Apuntes de
Introducción a la Probabilidad MA-1030

Ghael Solórzano Ramírez

October 2024

Contents

1 Variables Aleatorias Continuas (v.a.c.)

Definición: Variable aleatoria continua

Dado un espacio de probabilidad $(\Omega, A, \mathbb{P})$ una variable aleatoria

$$\text{X: } \Omega \to \mathbb{R}$$

se dice ser continua si y solo si R_x = X(r): r $\in \Omega$ es infinito no numerable.

Función de densidad

Sea (Ω, A, $\mathbb{P}$) un espacio de probabilidad y X: $\Omega \to \mathbb{R}$ una v.a.c
Una **función de densidad o función de densidad de probabilidad PDF** si y solo si se cumple:

1. $\displaystyle\int_{\mathbb{R}} \text{f(T)dt} = \int_{-\infty}^{+\infty} \text{f(T)dT} = 1$

2. Para a y b $\in \mathbb{R}$ con a $\leq$ b: $\mathbb{P}$(a$\leq$X$\leq$b) = $\displaystyle\int_a^b \text{f(T)dT}$

Nota: Sea $(\Omega, A, \mathbb{P})$ un espacio de proba.

1. Si $f_x : \mathbb{R} \to$ [0,+∞[es una función de densidad para una v.a.c X: $\Omega \to \mathbb{R}$ entonces se suele denotar f_x = f

2. La función de proba puntual es 0

Teorema: Función de densidad de una función de v.a.c

Sea (Ω, A, $\mathbb{P}$) un espacio de probabilidad, X: $\Omega \to \mathbb{R}$ una v.a.c con PDF $f_x : \mathbb{R} \to$ [0,+∞[.
Sea $g : R_x \to \mathbb{R}$ una función estrictamente monótona y derivable. Entonces, $Y = g_o X = g(X)$ es una nueva v.a.c. y esta cumple que su PDF es la función $f_y : \mathbb{R} \to [0, +\infty[$ dada por:

$$f_y(T) = \begin{cases} f_x(g^{-1}(T)) \cdot |\frac{dg^{-1}(T)}{dT}| & \text{si } T \in R_y \\ 0 & \text{en caso contrario.} \end{cases}$$

Función de distribución acumulada

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad y sea X: $\Omega \to \mathbb{R}$ con PDF f_x: $\mathbb{R} \to$ [0,+∞[
La **función de distribución acumulada** de X es la función $F_X : \mathbb{R} \to [0, 1]$ dada por:

$$F_x(\lambda) = \int_{-\infty}^{\lambda} f_x(T)dT = \mathbb{P}(\{X \leq \lambda\}) = \mathbb{P}(\{X < \lambda\})$$

unicamente en caso de que F_x sea continua por derecha, esto es: $\lim_{\lambda \to \lambda_0^+} F_x(\lambda) = F_x(\lambda_0)$ con $\lambda_0 \in \mathbb{R}$

Teorema: Propiedades de la función de distribución acumulada

Sea $F_x : \mathbb{R} \to [0, +\infty[$ la función de distribución acumulada de una v.a.c X: $\Omega \to \mathbb{R}$ con PDF f_x: $\mathbb{R} \to$ [0,+∞[, entonces:

1. F_x es creciente y continua por la derecha, tal que $F_x(a) \leq F_X(b)$ con $a \leq b$ y a,b $\in \mathbb{R}$

2. $\lim_{\lambda \to -\infty} F_x(\lambda) = 0$ y $\lim_{\lambda \to +\infty} F_x(\lambda) = 1$

3. Para a,b $\in \mathbb{R}$ con $a \leq b$: $\mathbb{P}(\{a \leq X \leq b\}) = \mathbb{P}(\{a < X < b\}) = F_x(b) - F_x(a)$

4. Para cada $\lambda \in \mathbb{R}$ F_x es derivable en λ y: $F_x^{'} = f_x(\lambda)$

5. $R_x = \{T \in \mathbb{R} : f_x(T) > 0\}$

Esperanza y varianza de una v.a.c.

Sea (Ω, A, ℙ) un espacio de probabilidad, X: $\Omega \to \mathbb{R}$ una v.a.c con PDF $f_x : \mathbb{R} \to [0,+\infty[$.

