Variables aleatorias y distribuciones de probabilidad

Mg. Ciro Ivan Machacuay Meza

Facultad de Economía

2 de marzo de 2025

Índice de contenidos

Definición

Una variable estadística es una característica (cualitativa o cuantitativa) que se mide u observa en una población. Si la población es aleatoria y la característica es cuantitativa, entonces, la variable estadística es denominada variable aleatoria. Por lo tanto, la variable aleatoria es un concepto particular del concepto general que es la variable estadística.

Por consiguiente, se denomina variable aleatoria a una variable estadística cuantitativa definida en un espacio muestral Ω. Esto es, una variable aleatoria X es una función definida en Ω tal que a cada elemento $w \in \Omega$ le asocia el número real $x = X(w)$.
El dominio de la variable aleatoria X es el espacio muestral Ω y su rango es el conjunto de todos sus valores posibles que denotaremos por R_x, esto es:

$$R_x = \{x \in \mathbb{R} / x = X(w), w \in \Omega\}$$

EJEMPLO 1

Sea Ω el espacio muestral que resulta del experimento aleatorio de lanzar al aire una moneda tres veces consecutivas y observar la cara superior. Entonces:

$$\Omega = \{SSS, SSC, SCS, CSS, SCC, CSC, CCS, CCC\}$$

Si se define en Ω la variable $X=$ "número de caras obtenidas", entonces, X es una variable aleatoria cuyo rango es el conjunto de sus valores posibles: $R_x = \{0, 1, 2, 3\}$. En efecto:

- $X = 0 \Longrightarrow \{SSS\}$
- $X = 1 \Longrightarrow \{SSC\}, \{SCS\}, \{CSS\}$
- $X = 2 \Longrightarrow \{SCC\}, \{CSC\}, \{CCS\}$
- $X = 3 \Longrightarrow \{CCC\}$

EJEMPLO 2

Sea Ω el espacio muestral que resulta de lanzar un dado sucesivamente hasta obtener el primer dos. Entonces:

$$\Omega = \{E, FE, FFE, FFFE, \dots\}$$

Donde, el éxito es E="Sale dos" y el fracaso es F="No sale dos".
Si se define en Ω la variable Y = "Número de lanzamientos realizados hasta que ocurra el primer éxito", entonces Y es una variable aleatoria, cuyo rango es el conjunto de sus valores posibles $R_y = \{1, 2, \dots\}$. Esta variable ejecuta:

$$Y(E) = 1, Y(FE) = 2, Y(FEE) = 3, \dots$$

EJEMPLO 3
Sea Ω el espacio muestral que resulta del control de la vida útil de un producto. Entonces $\Omega = \{w \in \mathbb{R}/w \geq 0\}$. Si se define en Ω la variable X="Tiempo de vida útil del producto", entonces, X es la variable aleatoria identidad cuyo rango es el conjunto:

$$R_X = \Omega = \{x \in \mathbb{R}/x \geq 0\}$$

Por consiguiente, la variable aleatoria es una función que atribuye a cada evento elemental de Ω un *número que no es aleatorio o imprevisible, si no fijo y predeterminado.*

Esta función puede estar definido por una ley o una regla o por un conjunto de números reales atribuidos mediante dicha ley o regla. En la mayoría de las aplicaciones, se está más interesado en los valores que toma la variable, ignorando completamente el espacio muestral en el que se define.

Las variables aleatorias son variables cuantitativas, por lo tanto, se clasifican en **discretas** y **continuas**. La variable aleatoria discreta es aquella entre cuyos valores posibles no admite otros. Su rango es un conjunto finito o infinito numerable de valores. Si la variable aleatoria X es discreta, su rango se expresará en general por:

$$R_X = \{x_1, x_2, \dots\}$$

La variable aleatoria continua es quella cuyo rango es un intervalo en $\mathbb{R}$ y por lo tanto es conjunto infinito (no numerable) de valores reales.

