-1}^2$. Por tanto, si $n=10$ y $\sigma=4$, es $10.S^2/4^2 \approx \chi_9^2$, y buscamos "c" ($c > 0$) tal que $P(S > c) = 0'99$:

$$\begin{aligned} P(S > c) = 0'99 &\Rightarrow P(S^2 > c^2) = 0'99 \Rightarrow P(10.S^2/4^2 > 10.c^2/4^2) = 0'99 \Rightarrow \\ &\Rightarrow P(\chi_9^2 > 10.c^2/4^2) = 0'99 \Rightarrow 10.c^2/4^2 = 2'09 \Rightarrow c = 1'82 \end{aligned}$$



¿TE GUSTAN LAS FLORES?
Con los **números** el cerebro
se puebla de las más
hermosas florecillas:
amapolas de
disciplina mental,
corazoncillos de
capacidad de abstracción
y razonamiento,
maripepas de **rigor,**
alarises de **seriedad,**
azaleas de **profesionalidad,**
amarantos de **mejora**
permanente
y tomasillos de **gusto por**
el trabajo bien hecho...
o sea, todo ventajas.

FONEMATO 2.6.9

En el muestreo aleatorio simple de tamaño "n" de una población normal, determínese la función de densidad de probabilidad del estadístico varianza muestral.

SOLUCIÓN

Siendo "Z" el estadístico varianza muestral, en el m.a.s de tamaño "n" de una población $N(\mu; \sigma)$, es $W = n.Z/\sigma^2 \approx \chi_{n-1}^2 \equiv G((n-1)/2; 1/2)$.

- Función de distribución de la varianza muestral "Z":

$$F_Z(z) = P(Z \leq z) = P(\sigma^2.W/n \leq z) = P(W \leq n.z/\sigma^2) = F_W(n.z/\sigma^2)$$

$$W = n.Z/\sigma^2 \Rightarrow Z = \sigma^2.W/n$$

siendo F_W la función de distribución de $W \approx \chi_{n-1}^2$

- Función de densidad de probabilidad de la varianza muestral "Z":

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{dF_W(n.z/\sigma^2)}{dz} = \frac{dF_W(n.z/\sigma^2)}{d(n.z/\sigma^2)} \cdot \frac{d(n.z/\sigma^2)}{dz} =$$

siendo f_W la función de distribución de $W \approx \chi_{n-1}^2$, es:

VENTANA

$$\frac{dF_W(n.z/\sigma^2)}{d(n.z/\sigma^2)} = f_W(n.z/\sigma^2)$$

$$= \frac{n}{\sigma^2} \cdot f_W(n.z/\sigma^2) = \frac{n}{\sigma^2} \cdot \frac{(1/2)^{(n-1)/2}}{\Gamma((n-1)/2)} \cdot (n.z/\sigma^2)^{(n-3)/2} \cdot e^{-n.z/2\sigma^2} =$$

$$W \approx \chi_{n-1}^2 \equiv G\left(\frac{n-1}{2}; \frac{1}{2}\right) \Rightarrow f_W(w) = \frac{(1/2)^{(n-1)/2}}{\Gamma((n-1)/2)} \cdot w^{(n-3)/2} \cdot e^{-w/2}, w > 0$$

$$= \frac{n}{\sigma^2} \cdot \frac{(1/2 \cdot \sigma^2)^{(n-1)/2}}{\Gamma((n-1)/2)} \cdot z^{(n-1)/2} \cdot e^{-n.z/2\sigma^2}, z > 0$$

que corresponde a una variable con distribución $G((n-1)/2; n/2\sigma^2)$.



