ASIGNATURA FST41: Muestreo en Distribuciones Normales

Cirilo Alvarez Rojas

Universidad Nacional De Ingeniería

Facultad De Ingeniería Económica,Ingeniería Estadística

Y Ciencias Sociales

ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA

Esta lección se estudia las características de muestras aleatorias de poblaciones normales; estas distribuciones de muestreo se utilizarán más adelante en pruebas de hipótesis y estimación de intervalos.

El muestreo de una población normal conduce a muchas propiedades útiles de estadística de la muestra y también a muchas distribuciones de muestreos conocidos.

Para estudiar las características de la distribución muestral se utilizará principalmente la función característica y sus propiedades.

Muchos fenómenos que se observa en la realidad tienen distribuciones de frecuencias relativas que al ser representadas tiene una forma parecida a la distribución normal, razón por la cual se puede suponer que la mayoría de las poblaciones con las que nos encontramos serán normales.

Propiedades de la Media y Varianza muestral

Ya hemos visto como calcular las medias y al varianzas muestrales, $\overline{X}$ y S^2 en general.

Ahora, bajo la suposición adicional de que la muestra proviene de una población normal, podemos derivar sus distribuciones muestrales completas, y otras características. Las propiedades de $\bar{X}$ y S^2 muestrales y de $\frac{(n-1)S^2}{\sigma^2}$s on resumidas en el siguiente Teorema.

Teorema 1

Sea $X_1, \ldots, X_n, n \geq 2$, una m.a.s. de una distribución normal $\mathcal{N}(\mu, \sigma^2)$, y sean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ la media y varianza muestral, respectivamente. Entonces

(a) $\bar{X}$ y S^2 son independientes; es decir la media muestral y la varianza muestral so independientes.

(b) $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}.\right)$

(c) $\dfrac{(n-1)S^2}{\sigma^2}$ tiene distribución Ji-cuadrado con (n-1) grados de libertad; esto es, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

(d) $t = \dfrac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$

Antes de demostrar el ítem (a) del Teorema 1, probemos la siguiente proposición.

Proposición 1

Las variables aleatorias X y Y son independientes si y solo si, la función característica del vector (X, Y) se puede factorizar;

$$X, Y \text{ son independientes} \Longleftrightarrow \Phi_{X,Y}(s,t) = \Phi_X(s)\Phi_Y(t) \ \forall (s,t) \in \mathbb{R}^2$$

Prueba.

$$\Phi_{X,Y}(s,t) = \mathrm{E}[e^{isX+itY}] = \mathrm{E}[e^{isX}e^{itY}] = \mathrm{E}[e^{isX}]\mathrm{E}[e^{itY}] = \Phi_X(s)\Phi_Y(t)$$

por independencia de X y Y.

$\square$

Recíprocamente, si $\Phi_{X,Y}(s,t) = \Phi_X(s)\Phi_Y(t)$, entonces, en el caso continuo, se tiene

$$\int\int e^{isx+ity} f(x,y)dxdy = \left[\int e^{isx} f(x)dx\right]\left[\int e^{ity} f(y)dy\right]$$

esto es,

$$\int\int e^{isx+ity} f(x,y)dxdy = \int e^{isx+ity} f(x)f(y)dxdy.$$

Por teorema de la unicidad de la función característica se debe tener

$$f(x,y) = f(x)f(y) \quad \text{para todo } (x,y) \in \mathbb{R}^2$$

Resulta que X y Y son independientes. Una prueba similar se da en el caso donde (X,Y) es de tipo discreto.