Variable aleatoria discreta: $p(x)$ y $P(x)$

Una variable aleatoria discreta X asume cada uno de sus valores con cierta probabilidad que denotaremos por P_X, pues lo adquiere de Ω vía X. El valor P_X se denomina *probabilidad inducida* por X y está definida en R_X (espacio muestral inducido por X). En efecto, si el rango de la variable aleatoria X es el conjunto finito de números, $R_X = \{x_1, x_2, x_3, \dots\}$, y si $B = \{x_i\}$ es un evento unitario en R_X, entonces:

$$P_X(\{x_i\}) = P(\{w \in \Omega / X(w) = x_i\})$$

Con frecuencia, se utiliza la expresión $P[X = x_i]$ para denotar la probabilidad $P_X(\{x_i\})$. Por ejemplo, si la variable aleatoria X es el número de caras que resultan al lanzar una moneda 3 veces, entonces, el rango de X es $R_X = \{0, 1, 2, 3\}$. Por tanto, en R_X se tiene:

- $P[X = 0] = P(\{SSS\}) = 1/8$
- $P[X = 1] = P(\{SSC, SCS, CSS\}) = 3/8$
- $P[X = 2] = P(\{SCC, CSC, CCS\}) = 3/8$
- $P[X = 3] = P(\{CCC\}) = 1/8$

- El conjunto de pares $(x_i, P[X = x_i])$ constituye la **distribución de probabilidades** de la variable aleatoria X. Esta distribución es **similar a una distribución de frecuencias relativas**. Por esta razón, una distribución de frecuencias relativas de una población de datos producidos por algún proceso.
- Las probabilidades $p_i = P[X = x_i]$ satisfacen las propiedades:

$$a) p_i \geq 0, \forall x_i \in R_X \qquad b) \sum_{x_i \in R_X} p_i = 1$$

- Por extensión, para todo número real $x \neq x_i$, donde $x_i \in R_X$:

$$P[X = x] = P(\phi) = 0$$

Función de probabilidad

Se denomina función (ley,modelo o distribución) de probabilidad de X a la función $f(x)$ definida por $f(x) = P[X = x]$ en todo x número real y que satisface:

$$i) f(x) \geq 0, \forall x \in \mathbb{R} \qquad ii) \sum_{x_i \in R_X} f(x_i) = 1$$

La función de probabilidad de una v.a X se puede describir como un conjunto de pares $\{x_i, p_i / p_i = f(x_i), x_i \in R_X\}$ en una tabla:

Valores x_i de X	x_1	x_2	x_3	$\dots$	x_n
Probabilidad $p_i = P[X = x_i]$	p_1	p_2	p_3	$\dots$	p_n

Ejercicio

Sea X la variable aleatoria definida como el número de caras que ocurre al lanzar una moneda 4 veces.

1. Obtenga la distribución de probabilidad de X y grafique.
2. Calcule la probabilidad $P[X > 2]$
3. Calcule la probabilidad $P[X \leq 1]$
4. Calcule la probabilidad $P[X < 3]$
5. Calcule la probabilidad $P[0 < X \leq 2]$

Función de distribución acumulada

La función de distribución acumulada de probabilidades de la variable aleatoria discreta X, cuya función de probabilidad es $f(x)$ se define en todos los números reales por:

$$F(x) = P[X \leq x] = \sum_{k \leq x} P[X = k] = \sum_{k \leq x} f(k), -\infty < x < \infty$$

En general, la función de distribución acumulada de X es:

$$F(x) = \begin{cases} 0 & \text{si } -\infty < x < x_i \\ \sum f(x_i) & \text{si } x_i \leq x < x_{i+1}, i = 1, 2, \ldots, n-1 \\ 1 & \text{si } x_n \leq x < +\infty \end{cases}$$

Ejercicio

Sea X la variable aleatoria definida como el número de caras que ocurre al lanzar una moneda 4 veces.