FONEMATO 2.6.10

En una m.a.s. X_1, \dots, X_n de una población $N(\mu; \sigma)$, siendo \bar{X} y S^2 las correspondientes media y varianza muestrales, y siendo $X_{n+1} \approx N(\mu; \sigma)$ independiente de X_1, \dots, X_n , determínese la distribución de probabilidad de "T", siendo:

$$T = \frac{X_{n+1} - \bar{X}}{S} \cdot \sqrt{\frac{n-1}{n+1}}$$

SOLUCIÓN

En el m.a.s de tamaño "n" de una población $N(\mu; \sigma)$ sucede que:

$$\bar{X} \approx N(\mu; \sigma/\sqrt{n}) ; n.S^2/\sigma^2 \approx \chi_{n-1}^2$$

Como las variables X_{n+1} y \bar{X} son independientes (pues X_{n+1} es independiente de X_1, \dots, X_n), la distribución de probabilidad de $X_{n+1} - \bar{X}$ es normal, siendo:

$$E(X_{n+1} - \bar{X}) = E(X_{n+1}) - E(\bar{X}) = \mu - \mu = 0$$

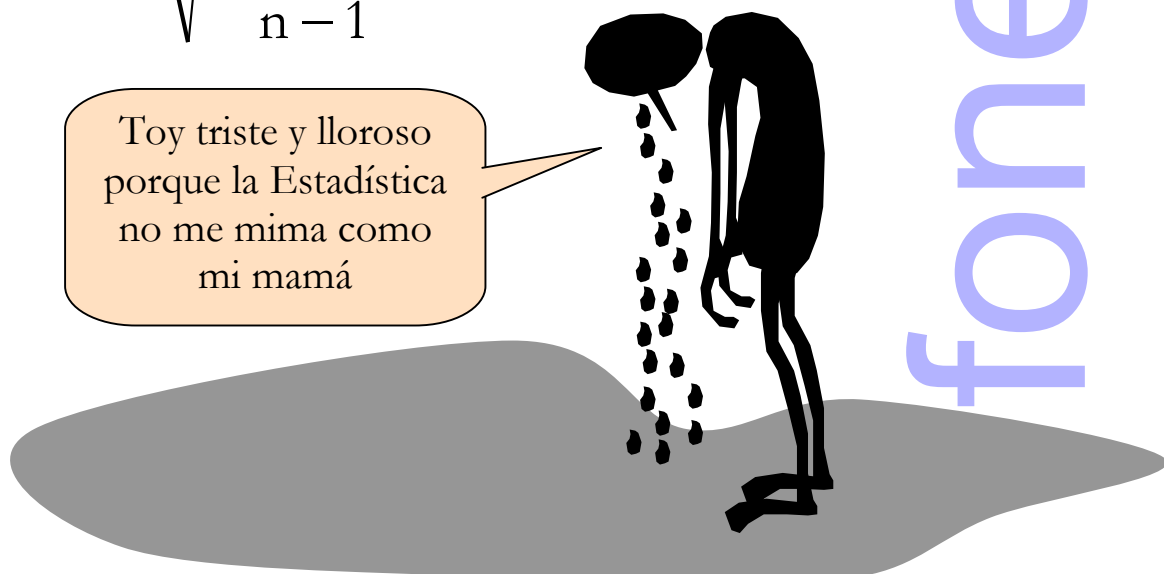
$$V(X_{n+1} - \bar{X}) = V(X_{n+1}) + V(\bar{X}) = \sigma^2 + \frac{\sigma^2}{n} = \frac{n+1}{n} \cdot \sigma^2$$

Por tanto, siendo $X_{n+1} - \bar{X} \approx N(0; \sigma \cdot \sqrt{(n+1)/n})$, es $\frac{X_{n+1} - \bar{X}}{\sigma \cdot \sqrt{(n+1)/n}} \approx N(0;1)$.

Recordando que siendo independientes $Y \approx N(0;1)$ y $U \approx \chi_k^2$ es $\frac{Y}{\sqrt{U/k}} \approx t_k$, entonces,

tomando como $N(0;1)$ la variable $\frac{X_{n+1} - \bar{X}}{\sigma \cdot \sqrt{(n+1)/n}}$ y recordando que $n.S^2/\sigma^2 \approx \chi_{n-1}^2$, resulta evidente que:

$$T = \frac{\frac{X_{n+1} - \bar{X}}{\sigma \cdot \sqrt{(n+1)/n}}}{\sqrt{\frac{n.S^2/\sigma^2}{n-1}}} = \frac{X_{n+1} - \bar{X}}{S} \cdot \sqrt{\frac{n-1}{n+1}} \approx t_{n-1}$$



FONEMATO 2.6.11

De una población $X \approx N(\mu; 2)$ se toma m.a.s de tamaño 9. Determinése un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre la varianza muestral. Repita lo anterior si $n = 101$.