Prueba (a)

Primero demostremos que $\bar{X}$ es independiente de $(X_j - \bar{X})$ para todo j. Para esto calculamos la función característica conjunta de $\bar{X}$ y de las desviaciones $(X_1 - \bar{X}), (X_2 - \bar{X}), \ldots, (X_n - \bar{X})$; tenemos

$$\Phi_{\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}}(t, t_1, t_2, \ldots, t_n) = E\left(e^{it\bar{X} + i\sum_{j=1}^{n} t_j (X_j - \bar{X})}\right) \quad (1)$$

Para continuar, desarrollemos el exponente de e en la ecuación (1), así tenemos,

$$\begin{aligned} it\bar{X} + i\sum_{j=1}^{n} t_j(X_j - \bar{X}) &= i\frac{t}{n}\sum_{J=1}^{n} X_j + i\sum_{J=1}^{n} t_j X_j - i\sum_{j=1}^{n} t_j \frac{1}{n}\sum_{J=1}^{n} X_j \\ &= i\frac{t}{n}\sum_{J=1}^{n} X_j + i\sum_{J=1}^{n} t_j X_j - \bar{t}i\left(\frac{1}{n}\right)\sum_{J=1}^{n} X_j \\ &= i\frac{t}{n}\sum_{J=1}^{n} X_j + i\sum_{J=1}^{n} X_j(t_j - \bar{t}) \end{aligned}$$

$$= i \sum_{J=1}^{n} X_j \left\{ \frac{t}{n} + (t_j - \bar{t}) \right\}$$

luego la ecuación (1) queda como sigue

$$\Phi_{\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}}(t, t_1, t_2, \ldots, t_n) = E\left[e^{i \sum_{J=1}^{n} X_j \left\{ \frac{t}{n} + (t_j - \bar{t}) \right\}} \right]$$

$$= E\left[\prod_{j=1}^{n} e^{i X_j \left\{ \frac{t}{n} + (t_j - \bar{t}) \right\}} \right]$$

como las variables X_j son iid (variables muestrales) resulta

$$\Phi_{\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}}(t, t_1, t_2, \ldots, t_n) = \prod_{j=1}^{n} E\left[e^{i X \left\{ \frac{t}{n} + (t_j - \bar{t}) \right\}} \right]$$

$$= \prod_{j=1}^{n} \Phi_X \left(\frac{t}{n} + (t_j - \bar{t}) \right)$$

y como $X \sim \mathcal{N}(\mu, \sigma^2)$ tenemos

$$\Phi_{\bar{X},X_1-\bar{X},\ldots,X_n-\bar{X}}(t, t_1, t_2, \ldots, t_n) = \prod_{j=1}^{n} e^{i\left(\frac{t}{n}+(t_j-\bar{t})\right)\mu-\frac{\sigma^2}{2}\left(\frac{t}{n}+(t_j-\bar{t})\right)^2}$$

$$= e^{\sum_{j=1}^{n}\left[i\left(\frac{t}{n}+(t_j-\bar{t})\right)\mu-\frac{\sigma^2}{2}\left(\frac{t}{n}+(t_j-\bar{t})\right)^2\right]}$$

$$\Phi_{\bar{X},X_1-\bar{X},\ldots,X_n-\bar{X}}(t, t_1, t_2, \ldots, t_n) = e^{it\mu-\frac{\sigma^2}{n}\frac{t^2}{2}-\frac{\sigma^2}{2}\sum_{j=1}^{n}\left(t_j-\bar{t}\right)^2}$$

$$\Phi_{\bar{X},X_1-\bar{X},\ldots,X_n-\bar{X}}(t, t_1, t_2, \ldots, t_n) = \underbrace{e^{it\mu-\frac{\sigma^2}{n}\frac{t^2}{2}}}_{\mathit{\Phi}_{\bar{X}}(t)} \underbrace{e^{-\frac{\sigma^2}{2}\sum_{j=1}^{n}\left(t_j-\bar{t}\right)^2}}_{\Psi_{X_1-\bar{X},\ldots,X_n-\bar{X}}(t_1,t_2,\ldots,t_n)}$$

es decir, la función característica conjunta se factoriza

$$\Phi_{\bar{X},X_1-\bar{X},\ldots,X_n-\bar{X}}(t, t_1, t_2, \ldots, t_n) = \mathit{\Phi}_{\bar{X}}(t).\Psi_{X_1-\bar{X},\ldots,X_n-\bar{X}}(t_1, t_2, \ldots, t_n)$$

y como el primer factor de lado derecho de la última igualdad es la función característica de la media muestral de la muestra que proviene de la $\mathcal{N}(\mu, \sigma^2)$, y el otro factor es función característica del vector de las desviaciones de cada variable respecto de la media muestral, $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$, entonces la media muestral, $\bar{X}$ y las desviaciones $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ son independientes.