1. La esperanza o valor esperado de X es

$$\int_{-\infty}^{+\infty} T \cdot f_x(T)dT \quad \text{versión continua}$$

siempre que esa integral converja

2. Para $g : \mathbb{B} \to \mathbb{R}$ función se define

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(T) \cdot f_x(T)dT$$

siempre que la integral converja

3. La varianza de X es

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{+\infty} T^2 \cdot f_x(T)dT - \left(\int_{-\infty}^{+\infty} T \cdot f_x(T)dT \right)^2$$

Nota: Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad, X: $\Omega \to \mathbb{R}$ una v.a.c con PDF $f_x : \mathbb{R} \to [0,+\infty[$. Sea Para a,b $\in \mathbb{R}$:

1. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
2. $V(aX + b) = a^2 Var(X)$

Momentos de una v.a.c

Sea (Ω, A, ℙ) un espacio de probabilidad, X: $\Omega \to \mathbb{R}$ una v.a.c con PDF $f_x : \mathbb{R} \to [0,+\infty[$.
Para cada m **numero entero positivo** se define el m-esimo momento de X como:

$$\mathbb{E}[X^m] = \int_{-\infty}^{+\infty} T^m \cdot f_x(T)dT$$

siempre que esta integral exista, o sea $\mathbb{E}[X^m] \neq +\infty$

Nota:

1. A los momentos $\mathbb{E}[X^m]$ se les dice **momentos centrados en 0**
2. A los momentos $\mathbb{E}[(X - \mu)^m]$ con $\mu = \mathbb{E}[X^m]$ se les llama **momentos centrados en la media**
3. $Var(X) = \mathbb{E}[(X - \mu)^2]$ donde $\mu = \mathbb{E}[X]$
4. Si m es un entero positivo tal que $\mathbb{E}[|X|^m]$ existe y es finito, entonces los momentos $\mathbb{E}[X], \mathbb{E}[X^2], ..., \mathbb{E}[X^{m-1}], \mathbb{E}[X^m]$ existen y son finitos.

Función generadora de momentos

Sea X una v.a.c con PDF $f_x : \mathbb{R} \to [0,+\infty[$.
Se define la **función generadora de momentos** de X mediante

$$\mu_x(\lambda) = \mathbb{E}[e^{\lambda \cdot x}] = \int_{-\infty}^{+\infty} e^{\lambda \cdot T} \cdot f_x(T)dT$$

Siempre que $\mu_x(\lambda)$ converja para λ en algún intervalo abierto que contiene a 0.

Teorema Sea X una v.a.c tal que μ_x es infinitamente derivable en algún intervalo abierto que contiene a 0 (y también existe ahí). Entonces para cada entero positivo n se tiene que:

$$\frac{d^n M_x}{d\lambda^n}(0) = \mathbb{E}[X^n]$$

Variables aleatorias continuas igualmente distribuidas

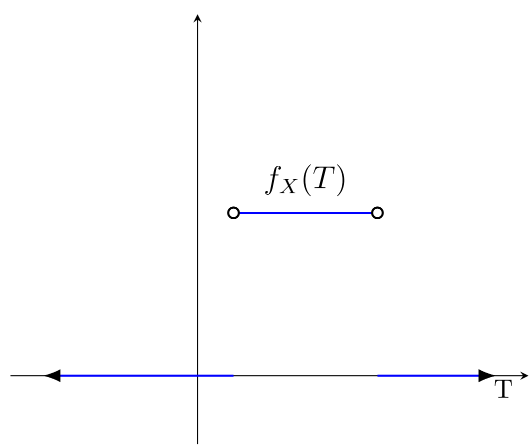
Sean X, Y dos v.a.c., estas se dicen estar **igualmente distribuidas**, y se denota X Ỹ o bien X $\overset{d}{=}$ Y, si y solamente si X,Y tienen la misma función de distribución acumulada, o sea $F_x = F_y$.

Teorema Sean X,Y v.a.c, si existe $\delta \in]0, +\infty[$ tal que $M_x(\lambda)$ y $\mu_y(\lambda)$ existen y son iguales para $\lambda \in]-\delta, \delta[$ entonces X,Y están igualmente distribuidas, tal que $F_x = F_y$ para $T \in \mathbb{R}$

Variable aleatoria uniforme

Sea X una v.a.c, X se dice seguir una **distribución de v.a.c uniforme** si y solo si para algunos números reales a y b se cumple que su PDF es:

$$f_x(T) = \begin{cases} \frac{1}{b-a} & \text{si } a < T < b \\ 0 & \text{en caso contrario} \end{cases}$$



En tal caso se denota $X \sim U_{nif}(a, b)$ o $X \sim U(a, b)$, ademas:

- $\mathbb{E}[X] = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$
- $F_x(\lambda) = \displaystyle\int_{-\infty}^{\lambda} f_x(T)dT =$