SOLUCIÓN

En el m.a.s de tamaño "n" de una población $N(\mu; \sigma)$ es $n.S^2/\sigma^2 \approx \chi_{n-1}^2$.

- Siendo $n = 9$ y $\sigma = 2$, es $9.S^2/4 \approx \chi_8^2$. En la distribución de probabilidad de la variable $9.S^2/4 \approx \chi_8^2$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. Como la distribución χ_8^2 no es simétrica respecto a ningún punto, por comodidad, elegimos el intervalo que reparte la probabilidad $1 - 0'95 = 0'05$ en partes iguales entre las dos colas (izquierda y derecha) de dicha χ_8^2 (puede demostrarse que éste no es el de menor amplitud, pero del de menor amplitud es tan petardo de calcular que pasamos de él); y la tabla de la función de distribución de $9.S^2/4 \approx \chi_8^2$ indica que:

$$P(2'18 \leq 9.S^2/4 \approx \chi_8^2 \leq 17'53) = 0'95$$

Por tanto, es:

$$P\left(\frac{2'18.4}{9} \leq S^2 \leq \frac{17'53.4}{9}\right) = 0'95$$

o sea:

$$P(0'96 \leq S^2 \leq 7'79) = 0'95$$

En definitiva, tomando sucesivas muestras de tamaño 9 de una población $X \approx N(\mu; 2)$, el 95 % de las veces sucederá que el estadístico varianza muestral se concreta en algún punto del intervalo (0'96; 7'79).

- Siendo $n = 101$ y $\sigma = 2$, es $101.S^2/4 \approx \chi_{100}^2$. Procediendo como en el caso anterior, la tabla de la función de distribución de $101.S^2/4 \approx \chi_{100}^2$ indica que:

$$P(74'2 \leq 101.S^2/4 \approx \chi_{100}^2 \leq 129'6) = 0'95$$

Por tanto, es $P\left(\frac{74'2.4}{101} \leq S^2 \leq \frac{129'6.4}{101}\right) = 0'95$; o sea:

$$P(2'93 \leq S^2 \leq 5'13) = 0'95$$

En definitiva, tomando sucesivas muestras de tamaño 101 de una población $X \approx N(\mu; 2)$, el 95 % de las veces sucederá que el estadístico varianza muestral se concreta en algún punto del intervalo (2'93; 5'13).

FONEMATO 2.6.12

De una población normal de varianza 4'5 se toman dos muestras independientes de tamaño "n". Determinése "n" de modo que la probabilidad de que las medias muestrales difieran en no más de 2 unidades sea al menos 0'95.

SOLUCIÓN

Siendo $X \approx N(\mu; \sqrt{4'5})$, si $(X_1; \dots; X_n)$ e $(Y_1; \dots; Y_n)$ son las respectivas variables muestrales que expresan los resultados que pueden presentarse al tomar las dos muestras de tamaño "n" indicadas, es:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i \approx N(\mu; \sqrt{4'5/n})$$

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i \approx N(\mu; \sqrt{4'5/n})$$

Por tanto, siendo independientes los estadísticos medias muestrales \bar{X} e \bar{Y} , es:

$$\bar{X} - \bar{Y} \approx N(0; \sqrt{9/n})$$

$$\begin{aligned} E(\bar{X} - \bar{Y}) &= E(\bar{X}) - E(\bar{Y}) = \mu - \mu = 0 \\ V(\bar{X} - \bar{Y}) &= V(\bar{X}) + V(\bar{Y}) = 9/n \end{aligned}$$

Debemos determinar "n" de modo que $P(|\bar{X} - \bar{Y}| \leq 2) \geq 0'95$:

$$\begin{aligned} P(|\bar{X} - \bar{Y}| \leq 2) &= 0'95 \Rightarrow P\left(|N(0;1)| \leq \frac{2}{\sqrt{9/n}}\right) \geq 0'95 \Rightarrow \\ \Rightarrow 2 \cdot P\left(N(0;1) \leq \frac{2}{\sqrt{9/n}}\right) - 1 &\geq 0'95 \Rightarrow P\left(N(0;1) \leq \frac{2}{\sqrt{9/n}}\right) \geq 0'975 \Rightarrow \\ \Rightarrow \frac{2}{\sqrt{9/n}} &\geq 1'96 \Rightarrow n \geq 9 \end{aligned}$$