El resultado es consecuencia de la proposición 1.

Corolario 1

$\overline{X}$ y S^2 son independientes

Prueba del corolario 1

Utilizando la función característica tenemos: sabemos que , si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces su función característica es $\Phi_X(t) = e^{\mu it - \frac{\sigma^2 t^2}{2}}$. Luego la función característica de la media muestral será

$$\Phi_{\bar{X}}(t) = \left[e^{\mu i \frac{t}{n} - \frac{\sigma^2 t^2}{2n^2}} \right]^n = \left[e^{it\mu - \frac{\sigma^2}{n} \frac{t^2}{2}} \right]$$

y por el teorema de la unicidad de la función característica concluimos que

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}).$$

Corolario 2

$$Y = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim \mathcal{N}(0,1)$$

Prueba La función característica de Y es

$$\Phi_Y(t) = \mathrm{E}\left[e^{itY}\right] = \mathrm{E}\left[e^{it\left\{\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right\}}\right] = \mathrm{E}\left[e^{\left\{\frac{\sqrt{n}\overline{X}}{\sigma}it\right\} - \left\{\frac{\mu}{\sigma}it\right\}}\right]$$

$$= e^{-\frac{\sqrt{n}\mu}{\sigma}it}\Phi_{\overline{X}}\left(\frac{\sqrt{n}}{\sigma}t\right) = e^{-\frac{\sqrt{n}\mu}{\sigma}it}e^{\frac{\sqrt{n}\mu}{\sigma}it - \frac{\sigma^2}{n}\frac{n}{\sigma^2}\frac{t^2}{2}}$$

$$\Phi_Y(t) = e^{-\frac{t^2}{2}}$$

Por lo tanto, por el Teorema de la unicidad de la función característica, Y es una variable aleatoria estándar; equivalentemente, $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Además, para todo $a, b \in \mathbb{R} : a \leq b$ se tiene

$$P(a \leq \overline{X} \leq b) = P(a < \overline{X} \leq b) = P(a \leq \overline{X} < b) = P(a \leq \overline{X} \leq b) =$$

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

Ejemplo

Sea $X_1, X_2, \ldots, X_{16}$ una muestra aleatoria de una distribución normal, $\mathscr{N}(77, 25)$: Calcular:(a) $P(77 < \overline{X} < 79{,}5)$, (b) $P(74{,}2 < \overline{X} < 78{,}4)$

Solución: $P(77 < \overline{X} < 79{,}5) = \Phi\left(\frac{\sqrt{16}(79{,}5-77)}{5}\right) - \Phi\left(\frac{\sqrt{16}(77-77)}{5}\right)$

$$= \Phi(2) - \Phi(0)$$

$$= 0{,}97725 - 0{,}50000 = 0{,}47725$$

Ejemplo 1.

El número de libros encuadernados diariamente por una máquina automática es una variable aleatoria cuya distribución se desconoce, con una desviación típica de 16 libros por día. Si se selecciona una muestra aleatoria de 49 días, hallar la probabilidad de que el número medio de libros encuadernados durante esos días (la media muestral) se encuentre a lo más 3 libros de la verdadera media poblacional.

Solución 1.

