

1- Probabilidad

1.1 - Espacios muestrales y eventos.

Propiedades sobre el álgebra de conjuntos:

1- Leyes de idempotencia

a) $A \cup A = A$

b) $A \cap A = A$

2- Leyes asociativas

a) $A \cup (B \cap C) = (A \cup B) \cap C$

b) $A \cap (B \cup C) = (A \cap B) \cup C$

3- Leyes conmutativas

a) $A \cup B = B \cup A$

b) $A \cap B = B \cap A$

4- Leyes distributivas

a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

5- Leyes de identidad

a) $A \cup \emptyset = A$

b) $A \cap \emptyset = \emptyset$

c) $A \cup U = U$

d) $A \cap U = A$

6- Leyes de complemento

a) $A \cup A^c = U$

b) $A \cap A^c = \emptyset$

c) $(A^c)^c = A$

d) $U^c = \emptyset, \quad \emptyset^c = U$

7- Leyes de De Morgan

a) $(A \cup B)^c = A^c \cap B^c$

b) $(A \cap B)^c = A^c \cup B^c$

Si A y B son dos eventos tales que $A \cap B = \emptyset$, se dice que son disjuntos o **mutuamente excluyentes**.

Dado un evento A asociado a un experimento aleatorio \mathcal{E} . Supongamos que se repite n veces el experimento \mathcal{E} , y anotamos n_A al número de veces que ocurre A en la n repeticiones de \mathcal{E} . Se define la **frecuencia relativa de A** , y se simboliza f_A , al cociente $\frac{n_A}{n}$. Es decir que f_A es la **proporción de veces que ocurre A en las n repeticiones de \mathcal{E}** .

Definición axiomática de probabilidad.

Sea \mathcal{E} un experimento aleatorio y S un espacio muestral asociado con \mathcal{E} . Con cada evento A asociamos un número real llamado probabilidad de A , que anotamos $P(A)$, el cual satisface las siguientes propiedades básicas o *axiomas*

1- $0 \leq P(A) \leq 1$

2- $P(S) = 1$

3- Si A y B son eventos mutuamente excluyentes entonces $P(A \cup B) = P(A) + P(B)$

4- Si $A_1, A_2, \dots, A_n, A_{n+1}, \dots$ es una secuencia de eventos tales que

$$A_i \cap A_j = \emptyset \quad \text{si } i \neq j, \text{ entonces } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Propiedades de la probabilidad

1- $P(\emptyset) = 0$

2- Si A^c es el evento complementario de A , entonces $P(A^c) = 1 - P(A)$

3- Si $A \subset B$ entonces $P(A) \leq P(B)$

4- Si $A \subset B$, entonces $P(B - A) = P(B) - P(A) \Leftrightarrow P(B - A) = P(B) - P(A \cap B)$

5- Si A y B son dos eventos *cualesquiera*, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1.2 - El espacio muestral finito

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1 \quad (4)$$

Es decir: *la suma de las probabilidades de los eventos elementales es igual a 1*Si asumimos que $P(A_1) = P(A_2) = \dots = P(A_n) = p$, es decir todos los eventos elementales tienen la misma probabilidad de ocurrir, entonces reemplazando en (4):

$$\underbrace{p + p + \dots + p}_{n \text{ veces } p} = 1 \Rightarrow np = 1 \Rightarrow p = \frac{1}{n} = \frac{1}{\#S}$$

Donde el símbolo $\#S$ se lee: número de elementos de S , (en general $\#S$ indica el *cardinal* de S)Cuando $P(A_1) = P(A_2) = \dots = P(A_n) = p$, se dice que S es *equiprobable*.Es decir *si B es un evento de un espacio muestral equiprobable entonces* $P(B) = \frac{\#B}{\#S}$ **1.3 - Espacios muestrales infinitos**Sea S un espacio muestral *infinito numerable* es decir $S = \{a_1; a_2; a_3; \dots\}$. Como en el caso finito, a cada a_i asignamos un número $p_i = P(\{a_i\})$ tal que

a) $p_i \geq 0$ b) $\sum_{i=1}^{\infty} p_i = 1$

La probabilidad $P(A)$ de un evento A es entonces la suma de las probabilidades de los eventos elementales que lo componen.

Los únicos espacios muestrales *infinitos no numerables* que consideraremos son aquellos de medida geométrica finita $m(S)$ tales como longitud, área o volumen, y en los cuales un punto se selecciona al azar. La probabilidad de un evento A , esto es aquella en la que el punto seleccionado pertenece a A , es entonces la relación entre $m(A)$ a $m(S)$, es decir

$$P(A) = \frac{\text{longitud de } A}{\text{longitud de } S} \quad \text{ó} \quad P(A) = \frac{\text{área de } A}{\text{área de } S} \quad \text{ó} \quad P(A) = \frac{\text{volumen de } A}{\text{volumen de } S}$$

2 - Probabilidad condicional

Sean A y B dos eventos de un espacio muestral S . La probabilidad condicional de B dado A se define como

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) \neq 0$$

Análogamente

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) \neq 0$$

Observación: si A y B son eventos de un espacio muestral S *equiprobable*, entonces

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\#(A \cap B)}{\#A}$$

Observaciones:

- a) Si $A \subset B$ entonces $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$
 b) Si $A \cap B = \emptyset$ entonces $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(\emptyset)}{P(A)} = 0$

Teorema de la multiplicación

Si A y B son dos eventos entonces $P(B/A) = \frac{P(A \cap B)}{P(A)}$ si $P(A) \neq 0$

Por lo tanto

$$P(A \cap B) = P(B/A) P(A) \quad (6)$$

Análogamente de $P(A/B) = \frac{P(A \cap B)}{P(B)}$ si $P(B) \neq 0$, se deduce

$$P(A \cap B) = P(A/B) P(B) \quad (7)$$

(6) y (7) se conocen como *teorema de la multiplicación*.