FONEMATO 2.6.13

De dos poblaciones independientes $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$ se toman muestras de tamaños respectivos "n" y "m", siendo \bar{X} e \bar{Y} las respectivas medias muestrales y S_{c1}^2 y S_{c2}^2 las cuasivarianzas muestrales. Determinése la distribución de probabilidad de los estadísticos T_1 y T_2 :

$$T_1 = \frac{n-1}{\sigma_1^2} \cdot S_{c1}^2 + \frac{m-1}{\sigma_2^2} \cdot S_{c2}^2 \quad ; \quad T_2 = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)}} \cdot \frac{\sqrt{n+m-2}}{\sqrt{T_1}}$$

SOLUCIÓN

Para la muestra de tamaño "n" de la población $N(\mu_1; \sigma_1)$, siendo S_1^2 la correspondiente varianza muestral, es $(n-1) \cdot S_{c1}^2 / \sigma_1^2 \approx \chi_{n-1}^2$, pues $n \cdot S_1^2 / \sigma_1^2 \approx \chi_{n-1}^2$ y $S_1^2 = (n-1) \cdot S_{c1}^2 / n$.

Para la muestra de tamaño "m" de la población $N(\mu_2; \sigma_2)$, siendo S_2^2 la correspondiente varianza muestral, es $(m-1) \cdot S_{c2}^2 / \sigma_2^2 \approx \chi_{m-1}^2$, pues $m \cdot S_2^2 / \sigma_2^2 \approx \chi_{m-1}^2$ y $S_2^2 = (m-1) \cdot S_{c2}^2 / m$.

En consecuencia, debido a la reproductividad de la distribución ji-dos, es:

$$T_1 = \frac{n-1}{\sigma_1^2} \cdot S_{c1}^2 + \frac{m-1}{\sigma_2^2} \cdot S_{c2}^2 \approx \chi_{n+m-2}^2$$

Siendo $\bar{X} \approx N(\mu_1; \sigma_1/\sqrt{n})$ e $\bar{Y} \approx N(\mu_2; \sigma_2/\sqrt{m})$ las respectivas medias muestrales, es $\bar{X} - \bar{Y} \approx N(\mu_1 - \mu_2; \sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)})$; por tanto:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)}} \approx N(0;1)$$

En consecuencia, recordando que la variable t-Student con "k" grados de libertad es el cociente entre una variable Pepa $\approx N(0;1)$ y la raíz cuadrada de una variable Juana $\approx \chi_k^2$ dividida por sus grados de libertad (siendo independientes "Pepa" y "Juana"), si consideramos que

$$\text{Pepa} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)}} \approx N(0;1)$$

$$\text{Juana} = T_1 \approx \chi_{n+m-2}^2$$

es:

$$T_2 = \frac{\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)}}}{\sqrt{\frac{T_1}{n+m-2}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n) + (\sigma_2^2/m)}} \cdot \frac{\sqrt{n+m-2}}{\sqrt{T_1}} \approx t_{n+m-2}$$

FONEMATO 2.6.14

De dos poblaciones independientes con la misma varianza σ^2 se toman muestras de tamaños respectivos "n" y "m", siendo \bar{X} e \bar{Y} las respectivas medias muestrales. Si \bar{W} es la media aritmética de todas las observaciones realizadas, demuéstrese que $V(\bar{W} - \bar{X}) = m \cdot \sigma^2 / (n^2 + n \cdot m)$.