1. Obtenga la distribución acumulada de probabilidad de X y grafique.
2. Calcule la probabilidad $P[X > 2]$, $P[X \leq 1]$, $P[X < 3]$, $P[0 < X < 2]$

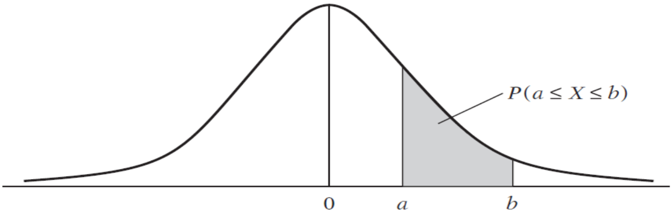
Variable aleatoria continua

Función de densidad de probabilidad

Sea X una **va** continua. Entonces, se dice que $f(x)$ es la FDP de X si satisface las siguientes condiciones:

$$i)\ f(x) \geq 0 \qquad ii)\ \int_{-\infty}^{+\infty} f(x)\,dx = 1 \qquad iii)\ \int_a^b f(x)\,dx = P(a \leq X \leq b)$$

Donde $f(x)dx$ se conoce como el **elemento probabilístico**(la probabilidad asociada a un pequeño intervalo de una variable continua) y donde $P(a \leq X \leq b)$ significa la probabilidad de que X se encuentre en el intervalo $[a, b]$. Geométricamente, tenemos



Ejercicio

La demanda mensual en miles de kilogramos de un producto es una variable aleatoria X cuya función de densidad es la función real de valor real:

$$f(x) = \begin{cases} cx^2 & \text{si } 0 \leq x \leq 2 \\ 0 & e.o.c \end{cases}$$

1. Hallar el valor de la constante c.
2. Si el stock mensual del producto es 1000 kg, ¿qué probabilidad hay de que la demanda no sea satisfecha en un mes cualquiera?

Función de distribución de probabilidad

La función de distribución acumulada de una variable aleatoria continua X cuya función de densidad es $f\left(x\right)$ se define en todo número real x por:

$$F\left(x\right) = P\left[X \leq x\right] = \int_{-\infty}^{x} f\left(t\right) dt, \; -\infty < x < +\infty$$

Ejercicios

1. La utilidad neta diaria en miles de dólares de las ventas de cueros para zapato de una empresa es una v.a X, cuya f.d.p es:

$$f(x) = \begin{cases} x/4 & \text{si } 1 \leq x \leq 3 \\ 0 & e.o.c \end{cases}$$

   1. Definir la función de distribución acumulada de X.
   2. Calcular la probabilidad que la utilidad neta se encuentre entre 1250 y 2500 dólares.
   3. ¿Cuánto es el cuarto superior de las utilidades netas diarias?

2. El ingreso monetario por hogar, en miles de soles, de una población es una v.a con la f.d.p :

$$f(x) = \begin{cases} cx & \text{si } 0 \leq x \leq 1 \\ c & \text{si } 1 \leq x \leq 2 \\ -cx + 3c & \text{si } 2 \leq x \leq 3 \\ 0 & e.o.c \end{cases}$$

   1. Determinar el valor de c y la función de distribución acumulada.
   2. ¿Qué porcentaje de hogares tienen ingresos inferiores a 1800 soles?

Media de una variable aleatoria

La media de una v.a X es un número real que se denota por μ_X. La media es denominada **valor esperado** o **esperanza** y se denota por $E\left(X\right)$.

El valor esperado de una variable aleatoria se define como:

$$E\left(X\right) = \sum_{x}^{\infty} x f\left(x\right), \qquad \textbf{vad}$$

$$E\left(X\right) = \int_{-\infty}^{+\infty} x f\left(x\right) dx\,, \qquad \textbf{vac}$$

Las propiedades del valor esperado son:

1.- El valor esperado de una constante es la constante misma.

2.- Si a y b son constantes, entonces :

$$E\left(aX + b\right) = aE\left(X\right) + b$$

Esto se puede generalizar. Si $X_1, X_2, \ldots, X_N$ son N variables aleatorias y $a_1, a_2, \ldots, a_N$ y b son constantes:

$$E\left(\sum_{i=1}^{n}(a_i X_i + b)\right) = E\left(\sum_{i=1}^{n} a_i X_i\right) + \sum_{i=1}^{n} b = \sum_{i=1}^{n} a_i E(X_i) + nb$$