El teorema de la multiplicación se puede generalizar a n eventos A_1, A_2, \dots, A_n :

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2 / A_1)P(A_3 / A_1 \cap A_2) \dots P(A_n / A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

2.2 - Teorema de la probabilidad total

Sean A_1, A_2, \dots, A_n eventos de un espacio muestral S que cumplen:

- a) $A_1 \cup A_2 \cup \dots \cup A_n = S$
- b) $A_i \cap A_j = \emptyset$ si $i \neq j$
- c) $P(A_i) > 0 \quad \forall i = 1, 2, \dots, n$ Se dice que A_1, A_2, \dots, A_n forman una **partición de S** Entonces para cualquier evento B de S

$$P(B) = P(B / A_1)P(A_1) + P(B / A_2)P(A_2) + \dots + P(B / A_n)P(A_n)$$

Teorema de Bayes

Sean A_1, A_2, \dots, A_n eventos de un espacio muestral S que cumplen:

- a) $A_1 \cup A_2 \cup \dots \cup A_n = S$
- b) $A_i \cap A_j = \emptyset$ si $i \neq j$
- c) $P(A_i) > 0 \quad \forall i = 1, 2, \dots, n$

Entonces para cualquier evento B de S tal que $P(B) > 0$

$$P(A_k / B) = \frac{P(B / A_k)P(A_k)}{\sum_{i=1}^n P(B / A_i)P(A_i)} \quad k = 1, \dots, n$$

2.3 - Independencia

Entonces, dos eventos A y B son **independientes** si $P(B / A) = P(B)$, y son **dependientes** de otro modo.

$$A \text{ y } B \text{ son independientes si y solo si } P(A \cap B) = P(A)P(B)$$

Es decir podemos usar (9) como definición de independencia de eventos.

Si dos eventos A y B son independientes entonces A y B^C son independientes

Independencia de más de dos eventos.

La noción de independencia de eventos se puede ampliar a n eventos de la siguiente manera:

Sean A_1, A_2, \dots, A_n eventos, se dice que son **independientes** si

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}) \quad k = 2, \dots, n \quad (11)$$

3 - VARIABLES ALEATORIAS

3.1- Generalidades

Definición: Sea \mathcal{E} un experimento aleatorio y S un espacio muestral asociado a él. Una **variable aleatoria** es una función que asigna a cada elemento de S un número real.

Dada una v.a. X a su imagen se la anota R_X y se la denomina **rango o recorrido de X**

Las variables aleatorias se clasifican según su rango.

Sea X es una v.a. con rango R_X . Si R_X es un conjunto **finito o infinito numerable** entonces se dice que X es una **v.a. discreta**. Si R_X es un conjunto **infinito no numerable** entonces X es una **v.a. continua**.

El rango R_X es considerado un nuevo **espacio muestral**, y sus subconjuntos son **eventos**.

siendo $A \subset S$ y $B \subset R_X$, A y B son **equivalentes** si $A = \{s \in S; X(s) \in B\}$

3.2 - Variables aleatorias discretas

Sea X una v.a. discreta. Anotamos su rango como $R_X = \{x_1, x_2, \dots, x_n\}$ si el rango es un conjunto finito de n elementos, y anotamos $R_X = \{x_1, x_2, \dots\}$ si el rango es un conjunto infinito numerable.

A cada x_i se le asigna un número $p(x_i) = P(X = x_i)$. Estos números deben satisfacer las condiciones siguientes

a) $p(x_i) \geq 0$ para todo i

b) $\sum_i p(x_i) = 1$

La función $p(x)$ que antes se definió, se llama **función de probabilidad o de frecuencia de la v.a. X** . El conjunto de pares $(x_i, p(x_i))$ $i = 1, 2, \dots$ es la **distribución de probabilidad de X** .

Observación:

Sea X una v.a. discreta con rango finito $R_X = \{x_1, x_2, \dots, x_n\}$, donde cada x_i es un número entero y x_{i+1} es el consecutivo de x_i .

Si $P(X = x_i) = \frac{1}{n}$ para cada i entonces se dice que X tiene **distribución uniforme discreta**.

Función de distribución acumulada

Sea X una v.a. con rango R_X . Se define la *función de distribución acumulada de X* (abreviamos F.d.a de X) como

$$F(x) = P(X \leq x) \quad -\infty < x < \infty \quad (12)$$

En el caso de ser X una v.a. discreta

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i) \quad -\infty < x < \infty$$

En general si X es una v.a. discreta cualquiera, su F.d.a. será una función escalonada.