SOLUCIÓN

Siendo $(X_1; \dots; X_n)$ e $(Y_1; \dots; Y_m)$ las respectivas variables muestrales que expresan los resultados que pueden presentarse al tomar las dos muestras indicadas, sean:

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i ; \quad \bar{Y} = \frac{1}{m} \cdot \sum_{j=1}^m Y_j$$

Siendo \bar{W} la media aritmética de todas las observaciones realizadas, es:

$$\bar{W} = \frac{1}{n+m} \cdot \left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j \right)$$

Por tanto:

$$V(\bar{W} - \bar{X}) = V \left(\frac{1}{n+m} \cdot \left(\sum_{j=1}^m Y_j - \frac{m}{n} \cdot \sum_{i=1}^n X_i \right) \right) =$$

VENTANA

$$\begin{aligned} \bar{W} - \bar{X} &= \frac{1}{n+m} \cdot \left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j \right) - \frac{1}{n} \cdot \sum_{i=1}^n X_i = \\ &= \left(\frac{1}{n+m} - \frac{1}{n} \right) \cdot \sum_{i=1}^n X_i + \frac{1}{n+m} \cdot \sum_{j=1}^m Y_j = \frac{1}{n+m} \cdot \left(\sum_{j=1}^m Y_j - \frac{m}{n} \cdot \sum_{i=1}^n X_i \right) \end{aligned}$$

$$\stackrel{\uparrow}{=} \frac{1}{(n+m)^2} \cdot \left(\sum_{j=1}^m V(Y_j) + \frac{m^2}{n^2} \cdot \sum_{i=1}^n V(X_i) \right) =$$

pues $X_1, \dots, X_n, Y_1, \dots, Y_m$ son independientes

$$\stackrel{\uparrow}{=} \frac{1}{(n+m)^2} \cdot \left(m \cdot \sigma^2 + \frac{m^2}{n^2} \cdot n \cdot \sigma^2 \right) = \frac{m \cdot \sigma^2}{n^2 + n \cdot m}$$

las variables $X_1, \dots, X_n, Y_1, \dots, Y_m$ tienen varianza σ^2 ,
como las dos poblaciones muestreadas

FONEMATO 2.6.15

La duración media de las bombillas tipo A es de 2500 horas con desviación típica 600 horas, y la duración media de las bombillas tipo B es de 2300 horas con desviación típica 800 horas. Se toman 300 bombillas tipo A y 200 tipo B.

- 1) Calcúlese la probabilidad de que la duración media de las bombillas tipo A no supere en más de 100 horas a la duración media de la del tipo B.
- 2) Calcúlese la probabilidad de que la duración media de las bombillas tipo A supere en más de 200 horas a la duración media de la del tipo B.
- 3) Determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre la diferencia de medias muestrales.

SOLUCIÓN

Siendo "X" la duración de las bombillas tipo A e "Y" la duración de las del tipo B, se nos dice que $E(X) = 2500$, $V(X) = 600^2$, $E(Y) = 2300$ y $V(Y) = 800^2$.

Siendo \bar{X} e \bar{Y} las respectivas medias muestrales correspondientes a las dos muestras indicadas, **por ser grandes los tamaños muestrales 300 y 200, podemos aproximar (Lindeberg-Levy) la distribución de probabilidad de \bar{X} mediante la normal con media y varianza las de \bar{X} , y la distribución de probabilidad de \bar{Y} mediante la normal con media y varianza las de \bar{Y} , que son:**

$$E(\bar{X}) = E(X) = 2500 ; V(\bar{X}) = V(X)/300 = 600^2/300$$

$$E(\bar{Y}) = E(Y) = 2300 ; V(\bar{Y}) = V(Y)/200 = 800^2/200$$

Como \bar{X} e \bar{Y} son independientes, la distribución de $\bar{X} - \bar{Y}$ es normal, siendo:

$$E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = 2500 - 2300 = 200$$

$$V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = (600^2/300) + (800^2/200) = 4400$$