Como no se especifica la distribución del número de libros y $n = 49 > 30$, podemos utilizar el TCL. para hallar la probabilidad solicitada. Tenemos

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

La probabilidad que pide es el siguiente

$$
\begin{aligned}
P(\left|\overline{X}-\mu\right| \leq 3) &= P(-3 \leq \overline{X}-\mu \leq 3) \\
&= P\left(\frac{-3}{16/\sqrt{49}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{3}{16/\sqrt{49}}\right) \\
&= \Phi(1{,}3125) - \Phi(-1{,}3125) \\
&= 0{,}81064
\end{aligned}
$$

Ejemplo 2.

Con referencia al ejemplo 1. Determinar el tamaño de la muestra para que la media muestral se encuentre a lo más a 3 libros de la media poblacional con una probabilidad de 0,95.

Solución 2.

Ahora se tiene que verificar que

$$P(|\overline{X}-\mu| \leq 3) = P(-3 \leq \overline{X}-\mu \leq 3) = 0{,}95$$

Estándarizando se tiene

$$\begin{aligned} P\left(\frac{-3}{16/\sqrt{49}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{3}{16/\sqrt{49}}\right) &= P(-0{,}1875\sqrt{n} \leq Z \leq 0{,}1857\sqrt{n}) \\ &= \Phi(0{,}1875\sqrt{n}) - \Phi(-0{,}1875\sqrt{n}) = 0{,}9 \\ 2\Phi(0{,}8175\sqrt{n}) - 1 &= 0{,}95 \\ \Phi(0{,}1875\sqrt{n}) &= 0{,}975 \\ 0{,}1875\sqrt{n} &= \Phi^{-1}(0{,}975) \\ 0{,}1875\sqrt{n} &= 1{,}95996 \\ n &\approx 110 \end{aligned}$$

Distribuciones Exactas:Chi-cuadrado, t, y F

(Los ítems (c) y (d) del Teorema 1 serán demostrados posteriormente.)

En esta sección se investiga las distribuciones que surgen en el muestreo de una población normal. Consideremos una muestra aleatoria simple $X_1, \ldots, X_n$ extraída de una población normal $\mathcal{N}(\mu, \sigma^2)$. Entonces sabemos que $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, probaremos que $\left\{\sqrt{n}(\bar{X} - \mu)/\sigma\right\}^2 \sim \chi^2(1)$.

La primera distribución de interés es la *distribución chi-cuadrado*, que también se puede definir como un caso especial de la distribución Gamma.

Definición 1

Una variable aleatoria X tiene una distribución chi-cuadrado (χ^2) con n *grados de libertad si su función de densidad de probabilidad está dado por*

$$f(x;\mathrm{n}) = \begin{cases} \dfrac{x^{\frac{\mathrm{n}}{2}-1}e^{-\frac{1}{2}x}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} & si \quad x > 0 \\ 0, & si \quad x \leq 0 \end{cases}$$

Notación: $X \sim \chi^2(\mathrm{n})$ *que se lee: la variable aleatoria X se distribuye como una v.a. Chi-cuadrada con* n *grados de libertad.*

Nota: *Si* $X \sim \chi^2(\mathrm{n})$, *entonces*

Media: $\mathrm{E}(X) = \mathrm{n}$

Varianza: $Var(X) = 2\mathrm{n}$

Asimetría: $\beta_1 = 2\sqrt{2/\mathrm{n}}$

Curtosis: $\beta_2 = 3 + \frac{12}{\mathrm{n}}$

Teorema 2

Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas y sea $S_n = \sum_{j=1}^{n} X_j$. Entonces

(a) $S_n \sim \chi^2(n) \Longleftrightarrow X_1 \sim \chi^2(1)$ y

(b) $X_1 \sim \mathcal{N}(0,1) \Longrightarrow \sum_{j=1}^{n} X_j^2 \sim \chi^2(n)$

Prueba del ítem (a)

$\Longrightarrow$) Supongamos que $S_n \sim \chi^2(n)$, entonces la función característica de S_n es $\Phi_{S_n}(t) = (1-2it)^{-\frac{n}{2}} = \prod_{J=1}^{n}(1-2it)^{-\frac{1}{2}}$, entonces cada factor es una función característica de una variable X_j, esto es $\Phi_{X_j}(t) = (1-2it)^{-\frac{1}{2}}$ para $j = 1, 2, \ldots, n$ y así por el teorema de la unicidad de la función característica $X_1 \sim \chi^2(1)$.