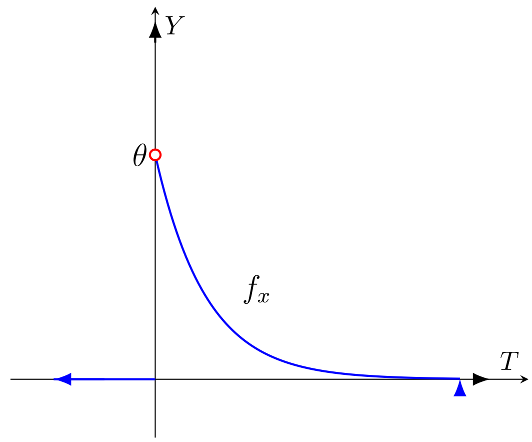
$$\begin{cases} 0 & \text{si } \lambda \leq a \\ \dfrac{\lambda - a}{b - a} & \text{si } a < \lambda < b \\ 1 & \text{si } b \leq \lambda \end{cases}$$

- $\mu_x(T) = \begin{cases} 1 & \text{si } T = 0 \\ \frac{e^{Tb} - e^{Ta}}{T(b-a} & \text{si } T \neq 0 \end{cases}$

Variable aleatoria exponencial

Sea X una v.a.c., X se dice seguir una **distribución exponencial** de parámetro θ si y solo si su función de densidad $f_x : \mathbb{R} \to [0,+\infty[$ cumple que:

$$f_x(T) = \begin{cases} \theta e^{-\theta T} & \text{si } 0 \leq T \\ 0 & \text{si } T < 0 \end{cases}$$



En tal caso se denota $X \sim exp(\theta)$.
Ademas, se cumple que:

- $\mathbb{E}[X] = \frac{1}{\theta}$,
- $Var(X) = \frac{1}{\theta^2}$
- $\mu_x(\lambda) = \frac{\theta}{\theta - \lambda}$
- $F_X(\lambda) = \begin{cases} 0 & \text{si } \lambda \leq 0 \\ 1 - e^{-\theta\lambda} & \text{si } 0 < \lambda \end{cases}$

Nota: Usualmente una v.a.c. $X \sim exp(\theta)$ se usa para representar la cantidad de tiempo necesario para que ocurra un hecho, y θ usualmente representa el promedio de sucesos que ocurren por unidad de tiempo.

Propiedad de perdida de memoria

Si $X \sim exp(\theta)$ entonces para S, T $\in \mathbb{R} > 0$

$$\mathbb{P}_{\{S<X\}}(\{S + T < X\}) = \mathbb{P}(\{T < X\})$$

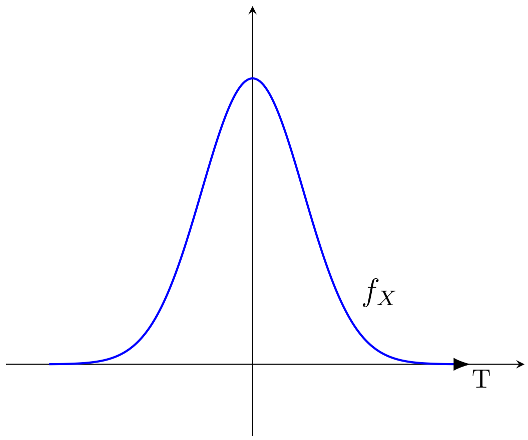
Esta es la única v.a.c con esta propiedad

Variable aleatoria normal (Gaussiana)

Sea X una v.a.c, X sigue una **distribución normal** de parámetros μ y σ^2 (esperanza y varianza de X) si y solo si su PDF f_x viene dada por

$$f_x(T) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \frac{-(T-\mu)^2}{2\sigma^2}$$

- $\mathbb{E}[X] = \mu$
- $Var(X) = \sigma^2$
- $\mu_x(\lambda) = e^{\mu \cdot \lambda + \frac{1}{2}\sigma^2\lambda^2}$ definida en $\lambda \in \mathbb{R}$



En tal caso, se escribe $X \sim N(\mu, \sigma^2)$. En la practica es común usar $\mu = 0$ y $\sigma^2 = 1$, de modo que $X \sim N(0, 1)$. En tal caso se dice que X esta distribuida como una **v.a.c normal estándar** y entonces:

$$f_x(T) = \frac{1}{\sqrt{2\pi}} e^{\frac{-T^2}{2}}$$

y en este caso particular:

- $\mathbb{E}[X] = 0$,
- $Var(X) = 1$,
- $\mu_x(\lambda) = e^{\frac{\lambda^2}{2}}$