Además si x_1, x_2, \dots, x_n son los valores del rango de X ordenados de menor a mayor entonces

$$\begin{aligned} P(X = x_1) &= F(x_1) \\ P(X = x_i) &= F(x_i) - F(x_{i-1}) \quad i = 2, \dots, n \end{aligned}$$

Es decir, se puede obtener la función de distribución de X a partir de su F.d.a.

Para números cualesquiera a y b

1- Si $a \leq b$ entonces $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$

2- Si $a < b$ entonces $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$

3- Si $a < b$ entonces $F(a) \leq F(b)$ (es decir $F(x)$ es una función creciente)

3.3 – Esperanza de una variable aleatoria discreta

Sea X una v.a. discreta con rango R_X . La *esperanza, valor medio o valor esperado de X* , lo anotamos $E(X)$, y se define como

$$E(X) = \sum_{x_i \in R_X} x_i P(X = x_i)$$

La sumatoria se hace sobre todos los posibles valores de X

Otra notación usual es μ_X o μ

Observaciones:

- 1- La esperanza de una v.a. no tiene que coincidir necesariamente con algún valor del rango de la variable
- 2- En el ejemplo 1 donde el rango es finito y equiprobable, la esperanza de X coincide con el **promedio de los valores del rango de X**
- 3- Se puede interpretar a la esperanza de una v.a. como un **promedio “pesado” o “ponderado” de los valores del rango de la variable**, donde el “peso” de cada x_i es la probabilidad $P(X = x_i)$

Teorema: Si X es una v.a. discreta con rango R_X y distribución de probabilidad $p(x)$, entonces la esperanza de cualquier función $h(X)$ es igual a

$$E(h(X)) = \sum_x h(x) p(x)$$

Propiedades de la esperanza

En el ejemplo anterior tenemos que Y es una función lineal de X , es decir $Y = aX + b$ con a y b números reales.

En este caso vale entonces la siguiente propiedad

$$E(aX + b) = aE(X) + b$$

Observaciones:

1- Para cualquier constante a , $E(aX) = aE(X)$

2- Para cualquier constante b , $E(X + b) = E(X) + b$

Varianza de una variable aleatoria

Sea X una v.a. discreta con rango R_X , función de distribución de probabilidad $p(x)$ y esperanza $E(X) = \mu$,

Entonces la **varianza de X** , que anotamos $V(X)$, σ^2 o σ_X^2 es

$$V(X) = E(X^2) - \mu^2$$

La **desviación estándar de X** es $\sigma_X = \sqrt{V(X)}$

Propiedades de la varianza

Las propiedades de la varianza de una v.a. son consecuencia de las propiedades de la esperanza de una v.a.

Si X es una v.a. discreta con rango R_X y distribución de probabilidad $p(x)$, entonces la varianza de cualquier función $h(X)$ es igual a

$$V(h(X)) = \sum_{x \in R_X} (h(x) - E(h(X)))^2 p(x)$$

Si $h(X)$ es una función lineal, entonces

$$V(aX + b) = a^2 V(X) \quad \text{y} \quad \sigma_{aX+b} = \sqrt{V(aX + b)} = |a| \sigma_X$$

Observaciones:

1- $V(aX) = a^2 V(X)$

2- $V(X + b) = V(X)$

3.4- Variables aleatorias discretas importantes

Distribución binomial

Sea \mathcal{E} un experimento aleatorio. Sea A un evento asociado a \mathcal{E} y anotamos $P(A) = p$.

Supongamos un experimento aleatorio \mathcal{E}_0 que cumple los siguientes requisitos:

- 1- se realizan n repeticiones *independientes* de \mathcal{E} , donde n se fija de antemano.
- 2- las repeticiones son idénticas, y en cada repetición de \mathcal{E} observamos *si ocurre A* o *no ocurre A* (cuando A ocurre se dice que se obtuvo un “éxito”, caso contrario se obtuvo un “fracaso”)
- 3- la probabilidad de éxito es constante de una repetición a otra de \mathcal{E} , y es igual a p

Se dice entonces que \mathcal{E}_0 es un *experimento binomial*

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \mathbb{E}(X) = np \quad \mathbb{V}(X) = np(1-p)$$

Notación: indicamos que X es una v.a. binomial *con parámetros n y p* con el símbolo $X \sim B(n, p)$

Observación: si $X \sim B(n, p)$ para calcular $P(X \leq k)$ en general se debe hacer

$$P(X \leq k) = \sum_{i=0}^k P(X = i) = P(X = 0) + P(X = 1) + \dots + P(X = k)$$

Notar que $P(X \leq k)$ es la F.d.a. de X evaluada en k , es decir $F(k) = P(X \leq k)$

Distribución geométrica

Sea \mathcal{E} un experimento aleatorio. Sea A un evento asociado a \mathcal{E} y anotamos $P(A) = p$.