- 1) $P(\bar{X} - \bar{Y} \leq 100) = P(N(200; \sqrt{4400}) \leq 100) = P\left(N(0;1) \leq \frac{100 - 200}{\sqrt{4400}}\right) = \dots$
- 2) $P(\bar{X} - \bar{Y} > 200) = P(N(200; \sqrt{4400}) > 200) = P\left(N(0;1) > \frac{200 - 200}{\sqrt{4400}}\right) = 0'5$
- 3) En la distribución de probabilidad de $\bar{X} - \bar{Y} \approx N(200; \sqrt{4400})$ hay infinidad de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, **como la variable $N(200; \sqrt{4400})$ es simétrica respecto al punto 200, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto, que es el de menor amplitud.** Así, debemos determinar $c > 0$ de modo que $P(|\bar{X} - \bar{Y} - 200| < c) = 0'95$:

$$\begin{aligned} P(|\bar{X} - \bar{Y} - 200| < c) = 0'95 &\Rightarrow P(|N(0;1)| < c/\sqrt{4400}) = 0'95 \Rightarrow \\ &\Rightarrow c/\sqrt{4400} = 1'96 \Rightarrow c = 130 \end{aligned}$$

Por tanto, si tomamos sucesivas muestras de tamaño 300 de las bombillas tipo A y de tamaño 200 de las del tipo B, la diferencia entre las medias muestrales se concretará en el intervalo $(200 - 130; 200 + 130)$ el 95 % de las veces.

FONEMATO 2.6.16

Sean S_1^2 y S_2^2 las respectivas varianzas muestrales correspondientes a dos muestras independientes de tamaños 5 y 4 de dos poblaciones normales con igual varianza.

- 1) Determínese la probabilidad del suceso $S_2^2/S_1^2 > 5'2$.
- 2) Con los mismos tamaños muestrales, si las poblaciones son $N(3;2)$ y $N(7;8)$, determínese un intervalo en el que, con probabilidad 0'95, se encuentre el aleatorio cociente de varianzas muestrales.

SOLUCIÓN

Es sabido que en el m.a.s de tamaño "n" de una población $N(\mu; \sigma)$, el estadístico $n.S^2/\sigma^2$ tiene distribución χ_{n-1}^2 . Por tanto, siendo S_1^2 y S_2^2 las respectivas varianzas muestrales correspondientes a dos muestras independientes de tamaños 5 y 4 de las poblaciones $X \approx N(\mu_1; \sigma_1)$ e $Y \approx N(\mu_2; \sigma_2)$, es:

$$\frac{5.S_1^2}{\sigma_1^2} \approx \chi_{5-1}^2 = \chi_4^2 ; \quad \frac{4.S_2^2}{\sigma_2^2} \approx \chi_{4-1}^2 = \chi_3^2$$

En consecuencia, siendo $\sigma_1 = \sigma_2$, es:

$$\frac{\frac{4.S_2^2/\sigma_2^2}{3}}{\frac{5.S_1^2/\sigma_1^2}{4}} = \frac{16.S_2^2}{15.S_1^2} \approx F_{3;4}$$

$$\begin{aligned} 1) \quad P(S_2^2/S_1^2 > 5'2) &= P\left(\frac{16.S_2^2}{15.S_1^2} > \frac{16}{15}.5'2\right) = P(F_{3;4} > \frac{16}{15}.5'2) = \\ &= 1 - P(F_{3;4} \leq \frac{16}{15}.5'2) = 1 - 0'928 \end{aligned}$$

interpolando entre $P(F_{3;4} \leq 4'19) = 0'9$ y $P(F_{3;4} \leq 6'59) = 0'95$

- 2) Si $X \approx N(3;2)$ e $Y \approx N(7;8)$, para tamaños muestrales respectivos 5 y 4, es:

$$\frac{5.S_1^2}{2^2} \approx \chi_{5-1}^2 = \chi_4^2 ; \quad \frac{4.S_2^2}{8^2} \approx \chi_{4-1}^2 = \chi_3^2$$

Por tanto, es:

$$\frac{\frac{5.S_1^2/2^2}{4}}{\frac{4.S_2^2/8^2}{3}} = \frac{15.S_1^2}{S_2^2} \approx F_{4;3}$$

