Universidad Nacional Agraria La Molina
Departamento de Estadística e Informática
Curso: Inferencia Estadística
Profesor: Clodomiro Fernando Miranda Villagómez

La Ley De Los Grandes Números

Se tiene la siguiente pregunta: ¿Pueden hacerse inferencias fiables acerca de E(X) de un número infinito de valores de X, utilizando únicamente un numero finito de valores de X (es decir una muestra aleatoria de tamaño n).

Respuesta: Si, mediante las leyes débil y fuerte de los grandes números.

Lo que dice la ley de los grandes números es que:

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[n \to \infty]{\text{Converge a}} E(X)$$

Cuando n es grande.

¿Qué tamaño de muestra será necesario para que la inferencia acerca de E(X) sea fiable?

Ejemplo 1: Se lanza una moneda balanceada n veces.

Se define la v.a. $X_i = \begin{cases} 1\,;\text{si resulta cara}; i = 1, 2, \ldots, n \\ 0\,;\text{si resulta sello} \end{cases}$

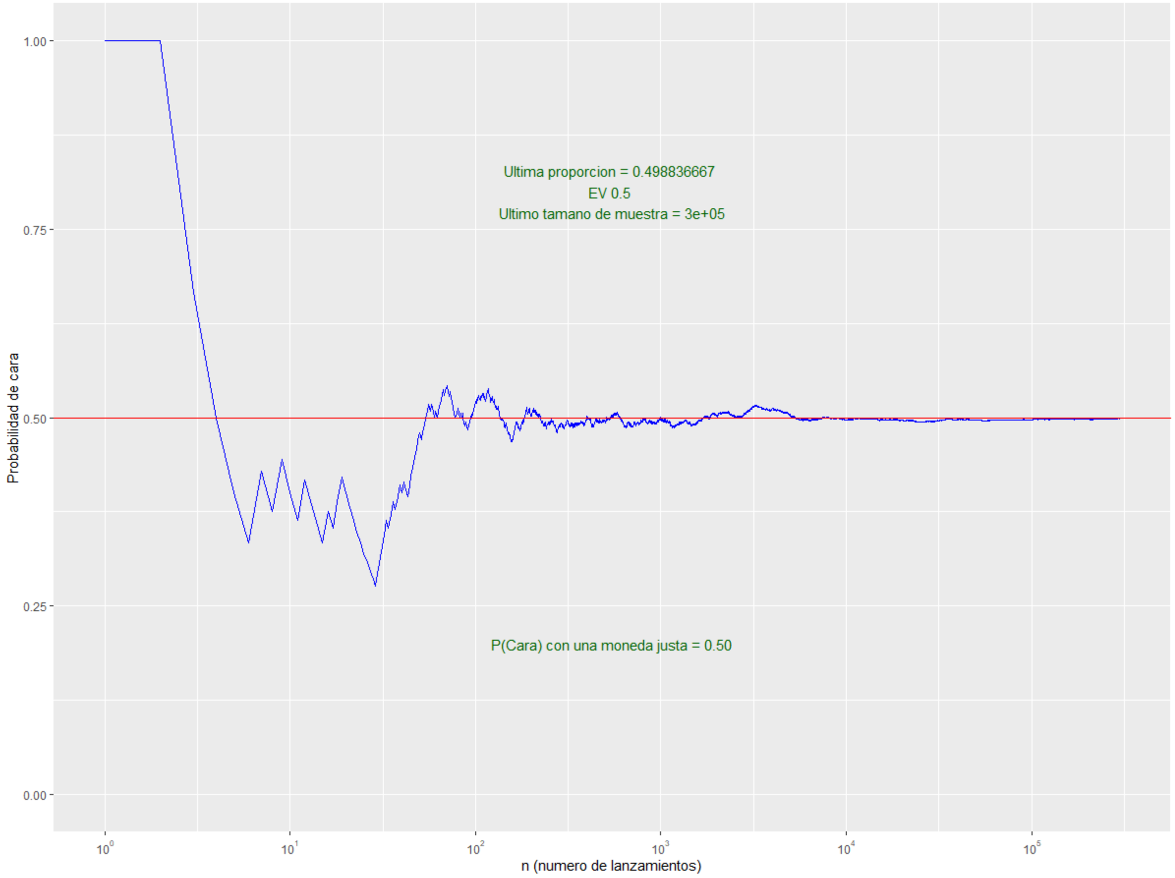
$$\frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[n \to \infty]{\text{Converge a}} E(X) = \frac{1}{2} = p$$

NOTA: El experimento aleatorio que consiste en lanzar una moneda balanceada n veces es equivalente a seleccionar una muestra aleatoria de tamaño n de una distribución Bernoulli que tiene probabilidad de éxito 0.5. De acuerdo a los casos la probabilidad de éxito puede tomar otro valor.

Por ejemplo, si con base a estudios anteriores se ha concluido que el 40 % de las personas de cierta población son empáticos. De esa población se toma una muestra aleatoria de tamaño 80 con la finalidad de utilizar la proporción muestral para verificar si la proporción poblacional sigue igual. En este caso la muestra aleatoria es $X_1, \ldots, X_{80}$, se cumple que $X_i \sim \text{Bernoulli}(\text{p}), i = 1, \ldots, 80$ y $\hat{\text{p}} = \frac{1}{80}\sum_{i=1}^{80} X_i = \frac{1+0+0+1+\ldots+0}{80} \xrightarrow[n\to\infty]{\text{Converje a}} E(X) = \text{p}$.

En este caso $n > 30$.

En forma gráfica se observa que a medida que aumenta el número de lanzamientos de la moneda (n), la proporción muestral converge a ½ que es la proporción (media) poblacional:



X_i se distribuye Binomial$\left(1,\dfrac{1}{2}\right)$; $\bar{X}$ converge a E(X)

¿Cómo Converge? ¿Qué tipo de convergencia?

Sean $X_1,\ldots,X_n,\ldots$ v.a definidas en un mismo espacio de probabilidades (Ω,Λ,P).

Convergencia en Probabilidad o convergencia débil:

X_n converge a μ en probabilidades si $\forall\varepsilon>0$

$$P\left(\left|X_n-\mu\right|>\varepsilon\right)\xrightarrow[n\to\infty]{}0$$

Cuando n es grande ($n\to\infty$)

Notación: $X_n \xrightarrow{P} X$

O también:

$$\lim_{n\to\infty}P\left(\left|X_n-\mu\right|>\varepsilon\right)=0$$

Convergencia cuadrática

Se dice que X_n converge a μ en el sentido cuadrático si $\lim_{n\to\infty} E\left(X_n-\mu\right)^2 = 0$, asumiendo que $E\left(X_n-\mu\right)^2$ existe (es finito) para cada n.

Convergencia con probabilidad 1 o convergencia fuerte (Casi Segura, c.s)

Se dice que X_n converge a μ c.s si $P\left(\lim_{n\to\infty} X_n = \mu\right) = 1$, donde $\lim_{n\to\infty} X_n = \mu$ define el conjunto donde la convergencia toma lugar, o equivalentemente $\lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} |X_m-\mu| > \varepsilon\right) = 0$.

La ley de los grandes números (LGN)

Si $X_1, X_2, \ldots$ son variables aleatorias no correlacionadas con la misma media $E\left(X_i\right) = \mu$ y varianza $Var\left(X_i\right) = \sigma^2$ entonces $\lim_{n\to\infty} \bar{X}_n = \mu$, la LGN débil. Si $X_1, X_2, \ldots$ son iid y tienen media finita μ, entonces $\lim_{n\to\infty} \bar{X}_n = \mu$ c.s, la LGN fuerte.

La ley débil de los grandes números de Khintchin

Sea $X_1, X_2, \ldots$ una sucesión de v.a. i.i.d. cada una con media finita μ y variancia finita σ^2. Si se define la sucesión de medias muestrales por:

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \qquad ; \; n = 1, 2, \ldots$$

entonces:

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon \right) = 0 \quad ; \quad \forall \varepsilon > 0$$

o equivalente:

$$\overline{X}_n \xrightarrow[n\to\infty]{P} \mu$$

Prueba:

$$\forall n \in \mathbb{N} \qquad \mu_{\bar{X}} = E\left[\bar{X}_n\right] = \mu \qquad \sigma^2_{\bar{X}} = V\left[\bar{X}_n\right] = \frac{\sigma^2}{n}$$

Según la desigualdad de Chebyschev, se tiene:

$$P\left[\left|\bar{X}_n - \mu_{\bar{X}}\right| > k\sigma_{\bar{X}}\right] \le \frac{1}{k^2} \quad ; \ \forall k > 0$$

$P\left[\left|\bar{X}_n - \mu_{\bar{X}}\right| > k\sigma_{\bar{X}}\right] \le \frac{1}{k^2}$; Haciendo $\varepsilon = k\frac{\sigma}{\sqrt{n}}$; entonces.