Supongamos un experimento aleatorio \mathcal{E}_0 que cumple los siguientes requisitos:

- 1- se realizan repeticiones independientes de \mathcal{E} , hasta que ocurre A por primera vez inclusive.
- 2- las repeticiones son idénticas, y en cada repetición de \mathcal{E} observamos *si ocurre A* o *no ocurre A* (cuando A ocurre se dice que se obtuvo un “éxito”, caso contrario se obtuvo un “fracaso”)
- 3- la probabilidad de éxito es constante de una repetición a otra de \mathcal{E} , y es igual a p

$$f(x) = p(1-p)^{x-1} \quad \mathbb{E}(X) = \frac{1}{p} \quad \mathbb{V}(X) = \frac{1-p}{p^2}$$

Notación: $X \sim G(p)$

La F.d.a. sería:

$$F(x) = \begin{cases} 1 - (1-p)^{[x]} & \text{si } x \geq 1 \\ 0 & \text{caso contrario} \end{cases}$$

Distribución binomial negativa

La distribución binomial negativa constituye una extensión de la distribución geométrica.

Sea r un entero positivo.

Sea \mathcal{E} un experimento aleatorio. Sea A un evento asociado a \mathcal{E} y anotamos $P(A) = p$.

Supongamos un experimento aleatorio \mathcal{E}_0 que cumple los siguientes requisitos:

- 1- se realizan repeticiones independientes de \mathcal{E} , hasta que ocurre A por r -ésima vez inclusive.
- 2- las repeticiones son idénticas, y en cada repetición de \mathcal{E} observamos **si ocurre A o no ocurre A** (cuando A ocurre se dice que se obtuvo un “éxito”, caso contrario se obtuvo un “fracaso”)
- 3- la probabilidad de éxito es constante de una repetición a otra de \mathcal{E} , y es igual a p

$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \mathbb{E}(X) = \frac{r}{p} \quad \mathbb{V}(X) = \frac{r(1-p)}{p^2}$$

Notación: $X \sim BN(r, p)$

Observación: la distribución geométrica puede verse como un caso particular de la distribución binomial negativa con $r = 1$

Distribución hipergeométrica

Supongamos que tenemos una población o conjunto de N objetos o individuos (es decir tenemos una población finita).

Clasificamos a los objetos de la población en dos categorías. Hay M objetos de una categoría y $N-M$ de la otra categoría. Se suele decir que tenemos M “éxitos” y $N-M$ “fracasos”.

Se extraen al azar y sin reemplazo n objetos de dicha población. Es decir se extrae una muestra de n objetos de la población, de manera tal que es igualmente probable que se seleccione cada subconjunto de tamaño n .

Consideramos la v.a. X : “número de éxitos en la muestra extraída”

Se dice que X tiene una distribución hipergeométrica con parámetros n , M y N

Notación: $X \sim H(n, M, N)$

$$f(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad \mathbb{E}(X) = n \frac{M}{N} \quad \mathbb{V}(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

Por lo tanto, para una fracción fija de defectuosos $\frac{M}{N} = p$ la función de probabilidad hipergeométrica converge a la función de probabilidad binomial cuando N se hace grande.

Distribución de Poisson

Una v.a. X con rango $R_X = \{0, 1, 2, \dots\}$ se dice tener **distribución de Poisson con parámetro λ** , si para algún $\lambda > 0$

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \mathbb{E}(X) = \lambda \quad \mathbb{V}(X) = \lambda$$

Proceso de Poisson

X : "número de eventos que ocurren en cualquier intervalo de longitud t ", tiene distribución Poisson con parámetro λt
Específicamente

$$P(X = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0, 1, 2, \dots$$

3.5 – Variables aleatorias continuas

Sea X una v.a.. Decimos que es **continua** si existe una función no negativa f , definida sobre todos los reales $x \in (-\infty, \infty)$, tal que para cualquier conjunto B de números reales

$$P(X \in B) = \int_B f(x) dx$$

O sea que la probabilidad de que X tome valores en B se obtiene al integrar la función f sobre el conjunto B .

A la función f la llamamos **función densidad de probabilidad (f.d.p.)**.

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x) dx$$

$$P(X \in B) = P(a \leq X \leq b) = \int_a^b f(x) dx$$

$$P(X = a) = \int_a^a f(x) dx = 0$$

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(x) dx$$

Función de distribución acumulada

Sea X una v.a. continua. Se define la **función de distribución acumulada de X** (abreviamos F.d.a de X) como

$$F(x) = P(X \leq x) \quad -\infty < x < \infty$$

Si X tiene f.d.p. $f(x)$ entonces

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad -\infty < x < \infty$$

Además

$$P(a \leq X \leq b) = \int_a^b f(x) dx = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = F(b) - F(a)$$

3.6 – Esperanza de una variable aleatoria continua

Para una v.a. discreta la $E(X)$ se definió como la suma de los $x_i p(x_i)$. Si X es una v.a. continua con f.d.p. $f(x)$, se define $E(X)$ sustituyendo la sumatoria por integración y $p(x_i)$ por $f(x)$.

La esperanza de una v.a. continua X con f.d.p. $f(x)$ se define como

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Teorema: Si X es una v.a. continua con f.d.p. $f(x)$ y $h(X)$ es cualquier función de X , entonces

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

Notar que de la misma forma que en el caso discreto, si $h(x) = ax + b$, es decir si h es una función lineal, aplicando las propiedades de linealidad de la integral tenemos

$$E(aX + b) = aE(X) + b$$

Varianza de una variable aleatoria continua

La interpretación de la varianza de una v.a. continua es la misma que para el caso discreto. Además sigue valiendo la igualdad

$$V(X) = E(X^2) - \mu^2$$

Pues en la demostración hecha para el caso discreto si sustituyen las sumatorias por integrales. Por la misma razón, también vale que

$$V(aX + b) = a^2V(X) \quad \text{y} \quad \sigma_{aX+b} = |a|\sigma_X$$

3.7 - Variables aleatorias continuas importantes

Distribución uniforme

La distribución continua más sencilla es análoga a su contraparte discreta.