En la distribución de probabilidad de la variable $15.S_1^2/S_2^2 \approx F_{4;3}$ hay infinitud de intervalos que "encierran" masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. Como la distribución $F_{4;3}$ no es simétrica respecto a ningún punto, por comodidad, elegimos el intervalo que reparte la probabilidad $1 - 0'95 = 0'05$ en partes iguales entre las dos colas (izquierda y derecha) de dicha $F_{4;3}$ (puede demostrarse que éste no es el de menor amplitud, pero del de menor amplitud es tan petardo que pasamos de él). Según la tabla de la función de distribución de la variable $F_{4;3}$ es $P(F_{4;3} \leq 15'1) = 0'975$; además:

$$\begin{aligned} P(F_{4;3} \leq c) &= 0'025 \Rightarrow \\ \Rightarrow P(1/F_{4;3} \geq 1/c) &= 0'025 \Rightarrow \\ \Rightarrow P(F_{3;4} \geq 1/c) &= 0'025 \Rightarrow \\ \Rightarrow 1 - P(F_{3;4} \leq 1/c) &= 0'025 \Rightarrow \\ \Rightarrow P(F_{3;4} \leq 1/c) &= 0'975 \Rightarrow \\ \Rightarrow 1/c &= 9'98 \Rightarrow \\ \Rightarrow c &= 0'1 \end{aligned}$$

Así, siendo $P(0'1 \leq 15.S_1^2/S_2^2 \approx F_{4;3} \leq 15'1) = 0'95$, es:

$$P(0'1/15 \leq S_1^2/S_2^2 \leq 15'1/15) = 0'95$$

Por tanto, si tomamos sucesivas muestras de tamaños 5 y 4 de las poblaciones $N(3;2)$ y $N(7;8)$, el 95 % de las veces ocurrirá que estadístico S_1^2/S_2^2 se concreta en el intervalo $(0'1/15; 15'1/15)$.



FONEMATO 2.6.17

Si de una población normal bidimensional (X;Y) cuyo coeficiente de correlación es $\rho = 0'5$ se toma muestra aleatoria simple de tamaño 19, determínese la probabilidad de que el coeficiente de correlación muestral "r" esté entre 0'45 y 0'55.

SOLUCIÓN

Siendo "r" el coeficiente de correlación muestral correspondiente a una m.a.s de tamaño "n" de la población normal bidimensional (X;Y) con coeficiente de correlación $\rho = \text{COV.}(X;Y)/\sqrt{V(X) \cdot V(Y)}$, según Fisher, es:

$$W = \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \approx N\left(\frac{1}{2} \cdot \text{Ln} \frac{1+\rho}{1-\rho}; \frac{1}{\sqrt{n-3}}\right)$$

Si $n = 19$ y $\rho = 0'5$, es

$$W \approx N\left(\frac{1}{2} \cdot \text{Ln} \frac{1+0'5}{1-0'5}; \frac{1}{\sqrt{19-3}}\right) \cong N(0'55; 0'25)$$

Por tanto:

$$\begin{aligned} P(0'45 < r < 0'55) &= P(0'484 < W < 0'618) = \\ &\quad * r = 0'45 \Rightarrow w = \frac{1}{2} \cdot \text{Ln} \frac{1+0'45}{1-0'45} = 0'484 \\ &\quad * r = 0'55 \Rightarrow w = \frac{1}{2} \cdot \text{Ln} \frac{1+0'55}{1-0'55} = 0'618 \\ &= P\left(\frac{0'484 - 0'55}{0'25} < N(0;1) < \frac{0'618 - 0'55}{0'25}\right) = 0'22 \end{aligned}$$



FONEMATO 2.6.18

Si de una población normal bidimensional (X;Y) cuyo coeficiente de correlación es $\rho = 0'8$ se toma muestra aleatoria simple de tamaño 28, determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre el coeficiente de correlación muestral "r".

SOLUCIÓN

Siendo "r" el coeficiente de correlación muestral correspondiente a una m.a.s de tamaño "n" de la población normal bidimensional (X;Y) con coeficiente de correlación $\rho = \text{COV.}(X;Y)/\sqrt{V(X) \cdot V(Y)}$, según Fisher, es:

$$W = \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \approx N\left(\frac{1}{2} \cdot \text{Ln} \frac{1+\rho}{1-\rho}; \frac{1}{\sqrt{n-3}}\right)$$

En nuestro caso, como $n = 28$ y $\rho = 0'8$, es:

$$W = \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \approx N\left(\frac{1}{2} \cdot \text{Ln} \frac{1+0'8}{1-0'8}; \frac{1}{\sqrt{28-3}}\right) \cong N(1'09; 0'2)$$