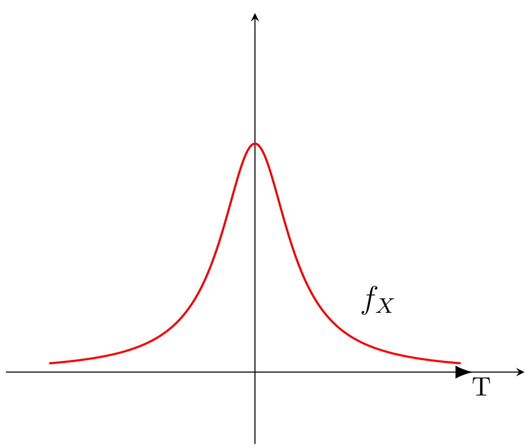
para todo n entero positivo:

- $\mathbb{E}[X^{2n+1}] = 0$
- $\mathbb{E}[x^{2n} = \frac{(2n)!}{2^n \cdot n!}]$

Variable aleatoria de Cauchy

Sea X una v.a.c.. X se dice estar **distribuida como una v.a.c de Cauchy** de parámetros α y β si y solo si su PDF $f_x : \mathbb{R} \to [0, +\infty[$ viene dada por:

$$f_x(T) = \frac{1}{\pi} [\frac{\beta}{(T-\alpha)^2 + \beta^2}]$$



Ademas

- $F_x(T) = \frac{1}{\pi} \arctan(\frac{T-\alpha}{\beta}) + \frac{1}{2}$,
- No posee función generadora de momentos ni $\mathbb{E}[X^n]$ para n entero positivo.

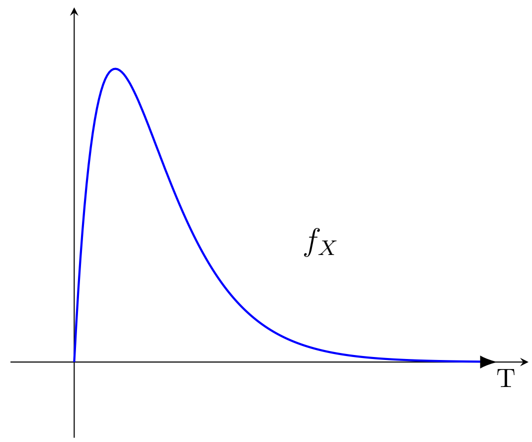
En tal caso se escribe $X \sim Cauchy(\alpha, \beta)$.
La forma usual de esta variable es con $\beta = 1$, en cuyo caso se denota $X \sim Cauchy(\alpha)$ y $f_x(T) = \frac{1}{\pi}[\frac{1}{(T-\alpha)^2+1}]$.
La forma estándar es con $\alpha = 0$ y $\beta = 1$, osea $X \sim Cauchy(0, 1)$, en cuyo caso $f_x(T) = \frac{1}{\pi}[\frac{1}{T^2+1}]$

Variable aleatoria con distribución Gamma

Sea X una v.a.c.. X se dice estar **distribuida como una v.a.c. Gamma** de parámetros α, β números reales positivos si y solo si su PDF $f_x : \mathbb{R} \to [0, +\infty[$ viene dado por:

$$f_x(T) = \begin{cases} \frac{\beta^{\alpha} \cdot T^{\alpha-1} \cdot e^{-\beta T}}{\Gamma(\alpha)} & \text{si } 0 < T \\ 0 & \text{si } T \leq 0 \end{cases}$$

Donde $\Gamma(\alpha) = \int_0^{+\infty} z^{\alpha-1} \cdot e^{-z} dz$ es la funcion Gamma que es continua y converge en su dominio.



En tal caso se escribe $X \sim Gamma(\alpha, \beta)$ o $X \sim \Gamma(\alpha, \beta)$. Ademas:

- $F_x(T) = \frac{\gamma(\alpha, \beta T)}{\Gamma(\alpha)}$
  donde $\gamma(\alpha, \beta T) = \int_0^{\beta T} z^{\alpha-1} e^{-z} dz$ es la funcion gamma incompleta inferior.
- $\mathbb{E}[X] = \frac{\alpha}{\beta}$
- $Var(X) = \frac{\alpha}{\beta^2}$
- $\mu_x(\lambda) = (\frac{1}{1-\frac{\lambda}{\beta}})^{\alpha}$ si $T \in]-\infty, \beta[$

Nota:

1. Para n entero positivo:
   $\Gamma(n) = (n-1)!$
2. Si $p \in]0, +\infty[$ entonces:
   $\int_0^{+\infty} T^{\alpha-1} e^{-pT} dT = \frac{\Gamma(\alpha)}{P^{\alpha}}$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
4. Si $\alpha = 1$ y $\beta = \theta$, entonces una v.a.c X con $X \sim \Gamma(\alpha, \beta) = \Gamma(\alpha, \theta)$ tambien cumple que $X \sim exp(\theta)$, y en este caso:
   $f_x(T) = \begin{cases} \theta e^{-\theta T} & \text{si } 0 < T \\ 0 & \text{si } T \leq 0 \end{cases}$