Una v.a. continua X se dice que tiene **distribución uniforme en el intervalo $[a, b]$** , con $a < b$, si tiene función de densidad de probabilidad dada por

$$\begin{array}{c} \text{f.d.p} \\ f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{c.c.} \end{cases} \end{array} \quad \begin{array}{c} \text{F.d.a} \\ F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x < b \\ 1 & \text{si } x \geq b \end{cases} \end{array} \quad \mathbb{E}(X) = \frac{a+b}{2} \quad \mathbb{V}(X) = \frac{(b-a)^2}{12}$$

Notación: $X \sim U[a, b]$

Distribución normal o gaussiana

Sea X una v.a. Decimos que tiene distribución normal con parámetros μ y σ si su *f.d.p.* es de la forma

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (13)$$

Donde $\mu \in \mathbb{R}$ y $\sigma > 0$

Notación: $X \sim N(\mu, \sigma^2)$

Si $\mu = 0$ y $\sigma = 1$ entonces se dice que X tiene **distribución normal estándar**. Se anota $X \sim N(0,1)$

En este caso la *f.d.p.* se simboliza con $\varphi(x)$, es decir

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$$

En este caso la gráfica de la densidad es simétrica con respecto al origen.

La *F.d.a.* de una v.a. normal estándar se anota $\Phi(x)$

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Notar que como la $\varphi(x)$ es simétrica con respecto al origen entonces

$$\Phi(-x) = P(X \leq -x) = P(X > x) = 1 - P(X \leq x) = 1 - \Phi(x)$$

Una propiedad importante de la distribución normal es que si $X \sim N(\mu, \sigma^2)$ entonces la v.a. $Y = aX + b$ con a y b números reales, $a \neq 0$, tiene también distribución normal pero con parámetros $a\mu + b$ y $a^2\sigma^2$, es decir

$$X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2) \quad (14)$$

Una consecuencia importante del resultado (14) es que

$$\text{si } X \sim N(\mu, \sigma^2) \text{ entonces } Y = \frac{X - \mu}{\sigma} \sim N(0,1) \quad (15)$$

$F(x)$ no puede expresarse en términos de funciones elementales y sólo hay tablas de la F.d.a. de la normal estándar.

Para calcular $F(x)$ procedemos de la siguiente forma

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Y \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

\nwarrow
 $Y \sim N(0,1)$

Esperanza y varianza de una variable aleatoria con distribución normal

Sea $X \sim N(\mu, \sigma^2)$ entonces $E(X) = \mu$ y $V(X) = \sigma^2$

Distribución exponencial

La distribución exponencial se utiliza algunas veces para modelar el tiempo que transcurre antes de que ocurra un evento. A menudo se lo llama *tiempo de espera*.

f.d.p	F.d.a		
$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{c.c.} \end{cases}$	$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{c.c.} \end{cases}$	$E(X) = \frac{1}{\lambda}$	$V(X) = \frac{1}{\lambda^2}$

Notación: $X \sim \text{Exp}(\lambda)$

Propiedades de la distribución exponencial

1- Relación entre la distribución exponencial y el proceso de Poisson.

Si los eventos siguen un proceso de Poisson con parámetro λ , y si T representa el tiempo de espera desde cualquier punto inicial hasta el próximo evento, entonces $T \sim \text{Exp}(\lambda)$

2- Propiedad falta de memoria

La distribución exponencial no “recuerda” cuánto tiempo se ha esperado.

En particular, si el tiempo de vida de un componente sigue una distribución exponencial, entonces la probabilidad de que un componente que tiene s unidades de tiempo dure t unidades de tiempo adicionales es la misma que la probabilidad de que un componente nuevo dure t unidades de tiempo.

si $X \sim \text{Exp}(\lambda)$ y t y s son números positivos, entonces $P(X > t + s \mid X > s) = P(X > t)$

5- VARIABLES ALEATORIAS BIDIMENSIONALES

5.1 – Generalidades

Sea ε un experimento aleatorio y S un espacio muestral asociado a él. Sean $X: S \rightarrow R, Y: S \rightarrow R$, que a cada resultado $s \in S$ le asignan el par de números reales (x, y)

Llamaremos a (X, Y) **variable aleatoria bidimensional**.

Si en lugar de dos variables aleatorias, tenemos n variables aleatorias X_1, X_2, \dots, X_n , llamaremos a (X_1, X_2, \dots, X_n) **variable aleatoria n -dimensional**

Al conjunto de valores que toma la variable aleatoria bidimensional (X, Y) lo llamaremos **recorrido** de la v.a. (X, Y) y lo indicaremos R_{XY} . En otras palabras $R_{XY} = \left\{ (x, y) : x = X(s) \text{ e } y = Y(s) \text{ con } s \in S \right\}$, es decir, es la imagen por (X, Y) del espacio muestral S .