3.- Si X e Y son variables aleatorias independientes:

$$E(XY) = E(X)\,E(Y)$$

4.- Si X e Y son v.a independientes y a y b son constantes reales, entonces:

$$E(aXbY) = aE(X)\,bE(Y) = abE(X)\,E(Y)$$

5.- Si $X_1, X_2, \ldots, X_n$ son n v.a independientes:

$$E(X_1 \cdot X_2 \cdot \cdots \cdot X_n) = E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

6.- En general, se tiene:

$$E(X \pm Y) = E(X) \pm E(Y)$$

Ejercicios

1. Calcular la media de la función de distribución de la v.a X que se define como el número de caras que resultan de lanzar cuatro monedas a la vez.
2. El número de llamadas recibidas en un teléfono celular, en un periodo dado, es una v.a X con función de probabilidad:
$$f(x) = \frac{e^{-\lambda}\lambda^{-x}}{x!}$$
Con $x = 1, 2, \ldots, \lambda > 0$. Calcular el valor esperado.
3. La vida útil, en miles de horas de un producto es una v.a cuya f.d.p es:
$$f(x) = \begin{cases} 1 - \frac{x}{2} & \text{si } 0 \leq x \leq 2 \\ 0 & e.o.c \end{cases}$$
   1. Calcular la esperanza de vida del producto
   2. Definir la función de distribución de probabilidad acumulada
   3. Calcular la mediana de la vida útil

Media de una función de variable aleatoria

Si X es una variable aleatoria, también lo es cualquier función $Y = H(X)$. Si tiene una distribución definida por $f(x)$:

$$E(H(X)) = \sum_{i=1}^{\infty} H(x_i) f(x_i), \quad \textbf{v.a.d}$$

$$E(H(X)) = \int_{-\infty}^{+\infty} H(x) f(x) dx, \quad \textbf{v.a.c}$$

Ejercicios

1. Sea X una variable cuya distribución de probabilidades es:

x_i	-1	0	1	2
$p(x_i)$	1/8	3/8	3/8	1/8

Calcular la media de: $a)Y = 2X - 2$, $b)Y = 2X^2 + 2$

2. Un juego consiste en lanzar un dado. El jugador gana \$2 cuando obtiene al menos 5 puntos y pierde el número obtenido en dólares en caso contrario. Calcular la utilidad esperada del juego.

Varianza de una variable aleatoria

Definición

Sea X una variable aleatoria y sea $E(X) = \mu$. La distribución o dispersión de los valores de X alrededor del valor esperado se mide por la varianza, la cual se define como:

$$Var(X) = \sigma_X^2 = E(X-\mu)^2$$

La raíz cuadrada positiva de σ_X^2, σ_X, se define como desviación estándar de X. Ambas indican qué tan cercanos o dispersos están los valores individuales de X respecto del valor de su media.

La varianza definida anteriormente se calcula de la siguiente forma:

$$Var(X) = \sum_x (X-\mu)^2 f(x), \quad \textbf{vad}$$

$$Var(X) = \int_{-\infty}^{+\infty} (X-\mu)^2 f(x)\, dx, \quad \textbf{vac}$$

Por conveniencia de cálculo, la varianza se expresa también como:

$$Var\left(X\right) = \sigma_X^2 = E\left(X - \mu\right)^2$$

$$Var\left(X\right) = E\left[X^2 - 2X\mu + \mu^2\right] = E\left(X^2\right) - 2\mu E\left(X\right) + \mu^2$$

$$Var\left(X\right) = E\left(X^2\right) - \mu^2 = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

Las propiedades más importantes de la varianza son:

1.-La varianza de una constante es cero.