2 Funciones conjuntas de distribución

Función de probabilidad puntual conjunta

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}$ y $Y : \Omega \to \mathbb{R}$ variables aleatorias discretas con rangos respectivos $R_X = \{x_1, x_2, ..., x_m\}$ y $R_Y = \{y_1, y_2, ..., y_M\}$, donde m, M son enteros positivos o infinito (numerable), es decir, $1 \leq m \leq +\infty$ y $1 \leq M \leq +\infty$.
Se define la **función de probabilidad puntual conjunta** de X, Y mediante:

$$P_{X,Y}(x_i, y_j) = \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) = \mathbb{P}(\{X = x_i\}, \{Y = y_j\}) = \mathbb{P}(X = x_i, Y = y_i)$$

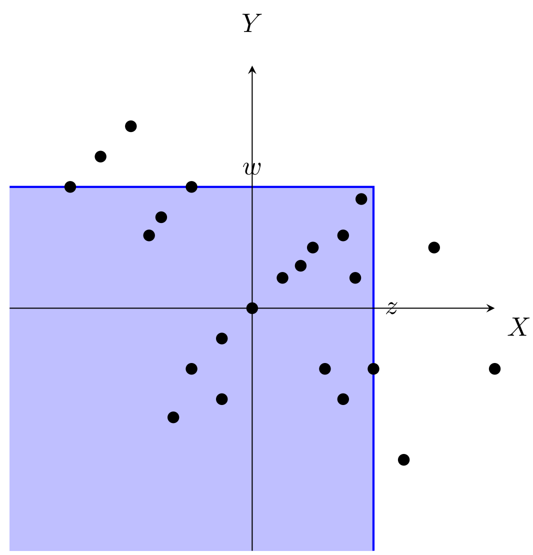
Para $x_i \in R_X$ y $y_j \in R_Y$.
Nota: Se tiene:

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{M} p_{X,Y}(x_i, y_j) \right) = \sum_{j=1}^{M} \left(\sum_{i=1}^{m} p_{X,Y}(x_i, y_j) \right) = 1.$$

Si además se define $p_{X,Y}(w, z) = 0$ para (w, z) un punto de $\mathbb{R}^2$ que no está en $R_X \times R_Y$, es decir, cuando $w \notin R_X$ o $z \notin R_Y$ o ambos, entonces para $A \subseteq \mathbb{R}^2$ se tiene:

$$\mathbb{P}(\{(X, Y) \in A\}) = \sum_{(w,z) \in A} p_{X,Y}(w, z).$$

Esto es igual a sumar todos los posibles valores de $p_{X,Y}(x_i, y_j)$ en los puntos (x_i, y_j) tales que $x_i \leq z$ y $y_j \leq w$. O sea, sumar $p(x_i, y_j)$ en los puntos del rectángulo $] - \infty, z]X] - \infty, w]$.



Nota: $F_{X,Y}$ se define sobre todo $\mathbb{R}$. De hecho $F_{X,Y} : \mathbb{R} \to [0, 1] yestaescreciente.$

Función de probabilidad marginal

Sean X, Y v.a.d con:

$$p_X(x_i) = \mathbb{P}(\{X = x_i\}) \qquad P_Y(y_j) = \mathbb{P}(\{Y = y_j\}) \qquad P_{X,Y}(x_i, y_j) = \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\})$$

¿Como se relacionan?

Las funciones de proba puntual de X, Y se pueden recuperar a partir de la conjunta:
1. Función de probabilidad (puntual) marginal de X Esto es igual a sumar en la tabla de $p_{X,Y}$ las entradas en la fila i

$$p_X(x_i) = \sum_{y_j \in R_y} (p_{X,Y}(x_i, y_j))$$
$$R_y = \{y_1, y_2, ..., y_M\} \implies = \sum_{j=1}^{M} (p_{X,Y}(x_i, y_j))$$

2. Función de probabilidad (puntual) marginal de Y Esto es igual a sumar en la tabla de $p_{X,Y}$ las entradas en la columna j

$$p_X(y_j) = \sum_{x_i \in R_x} (p_{X,Y}(x_i, y_j))$$
$$R_x = \{x_1, x_2, ..., x_m\} \implies = \sum_{i=1}^{m} (p_{X,Y}(x_i, y_j))$$

Observación: Recuperando esperanzas a partir de $p_{X,Y}$ Sean X, Y v.a.d. con rangos $R_x = \{x_1, x_2, ..., x_m\}$ y $R_y = \{y_1, y_2, ..., y_M\}$
Para f una función bien definida sobre R_X y g una función bien definida sobre R_Y, es posible calcular $\mathbb{E}[f(x)]$ y $\mathbb{E}[g(y)]$ a partir de $p_{X,Y}$.