$P\left[\left|\bar{X}_n - \mu_{\bar{X}}\right| > \varepsilon\right] \le \frac{\sigma^2}{\varepsilon^2 n}$ (1); luego:

$$\lim_{n\to\infty} P\left[\left|\bar{X}_n - \mu\right| > \varepsilon\right] \le \lim_{n\to\infty}\frac{\sigma^2}{\varepsilon^2 n} = 0 \to \lim_{n\to\infty} P\left[\left|\bar{X}_n - \mu\right| > \varepsilon\right] = 0 \quad \forall \varepsilon > 0$$

<u>Nota:</u>

-De (1) para n suficientemente grande $\frac{\sigma^2}{\varepsilon^2 n}$ es muy pequeño, luego existe, $\delta > 0$ (tan pequeño como se quiere) tal que:

$$P\left[\left|\bar{X}_n - \mu\right| > \varepsilon\right] \le \frac{\sigma^2}{\varepsilon^2 n} < \delta \qquad (2)$$

Del lado derecho de (2) se tiene $n > \frac{\sigma^2}{\varepsilon^2 \delta}$ (3)

(3) dice que para $n > \frac{\sigma^2}{\varepsilon^2 \delta}$ el teorema de la ley débil de los grandes números es verdadera.

De (2) se obtiene $P\left[\left|\bar{X}_n - \mu\right| \le \varepsilon\right] \ge 1 - \frac{\sigma^2}{\varepsilon^2 n} > 1 - \delta$

La Ley Fuerte de los Grandes Números de Kolmogorov

Si $X_1, X_2, \ldots$ (o $\{X_n\}_{n\in N}$) es una sucesión de variables aleatorias i.i.d con media común $\mu < \infty$ entonces $\lim_{n\to\infty} \bar{X}_n = \mu$ casi seguramente o $\left(\bar{X} \to \mu \quad \text{c.s} \quad \Rightarrow P\left[\bar{X} \to \mu\right] = 1\right)$.
La convergencia casi segura consiste en que, si existe un conjunto $A \subseteq \Omega$ tal que sí $a \in A$, $\lim_{n\to\infty} \bar{X}_n(a) \ne \mu \to P[A] = 0$. Lo anterior implica que $P\left[\lim_{n\to\infty} \bar{X}_n = \mu\right] = 1$.
Esta ley justifica la interpretación intuitiva del esperado de una variable aleatoria como "el promedio a largo plazo al realizarse muestreos repetitivos"

Ejemplo: La ley de los grandes números indica que si $X_1, X_2, \dots$ $\left(\text{o } \{X_n\}_{n\in N} \underset{iid}{\sim} X\right)$ son independientes e idénticamente distribuidas de acuerdo con la distribución de X, entonces:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{c.s} E(X)$$

$$\bar{X}_{n(2)} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{c.s} E\left(X^2\right)$$

En general: $\bar{X}_{n(k)} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{c.s} E\left(X^k\right)$

Ejemplo 2: Observación: Sean $\{X_n\}_{n\in N}$ i.i.d con distribución F_θ donde θ es un parámetro. Se considera la familia $\{T_n(X_1,\dots,X_n)\}_{n\in N}$ donde T_n es una función de los n datos (que cumple ciertas hipótesis). $T_n(X_1,\dots,X_n)$ se llama un estimador de θ. Se dice que un estimador es consistente sí $T_n(X_1,\dots,X_n) \to \theta$ c.s.

Sean $\{X_n\}_{n\in N}$ i.i.d tales que $E(X_i) = \mu$, y $V(X_i) = \sigma^2 < \infty$ $(\sigma > 0)$.

a. Demuestre que $\bar{X}_n$ es un estimador consistente de μ, esto es $\bar{X}_n \to \mu$ c.s

b. Demuestre que cada uno de los estimadores $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ y $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ de σ^2 convergen en forma casi segura a σ^2. También demuestre que cada una de las desviaciones estándar muestrales convergen en forma casi segura a σ. SUGERENCIA: $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\bar{X}_n\right)^2$ y utilice los siguientes resultados:

Si $X_n \to X$ c.s y $g:\mathbb{R}\to\mathbb{R}$ es continua $\Rightarrow$ $g(X_n) \to g(X)$ c.s

Si $X_n \to X$ c.s y $Y_n \to Y$ c.s y $g:\mathbb{R}^2\to\mathbb{R}$ es continua $\Rightarrow$ $g(X_n, Y_n) \to g(X, Y)$ c.s

Ejemplo 3: Si $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$ c.s $\Rightarrow \sum_{i=1}^{n} X_i \to n\mu \underset{\infty}{\to} \infty \left[\sum_{i=1}^{n} X_i \to \infty \text{ c.s}\right]$

Ejemplo 4: Si $\{X_i\}_{i\in N}$ son v.as iid $\sim$ Exponencial$\left(\text{con media } \theta\right)$, halle el límite casi seguro de $\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^2$.

De acuerdo con la sugerencia del ejemplo 2, $\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^2 = \left[\bar{X}_n\right]^2 \to \theta^2$ c.s. En este caso $g:\mathbb{R}\to\mathbb{R} / g(y) = y^2$.

Ejemplo 5: Si $\{X_i\}_{i\in N}$ son v.as iid $\sim \text{Exponencial}(\text{con media } \theta)$, halle el límite casi seguro de $\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right]$.

En este caso hacemos:

$$Z_i = X_i^2 \Rightarrow \left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] \to E\left(X^2\right) = V\left(X\right) + \left[E\left(X\right)\right]^2 = \theta^2 + [\theta]^2 = 2\theta^2 \text{ c.s}$$

Ejemplo 6: Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias independientes con la función de probabilidad dada abajo. Diga si se cumple la Ley de los Grandes Números.

a. $P\left(X_k = \pm 2^k\right) = \frac{1}{2}$, $k = 1, 2, 3, \ldots$

Solución

X_k	2^k	-2^k
$P(X_k = x_k)$	$\frac{1}{2}$	$\frac{1}{2}$

$$\mu = E\left(X_k\right) = \left(2^k\right)\left(\frac{1}{2}\right) + \left(-2^k\right)\left(\frac{1}{2}\right) = 0$$

$$E\left(X_k^2\right) = \left(2^k\right)^2\left(\frac{1}{2}\right) + \left(-2^k\right)^2\left(\frac{1}{2}\right) = 2^{2k}$$

$$\sigma^2 = Var\left(X_k\right) = E\left(X_k^2\right) - \left[E\left(X_k\right)\right]^2 = 2^{2k} - 0^2 = 2^{2k}$$

σ^2 depende de k, y k puede tomar valores que tienden al infinito.

$$\lim_{n\to\infty} P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \le \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = \lim_{n\to\infty} \frac{2^{2k}}{n\varepsilon^2} \ne 0$$

$\to$ no se cumple la Ley de los Grandes Números.

b. $P\left(X_k = \pm 2^k\right) = \frac{1}{2^{2k+1}}$, $P\left(X_k = 0\right) = 1 - \frac{1}{2^{2k}}$, $k = 1, 2, 3, \ldots$

Solución

Se cumple que:

$$E\left(X_k\right) = -2^k \times \frac{1}{2^{2k+1}} + 2^k \times \frac{1}{2^{2k+1}} + 0 \times \left[1 - \frac{1}{2^{2k}}\right] = 0$$

$$\mu = E\left(X_k\right) = 0$$

$\sigma^2 = Var\left(X_k\right) = 1$, no depende de k.

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|>\varepsilon\right)\leq \lim_{n\to\infty}\frac{\sigma^2}{n\varepsilon^2}=\lim_{n\to\infty}\frac{1}{n\varepsilon^2}=0$$

$\rightarrow$ se cumple la Ley de los Grandes Números.

Ejemplo 7: Con una muestra aleatoria de tamaño n de $f\left(x\right)=\dfrac{e^{-\frac{x}{\lambda}}}{\lambda}I_{(0,\infty)}\left(x\right)$, diga si se cumple la ley de los grandes números.

Solución

$$\mu=E\left(X\right)=\lambda$$

$$\sigma^2=Var\left(X\right)=\lambda^2$$

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|>\varepsilon\right)\leq \lim_{n\to\infty}\frac{\sigma^2}{n\varepsilon^2}=\lim_{n\to\infty}\frac{\lambda^2}{n\varepsilon^2}=0$$

$\rightarrow$ se cumple la Ley de los Grandes Números.

Ejemplo 8: En la estimación de μ con la $\overline{X}$, suponga que deseamos con un 95% de seguridad que la media muestral no se desvíe de la media poblacional en más de.

a. $\dfrac{\sigma}{5}$. Calcule el tamaño de muestra mínimo requerido.

Solución

$$P\left(\left|\overline{X}-\mu\right|<\frac{\sigma}{5}\right)>0.95$$

$$\varepsilon=\frac{\sigma}{5},\quad \delta=0.05$$

$$n>\frac{\sigma^2}{\left(\frac{\sigma}{5}\right)^2(0.05)}=500$$

b. $\dfrac{\sigma}{10}$. Calcule el tamaño de muestra requerido. (Queda como ejercicio)

c. $\dfrac{\sigma}{20}$. Calcule el tamaño de muestra requerido. (Queda como ejercicio)

Ejemplo 9: Una empresa encuestadora desea determinar la proporción p de votantes que están a favor del candidato EI, basada en una muestra de tamaño n. La empresa desea usar la proporción muestral como estimadora de p. Si la empresa quiere que el

estimador no difiera de p en más de 3% con probabilidad de por lo menos 0.95. ¿Cuál es el tamaño de muestra mínimo que debe tomar?
Solución

a. Use la Ley Débil de los Grandes Números.
Solución

Si $X_1,\ldots,X_n$ es una m.a de una distribución $\text{Binomial}(1,p) \rightarrow \frac{X}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ es la proporción de votantes a favor del candidato en la muestra. Según el problema:

$$P\left(\left|\frac{X}{n} - p\right| < 0.03\right) > 0.95$$

$\varepsilon = 0.03, \quad \delta = 0.05$

Trabajando con la máxima variancia de la distribución $\text{Binomial}(1,p)$:

Maximizando $\sigma^2 = p(1-p) = p - p^2 \rightarrow \frac{d\sigma^2}{dp} = 1 - 2p = 0 \rightarrow p = \frac{1}{2} \rightarrow \sigma^2 = \frac{1}{4}$

$$n > \frac{\sigma^2}{\varepsilon^2 \delta} = \frac{\frac{1}{4}}{(0.03)^2 (0.05)} = 5555.5 \rightarrow n \geq 5556$$

El tamaño de muestra mínimo es 5556.

b. Use la aproximación a la normal.
Solución

Considerando la máxima variancia de la distribución $\text{Binomial}(1,p)$.

$$P\left(\frac{\left|\frac{X}{n} - p\right|}{\frac{\sigma}{\sqrt{n}}} < \frac{0.03}{\frac{\sqrt{\frac{1}{4}}}{\sqrt{n}}}\right) \geq 0.95 \rightarrow P\left(|Z| < 0.06\sqrt{n}\right) \geq 0.95 = P\left(-0.06\sqrt{n} Z < 0.06\sqrt{n}\right) \geq 0.95$$

$$= F\left(0.06\sqrt{n}\right) - F\left(-0.06\sqrt{n}\right) \geq 0.95$$

$$\rightarrow 0.06\sqrt{n} = 1.96 \rightarrow n \geq 1067.1$$

El n mínimo es 1068.