2.- Si a y b son constantes:

$$Var\left(aX + b\right) = a^2 Var\left(X\right)$$

3.- Si X e Y son va independientes y a y b son constantes:

$$Var\left(aX + bY\right) = a^2 Var\left(X\right) + b^2 Var\left(Y\right)$$

4.- Si X e Y son variables aleatorias independientes:

$$Var\left(X \pm Y\right) = Var\left(X\right) + Var\left(Y\right)$$

Ejercicios

1. Calcular la varianza y la desviación estándar de la distribución de probabilidad de la variable aleatoria X que se define como el número de caras al lanzar cuatro monedas.
2. La vida útil, en miles de horas de un producto es una v.a cuya f.d.p es:

$$f(x) = \begin{cases} \beta e^{-\beta x} & \text{si } x \geq 0 \\ 0 & e.o.c \end{cases}$$

   Calcular la varianza y desviación estándar de X.
3. Sea X una v.a con fdp:

$$f(x) = \begin{cases} \frac{3}{4}x(2-x) & \text{si } 0 \leq x \leq 2 \\ 0 & e.o.c \end{cases}$$

   Calcular E(X), E(2X+1), Var(X), Var(2X+1)

Distribución de probabilidad de Bernoulli

Se denomina **prueba o ensayo de Bernoulli** a todo experimento aleatorio que consiste de solo dos resultados prosibles mutuamente excluyentes, generalmente llamados: **Éxito (E)** y **Fracaso (F)**.

Por ejemplo, son ensayos de Bernoulli, lanzar una moneda al aire con los resultados cara o sello, elegir al azar un objeto fabricado con los resultados defectuoso o no defectuoso, recibir una opinión con los resultados a favor o en contra, etc.

El espacio muestral asociado al experimento aleatorio de Bernoulli se puede escribir como el conjunto: $\Omega = \{E, F\}$.

La variable aleatoria X definida en Ω de manera que atribuye a E, el valor 1 y a F el valor 0, se denomina ***variable aleatoria de Bernoulli***.

Si al éxito E se le asigna $p = P[X = 1]$, donde $0 < p < 1$, entonces el fracaso F tiene probabilidad $P[X = 0] = 1 - p$.

Luego, la distribución de probabilidad de Bernoulli de parámetro p está definida por:

X	$x=0$	$x=1$
$f(x)=P[X=x]$	$1-p$	p

Y en forma resumida, por la ecuación o modelo, se tendría:

$$f(x)=P[X=x]=p^x(1-p)^{1-x}, \;\; x=0,1$$

Si X tiene distribución de Bernoulli de parámetro p, entonces:

$$a) E(X)=p \quad b) Var(X)=p(1-p)$$

Distribución de probabilidad binomial

Se denomina **experimento binomial** a un número n de repeticiones o pruebas sucesivas de un experimento aleatorio Bernoulli, y por lo tanto, se caracteriza por:

- Las n repeticiones son estadísticamente independientes.
- Cada repetición de Bernoulli tiene dos resultados mutuamente excluyentes, éxito (E) o fracado (F).
- La probabilidad p de éxito es invariante en cada una de las n repeticiones (probabilidad constante).

La probabilidad de que ocurran k éxitos en n ensayos de Bernoulli es:

$$P[X = k] = C_k^n p^k q^{n-k}, \qquad k = 0, 1, 2, \ldots, n$$

La variable aleatoria X definida como el número de éxitos que ocurren en n pruebas de Bernoulli independientes, tiene distribución binomial con parámetros n y p y se escribe : $X \sim B(n, p)$, con función de probabilidad:

$$f(k) = P[X = k] = C_k^n p^k q^{n-k}, \qquad k = 0, 1, 2, \ldots, n$$

La función de distribución acumulada de la variable aleatoria binomial es:

$$F(x) = P\left[X \leq x\right] = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{k=0}^{x} C_k^n p^k q^{n-k} & \text{si } 0 \leq x < n-1 \\ 1 & \text{si } x \geq n \end{cases}$$

Las características numéricas de la variable aleatoria X es:

$$E(X) = np \qquad\qquad Var(X) = np(1-p)$$

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes cada una con una distribución binomial B(1,p). Entonces, la suma de estas n variables aleatorias independientes es el número de éxitos en n ensayos de Bernoulli y su distribución es :

$$X = \sum_{i=1}^{n} X_i \sim B\left(n, p\right)$$

Distribución de probabilidad Poisson

Se dice que la variable aleatoria discreta X cuyos valores posibles son: 0,1,..., tiene distribución de Poisson con parámetro λ y se escribe $X \sim P(\lambda)$ si su función de probabilidad es:

$$f(x) = P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}, \quad \lambda > 0, x = 0, 1, 2, \ldots$$

La distribución de Poisson se aplica a problemas donde la variable aleatoria es el número de eventos independientes que ocurre en un intervalo de tiempo, o de longitud, etc., o en una medida unitaria plana o cúbica. Por ejemplo, son variables de Poisson:

- Número de llamadas que recibe una central telefónica en el periodo de un minuto.
- Número de accidentes de trabajo que ocurren en una fábrica durante una semana.
- Número de fallas en la superficie de una cerámica rectangular.
- Número de bacterias en un volumen de un m^3 de agua.