Inicialmente

$$\mathbb{E}[f(x)] = \sum_{i=1}^{m} (f(x_i) \cdot p_x(x_i))$$
$$\mathbb{E}[g(y)] = \sum_{j=1}^{M} (g(y_j) \cdot p_y(y_j))$$

Ahora también se puede escribir como

$$\mathbb{E}[f(x)] = \sum_{i=1}^{m} \sum_{j=1}^{M} (f(x_i) \cdot p_{X,Y}(x_i, y_j))$$
$$\mathbb{E}[g(y)] = \sum_{j=1}^{M} \sum_{i=1}^{m} (g(y_j) \cdot p_{X,Y}(x_i, y_j))$$

3 Propiedades de la esperanza y la covarianza

Esperanza de una v.a.d. que es función de dos v.a.d.

Sea $\Omega, A, \mathbb{P}$ un espacio de probabilidad y sea $X : \Omega \to \mathbb{R}$ y $Y : \Omega \to \mathbb{R}$ dos v.a.d. de rangos respectivos $R_x = \{x_1, x_2, ..., x_m\}$ y $R_y = \{y_1, y_2, ..., y_M\}$, con m, M enteros positivos o infinitos.
Sea h una función definida en algún subconjunto de $\mathbb{R}^2$ que contiene a $R_X x R_Y$, y que toma valores reales, entonces $Z = h(X, Y)$ es una nueva v.a.d y su esperanza es:

$$\mathbb{E}[Z] = \mathbb{E}[h(X, Y)] = \sum_{i=1}^{m} \left(\sum_{j=1}^{M} (h(x_i, y_j) \cdot p_{X,Y}(x_i, y_j)) \right)$$

siempre que esta suma sea infinita

Teorema:

Sean X, Y v.a.d y a un numero real

1. $\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$
2. $Var(aX + Y) = a^2 Var(x) + Var(Y) + 2a(\mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$

Covarianza

Sean X, Y v.a.d. tales que $\mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[x \cdot Y]$ son finitos
Se define la **covarianza** de X, Y como:

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] \\ &= \mathbb{E}[(X - \mu_x) \cdot (Y - \mu_y)] \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{M} ((x_i - \mu_x) \cdot (y_j - \mu_y) \cdot p_{X,Y}(x_i, y_j)) \right) \end{aligned}$$

donde $\mu_x = \mathbb{E}[X]$ y $\mu_y = \mathbb{E}[Y]$ que son números reales

Nota: No siempre $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
Observación: Si X=Y entonces $Cov(X, X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = Var(X)$

Teorema

Sean X, Y v.a.d y a, b, c, d números reales

1. $Cov\,(aX + b, cY + d) = a \cdot c \cdot Cov(X, Y)$
2. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Probabilidad condicional dado que una variable tiene valor fijo

Sean $X : \Omega \to \mathbb{R}$ y $Y : \Omega \to \mathbb{R}$ dos v.a.d. de rangos respectivos $R_x = \{x_1, x_2, ..., x_m\}$ y $R_y = \{y_1, y_2, ..., y_M\}$, con m, M enteros positivos o $+\infty$.
Sea $y_j \in R_Y$ y considérese el evento $E = \{Y = y_j\}$ y supongase que $\mathbb{P}(E) > 0$.
Se recuerda que hay una probabilidad condicional asociada al evento E y devuelve

$$\mathbb{P}_{\{Y=y_j\}}(F) = \mathbb{P}(F|\{Y = y_j\}) = \frac{\mathbb{P}(\{Y = y_j\} \cap F)}{\mathbb{P}(\{Y = y_j\})}$$

Tiene sentido considerar la función de probabilidad puntual de la v.a.d. X respecto a esta probabilidad condicional $\mathbb{P}_{\{Y=y_j\}}(\cdot)$, que se denota por $p_X^{(j)} = p_X|Y = y_j$, esta viene dada por:

$$\begin{aligned} p_X^{(j)}(x_i) &= p_X|Y = y_j(x_i) \\ &= \mathbb{P}_{Y=y_j}(\{X = x_i\}) \\ &= \frac{\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\})}{\mathbb{P}(\{Y = y_j\})} \\ &= \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} \end{aligned}$$