En la distribución de probabilidad del estadístico $W \approx N(1'09; 0'2)$ hay infinidad de intervalos que **encierran** masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, **como la variable $N(1'09; 0'2)$ es simétrica respecto al punto 1'09, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto, que es el de menor amplitud.**

Por tanto, debemos determinar $c > 0$ de modo que $P(|W - 1'09| < c) = 0'95$:

$$\begin{aligned} P(|W - 1'09| < c) &= 0'95 \Rightarrow \\ \Rightarrow P\left(|N(0;1)| < \frac{c}{0'2}\right) &= 0'95 \Rightarrow \\ \Rightarrow \frac{c}{0'2} &= 1'96 \Rightarrow c = 0'39 \end{aligned}$$

Así, si tomamos sucesivas muestras de tamaño 28, el estadístico "W" se concretará en el intervalo $(1'09 - 0'39; 1'09 + 0'39) \equiv (0'70; 1'48)$ el 95 % de las veces... y el 95 % de las veces "r" se concretará en el intervalo $(0'60; 0'90)$:

$$\begin{aligned} * \quad w = 0'70 &= \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \Rightarrow \frac{1+r}{1-r} = e^{1'4} \Rightarrow \dots \Rightarrow r = 0'60 \\ * \quad w = 1'48 &= \frac{1}{2} \cdot \text{Ln} \frac{1+r}{1-r} \Rightarrow \frac{1+r}{1-r} = e^{2'96} \Rightarrow \dots \Rightarrow r = 0'90 \end{aligned}$$

FONEMATO 2.6.19

Si de una población normal bidimensional $(X; Y)$ cuyo coeficiente de correlación es $\rho = 0$ se toma muestra aleatoria simple de tamaño 6, determínese la probabilidad de que el coeficiente de correlación muestral " r " esté entre 0'608 y 0'729. Determínese un intervalo en el que puede esperarse, con probabilidad 0'95, que se encuentre el coeficiente de correlación muestral " r ".

SOLUCIÓN

Siendo " r " el coeficiente de correlación muestral correspondiente a una m.a.s de tamaño " n " de la población normal bidimensional $(X; Y)$ con coeficiente de correlación $\rho = 0$, es:

$$Z = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \approx t_{n-2}$$

En nuestro caso, como $n = 6$, es $Z = 2 \cdot r / \sqrt{1-r^2} \approx t_4$; por tanto:

$$P(0'608 < r < 0'729) = P(1'531 < Z \approx t_4 < 2'129) =$$

$$\begin{aligned} * r = 0'608 &\Rightarrow z = 2 \cdot 0'608 / \sqrt{1-0'608^2} = 1'531 \\ * r = 0'729 &\Rightarrow z = 2 \cdot 0'729 / \sqrt{1-0'729^2} = 2'129 \end{aligned}$$

$$= P(t_4 < 2'129) - P(t_4 < 1'531) = 0'95 - 0'90 = 0'05$$

En la distribución de probabilidad del estadístico $Z \approx t_4$ hay infinidad de intervalos que **encierran** masa 0'95, y cualquiera de ellos "sirve" para resolver la papeleta. No obstante, **como la t_4 es simétrica respecto al origen, entre los infinitos intervalos citados elegiremos el simétrico respecto a dicho punto, que es el de menor amplitud.** Así, debemos determinar $c > 0$ de modo que $P(|t_4| < c) = 0'95$:

$$P(|t_4| < c) = 0'95 \Rightarrow c = 2'777$$

Por tanto, si tomamos sucesivas muestras de tamaño 6, el estadístico $Z \approx t_4$ se concretará en el intervalo $(-2'777; 2'777)$ el 95 % de las veces... y el 95 % de las veces " r " se concretará en el intervalo $(-0'811; 0'811)$:

$$* z = 2'777 = 2 \cdot r / \sqrt{1-r^2} \Rightarrow r = 0'811$$

$$* z = -2'777 = 2 \cdot r / \sqrt{1-r^2} \Rightarrow r = -0'811$$