Recíprocamente:

($\Longleftarrow$ si $X_1 \sim \chi^2(1)$, entonces la función característica de X_1 es $\Phi_{X_1}(t) = (1-2it)^{-\frac{1}{2}}$ y como las variables $X_1, \ldots, X_n$ son iid se tiene que,

$$\Phi_{S_n}(t) = \Phi_{\sum_{j=1}^{n} X_j}(t) = \prod_{j=1}^{n}(1-2it)^{-\frac{1}{2}} = (1-2it)^{-\frac{n}{2}}$$

y nuevamente por el teorema de la unicidad de la función característica se concluye que $S_n \sim \chi^2(n)$.

Prueba del ítem (b)

Si $X_1 \sim \mathcal{N}(0,1) \Longrightarrow X_1^2 \sim \chi^2(1)$, y como las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes e idénticamente distribuidas, por el ítem (a) se concluye que $\sum_{j=1}^{n} X_j^2 \sim \chi^2(n)$.

Corolario 3

Si $X_1, \ldots, X_n$ es una muestra aleatoria simple de tamaño n de una población con distribución normal, $\mathcal{N}(\mu, \sigma^2)$. Entonces la variable aleatoria

$$Z = \sum_{J=1}^{n} \left(\frac{X_j - \mu}{\sigma} \right)^2 = \frac{\sum_{J=1}^{n} (X_j - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

es decir tiene una distribución chi-cuadrado con n grados de libertad.

Prueba del ítem (c) del Teorema 1

Por el corolario anterior sabemos que:

$$\sum_{j=1}^{n} \frac{(X_j - \mu)^2}{\sigma^2} \sim \chi^2(n) \quad \text{y} \quad n \left(\frac{\bar{X} - \mu}{\sigma} \right)^2 \sim \chi^2(1)$$

y como $(\bar{X})$ y S^2 son independientes,resulta de

$$\frac{\sum_{J=1}^{n}(X_j-\mu)^2}{\sigma^2} = n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 + (n-1)\frac{S^2}{\sigma^2}$$

que

$$E\left\{\exp\left[it\sum_{j=1}^{n}\frac{(X_j-\mu)^2}{\sigma^2}\right\}\right] = E\left\{\exp\left[itn\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 + (n-1)\frac{S^2}{\sigma^2}it\right]\right\}$$
$$= E\left\{\exp\left[itn\left(\frac{\bar{X}-\mu}{\sigma}\right)^2\right]\right\} E\left\{\exp\left[(n-1)\frac{S^2}{\sigma^2}it\right]\right\},$$

esto es

$$(1-2it)^{-\frac{n}{2}} = (1-2it)^{-\frac{1}{2}} E\left\{\exp\left[(n-1)\frac{S^2}{\sigma^2}it\right]\right\},$$

y resulta que

$$E\left\{\exp\left[(\mathrm{n}-1)\frac{S^2}{\sigma^2}it\right]\right\} = (1-2it)^{-\frac{(\mathrm{n}-1)}{2}}.$$

Por por el teorema de la unicidad de la función característica resulta que

$$\frac{(\mathrm{n}-1)S^2}{\sigma^2} \sim \chi^2(\mathrm{n}-1)$$

La distribución $\chi^2(\mathrm{n})$ está tabulado para valores de $\mathrm{n} = 1, 2, 3, \ldots$. Las tablas por lo general están hechas por hasta $\mathrm{n} = 30$, puesto que para $\mathrm{n} > 30$ se puede usar la aproximación normal.