Nota: Símbolo para la función: $F \mapsto \mathbb{P}_{\{Y=y_j\}}(F)$
Esta función $p_X^{(j)} = p_X|Y = y_j$ tiene dominio $R_x = \{x_1, x_2, ..., x_m\}$ y se denomina **función de probabilidad puntual de X dado que** $Y = y_j$.
Similarmente, si $x_i \in R_x$ es tal que $\mathbb{P}(\{X = x_i\}) > 0$ entonces se tiene que:

$$\begin{aligned} p_Y^{(i)}(y_j) &= p_Y|X = x_i(y_j) \\ &= \mathbb{P}_{\{X=x_i\}}(\{Y = y_j\}) \\ &= \frac{p_{X,Y}(x_i, y_j)}{P_Y(x_i)} \end{aligned}$$

que es la **función de probabilidad de** Y **dado que** $X = x_i$, que esta definida en los $y_j \in R_Y$.
Ademas, dada $g : R_X \to \mathbb{R}$ función, trabajando respecto a la probabilidad condicional $\mathbb{P}_{\{Y=y_j\}} = \mathbb{P}_{\{Y=y_j\}}(\cdot) = \mathbb{P}(\cdot|\{Y = y_j\})$ se denota:
La esperanza condicional:

$$\mathbb{E}[g(x)|\{Y = y_j\}] = \sum_{i=1}^{m} \left(g(x_i) \cdot \mathbb{P}_{\{Y=y_j\}}(\{X = x_i\})\right) = \sum_{i=1}^{m} \left(g(x_i) \cdot p_{X|Y=y_i}(x_i)\right)$$

Varianza condicional

$$Var(X|\{Y = y_j\}) = \mathbb{E}[X^2|\{Y = y_j\}] - (\mathbb{E}[X|\{Y = y_j\}])^2$$

Variables aleatorias discreta independientes

Sean X, Y v.a.d. sobre un espacio de probabilidad $(\Omega, A, \mathbb{P})$ con rangos respectivos $R_X = \{x_1, x_2, ..., x_m\}$ y $R_Y = \{y_1, y_2, ..., y_M\}$ con m, M entonces positivos o infinito.
X, Y se dicen ser **independientes** bajo $\mathbb{P}$ si y solo si se cumple:

$$\mathbb{P}\left(\{X = x_i\} \cap \{Y = y_j\}\right) = \mathbb{P}\left(\{X = x_i\}\right) \cdot \mathbb{P}\left(\{Y = y_j\}\right)$$

para todos los $x_i \in R_X$ y $y_j \in R_Y$.
Esto equivale a que $p_{X,Y}(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j)$ para todos los $x_i \in R_X$ y $y_j \in R_Y$.

Teorema: Criterio de independencia

Sean X, Y v.a.d. sobre un espacio de probabilidad $(\Omega, A, \mathbb{P})$ con rangos respectivos $R_X = \{x_1, x_2, ..., x_m\}$ y $R_Y = \{y_1, y_2, ..., y_M\}$ con m, M entonces positivos o $+\infty$. Son equivalentes:

1. X, Y son independientes bajo $\mathbb{P}$
2. Para cada $x_i \in R_x$ y $y_j \in R_y$ se tiene:
   $p_{X,Y}(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j)$
3. Para todos los $\tilde{x}$ y $\tilde{y}$ se tiene:
   $F_{x,y}(\tilde{x}, \tilde{y}) = F_X(\tilde{x}) \cdot F_Y(\tilde{y})$

Nota: Si se encuentra un $x_i \in R_X$ y un $y_j \in R_Y$ tal que

$$p_{X,Y}(x_i, y_j) \neq p_X(x_i) \cdot p_Y(y_j)$$

entonces X, Y **no son independientes bajo** $\mathbb{P}$
Similarmente, si se encuentran $\tilde{x}, \tilde{y} \in \mathbb{R}$ tales que:

$$F_{X,Y}(\tilde{x}, \tilde{y}) \neq F_x(\tilde{x}) \cdot F_Y(\tilde{y})$$

entonces X, Y **no son independientes**

Teorema:

Sean X, Y v.a.d. sobre un espacio de probabilidad $(\Omega, A, \mathbb{P})$ con rangos respectivos $R_X = \{x_1, x_2, ..., x_m\}$ y $R_Y = \{y_1, y_2, ..., y_M\}$ con m, M entonces positivos o $+\infty$.
Si X, Y son independientes bajo P entonces:

1. Para $f : R_X \to \mathbb{R}$ y $g : R_Y \to \mathbb{R}$ se tiene:
   $\mathbb{E}[f(x) \cdot g(y)] = \mathbb{E}[f(x)] \cdot \mathbb{E}[g(y)]$
2. $M_{X+Y}(T) = M_X(T) \cdot M_Y(T)$ para T en algún intervalo que contiene a 0
3. $Var(aX + Y) = a^2 Var(X) + Var(Y)$ para algún $a \in \mathbb{R}$