EJERCICIOS

1. Demuestre que el momento muestral r en torno a cero converge en probabilidades al momento poblacional r en torno a cero.

   $$\lim_{n\to\infty} P\left(\left|M_r^{'} - \mu_r^{'}\right| > \varepsilon\right) \le \lim_{n\to\infty} \frac{\frac{1}{n}\left[\mu_{2r}^{|} - \left(\mu_r^{|}\right)^2\right]}{\varepsilon^2} = 0$$

   $\rightarrow$ se cumple la Ley de los Grandes Números.

2. ¿El tercer momento muestral en torno a cero converge en probabilidad al tercer momento poblacional en torno a cero? Justifique.

Convergencia en distribución

Se dice que X_n converge a la distribución de X si $\lim_{n\to\infty} P\left(X_n \le x\right) = P\left(X \le x\right)$ o $\lim_{n\to\infty} F_n\left(x\right) = F\left(x\right)$, donde F_n es la función de distribución acumulada (fda) de X_n, F es la fda de X, y x es el punto de continuidad de $F\left(x\right)$. Este tipo de convergencia se usa para formular el teorema del límite central, ver abajo. A veces se usa la notación $X_n \simeq F$ para indicar que X_n converge a una variable aleatoria con fda F.

Teorema del límite central

Si $X_1, X_2, X_3, \ldots$ son variables aleatorias iid con media $E\left(X_i\right) = \mu$ y varianza $Var\left(X_i\right) = \sigma^2$ entonces la distribución de la suma estandarizada converge a la distribución normal estándar: $\dfrac{\sum_{i=1}^{n}\left(X_i - \mu\right)}{\sigma\sqrt{n}} \rightarrow \text{Normal}(0,1),\ n \rightarrow \infty$. (Se dice que converge en distribución a la distribución normal estándar).

Prueba:

$$\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i$$

donde las: $Z_i = \dfrac{X_i - \mu}{\sigma}$ son variables aleatorias estandarizadas iid

alrededor de $t = 0$.

$$\Psi_Z\left(\frac{t}{\sqrt{n}}\right) = E\left(e^{\frac{t}{\sqrt{n}}Z}\right) = 1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2!} + 0\left[\left(\frac{t}{\sqrt{n}}\right)^2\right]$$

Así, si t se fija, se obtiene

$$\lim_{n\to\infty}\Psi(t)=\lim_{n\to\infty}\left\{1+\frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2!}+0\left[\left(\frac{t}{\sqrt{n}}\right)^2\right]\right\}^n=\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n$$

donde $a_n=\frac{t^2}{2}+n\times 0\left[\left(\frac{t}{\sqrt{n}}\right)^2\right]\Rightarrow\lim_{n\to\infty}a_n=\frac{t^2}{2}$. Del cálculo se conoce que:

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n=\left[\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^{\frac{n}{a_n}}\right]^{a_n}=\lim_{n\to\infty}e^{a_n}=e^{\frac{t^2}{2}}$$

$\therefore$ la fgm de $\dfrac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{\sigma}=\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i$ converge a $e^{\frac{t^2}{2}}$ que es la fgm de de la distribucion normal estándar.

Ejemplo 10: Si $U_1,\ldots,U_n$ es una muestra aleatoria de una distribución $\text{Uniforme}(0,1)$ entonces para n suficientemente grande el Teorema del Límite Central dice que:

$$\frac{\sum_{i=1}^{n}U_i-E\left(\sum_{i=1}^{n}U_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n}U_i\right)}}=\frac{\sum_{i=1}^{n}U_i-\sum_{i=1}^{n}E(U_i)}{\sqrt{Var\left(\sum_{i=1}^{n}U_i\right)}}=\frac{\sum_{i=1}^{n}U_i-\sum_{i=1}^{n}E(U_i)}{\sqrt{\sum_{i=1}^{n}Var(U_i)}}=\frac{\sum_{i=1}^{n}U_i-\frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

tiene una distribución aproximadamente $N(0,1)$.

Suponga que: 0.25, 0.88, 0.96, 0.18, 0.29, 0.72, 0.48, 0.62, 0.55, 0.44, 0.07, 0.09, 0.16, 0.22, 0.35, 0.94, 0.63, 0.12, 0.08, 0.21, 0.36, 0.49, 0.51, 0.53, 0.62, 0.74, 0.05, 0.91, 0.88, 0.24, 0.33, 0.32 es una muestra aleatoria de tamaño 32 de una distribución $\text{Uniforme}(0,1)$.

Halle una muestra aleatoria de tamaño 1 de una distribución $N(0,1)$.

Solución

$$Z=\frac{\sum_{i=1}^{n}U_i-\frac{n}{2}}{\sqrt{\frac{n}{12}}}=\frac{14.22-\frac{32}{2}}{\sqrt{\frac{32}{12}}}=-1.09002$$

Ejemplo 11: Un antropólogo quiere estimar la estatura promedio de los hombres Masáis. Si supone que la desviación estándar de la población es de 1.25 cm y se selecciona al azar a 100 hombres Masáis, encuentre la cota de la probabilidad de que la diferencia entre la media de la muestra y la media verdadera de la población no exceda de 0.25 cm.

a. Utilice el teorema del límite central.

$n = 100$ se considera grande por ser mayor que 30.

$$\bar{X} \sim N\left(\mu \ , \ \frac{\sigma^2}{n} = \frac{1.25^2}{100} = 0.015625\right)$$

$$P\left(\left|\bar{X}-\mu\right| < 0.25\right) = P\left(|Z| < \frac{0.25}{\sqrt{0.015625}}\right) = P\left(|Z| < 2\right) = P\left(-2 < Z < 2\right) = 0.9544$$

b. Utilice la ley de los grandes números.
$n = 100$ se considera grande por ser mayor que 30.

$$P\left(\left|\bar{X}-\mu\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2 n} \rightarrow P\left(\left|\bar{X}-\mu\right| < 0.25\right) \geq 1 - \frac{1.25^2}{0.25^2 \times 100} \geq 0.75$$

Ejemplo 12: Halle la distribución exacta y la distribución límite de $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, si la muestra aleatoria $X_1, \ldots, X_n$ se tomó de una distribución $\text{Binomial}(1, p)$.

Determinación de la distribución exacta
Se sabe que:

$$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p) \rightarrow \text{la distribución exacta de } \bar{X} = \frac{Y}{n} \text{ es:}$$

$$f_{\bar{X}}\left(\bar{x}\right) = P\left(\bar{X} = \bar{x}\right) = P\left(\frac{Y}{n} = \bar{x}\right) = P\left(Y = n\bar{x}\right) = \binom{n}{n\bar{x}} p^{n\bar{x}} \left(1-p\right)^{n-n\bar{x}} I_{\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}}\left(\bar{x}\right)$$

Determinación de la distribución límite

$$\mu_{\bar{X}} = E\left(\frac{Y}{n}\right) = \frac{1}{n} E\left(Y\right) = \frac{np}{n} = p$$

$$\sigma^2_{\bar{X}} = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2} Var\left(Y\right) = \frac{np\left(1-p\right)}{n^2} = \frac{p\left(1-p\right)}{n}$$

$$\Rightarrow \bar{X} \underset{\text{aprox}}{\sim} \text{Normal}\left(\mu = p, \frac{\sigma^2}{n} = \frac{p\left(1-p\right)}{n}\right), \text{ o también:}$$

$$\rightarrow \text{La distribución límite será: } Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - p}{\sqrt{\frac{p\left(1-p\right)}{n}}} \sim N\left(0,1\right)$$

Ejemplo 13: Demuestre que el momento muestral r en torno a cero converge en distribución a la distribución normal.

$$M_r^{'} = \frac{1}{n}\sum_{i=1}^{n} X_i^r \text{ y } \mu_r^{'} = E\left(X_i^r\right)$$

$$nM_r^{'} = \sum_{i=1}^{n} X_i^r$$

$$E\left(nM_r^{'}\right) = nE\left(M_r^{'}\right) = n\mu_r^{'}$$

$$Var\left(nM_r^{'}\right) = n^2 Var\left(M_r^{'}\right) = n^2\left[\frac{\mu_{2r}^{'} - \left(\mu_r^{'}\right)^2}{n}\right]$$