Definición 2

Sean $X_1, \ldots, X_{\mathsf{n}}$ *variables aleatorias normales independientes con* $E(X_j) = \mu_j$ *y varianza* $var(X_j) = \sigma^2, j = 1, 2, \ldots, \mathsf{n}$*. También, sea* $Y = \sum_{j=1}^{\mathsf{n}} X_i/\sigma_j^2$*. La variable aleatoria* Y *se dice que se distribuye en forma chi-cuadrado no-central con parámetro de no-centralidad* $\sum_{j=1}^{\mathsf{n}} \mu_j/\sigma^2$ *y* n *grados de libertad. Se escribe como* $Y \sim \chi^2(\mathsf{n}, \delta)$*, donde* $\delta = \sum_{j=1}^{\mathsf{n}} \mu_j/\sigma^2$*. La función de densidad de probabilidad de una variable aleatoria* $\chi^2(\mathsf{n}, \delta)$ *se puede demostrar que ses igual a*

$$f_{\mathsf{n}}(y, \delta) = \begin{cases} \pi^{-\frac{1}{2}} 2^{-\mathsf{n}/2} \exp\left\{-\frac{1}{2}(\delta + y)\right\} y^{(\mathsf{n}-2)/2} & \\ \sum_{j=0}^{\infty} \frac{(\delta y)^j (\Gamma(j+1/2))}{(2j)!\Gamma(j1/2n)} & ; \quad y > 0 \\ 0 & ; \quad y \leq 0 \end{cases} \tag{2}$$

donde $\delta = \sum_{j=1}^{\mathsf{n}} \mu_j/\sigma^2$*. Haciendo* $\delta = 0$*, observamos que ecuación* (2) *se reduce a la función de densidad de probabilidad* $\chi^2(\mathsf{n})$ *central.*

Definición 3

Considere que la variable aleatoria X se distribuye en forma normal estándar ($X \sim \mathcal{N}(0,1)$) y la variable aleatoria Y se distribuye en forma de distribución chi-cuadrado con n grados de libertad ($Y \sim \chi^2(\mathrm{n})$) y además, las variables X y Y son independientes. Entonces el estadístico

$$T = \frac{X}{\sqrt{Y/\mathrm{n}}} \tag{3}$$

se dice que tiene una distribución-t con n grados de libertad y se escribe $T \sim t(\mathrm{n})$.

Distribución t de "student"

Definición 4

Una variable aleatoria X tiene distribución t de "student" si su función de densidad de probabilidad está dada por

$$f_X(x) = \frac{\Gamma((\frac{n+1}{2}))}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

donde n *es el parámetro de distribución y recibe el nombre de "número de grados de libertad".*

Notación $X \sim t(\mathrm{n})$

Propiedades

Si $X \sim t(\mathrm{n})$, entonces tiene las siguientes propiedades:

(a) La distribución de t es simétrica respecto de $x = 0$.

(b) $\max f_X$ se da en $x = 0$ y es igual a $\frac{\Gamma((\frac{\mathrm{n}+1}{2}))}{\Gamma(\frac{\mathrm{n}}{2})\sqrt{\mathrm{n}\pi}}$

(c) Función de distribución:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{\Gamma((\frac{\mathrm{n}+1}{2}))}{\Gamma\left(\frac{\mathrm{n}}{2}\right)\sqrt{\mathrm{n}\pi}} \left(1 + \frac{t^2}{\mathrm{n}}\right)^{-\frac{\mathrm{n}+1}{2}} dt.$$

(d) Por la simetría se tiene $F_X(-a) = 1 - F_X(a)$

(e) $\mathrm{E}(X) = 0$ para todo $\mathrm{n} \in \mathbb{N}$ y $\mathrm{V}ar(X) = \frac{\mathrm{n}}{\mathrm{n}-2}$ para $n > 2$.

(f) Cuando $n \to \infty$ la distribución t se aproxima para para la distribución normal estándar.