Sea (X, Y) una variable aleatoria bidimensional discreta y sea R_{XY} su recorrido (numerable). Sea $p: R_{XY} \rightarrow R$ una función que a cada elemento (x_i, y_j) le asigna un número real $p(x_i, y_j)$ tal que

$P\left(X = x_i, Y = y_j\right) = p(x_i, y_j) \quad \forall (x_i, y_j) \in R_{XY}$ y que verifica.

$$a) \quad p(x_i, y_j) \geq 0 \quad \forall (x_i, y_j) \in R_{XY}$$

$$b) \quad \sum_{(x_i, y_j) \in R_{XY}} p(x_i, y_j) = \sum_i \sum_j p(x_i, y_j) = 1$$

A esta función la llamaremos **función de probabilidad puntual conjunta** de la variable aleatoria bidimensional (X, Y) . En forma abreviada la designaremos **fdp conjunta**.

5.2 - Funciones de distribución marginales de una v.a. (X, Y) discreta

Sea (X, Y) discreta y sea $p(x_i, y_j)$ ($i=1, 2, \dots, n, j=1, 2, \dots, m$) su función de probabilidad conjunta (Eventualmente n y/o m pueden ser ∞).

La función de probabilidad marginal de X es

$$p(x_i) = P(X = x_i) = \sum_{j=1}^m p(x_i, y_j) \quad (i=1, 2, \dots, n)$$

La función de probabilidad marginal de Y es

$$q(y_j) = P(Y = y_j) = \sum_{i=1}^n p(x_i, y_j) \quad (j=1, 2, \dots, m)$$

5.3 - Funciones de probabilidades condicionales

En general definimos la *función de probabilidad puntual de X condicional a Y* como sigue:

$p(x_i|y_j) = P(X = x_i|Y = y_j) = \frac{p(x_i, y_j)}{q(y_j)}$, es decir como el cociente de la función de probabilidad conjunta de (X, Y) y la función de probabilidad puntual marginal de Y .

Análogamente, definimos la *función de probabilidad puntual de Y condicional a X*:

$q(y_j|x_i) = P(Y = y_j|X = x_i) = \frac{p(x_i, y_j)}{p(x_i)}$, es decir como el cociente de la función de probabilidad puntual conjunta de (X, Y) y la función de probabilidad puntual marginal de X .

5.4- Variables aleatorias independientes

Sea (X, Y) una variable aleatoria bidimensional discreta. Sea $p(x_i, y_j)$ su *fdp* conjunta y $p(x_i)$ y $q(y_j)$ las correspondientes *fdp* marginales de X e Y . Decimos que X e Y son **variables aleatorias independientes** si y sólo si

$$p(x_i, y_j) = p(x_i)q(y_j) \quad \forall (x_i, y_j) \in R_{XY}$$

Teorema

Sea (X, Y) una variable aleatoria bidimensional discreta cuyas *fdp* conjunta, condicionales y marginales son, respectivamente, $p(x_i, y_j)$; $p(x_i|y_j)$, $q(y_j|x_i)$ y $p(x_i)$, $q(y_j)$.

Entonces, X e Y son variables aleatorias independientes si y sólo si

- 1) $p(x_i|y_j) = p(x_i) \quad \forall (x_i, y_j) \in R_{XY}$, o
- 2) $q(y_j|x_i) = q(y_j) \quad \forall (x_i, y_j) \in R_{XY}$, que es equivalente a lo anterior

5.5 - Función de una variable aleatoria bidimensional

Algunas variables aleatorias que son función de variables aleatorias bidimensionales son $Z = X + Y$, $Z = X \cdot Y$, $Z = X / Y$, $Z = \min(X, Y)$, $Z = \max(X, Y)$, etc.

Lo anterior se puede generalizar si en lugar de dos variables aleatorias tenemos n variables aleatorias X_1, X_2, \dots, X_n , y $z = H(x_1, x_2, \dots, x_n)$ es una función de n variables a valores reales.

Esperanza de una v.a. que es función de una v.a. bidimensional

Teorema Sea (X, Y) una variable aleatoria bidimensional y sea $Z = H(X, Y)$ una variable aleatoria que es función de (X, Y) .

Si Z es variable aleatoria discreta que proviene de la variable aleatoria bidimensional discreta (X, Y) cuyo recorrido es R_{XY} y su *fdp* conjunta es $p(x_i, y_j)$, entonces:

$$E(Z) = E[H(X, Y)] = \sum_{(x_i, y_j) \in R_{XY}} H(x_i, y_j) p(x_i, y_j)$$

Esperanza de una suma de variables aleatorias

Sean X e Y dos variables aleatorias arbitrarias. Entonces $E(X + Y) = E(X) + E(Y)$.

Sean X_1, X_2, \dots, X_n n variables aleatorias arbitrarias. Entonces:

$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$ o, en notación más concentrada:

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

Observación: se deduce que la esperanza verifica la **propiedad lineal**:

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

Si (X, Y) es una variable aleatoria bidimensional tal que X e Y son variables aleatorias **independientes**, entonces: $E(X \cdot Y) = E(X) \cdot E(Y)$

Varianza de una suma de variables aleatorias

$$V(X + Y) = V(X) + V(Y) + 2\sigma_{XY} \quad \text{con} \quad \sigma_{XY} = E(X \cdot Y) - E(X) \cdot E(Y)$$

$\sigma_{XY} = E(X \cdot Y) - E(X) \cdot E(Y)$ se la llama la **covarianza de X e Y** .