Aplicando el teorema conocido como el Teorema de Lindeberg-Levy, resulta:

$$\frac{nM_r^{'} - E\left(nM_r^{'}\right)}{\sqrt{Var\left(nM_r^{'}\right)}} = \frac{n\left[M_r^{'} - E\left(M_r^{'}\right)\right]}{n\sqrt{Var\left(M_r^{'}\right)}} = \frac{M_r^{'} - E\left(M_r^{'}\right)}{\sqrt{Var\left(M_r^{'}\right)}}$$

$$\rightarrow \lim_{n\to\infty} \frac{M_r^{'} - E\left(M_r^{'}\right)}{\sqrt{Var\left(M_r^{'}\right)}} = N\left(0,1\right)$$

Es decir $M_r^{'}$ converge en distribución a: $N\left(\mu_r^{'}, \frac{\mu_{2r}^{'} - \left(\mu_r^{'}\right)^2}{n}\right)$

Ejemplo Si $X_1,\ldots,X_n$ es una muestra aleatoria de una distribución $\text{Uniforme}\left(0,1\right)$, entonces la siguiente variable aleatoria estandarizada $Z_n = \frac{Y_n - E\left(Y_n\right)}{\sqrt{Var\left(Y_n\right)}}$ no converge a la distribución normal estándar cuando $n \to \infty$, porque $Y_n = \max\left(X_1,\ldots,X_n\right)$ no es una suma.

<u>EJERCICIOS:</u>

1.) Hallar la distribución exacta y la distribución límite de $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, si la muestra aleatoria: $X_1,\ldots,X_n$ proviene de la distribución: a) $\text{Binomial}\left(r,p\right)$ b) $\text{Poisson}\left(\lambda\right)$ c) $\text{Exponencial}\left(\text{con media } \frac{1}{\lambda}\right)$.

Poisson

$$Y = \sum_{i=1}^{n} X_i \sim \text{Poisson}\left(n\lambda\right) \rightarrow \text{la distribución exacta de } \overline{X} = \frac{Y}{n} \text{ es:}$$

$$f_{\overline{X}}\left(\overline{x}\right) = P\left(\overline{X} = \overline{x}\right) = P\left(\frac{Y}{n} = \overline{x}\right) = P\left(Y = n\overline{x}\right) = \frac{e^{-n\lambda}\left(n\lambda\right)^{n\overline{x}}}{\left(n\overline{x}\right)!} I_{\left\{0,\frac{1}{n},\frac{2}{n},\ldots,\infty\right\}}\left(\overline{x}\right)$$

Determinación de la distribución límite

$$\mu_{\bar{X}} = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{n\lambda}{n} = \lambda$$

$$\sigma^2_{\bar{X}} = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2}Var(Y) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

$\Rightarrow \bar{X} \underset{\text{aprox}}{\sim} \text{Normal}\left(\mu = \lambda, \frac{\sigma^2}{n} = \frac{\lambda}{n}\right)$, o también:

$\rightarrow$ La distribución límite será: $Z = \dfrac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \dfrac{\bar{X} - \lambda}{\sqrt{\dfrac{\lambda}{n}}} \sim N(0,1)$

Exponencial

La distribución exacta $\bar{X} \sim \text{Gamma}\left[r = n, \lambda = \lambda n\right]$

La distribución límite $\bar{X} \underset{\text{aprox}}{\sim} \text{Normal}\left[\mu = \frac{1}{\lambda}, V\left(\bar{X}\right) = \frac{1}{n\lambda^2}\right]$

2.) Según la Ingeniera Chabuca Granda el rendimiento de jugo de la variedad lima dulce tiene distribución $\text{Gamma}(r = 90, \lambda = 2)$. Si $X_1, \ldots, X_n$ es una muestra aleatoria de esa distribución, primero halle la distribución exacta de $\bar{X}$ y luego la distribución límite. Justifique.

3.) Suponga que con la variedad lima naranja: En el laboratorio A las unidades experimentales (10 frutos) se ponen en recipientes para posteriormente ser analizadas. Se estima que el 85% de las unidades experimentales tienen un rendimiento de jugo mayor de 44%. Diariamente se seleccionan al azar una muestra de 4 unidades experimentales de un recipiente para ser analizadas. Considere que la variable aleatoria X se define como el número de unidades experimentales, en la muestra, con un rendimiento de jugo mayor de 44%. En el laboratorio B se estima que el 85% de las unidades experimentales (10 frutos) tienen un rendimiento de jugo mayor de 44% y diariamente se selecciona al azar una muestra de 6 unidades experimentales de un recipiente para ser analizadas. Considere que la variable aleatoria Y se define como el número de unidades experimentales, en la muestra, con un rendimiento de jugo mayor de 44%. En el laboratorio C se estima que el 85% de las unidades experimentales tienen un rendimiento de jugo mayor de 44% y diariamente se selecciona al azar una muestra de 5 unidades experimentales de un recipiente para ser analizadas. Considere que la variable aleatoria Z se define como el número de unidades experimentales, en la muestra, con un rendimiento de jugo mayor de 44%. Para un día en particular determine la función de probabilidad exacta del promedio de unidades experimentales con un rendimiento de jugo mayor de 44%.

4.) Los tiempos de espera para los clientes que pasan por una caja registradora a la salida de una tienda de mercado son v.a independientes con una media de 1.5 minutos y una varianza de 1.0, aproxime la probabilidad de que se pueda atender a 100 clientes en menos de 2 horas.

5.) Se conectan 32 focos de luz infrarroja en un invernadero, de tal manera que si falla un foco, otro se enciende inmediatamente. (Se enciende sólo un foco a la vez) Los focos funcionan independientemente y cada uno tiene una vida media de 50 horas y una desviación estándar de 4 horas. Si no se inspecciona el invernadero durante 1664 horas después de encender el sistema de focos ¿Cuál es la probabilidad de que un foco esté encendido al final del periodo de 1664 horas?

6.) Supóngase que $X_1,\ldots,X_n$ son v.a independientes, cada una con media μ_1 y varianza σ_1^2. Supóngase que $Y_1,\ldots,Y_n$ son también v.a independientes cada uno con media μ_2 y varianza σ_2^2. Demuestre que la v.a:

$$Un = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{n}}}$$

Satisface las condiciones del teorema del límite central.

DISTRIBUCIONES ASOCIADAS A LA NORMAL

La Distribución Ji-Cuadrado

Si $X \sim \text{Gamma}\left(\frac{k}{2},\frac{1}{2}\right)$ se dice que $X \sim \text{Ji Cuadrado}(k)$ (X tiene distribución Ji Cuadrado con k grados de libertad) y su densidad es $f(x) = \dfrac{\left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-\frac{1}{2}x}}{\Gamma\left(\frac{k}{2}\right)} I_{[0,\infty)}(x)$.

Teorema

Si $X \sim \text{Ji Cuadrado}(k)$ entonces $E(X) = k, Var(X) = 2k$ y $\Psi_X(t) = \left(\dfrac{1}{1-2t}\right)^{\frac{k}{2}}$.

Teorema

Si $X_1,\ldots,X_n$ son variables aleatorias independientes tales que $X_i \sim \chi^2(k_i)$, $i = 1,2,\ldots,n$ entonces $\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right)$.

Corolario 1

Si $Z \sim N(0,1) \rightarrow Z^2 \sim \chi^2(1)$

Corolario 2

Si $Z_1,\ldots,Z_n$ es una muestra aleatoria de una distribución $N(0,1)$ entonces $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$.

Nota: Ver el Script

Corolario 3

Sean las variables independientes $X_i \sim N\left(\mu_i,\sigma_i^2\right),\ i=1,2,\ldots,n$ entonces $\sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2(n)$.

Corolario 4

Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución $N\left(\mu,\sigma^2\right)$ entonces $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$.

Ejemplo: Si $X \sim \chi^2_{k_1}$ e $Y \sim \chi^2_{k_2}$ son independientes entonces $X-Y \sim \chi^2_{k_1-k_2}$ siempre que $k_1-k_2>0$.