Prueba del ítem (d) del Teorema 1

Como $\frac{n(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$, y $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ y puesto que $\bar{X}$ y S^2 son independiente , por definición anterior resulta que,

$$\frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1).$$

Definición 5

Sean X y Y variables aleatorias independientes chi-cuadras com m y n grados de libertad respectivamente. Entonces se dice que la variable aleatoria

$$F = \frac{X/m}{Y/n} \tag{4}$$

tiene distribución-F con (m, n) grados de libertad, y se escribe $F \sim F(m, n)$.

Teorema 3

La función de densidad de probabilidad del estadístico-F dada en al ecuación (4) está dada por,

$$g(f) = \begin{cases} \frac{\Gamma\left[\frac{(m+n)}{2}\right]}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right) \left(\frac{m}{n} f\right)^{\frac{m}{2}-1} \left(1 + \frac{m}{n} f\right)^{-\frac{(m+n)}{2}} & f > 0, \\ 0, & f \leq 0 \end{cases} \tag{5}$$

Comentario 1

Si $X \sim F(m, n)$*, entonces* $1/X \sim F(n, m)$*. Si hacemos* $m = 1$*, entonces* $F = [t(n)]^2$ *de modo que* $F(1, n)$ *y* $t^2(n)$ *tienen la misma distribución. También se continúa, que si* $Z \sim \mathscr{C}(0, 1)$ *[la cual es la misma como* $t(1)$*],* $Z^2 \sim F(1, 1)$.

Comentario 2

Como es natural, escribimos $F(m, n; \alpha)$ *para el porcentaje de puntos supriores de la distribución* $F(m, n)$*, esto es,*

$$P\left(F(m, n) > F(m, n; \alpha)\right) = \alpha \tag{6}$$

De la observación 1, se tiene la siguiente relación

$$F(m, n; 1 - \alpha) = \frac{1}{F(n, m; \alpha)} \tag{7}$$

Por tanto es suficiente tabular valores de F que sean mayores o iguales a 1.

Teorema 4

Si $X \sim F(m, n)$. Entonces, para $k > 0$, entero,

$$E(X^k) = \left(\frac{n}{m}\right)^k \frac{\Gamma\left[k + \frac{m}{2}\right]\Gamma\left[\frac{n}{2} - k\right]}{\Gamma\left[\frac{m}{2}\right]\Gamma\left[\frac{n}{2}\right]} \quad n > 2k \tag{8}$$

En particular,

$$E(X) = \frac{n}{n-2}, \quad n > 2, \tag{9}$$

y

$$Var(X) = \frac{n^2(2m + 2n - 4)}{m(n-2)^2(n-4)}, \quad n > 4. \tag{10}$$

Teorema 5

Si $X \sim F(m, n)$, entonces, $Y = 1/[1 + (m/n)X]$ es $B(n/2, m/2)$. Consecuentemente, para cada $x > 0$

$$F_X(x) = 1 - F_Y\left[\frac{1}{1 + (m/n)x}\right]$$

Demostración.

Se deja como ejercicio □

Si en la definición (3) hacemos que $X \sim \chi^2(n, \delta)$ se obtiene la variable aleatoria F no-central.

Definición 6

Sea $X \sim \chi^2(n, \delta)$ *y* $Y \sim \chi^2(n)$, *y sean* X *y*Y *independientes. Entonces la variable aleatoria*

$$F = \frac{X/m}{Y/n} \tag{11}$$

se dice que tiene un a distribución-F *no-central con grados de libertad* (m, n) *y parámetro de no-centralidad* δ.