Observaciones:

1- Teniendo presente la definición de la desviación estándar de una v.a. X : $\sigma_X = \sqrt{V(X)}$, vemos que a la propiedad anterior la podemos escribir:

$$V(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

2- Análogamente se prueba que $V(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$

3- X e Y son independientes, entonces $V(X + Y) = V(X - Y) = V(X) + V(Y)$

Esto es porque si las variables aleatorias X e Y son independientes, entonces $E(X \cdot Y) = E(X) \cdot E(Y)$.

Por lo tanto la covarianza vale cero: $\sigma_{XY} = E(X \cdot Y) - E(X) \cdot E(Y) = 0$.

4- Podemos generalizar, usando el principio de inducción completa, al caso de n variables aleatorias independientes:

Si X_1, X_2, \dots, X_n son n variables aleatorias independientes entonces:

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n) \quad \text{o, en forma más compacta,} \quad V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i).$$

5- Vemos que la esperanza de la suma de dos variables aleatorias X e Y es igual a la suma de las esperanzas $E(X + Y) = E(X) + E(Y)$ cualesquiera sean X e Y . En cambio la varianza de la suma de las variables aleatorias X e Y es, en general, igual a la suma de las varianzas, $V(X + Y) = V(X) + V(Y)$, sólo si X e Y son variables independientes.

5.6 - Covarianza

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

Notación: la notación usual para la covarianza de X e Y es σ_{XY} o $\text{Cov}(X, Y)$

Si X e Y son variables aleatorias independientes, entonces $\text{Cov}(X, Y) = 0$.

Propiedades de la covarianza

Las siguientes propiedades son útiles y su verificación se deja como ejercicio

1- $\text{Cov}(a + bX, c + dY) = bd\text{Cov}(X, Y)$

2- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

3- $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$

4- $\text{Cov}(X, X) = V(X)$

5.7 - Coeficiente de correlación lineal.

Sea (X, Y) una variable aleatoria bidimensional. Definimos el *coeficiente de correlación lineal entre X e Y* como $\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$

Propiedad 1

Si X e Y son variables aleatorias independientes entonces $\rho_{XY} = 0$.

Propiedad 2 :

$$-1 \leq \rho_{XY} \leq 1$$

Propiedad 3 :

Si $\rho_{XY}^2 = 1$, entonces con probabilidad 1 es $Y = AX + B$ donde A y B son constantes.

Propiedad 4 :

Si X e Y son dos variables aleatorias tales que $Y = AX + B$, donde A y B son constantes, entonces $\rho_{XY}^2 = 1$. Si $A > 0$ es $\rho_{XY} = 1$ y si $A < 0$ es $\rho_{XY} = -1$.

6- SUMA DE VARIABLES ALEATORIAS Y TEOREMA CENTRAL DEL LÍMITE

6.1 – Suma de variables aleatorias independientes

1- Suma de variables aleatorias independientes con distribución Poisson

$$X \sim P(\lambda_1) ; Y \sim P(\lambda_2) ; X \text{ y } Y \text{ independientes} \Rightarrow X + Y \sim P(\lambda_1 + \lambda_2)$$

2- Suma de variables aleatorias binomiales independientes

$$X \sim B(n_1, p) ; Y \sim B(n_2, p) ; X \text{ y } Y \text{ independientes} \Rightarrow X + Y \sim B(n_1 + n_2, p)$$

Observación: en los dos casos anteriores se puede generalizar el resultado a n variables aleatorias independientes, usando el principio de inducción completa, es decir

1- Si X_1, X_2, \dots, X_n son n variables aleatorias independientes donde $X_i \sim P(\lambda_i)$ para todo

$$i = 1, 2, \dots, n \text{ entonces } \sum_{i=1}^n X_i \sim P\left(\sum_{i=1}^n \lambda_i\right)$$

2- Si X_1, X_2, \dots, X_n son n variables aleatorias independientes donde $X_i \sim B(n_i, p)$ para todo

$$i = 1, 2, \dots, n \text{ entonces } \sum_{i=1}^n X_i \sim B\left(\sum_{i=1}^n n_i, p\right)$$

Suma de variables aleatorias normales independientes

Si X e Y son variables aleatorias independientes donde $X \sim N(\mu_1, \sigma_1^2)$ y $Y \sim N(\mu_2, \sigma_2^2)$ entonces $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Por inducción completa se puede generalizar este resultado a n variables:

Si X_1, X_2, \dots, X_n son n variables aleatorias independientes donde $X_i \sim N(\mu_i, \sigma_i^2)$ para todo $i = 1, 2, \dots, n$ entonces $\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$

De lo anterior y del hecho que $X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$ tenemos:

Si X_1, X_2, \dots, X_n son n variables aleatorias independientes donde $X_i \sim N(\mu_i, \sigma_i^2)$ para todo $i = 1, 2, \dots, n$ entonces $\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$ donde a_1, a_2, \dots, a_n son números reales

Se dice que $\sum_{i=1}^n a_i X_i$ es una **combinación lineal de variables aleatorias**.