TEOREMA

Si $X_1,\ldots,X_n$ es una muestra aleatoria de una distribución $N\left(\mu,\sigma^2\right)$ entonces:

1. $\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$
2. Las estadísticas: $\overline{X}$ y $\sum_{i=1}^{n}(X_i-\overline{X})^2$ son independientes
3. Y = $\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

PRUEBA:

1. Ya se demostró.

2. $\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$; dividiendo entre σ^2 y reescribiendo convenientemente:

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

$$U \quad = \quad Y \quad + \quad V$$

Como $U\sim\chi^2_{(n)}$; $V\sim\chi^2_{(1)}$ pues : $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim \text{Normal}(0,1)$

También Y y V son independientes (parte 2) entonces:

$$\Psi_U(t)=\Psi_{Y+V}(t)=\Psi_Y(t)\Psi_V(t)$$

$$\Psi_Y(t)=\frac{\Psi_U(t)}{\Psi_V(t)}=\frac{\left(\frac{1}{1-2t}\right)^{n/2}}{\left(\frac{1}{1-2t}\right)^{1/2}}=\left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}} \quad t<\tfrac{1}{2}\ ,\ \therefore\ Y=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}\ .$$

Ejemplo 14: Sea $X_1,\ldots,X_n$ una muestra aleatoria de una distribución $N\left(\mu,\sigma^2\right)$. Se sabe que $E\left(S^2\right)=\sigma^2$. Obtenga $E(S)$.

$$S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}\rightarrow S^2=\frac{\sigma^2}{n-1}Y\rightarrow S=\frac{\sigma}{\sqrt{n-1}}\sqrt{Y}\rightarrow$$

$$E(S)=\frac{\sigma}{\sqrt{n-1}}E\left(\sqrt{Y}\right)=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty y^{\frac{1}{2}}\left[\frac{2^{-\frac{(n-1)}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}y^{\frac{n-1}{2}-1}e^{-\frac{1}{2}y}\right]dy=$$

$$E(S)=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty\left[\frac{2^{-\frac{(n-1)}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}y^{\frac{n-1}{2}-\frac{1}{2}}e^{-\frac{1}{2}y}\right]dy=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty\left[\frac{2^{-\frac{n}{2}+\frac{1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}\right]dy=$$

$$E(S)=\frac{\sigma}{\sqrt{n-1}}\left[\frac{2^{-\frac{n}{2}+\frac{1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\right]\int_0^\infty y^{\frac{n}{2}-1}e^{-\frac{1}{2}y}dy=\frac{\sigma}{\sqrt{n-1}}\left[\frac{2^{-\frac{n}{2}+\frac{1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\right]\left[\frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{1}{2}\right)^{\frac{n}{2}}}\right]=$$

$$E(S)=\frac{\sigma}{\sqrt{n-1}}\left(2^{-\frac{n}{2}+\frac{1}{2}}\times 2^{\frac{n}{2}}\right)\left[\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right]=\frac{\sqrt{2}}{\sqrt{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\sigma\neq\sigma$$

Ejemplo 15: Demuestre que si $X_1,\ldots,X_n$ una m.a. de una distribución $N(\mu,\sigma^2)$ entonces las variables $\bar{X}$ y $X_i-\bar{X}$ son independientes para todo i.

Como las dos variables son normales entonces es suficiente verificar que su covariancia es cero (esto indica que no están correlacionadas entonces son independientes).

$$Cov\left(\bar{X},X_i-\bar{X}\right)=E\left[\bar{X}\left(X_i-\bar{X}\right)\right]-E\left(\bar{X}\right)\underbrace{E\left(X_i-\bar{X}\right)}_{\text{cero}}=E\left(\bar{X}X_i-\bar{X}^2\right)=$$

$$=\underbrace{E\left[\left(\frac{X_1}{n}+\frac{X_2}{n}+\cdots+\frac{X_i}{n}+\cdots\frac{X_n}{n}\right)X_i\right]}_{A}-\underbrace{E\left(\bar{X}^2\right)}_{B}=A-B$$

$$A=\frac{n-1}{n}E\left(X_1X_i\right)+\frac{1}{n}E\left(X_i^2\right)=\frac{n-1}{n}E\left(X_1\right)E\left(X_i\right)+\frac{1}{n}\left\{Var\left(X_i\right)+\left[E\left(X_i\right)\right]^2\right\}=$$

Nota: $V(T)=E\left(T^2\right)-\left[E(T)\right]^2\Rightarrow E\left(T^2\right)=V(T)+\left[E(T)\right]^2$

$$A=\frac{n-1}{n}\mu^2+\frac{1}{n}\left(\sigma^2+\mu^2\right)=\mu^2+\frac{\sigma^2}{n}$$

$$B=E\left(\bar{X}^2\right)=Var\left(\bar{X}\right)+\left[E\left(\bar{X}\right)\right]^2=\frac{\sigma^2}{n}+\mu^2$$

$$\therefore Cov\left(\bar{X},X_i-\bar{X}\right)=A-B=\left(\mu^2+\frac{\sigma^2}{n}\right)-\left(\frac{\sigma^2}{n}+\mu^2\right)=0$$

Ejemplo 16: De la distribución $N\left(\mu=4,\sigma^2=2\right)$ se extrae una muestra aleatoria de tamaño 8. Calcule la $P\left(\bar{X}>3.5\ ;\ S^2>2.2\right)$.

Se sabe que $\bar{X}\sim N\left(4,\frac{2}{8}\right)$ y $\frac{(n-1)S^2}{\sigma^2}=\frac{7S^2}{2}\sim\chi^2_{(7)}\rightarrow$

Como $\bar{X}$ y S^2 son independientes $\rightarrow$

$$P\left(\bar{X} > 3.5 \;\; ; \;\; S^2 > 2.2\right) = P\left(\bar{X} > 3.5\right)P\left(S^2 > 2.2\right) = P\left(Z > \frac{(3.5-4)\sqrt{8}}{\sqrt{2}}\right)P\left(\chi^2_{(7)} > \frac{2.2\times 7}{2}\right)$$

$$P\left(\bar{X} > 3.5 \;\; ; \;\; S^2 > 2.2\right) = P\left(Z > -1\right)P\left(\chi^2_{(7)} > 7.7\right) = 0.8413\times 0.3598 = 0.3027$$

La distribución t de Student

Teorema

Si las v.as $Z \sim N(0,1)$ y $V \sim \chi^2_{(m)}$ son independientes, entonces la v.a $X = \dfrac{Z}{\sqrt{\dfrac{V}{m}}} \sim t_{(m)}$.

Una v.a X tiene distribución t con m grados de libertad si su función de probabilidad es:

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

Prueba

Por independencia:

$$f_{Z,V}(z,v) = \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\right]\left[\frac{(1/2)^{\frac{m}{2}} v^{\frac{m}{2}-1} e^{-\frac{v}{2}}}{\Gamma\left(\dfrac{m}{2}\right)}\right], -\infty < z < \infty, 0 < v < \infty.$$

Luego

$$X = \frac{Z}{\sqrt{\dfrac{V}{m}}} \qquad V = Y$$

La función $\Phi(x,y) = (\Phi_1(x,y), \Phi_2(x,y)) = \left(x\sqrt{y/m}, y\right) = (z,v)$ define una transformación uno a uno de la región $B = \{(x,y) / -\infty < x < \infty,\, 0 < y < \infty\}$ hacia la región $A = \{(z,v) / -\infty < z < \infty,\, 0 < v < \infty\}$. El jacobiano de la transformación está dado por

$$J = \begin{vmatrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} \sqrt{\dfrac{y}{m}} & \dfrac{x}{2}\dfrac{1}{\sqrt{my}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y}{m}}$$

la distribución conjunta es:

$$g_{X,Y}(x,y)= f_{Z,V}\left(x\sqrt{\frac{y}{m}},y\right)|J|$$

$$g_{X,Y}(x,y)=\frac{1}{\sqrt{2\pi}\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}}y^{\frac{m}{2}-1}e^{\left[-\frac{y}{2}\left(1+\frac{x^2}{m}\right)\right]}\sqrt{\frac{y}{m}},-\infty<x<\infty,0<y<\infty.$$

Por lo tanto, la densidad de X es:

$$g_X(x)=\int_0^\infty g_{X,Y}(x,y)\,dy=\int_0^\infty \frac{1}{\sqrt{2\pi}\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}}y^{\frac{m}{2}-1}e^{\left[-\frac{y}{2}\left(1+\frac{x^2}{m}\right)\right]}\sqrt{\frac{y}{m}}dy.$$

$$g_X(x)=\frac{1}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)2^{\frac{m+1}{2}}}\underbrace{\int_0^\infty y^{\frac{m+1}{2}-1}e^{\left[\frac{-y}{2\left(1+\frac{x^2}{m}\right)^{-1}}\right]}dy}_{\text{Gamma}}=\frac{1}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)2^{\frac{m+1}{2}}}.\Gamma\left(\frac{m+1}{2}\right)\left[2\left(1+\frac{x^2}{m}\right)^{-1}\right]^{\frac{m+1}{2}}$$

$$g_X(x)=\frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)}.\frac{1}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}I_{(-\infty,\infty)}(x)$$

Por lo tanto, X tiene una distribución t con m grados de libertad

Teorema

Si $X\sim t_{(m)}$, entonces $\mu_X=0$ y $\sigma_X^2=\frac{m}{m-2}$.

Ejemplo 17: De la distribución $N\left(\mu=-2,\sigma^2=\text{Desconocida}\right)$ se extrae una muestra aleatoria de tamaño 8 (-4, 0, 5, -6, 1, 2, -5, 1). Calcule $P\left(\bar{X}<-0.6\right)$.

$$Z=\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{\sigma}\qquad V=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)\to X=\frac{Z}{\sqrt{\frac{V}{n-1}}}=\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{S}\sim t(n-1)$$

$$\frac{\left(\bar{X}-\mu\right)\sqrt{n}}{S}=\frac{\left(\bar{X}+2\right)\sqrt{8}}{\sqrt{14.78571}}\sim t(n-1=7)$$

$$P\left(\bar{X}<-0.6\right)=P\left[t(7)<\frac{(-0.6+2)\sqrt{8}}{\sqrt{14.78571}}\right]=P\left[t(7)<1.029798\right]=0.8313164$$

La distribución t no central

Si las v.as $Z \sim N\left(\mu,\sigma^2\right)$ y $V \sim \chi^2_{(m)}$ son independientes, entonces la v.a $X = \dfrac{\frac{Z}{\sigma}}{\sqrt{\frac{V}{m}}} \sim t_{\left(m,\frac{\mu}{\sigma}\right)}$.

Se dice que X tiene distribución t no central con m grados de libertad y parámetro de no centralidad μ.