Corolario 4

Si $X_1, X_2, \ldots, X_m$ son variables aleatorias iid. $\mathcal{N}(\mu_X, \sigma_X^2)$, $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias iid. $\mathcal{N}(\mu_Y, \sigma_Y^2)$ y las dos muestras son tomadas independientemente, entonces $(S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2) \sim F(m-1, n-1)$. Si en particular, $\sigma_X^2 = \sigma_Y^2$, entonces $S_X^2/S_Y^2 \sim F(m-1, n-1)$

Corolario 5

Sean $X_1, X_2, \ldots, X_m$ y $Y_1, Y_2, \ldots, Y_n$, respectivamente, muestras independientes de $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Entonces

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\left\{\left[(m-1)S_X^2/\sigma_X^2\right] + \left[(n-1)S_Y^2/\sigma_Y^2\right]\right\}^{1/2}} \sqrt{\frac{m+n-2}{\sigma_X^2/m + \sigma_Y^2/n}} \sim t(m+n-2)$$

En particular, si $\sigma_X^2 = \sigma_Y^2$, entonces

$$\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\left[(m-1)S_X^2+(n-1)S_Y^2\right]}}\sqrt{\frac{mn(m+n-2)}{m+n}} \sim t(m+n-2)$$

El corolario 5 resulta, como

$$\overline{X}-\overline{Y} \sim \mathscr{N}\left(\mu_X-\mu_Y, \frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right) \text{ y } \frac{(m-1)S_X^2}{\sigma_X^2}+\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(m+n-2$$

y los dos estadísticos son independientes.

Ejercicios propuestos

1. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de $X \sim \mathcal{N}(\mu, \sigma^2)$ y $\overline{X}$ y S^2, respectivamente, la media muestral y la varianza muestral. Sea $X_{n+1} \sim X \sim \mathcal{N}(\mu, \sigma^2)$, y suponga que $X_1, X_2, \ldots, X_n$, X_{n-1} son independientes. Encuentre la distribución de muestreo de $[\frac{(X_{n+1}-\overline{X})}{S/\sqrt{\mathrm{n}/(n+1)}}]$
2. Sean $X_1, X_2, \ldots, X_\mathrm{m}$ y $Y_1, Y_2, \ldots, Y_n$ muestras aleatorias independientes de $\mathcal{N}(\mu_X, \sigma_X^2)$ y $\mathcal{N}(\mu_Y, \sigma_Y^2)$, respectivamente. Además, sean α y β dos números reales fijos. Si $\overline{X}$, $\overline{Y}$ denotan las medias muestrales correspondientes, ¿cuál es la distribución muestral de
$$\frac{\alpha(\overline{X}-\mu_X)+\beta(\overline{Y}-\mu_Y)}{\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}}$$

donde S_X^2 y S_Y^2respectivamente denotan las varianzas muestrales de ls $X's$ y de las Y'.

3. Sea $X_1, X_2, \ldots, X_\mathrm{n}$ una muestra aleatoria de $X \sim \mathcal{N}(\mu, \sigma^2)$ y sea k es un número entero entero positivo. Encuentra $\mathrm{E}(S^{2k})$. En particular, encuentre $\mathrm{E}(S^2)$ y $\mathrm{V}ar(S^2)$.
4. Se toma una muestra aleatoria de 5 de una población normal con una media de 2,5 y una varianza $\sigma^2 = 36$.
   (a) Encuentre la probabilidad de que la varianza muestral se encuentre entre 30 y 44.
   (b) Encuentre la probabilidad de que la media muestral se encuentre entre 1.3 y 3.5, mientras que la varianza muestral se encuentra entre 30 y 44.

5. Se observó que la vida media de una muestra de 10 bombillas era de 1327 horas con una desviación estándar de 425 horas. Una segunda muestra de 6 bombillas elegidas de un lote diferente mostró una vida media de 1215 horas con una desviación estándar de 375 horas. Si se supone que las medias de los dos lotes son las mismas, ¿qué tan probable es la observación? diferencia entre las dos medias de muestra?
6. Sean S_X^2 y S_Y^2 las varianzas muestrales de dos muestras independientes de tamaños $n_x = 5$ y $n_Y = 4$ de dos poblaciones que tienen la misma varianza desconocida σ^2. Encuentre (aproximadamente) la probabilidad de que $\frac{S_X^2}{S_Y^2} < 1/5{,}2$ o $> 6{,}25$.