Promedio de variables aleatorias normales independientes

Si X_1, X_2, \dots, X_n son n variables aleatorias independientes donde $X_i \sim N(\mu, \sigma^2)$ para todo

$i = 1, 2, \dots, n$ entonces la v.a. $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ tiene distribución normal con

media μ y varianza $\frac{\sigma^2}{n}$

Observación: a \bar{X} se lo llama **promedio muestral** o **media muestral**

6.2 - Teorema central del límite

Teorema central del límite (T.C.L.):

Sean X_1, X_2, \dots, X_n variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \dots, n$, es decir **independientes idénticamente distribuidas**

Sea la v.a. $S_n = \sum_{i=1}^n X_i$ y sea $Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$.

Entonces $\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$, esto es $\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$

Observaciones:

1- Notar que $E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu$ y $V(S_n) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = n\sigma^2$

Por lo tanto $Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ es la v.a. S_n **estandarizada**

2- Notar que $P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = P\left(\frac{\frac{S_n - n\mu}{n}}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right)$, por lo tanto también se puede

enunciar el Teorema central del límite de la siguiente forma

Sean X_1, X_2, \dots, X_n variables aleatorias independientes con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \dots, n$, es decir **independientes idénticamente distribuidas**

Sea la v.a. promedio muestral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ y sea $Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$.

Entonces $\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$, esto es $\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$

Donde $Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ es el **promedio muestral estandarizado**

3- Aunque en muchos casos el T.C.L. funciona bien para valores de n pequeños, en particular donde la población es continua y simétrica, en otras situaciones se requieren valores de n mas grandes, dependiendo de la forma de la distribución de las X_i . En muchos casos de interés práctico, si $n \geq 30$, la aproximación normal será satisfactoria sin importar cómo sea la forma de la distribución de las X_i . Si $n < 30$, el T.C.L. funciona si la distribución de las X_i no está muy alejada de una distribución normal

Aplicaciones del Teorema central del límite

Aproximación normal a la distribución binomial

El Teorema central del límite se puede utilizar para aproximar las probabilidades de algunas variables aleatorias discretas cuando es difícil calcular las probabilidades exactas para valores grandes de los parámetros.

Supongamos que X tiene una distribución binomial con parámetros n y p . Para calcular $P(X \leq k)$

debemos hacer la suma $P(X \leq k) = \sum_{i=0}^k P(X = i)$ o recurrir a las tablas de la F.d.a., pero para valores de n grandes no existen tablas, por lo tanto habría que hacer el cálculo en forma directa y muchas veces es laborioso.

Como una opción podemos considerar a X como suma de variables aleatorias más simples, específicamente, si definimos

$$X_i = \begin{cases} 1 & \text{si en la } i\text{-ésima repetición de } \varepsilon \text{ ocurre éxito} \\ 0 & \text{caso contrario} \end{cases} \quad i = 1, 2, \dots, n$$

entonces cada X_i se la puede considerar $B(1, p)$, y además X_1, X_2, \dots, X_n son independientes

Podemos escribir $X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$ y *si n es grande* entonces X tendrá *aproximadamente* una distribución normal con parámetros np y $np(1-p)$, es decir

$$Z_n = \frac{X - n\mu}{\sqrt{n\sigma^2}} = \frac{X - n.p}{\sqrt{n.p(1-p)}} \approx N(0,1) \quad \text{si } n \text{ es lo suficientemente grande}$$

Observaciones:

1- La aproximación normal a la distribución binomial funciona bien aun cuando n no sea muy grande si p no está demasiado cerca de cero o de uno. En particular la aproximación normal a la binomial es buena si n es grande, $np > 5$ y $n(1-p) > 5$, *pero es más efectivo aplicar esta aproximación cuando $np > 10$ y $n(1-p) > 10$*

2- Corrección por continuidad.

$$P(X \leq k) \approx P\left(X \leq k + \frac{1}{2}\right)$$

$$\text{Y en lugar de } P(X \geq k) \approx P\left(X \geq k - \frac{1}{2}\right)$$

Aproximación normal a la distribución Poisson

Se puede probar aplicando Teorema central del límite que

Si $X \sim P(\lambda)$ entonces para λ suficientemente grande $\frac{X - \lambda}{\sqrt{\lambda}}$ tiene aproximadamente distribución $N(0,1)$

Es decir para λ suficientemente grande $\frac{X - \lambda}{\sqrt{\lambda}} \approx N(0,1)$

En la práctica si $\lambda \geq 30$ la aproximación es buena.

Observación: la demostración es sencilla si λ es igual a un número natural n pues, si consideramos las variables aleatorias $X_i \sim P(1)$ con $i = 1, 2, \dots, n$ independientes, entonces ya sabemos que

$$\sum_{i=1}^n X_i \sim P\left(\sum_{i=1}^n 1\right), \text{ es decir } \sum_{i=1}^n X_i \sim P(n)$$

Pero además por T.C.L. si n es grande $\sum_{i=1}^n X_i$ tiene aproximadamente distribución normal con parámetros $n\mu = n \times 1 = n$ y $n\sigma^2 = n \times 1 = n$

O sea la distribución de $\sum_{i=1}^n X_i$ que es exactamente Poisson con parámetro n , se puede aproximar con una $N(n, n)$, por lo tanto $\frac{X - n}{\sqrt{n}} \approx N(0,1)$ aproximadamente para valores de n suficientemente grandes