Distribución de probabilidad normal

Se dice que la variable aleatoria continua X tiene distribución de probabilidad normal con parámetros μ y σ^2 y se denota por $X \sim N(\mu, \sigma^2)$, si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < +\infty, \;\; -\infty < \mu < +\infty$$

La distribución normal es el modelo de probabilidad fundamental de la Estadística pues, tiene diversas aplicaciones y sirve como una buena aproximación de otras distribuciones, sean estas simétricas o asimétricas o sean estas discretas o continuas.

Si la variable aleatoria se estandariza, tal como $Z = (X - \mu)/\sigma$ tiene una distribución normal $N(0, 1)$. Se denotará por Φ a la función de distribución acumulada de la variable normal estándar Z:

$$\Phi(z) = P[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Algunas propiedades de la función de distribución acumulada son:

$$1)\quad P\left[a \leq Z \leq b\right] = \Phi\left(b\right) - \Phi\left(a\right)$$

$$2)\quad \Phi\left(-Z\right) = 1 - \Phi\left(Z\right)$$

$$3)\quad P\left[-a \leq Z \leq a\right] = \Phi\left(a\right) - \Phi\left(-a\right) = \Phi\left(a\right) - \left(1 - \Phi\left(a\right)\right) = 2\Phi\left(a\right) - 1$$

$$4)\quad P\left[a \leq X \leq b\right] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Si $X_i \sim N(\mu, \sigma^2)$ y se define una v.a $Y = X_1 + X_2 + \cdots + X_n$ entonces:

$$E\left(Y\right) = \mu + \mu + \cdots + \mu = \sum_{i=1}^{n} \mu = n\mu$$

$$Var\left(Y\right) = \sigma^2 + \sigma^2 + \cdots + \sigma^2 = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$$

De donde, por **propiedad reproductiva de la normal**, se obtiene $Y \sim N(n\mu, n\sigma^2)$.

Distribución de probabilidad Chi-cuadrado

Se dice que la variable aleatoria continua X tiene una distribución chi-cuadrado con r grados de libertad y se denota por $X \sim \chi^2(r)$, si su función de densidad es:

$$f(x) = \frac{2^{-\frac{r}{2}}}{\Gamma\left(\frac{r}{2}\right)} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}, \;\; x \geq 0, r > 0$$

Además, si $X \sim \chi^2(r)$, se tiene que:

$$E(X) = r \qquad\qquad Var(X) = 2r$$

Algunas propiedades de la distribución chi-cuadrado:

- $Z \sim N(0,1) \Longrightarrow Z^2 \sim \chi^2(1)$
- $Z_i \sim N(0,1) \Longrightarrow \sum_{i=1}^{r} Z_i^2 \sim \chi^2(r)$
- $X_i \sim \chi^2(r_i) \Longrightarrow \sum_{i=1}^{k} X_i \sim \chi^2(r_1 + r_2 + \cdots + r_k)$

Distribución de probabilidad t-Student

Se dice que la variable aleatorio continua T se distribuye según el modelo de probabilidad t-Student con r grados de libertad y se denota por $T \sim t(r)$, si su función de densidad está dada por:

$$f(t) = \frac{\Gamma\left[\frac{r+1}{2}\right]}{\Gamma\left(\frac{r}{2}\right)\sqrt{r(\pi)}}\left(1+\frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \qquad -\infty < t < +\infty$$

Si Z y V son dos variables aleatorias independientes tales que $Z \sim N(0,1)$ y $V \sim \chi^2(r)$, entonces se cumple:

$$T = \frac{Z}{\sqrt{\frac{V}{r}}} \sim t(r)$$

Si $T \sim t(r)$, se tiene que $E(T) = 0$ y $Var(T) = \frac{r}{r-2}$. La varianza es mayor que la de la distribución $N(0,1)$, pero tiende a 1 cuando el número de libertad $r \rightarrow +\infty$. Además, la distribución t-Student se aproxima a una distribución $N(0,1)$ cuando $r \rightarrow +\infty$. La aproximación es buena si $r \geq 30$.