Nota: En pruebas de hipótesis para una media de una distribución normal el parámetro de no centralidad cambia y en pruebas de hipótesis para diferencia de medias de dos distribuciones normales el parámetro de no centralidad es otro.

La distribución F

Teorema

Si $U \sim \chi^2_{(n)}$ y $V \sim \chi^2_{(m)}$, son variables aleatorias independientes, entonces la variable aleatoria $X = \dfrac{\frac{U}{n}}{\frac{V}{m}} = \dfrac{U}{V}\dfrac{m}{n} \sim F\left(n,m\right)$

Una variable aleatoria X tiene una distribución F con n y m grados de libertad si su función de probabilidad es:

$$f\left(x\right) = \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)\left[1+\frac{n\,x}{m}\right]^{\frac{n+m}{2}}} \quad , \quad x > 0$$

Prueba

Por independencia:

$$f_{U,V}\left(u,v\right) = \frac{1}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)2^{(n+m)/2}} u^{\frac{n}{2}-1} v^{\frac{m}{2}-1} e^{-\frac{(u+v)}{2}} ,0 < u < \infty, 0 < v < \infty.$$

Luego

$$X = \frac{\frac{U}{n}}{\frac{V}{n}} \qquad\qquad V = Y$$

La función $\Phi\left(x,y\right) = \left(\Phi_1(x,y),\Phi_2\left(x,y\right)\right) = \left(\left(nxy\right)/m, y\right) = \left(u,v\right)$ define una transformación uno a uno de $B = \left\{\left(x,y\right)/\, 0 < x < \infty, 0 < y < \infty.\right\}$ hacia la región $A = \left\{\left(u,v\right)/\, 0 < u < \infty, 0 < v < \infty\right\}$. El jacobiano de la transformación está dado por

$$J = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{n}{m} y & \frac{n}{m} x \\ 0 & 1 \end{vmatrix} = \frac{n}{m} y$$ la distribución conjunta es:

$$g_{X,Y}(x,y) = f_{U,V}\left(\frac{n}{m} xy, y\right)|J|$$

$$g_{X,Y}(x,y) = \frac{1}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)2^{(n+m)/2}}\left(\frac{n}{m} xy\right)^{\frac{n}{2}-1} y^{\frac{m}{2}-1} e^{-\frac{y}{2}\left(\frac{n}{m}x+1\right)} \frac{n}{m} y, 0<x<\infty, 0<y<\infty$$

$$g_{X,Y}(x,y) = \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)2^{\frac{(n+m)}{2}}} . y^{\frac{n+m}{2}-1} e^{-\frac{y}{2}\left(\frac{n}{m}x+1\right)}, \quad 0<x<\infty, 0<y<\infty$$

Por lo tanto, la densidad de X es:

$$g_X(x) = \int_0^\infty g_{X,Y}(x,y)\,dy = \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)2^{(n+m)/2}} \underbrace{\int_0^\infty y^{\frac{(n+m)}{2}-1} e^{\frac{-y}{2\left(\frac{n}{m}x+1\right)^{-1}}} dy}_{\text{Gamma}}$$

$$g_X(x) = \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)2^{\frac{n+m}{2}}} \Gamma\left(\frac{n+m}{2}\right)\left[2\left(\frac{n}{m}x+1\right)^{-1}\right]^{\frac{n+m}{2}}$$

$$g_X(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} . \frac{x^{\frac{n}{2}-1}}{\left[\frac{n}{m}x+1\right]^{\frac{n+m}{2}}} I_{(0,\infty)}(x)$$

Por lo tanto, X tiene una distribución F con n grados de libertad en el numerador y m en el denominador

Corolario Si $U \sim \chi^2_{(n)}$ y $V \sim \chi^2_{(m)}$, son variables aleatorias independientes, entonces la variable aleatoria $\frac{1}{X} = \frac{\frac{V}{m}}{\frac{U}{n}} = \frac{V}{U}\frac{n}{m} \sim F(m,n)$

Teorema

Si $X \sim F(n,m)$ entonces $E(X) = \frac{m}{m-2}$, $m > 2$ y $Var(X) = \frac{2m(n+m-2)}{n(m-2)^2(m-4)}, m > 4$

Teorema

Si $X \sim t(m) \to X^2 \sim F(1,m)$.

Ejemplo 18: Si $A \sim F(1,10)$ y $P(1.4 < A < 10.8) = 0.2558886$ entonces $X = \pm\sqrt{A} \sim t(10)$.
Se verifica que

$$P(1.4 < A < 10.8) = P(A < 10.8) - P(A < 1.4) = P(X^2 < 10.8) - P(X^2 < 1.4) =$$

$$= P(0 < X^2 < 10.8) - P(0 < X^2 < 1.4) = P(-\sqrt{10.8} < X < 0) + P(0 < X < \sqrt{10.8})$$

$$-\left[P(-\sqrt{1.4} < X < 0) + P(0 < X < \sqrt{1.4})\right] == 2\times 0.4958997 - 2\times 0.3679554 = 0.2558886$$

Ejemplo 19: La velocidad de las moléculas de cierto gas es $V = \sqrt{X^2 + Y^2 + Z^2}$ donde X, Y y Z son variables aleatorias independientes con distribución $N(0,\sigma^2)$. Sean A y B dos moléculas que se mueven independientemente una de otra. Calcule la probabilidad de que B se desplace con una velocidad 5.4278 veces o más que la de A.

$$\frac{V^2}{\sigma^2} = \frac{X^2}{\sigma^2} + \frac{Y^2}{\sigma^2} + \frac{Z^2}{\sigma^2} = \frac{(X-\mu)^2}{\sigma^2} + \frac{(Y-\mu)^2}{\sigma^2} + \frac{(Z-\mu)^2}{\sigma^2} \sim \chi^2_{(3)}$$

Sean V_A y V_B las velocidades de las moléculas A y B.

Como estas velocidades son independientes $\to$

$$F = \frac{\frac{V_B^2}{\sigma^2}\Big/3}{\frac{V_A^2}{\sigma^2}\Big/3} = \frac{V_B^2}{V_A^2} \sim F(3,3) \to$$

$$P(V_B \geq 5.4278 V_A) = P\left(\frac{V_B}{V_A} \geq 5.4278\right) = P\left(\frac{V_B^2}{V_A^2} \geq 29.46\right) = P\left[F(3,3) \geq 29.46\right] = 0.009998383$$

Ejemplo 20: Suponga que $X_1,...,X_{18}$ y $Y_1,...,Y_{20}$ son muestras aleatorias independientes de las distribuciones $N(8,4)$ y $N(8,5)$ respectivamente, y sean $\bar{X}, \bar{Y}, S_X^2$ y S_y^2 las respectivas medias y variancias muestrales. Si $P\left(\frac{\bar{X}-\bar{Y}}{S_1} \le k\right) = 0.95$, halle el valor de k

Solución:

$X_1,...,X_{18} \sim N(8,4) \quad \bar{X} \quad S_1^2$

$Y_1,...,Y_{20} \sim N(8,5) \quad \bar{Y} \quad S_2^2$

Si $P\left(\frac{\bar{X}-\bar{Y}}{S_1} \le k\right) = 0.95$

$W = \bar{X} - \bar{Y}$

$\mu_W = 8 - 8 = 0$

$\sigma_W^2 = \frac{4}{18} + \frac{5}{20} = \frac{17}{36}$

También se sabe que: $\sigma_1^2 = 4$

$$P\left(\frac{\bar{X}-\bar{Y}-0}{\sqrt{\frac{17}{36}}S_1} \le \frac{k-0}{\sqrt{\frac{17}{36}}}\right) = P\left(\frac{Z}{\sqrt{S_1^2}} \le \frac{6k}{\sqrt{17}}\right) = P\left(\frac{Z}{\sqrt{\frac{17S_1^2}{4}}} \le \frac{12k}{17}\right) = P\left(\frac{Z}{\sqrt{\frac{\chi^2_{(17)}}{17}\times 17}} \le \frac{12k}{17}\right) = 0.95$$

$$P\left(t(17) \le \frac{12\sqrt{17}k}{17}\right) = 0.95 \rightarrow \frac{12\sqrt{17}k}{17} = 1.7396067 \rightarrow k = 1.7396067 \times \frac{17}{12\sqrt{17}}$$

Ejemplo 21: Los adelantos o atrasos de las agencias de transportes X e Y, que llevan pasajeros se Lima a Ayacucho, tienen distribución N(0,16) y N(0,32) respectivamente. Suponga que $X_1,...,X_{18}$ y $Y_1,...,Y_{19}$ son muestras aleatorias independientes de estas distribuciones y $\bar{X}, \bar{Y}, S_X^2$ y S_y^2 denotan las respectivas medias y variancias muestrales. Halle el valor de k para la variable aleatoria $A = k\frac{\bar{X}-\bar{Y}}{\sqrt{68S_X^2 + 36S_Y^2}}$ tenga una distribución conocida.