Distribución de probabilidad F de Snedecor

Se dice que la variable aleatoria continua X se distribuye según el modelo de probabilidad F con r_1 y r_2 grados de libertad y se representa por $X \sim F(r_1, r_2)$, si su función de densidad es:

$$f(x) = \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} \Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \cdot \frac{x^{\frac{r}{2}-1}}{\left(1+\frac{xr_1}{r_2}\right)^{\frac{r_1+r_2}{2}}}, \qquad 0 \leq x < +\infty$$

Donde r_1 y r_2 son números enteros positivos. Si U y V son dos variables aleatorias independientes tales que $U \sim \chi^2(r_1)$ y $V \sim \chi^2(r_2)$, entonces se cumple la siguiente relación:

$$X = \frac{U/r_1}{V/r_2} \sim F(r_1, r_2)$$

Si X tiene distribución F con r_1 y r_2 gl, entonces $1/X$ cumple:

$$F_{1-\alpha, r_1, r_2} = \frac{1}{F_\alpha, r_1, r_2}$$

Teorema del límite central (TLC)

Si $X_1, X_2, \ldots, X_n$ son n variables aleatorias independientes idénticamente distribuidas con media μ y varianza σ^2, entonces, cuando n tiende al infinito, la distribución de la variable aleatoria estándar:

$$Z_n = \frac{Y_n - \mu_{Y_n}}{\sigma_{Y_n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim N\left(0, 1\right)$$

La aproximación a la distribución normal es buena si n es grande $(n \geq 30)$ sin importar si la distribución es discreta o continua, o si la distribución es de forma simétrica o asimétrica.

Una consecuencia inmediata del TLC es que cuando n es grande, se tiene que la variable aleatoria total:

$$Y_n \underset{TLC}{\sim} N\left(n\mu, n\sigma^2\right)$$

Aproximación de la binomial a la normal

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias independientes e idénticamente distribuidas según el modelo de probabilidad Bernoulli de parámetro p. Cada variable presenta la siguiente distribución de probabilidad $X_i \sim B(1, p)$, es decir, $X_i = 1$ si ocurre E(éxito) con probabilidad p y $X_i = 0$ si ocurre F(fracaso), con probabilidad $1 - p$.

Además, cada X_i tiene media $\mu_{X_i} = p$ y varianza $\sigma^2_{X_i} = p(1-p)$.

Por otro lado, la variable aleatoria X, considerada como el número total de éxitos de n, se distribuye como una binomial:

$$X = \sum_{i=1}^{n} X_i \sim B\left(n, p\right)$$

Con media $\mu_X = np$ y su varianza $\sigma^2_X = np(1-p)$. Para $n \geq 30$ se tiene:

$$Z = \frac{X - np}{\sqrt{np\left(1-p\right)}} \sim N\left(0, 1\right)$$

Aproximación de la hipergeométrica a la normal

Sea X la v.a hipergeométrica, con las siguientes características numéricas:

$$a) E(X) = np \qquad b) Var(X) = np(1-p) \times \frac{N-n}{N-1}$$

Cuando n es grande se cumple:

$$Z = \frac{X - np}{\sqrt{np(1-p) \times \frac{N-n}{N-1}}} \sim N(0,1)$$

Aproximación de Poisson a la normal

Sea X la v.a Poisson, con las siguientes características numéricas:

$$a) E(X) = \lambda \qquad b) Var(X) = \lambda$$

Cuando n es grande se cumple:

$$Z = \frac{X - n\lambda}{\sqrt{n\lambda}} \sim N(0,1)$$