Solución:

$X \sim N(0,16) \quad X_1,...,X_{18} \quad \bar{X} \quad S_X^2 \quad n_X - 1 = 17$

$Y \sim N(0,32) \quad Y_1,...,Y_{19} \quad \bar{Y} \quad S_Y^2 \quad n_Y - 1 = 18$

$W = \bar{X} - \bar{Y} \qquad \sigma_W^2 = \frac{16}{18} + \frac{32}{19} = \frac{440}{171}$

$\mu_W = 0 - 0 = 0$

$$A=\frac{\dfrac{k(\bar{X}-\bar{Y}-0)}{\sqrt{\dfrac{440}{171}}}}{\dfrac{\sqrt{68S_X^2+36S_Y^2}}{\sqrt{\dfrac{440}{171}}}}=\frac{kZ}{\sqrt{4\times 17S_X^2+2\times 18S_Y^2}}\sqrt{\frac{440}{171}}=\frac{kZ}{\sqrt{4\times\dfrac{17}{16}\times 16S_X^2+2\times\dfrac{18}{32}\times 32S_Y^2}}\sqrt{\frac{440}{171}}$$

$$A=\frac{kZ}{\sqrt{64\chi^2_{(17)}+64\chi^2_{(18)}}}\sqrt{\frac{440}{171}}=\frac{kZ}{8\sqrt{\chi^2_{(35)}}}\sqrt{\frac{440}{171}}=\frac{k\sqrt{440}}{8\sqrt{171\times 35}}\frac{Z}{\sqrt{\dfrac{\chi^2_{(35)}}{35}}}=\frac{k\sqrt{440}}{8\sqrt{171\times 35}}t\left(35\right)\rightarrow$$

$$\frac{k\sqrt{440}}{8\sqrt{171\times 35}}=1\rightarrow k=\frac{8\sqrt{171\times 35}}{\sqrt{440}}=29.50500728$$

De otro modo:

$$A=\sqrt{\frac{Z^2}{x^2_{(35)}}}\frac{k}{8}\sqrt{\frac{440}{171}}=\sqrt{\frac{x^2_{(1)}}{x^2_{(35)}}}\frac{k}{8}\sqrt{\frac{440}{171}}=\sqrt{\frac{35x^2_{(1)}}{x^2_{(35)}}}\frac{k}{8}\sqrt{\frac{440}{171\times 35}}=\sqrt{F(1,35)}\frac{k}{8}\sqrt{\frac{440}{171\times 35}}\rightarrow$$

$$\frac{k\sqrt{440}}{8\sqrt{171\times 35}}=1\rightarrow k=\frac{8\sqrt{171\times 35}}{\sqrt{440}}=29.50500728$$

Ejemplo 22: Si $X\sim \text{Beta}\left(\frac{k_1}{2},\frac{k_2}{2}\right)$ demuestre que la variable aleatoria $F=\frac{k_2X}{k_1\left(1-X\right)}\sim F\left(k_1,k_2\right)$ con $k_1,k_2=1,2,\ldots$.

$$f\left(x\right)=\frac{x^{\frac{k_1}{2}-1}\left(1-x\right)^{\frac{k_2}{2}-1}}{B\left(\dfrac{k_1}{2},\dfrac{k_2}{2}\right)}I_{(0,1)}\left(x\right)$$

$$f=\frac{k_2x}{k_1\left(1-x\right)}\rightarrow x=\frac{fk_1}{k_2+fk_1}\rightarrow J=\frac{dx}{df}=\frac{k_1k_2}{\left(k_2+fk_1\right)^2}$$

$$g\left(f\right)=f\left(\frac{fk_1}{k_2+fk_1}\right)\left|J\right|=\frac{1}{B\left(\dfrac{k_1}{2},\dfrac{k_2}{2}\right)}\left(\frac{fk_1}{k_2+fk_1}\right)^{\frac{k_1}{2}-1}\underbrace{\left(1-\frac{fk_1}{k_2+fk_1}\right)^{\frac{k_2}{2}-1}}_{\left(\frac{k_2}{k_2+fk_1}\right)^{\frac{k_2}{2}-1}}\left|\frac{k_1k_2}{\left(k_2+fk_1\right)^2}\right|$$

pero $B\left(\frac{k_1}{2},\frac{k_2}{2}\right)=\frac{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}+\frac{k_2}{2}\right)}=\frac{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}{\Gamma\left(\frac{k_1+k_2}{2}\right)}\rightarrow$

$$g(f)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\frac{k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}f^{\frac{k_1}{2}-1}}{\left(k_2+fk_1\right)^{\frac{k_1}{2}+\frac{k_2}{2}}}=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\frac{k_1^{\frac{k_1}{2}}k_2^{\frac{k_2}{2}}f^{\frac{k_1}{2}-1}}{k_2^{\frac{k_1}{2}+\frac{k_2}{2}}\left(1+\frac{k_1}{k_2}f\right)^{\frac{k_1}{2}+\frac{k_2}{2}}}$$

$$g(f)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}\frac{f^{\frac{k_1-2}{2}}}{\left(1+\frac{k_1}{k_2}f\right)^{\frac{k_1+k_2}{2}}}I_{(0,\infty)}(f)\quad ;\ k_1,k_2=1,2,\ldots$$

Distribución de Muestreo del Coeficiente de Correlación Lineal (r)

Teorema
Si de una distribución normal bivariada se toma una muestra bidimensional de tamaño n $(x_1,y_1),\ldots,(x_n,y_n)$ se demuestra que:

$$g(r)=\frac{n-2}{\pi}\left(1-\rho^2\right)^{\frac{n-1}{2}}\left(1-r^2\right)^{\frac{n-4}{2}}\int_0^1\frac{x^{n-2}}{(1-\rho r x)^{n-1}\sqrt{1-x^2}}dx$$

Para valores moderados de *n*, y cuando $n\rightarrow\infty$ la v.a:

$$C=\underbrace{\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)}_{\text{Arctgh}(r)}\text{ es aproximadamente } N\left[\underbrace{\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}_{\text{Arctgh}(\rho)}+\frac{\rho}{2(n-1)};\frac{1}{n-3}\right]$$

Pero es más frecuente usar:

$$C=\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)\text{ es aproximadamente } N\left[\mu_C=\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right);\sigma_C^2=\frac{1}{n-3}\right]$$

El error de prescindir de $\frac{\rho}{2(n-1)}$ es despreciable con muestras grandes.

Teorema
Si en la distribución normal bivariada se conoce que $\rho=0$ entonces:

$$g(r)=\frac{n-2}{\pi}\left(1-r^2\right)^{\frac{n-4}{2}}\int_0^1\frac{x^{n-2}}{\sqrt{1-x^2}}dx=\frac{1}{B\left(\frac{n-2}{2},\frac{1}{2}\right)}\left(1-r^2\right)^{\frac{n-4}{2}}$$

Haciendo el cambio de variable $t=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ se llega a:

$$g(t)=\frac{1}{\sqrt{n-2}\,B\left(\frac{n-2}{2},\frac{1}{2}\right)}\left[1+\frac{t^2}{n-2}\right]^{-\left(\frac{n-1}{2}\right)}$$

$$\therefore\quad t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)$$

Ejemplo 23: El contenido de vitamina C y vitamina D en mg/100 g de chirimoya tienen distribución normal bivariada con una correlación de 0.48. Si se toma una muestra aleatoria de 200 porciones de 100 g de chirimoya calcule la probabilidad de que la correlación muestral sea mayor que 0.42 pero menor que 0.51.

$$C=\text{Arctgh}(r)=\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)\sim N\left[\mu_C=\frac{1}{2}\ln\left(\frac{1+0.48}{1-0.48}\right)=0.52298;\sigma_C^2=\frac{1}{197}\right]$$

$$P(0.42<r<0.51)=P(\text{Arctgh}(0.42)<C<\text{Arctgh}(0.51))=P(0.4477<C<0.5627)$$

$$P(0.42<r<0.51)=P(-1.06<Z<0.56)=0.5661$$

Ejemplo 24: La productividad de aguacate de cierta variedad en Kg/parcela y la cantidad de abono utilizada en Kg/parcela se distribuye en forma independiente y de acuerdo a una distribución normal bivariada. Con una muestra aleatoria de 8 parcelas calcule la probabilidad de que la correlación muestral tome un valor menor que 0.12.

Como $\rho=0\rightarrow t=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)\sim t(6)$

$$P(r<0.12)=P\left[t(6)<\frac{0.12\sqrt{8-2}}{\sqrt{1-0.12^2}}\right]=P\left[t(6)<0.296\right]=0.6113962$$

<u>EJERCICIOS</u>

1. Enuncie y demuestre la ley débil de los grandes números.
2. El año pasado, en Villa el Salvador, se encuestó a 1000 hombres separados y entre otras cosas se encontró que su ingreso mensual, en cientos de soles, X tiene densidad $f(x)=\dfrac{40-x}{450}I_{(10,40)}(x)$. Si $X_1,\ldots,X_n$ es una muestra aleatoria de esa distribución diga si se cumple la ley de los grandes números.
   Suponga que este año la distribución del ingreso mensual de los hombres separados de Villa el Salvador ha cambiado. Si se quiere estimar a μ con la $\overline{X}$ con un 95% de seguridad de que la media muestral no se desvíe de la media poblacional en más de $\dfrac{\sigma}{25}$,

¿qué tamaño de muestra mínimo se debe utilizar? Utilice dos criterios para responder esta subpregunta.

3. Un antropólogo quiere estimar la estatura promedio de los hombres de cierta raza. Se supone que la desviación estándar de la población es 5.8 cm. Si el antropólogo quiere que la media de la muestra no difiera de la media poblacional en más de 0.9 cm , con una probabilidad de 0.95. ¿Cuántos hombres tendría que seleccionar para alcanzar su objetivo?
4. El ingeniero Thomas Bayes encontró que el nivel de pH en el suelo para el crecimiento óptimo de lechuga tiene distribución $\text{Desconocida}\left(\mu = 6.5, \sigma^2 = 0.19^2\right)$.

   a. Si $X_1, \ldots, X_n$ es una muestra aleatoria de esa distribución desconocida deduzca la media y la variancia de $\bar{X} = \sum_{i=1}^{n} X_i$ y discuta brevemente si se puede conocer la distribución exacta y la distribución límite de $\bar{X}$.

   b. Si $X_1, \ldots, X_n$ es una muestra aleatoria de esa distribución desconocida obtenga la $E\left(S^2\right)$.

   c. El ingeniero Pierre Simon de Laplace considera que el nivel de pH en el suelo para el crecimiento óptimo de lechuga tiene distribución $\text{Desconocida}\left(\mu, \sigma^2 = 0.17^2\right)$. Laplace quiere estimar a μ con la $\bar{X}$ con un 95% de seguridad que la media muestral no se desvíe de la media poblacional en más de $\frac{0.17}{4}$. Calcule de dos maneras el tamaño de muestra que necesita Pierre.

5. Una empresa encuestadora desea determinar la proporción de de estudiantes universitarios de Lima Metropolitana que conocen Chan Chan, basándose en una muestra aleatoria de tamaño n. La encuestadora desea usar la proporción muestral como estimadora de p. La empresa quiere un margen de error de estimación de 3% y un nivel de confianza de por lo menos 0.95. En una encuesta piloto se encontró que de 120 estudiantes 40 conocen Chan Chan. ¿Cuál debe ser el tamaño de muestra mínimo que la empresa debe tomar? Utilice los dos criterios para hallar n y diga cuál recomendaría ud. ¿Por qué razón?
6. 0.56, 0.08, 0.24, 0.97, 0.88, 0.42, 0.18, 0.03, 0.54, 0.62, 0.31, 0.95, 0.28, 0.19, 0.88, 0.44, 0.12, 0.64, 0.87, 0.38, 0.29, 0.51, 0.68, 0.54, 0.34, 0.22, 0.88, 0.22, 0.92, 0.09, 0.27, 0.54, 0.18 es una muestra aleatoria de una distribución $Unif\,(0,1)$. Halle una muestra aleatoria unitaria de una distribución $Normal\left(-3, 2\right)$.
7. Si $f(x) = \frac{1}{x^2} I_{[1,\infty)}(x)$. Diga si se cumple la ley de los grandes números. Justifique.

8. El tiempo de duración en años de un componente de TV marca SONY tiene como densidad: $f(x) = \frac{2(\theta - x)}{\theta^2} I_{(0,\theta)}(x)$. Con una muestra aleatoria de tamaño n de dicha distribución diga si se cumple la ley débil de los grandes números y justifique.
9. Enuncie y demuestre el Teorema del Límite Central.
10. En cierta reserva nacional A las ardillas grandes bebés se agrupan de a cuatro. Sea la variable aleatoria X definida como el número de hembras por grupo de cuatro que tiene la siguiente función de probabilidad:

X	0	1	2	3	4
P(X = x)	0.07	0.16	0.62	0.14	0.01

De los archivos de la reserva nacional A se toman 50 grupos al azar, y si el número total de hembras es a lo más 101 se considera que la distribución de sexos en A es "normal" ¿Cuál es la probabilidad de que en la reserva nacional A la distribución de sexos sea "normal"?

11. En un yacimiento de fósiles de micromamíferos se recogen aleatoriamente 15 sacos de sedimento. Después de lavarse el sedimento se define la v.a X como el peso en gramos del material óseo contenido en cada saco, que tiene distribución $\text{Normal}(42, 36)$. De otro lado, hay en el mercado tres tipos de cajas especiales para transportar fósiles de micromamíferos con capacidades de 555, 655 y 700 gramos, respectivamente. Si se debe elegir el tipo de caja más pequeño capaz de contener el total de las 15 extracciones de material óseo ¿Qué tipo de caja recomendaría usted para transportar el material óseo obtenido? Justifique su respuesta.

12. Se conectan 35 focos de luz infrarroja en un invernadero, de tal manera que si falla un foco otro se enciende inmediatamente. (Se enciende sólo un foco a la vez) Los focos funcionan independientemente y cada uno tiene una vida media de 50 horas y una desviación estándar de 5 horas. Si no se inspecciona el invernadero durante 1820 horas después de encender el sistema de focos ¿Cuál es la probabilidad de que un foco esté encendido al final del periodo de 1820 horas?

13. Si X_1, X_2 es una m.a de una distribución $\text{Exponencial}\left(\text{con media } \frac{1}{\theta}\right)$. Primero encuentre la distribución exacta de $\bar{X} = \frac{1}{2}\sum_{i=1}^{2} X_i$ y luego diga cuál sería la distribución límite de $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
14. En cierta región A se sabe que la probabilidad de obtener una temperatura de suelo mayor a 20°C es 0.25. X es el número de "intentos" hasta obtener por primera vez una temperatura de suelo mayor a 20°C. Se extrae una muestra aleatoria de tamaño 50 del rango de X. Calcule la probabilidad de que el promedio muestral de intentos sea a lo más 4.2, utilizando primero la distribución exacta y luego la distribución límite.

15. El ingreso mensual de un padre de familia que vive en cierto distrito tiene densidad $f(x) = \dfrac{4x^3}{\theta^4} I_{(0,\theta)}(x)$. Sea $X_1, \ldots, X_n$ una muestra aleatoria de esa distribución. En primer lugar asumiendo que n = 2 y $\theta = 1$ halle la distribución exacta de $\bar{X} = \dfrac{\sum_{i=1}^{2} X_i}{2}$, y después responda a lo siguiente ¿Cuál sería la distribución límite de $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$?

16. Si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución $N(\mu, \sigma^2)$, asumiendo que la media muestral y la variancia muestral son independientes demuestre que $\dfrac{(n-1)S^2}{\sigma^2}$ tiene distribución conocida. Justifique.

17. Un estudiante dice: "Si $X_1, \ldots, X_n$ es una m.a de una distribución $N(\mu, \sigma^2)$ entonces $\bar{X}$ y S^2 no son independientes por que S^2 está en función de $\bar{X}$". ¿Cuál es su opinión acerca de esta afirmación? Justifique su respuesta.

18. De una distribución $N(50, \sigma^2)$ se extraen m.as independientes entre si de tamaños 6 y 8:

Muestra 1	49	53	54	45	56	52		
Muestra 2	52	58	44	48	52	55	50	52

Calcule $P(\bar{X}_1 - \bar{X}_2 < -2)$.

19. Suponga que $X_1, \ldots, X_6$ y $Y_1, \ldots, Y_9$ son muestras aleatorias independientes de las distribuciones $N(0,2)$ y $N(0,3)$ respectivamente, y sean $\bar{X}, \bar{Y}, S_1^2$ y S_2^2 las respectivas medias y variancias muestrales.
   a. Si $P\left(\dfrac{\sum_{i=1}^{6} X_i^2}{\sum_{i=1}^{9} Y_i^2} \le k\right) = 0.05$, halle el valor de k. Respuesta $k = \dfrac{4}{9 \times 4.1}$.
   b. Si $P\left(\dfrac{\bar{X} - \bar{Y}}{S_1} \le k\right) = 0.90$, halle el valor de k. Respuesta $k = \dfrac{1.48}{\sqrt{6}}$.
   c. Encontrar los valores de k y gl de tal manera que la variable aleatoria $V = \dfrac{k(\bar{X} - \bar{Y})}{\sqrt{45S_1^2 + 48S_2^2}} \sim t(gl)$. Respuesta $k = \sqrt{351}$ y $gl = 13$.

20. Suponga que $E \sim N(0,1)$; $W_1,\ldots,W_{15}$ son i.i.d exponenciales con media 5 y todas las variables son independientes. Sean $\overline{W} = \frac{1}{15}\sum_{i=1}^{15} W_i$ y $A = k\frac{E}{\sqrt{\overline{W}}}$. Para qué valor de k, A tiene distribución conocida.
21. Suponga que $X_1,\ldots,X_n$ es una muestra aleatoria de una distribución $N(\mu,\sigma^2)$. Alguien afirma que bajo estas condiciones S^2 es un estimador insesgado de σ^2 por lo tanto S es un estimador insesgado de σ. ¿Está Ud. de acuerdo con esta afirmación? Justifique su respuesta.
22. Si $X_1,\ldots,X_n$ son variables aleatorias independientes tales que $X_i \sim \chi^2(k_i)$, $i = 1,2,\ldots,n$. Obtenga la distribución de $\sum_{i=1}^{n} X_i$.
23. El tiempo en minutos que demora un nadador novato en recorrer 250 m tiene una distribución $N(10,3)$ Sea $X_1,\ldots,X_{10}$ una m.a de esa distribución. Halle el valor de c tal que la v.a $Y = \frac{c(X_1 - 2X_3 + 4X_2 - 30)}{\left[\sum_{i=4}^{10}(X_i - 10)^2\right]^{0.5}}$ tenga una distribución conocida.
24. En una población normal bidimensional $\rho = 0$. Se toma una m.a. de tamaño 18. Calcule la probabilidad de que la correlación muestral sea mayor que -0.400 si se sabe que es menor que 0.5425.
25. El contenido de colesterol y el contenido de triglicéridos de las personas de cierta población tiene distribución normal bidimensional con un coeficiente de correlación igual a 0.88. Se toma una muestra aleatoria de 400 personas. Determine el valor de la probabilidad de que la correlación muestral esté comprendida entre 0.864 y 0.898.