Teorema

Sean X, Y v.a.d., entonces:

1. La v.a.d. $Z = X + Y$ cumple:
   $p_Z(Z_x) = p_{X+Y}(Z_k) = \mathbb{P}(\{X + Y = Z_k\}) = \sum_{x_i \in R_X} (p_{X,Y}(x_i, Z_k - x_i)$
   para cada $Z_k \in R_Z = R_{X+Y}$
2. Si X, Y son independientes entonces:
   $p_{X+Y}(Z_k) = \mathbb{P}(\{X + Y = Z_k\}) = \sum_{x_i \in R_x} (p_X(x_i) \cdot p_Y(Z_k = x_i))$
   para cada $Z_k \in R_{X+y}$
3. Si X,Y son independientes:
   $\mathbb{P}(\{X \leq Y\}) = \sum_{x_i \in R_X} (\mathbb{P}(\{x_i \leq Y\}) \cdot p_x(x_i))$

Teorema:

Sean X, Y v.a.d. independientes respecto a $\mathbb{P}$.
Si $X \sim Bin(n, p)$ y $Y \sim Bin(m, p)$ entonces $X + Y \sim Bin(n + m, p)$, y por ende:

$$\mathbb{P}(\{X + Y = K\}) = \binom{n+m}{k} \cdot (1 - p)^{n+m-k} \cdot p^k$$

para cada k en $\{0, 1, ..., n + m\}$

Teorema

Si $X_1 \sim Ber(p), ..., X_n \sim Ber(p)$ son variables independientes respecto a $\mathbb{P}$, entonces $X_1 + ... + X_N \sim Bin(n, p)$ y por ende

$$\mathbb{P}(\{X_1 + ... + X_n = k\}) = \binom{n}{k} \cdot (1 - p)^{n-k} \cdot p^k$$

para k en $\{0, 1, ..., n\}$.

Teorema

Sean $X \sim Poisson(\lambda_1)$ y $Y \sim Poisson(\lambda_2)$ v.a.d. independientes, entonces $X + Y \sim Poisson(\lambda_1 + \lambda_2)$ y por ende

$$\mathbb{P}(\{X + Y = k\}) = e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1 + \lambda_2)^k}{k!}$$

para k en $\{0, 1, 2, 3, ...\}$

4 Distribuciones conjuntas de v.a.c.

Función de densidad conjunta

Sean X, Y v.a.c., una función $f : \mathbb{R}^2 \to \mathbb{R}$ es una **función de densidad conjunta** de X, Y si y solo si se cumplen:

1. Para cada $(x, y) \in \mathbb{R}^2$ se tiene $f(x, y) \geq 0$
2. Para cada C una región del plano $\mathbb{R}^2$ se tiene:

   $$\mathbb{P}(\{(X, Y) \in C\}) = \iint_C f(x, y)dA = \iint_C f(x, y)dxdy$$

3. $\iint_{\mathbb{R}^2} f(x, y)dxdy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)dxdy = 1$

Se suele denotar $f_{X,Y} = f$

Nota: (X, Y) es una variable bidimensional:

$$(X, Y) : \Omega \to \mathbb{R}^2$$

Toma un $r \in \Omega$ y devuelve un punto $\left(x_{(r)}, y_{(r)}\right)$ en $\mathbb{R}^2$. El rango de (X, Y) es por definición:

$$R_{X,Y} = \{(x, y) \in \mathbb{R}^2 : f(x, y) \neq 0\}$$

Función de distribución acumulada conjunta

Sean X, Y v.a.c., una función $F : \mathbb{R}^2 \to [0, 1]$ se dice ser **función de distribución acumulada de** X, Y si y solo si

$$F(x, y) = \mathbb{P}\left(\{X \leq x\} \cap \{Y \leq y\}\right)$$

se denota $F_{X,Y} = F$

Teorema

Si $f : \mathbb{R}^2 \to [0, +\infty[$ es la función de densidad conjunta de dos v.a.c. X, Y, entonces

1. $F_{X,Y}(\tilde{X}, \tilde{y}) = \displaystyle\int_{-\infty}^{\tilde{x}} \int_{-\infty}^{\tilde{y}} f(x, y)dydx$ para $\tilde{x}, \tilde{y} \in \mathbb{R}$ siempre que las integrales converjan
2. Para A, B subconjuntos de $\mathbb{R}$ se tiene

   $$\mathbb{P}\left(\{(X, Y) \in A \times B\}\right) = \mathbb{P}\left(\{X \in A\} \cap \{Y \in B\}\right) = \int_B \int_A f(x, y)dxdy$$

   Nota: $A \times B = \{(a, b) : a \in A, b \in B\}$
3. Si F es la distribución conjunta acumulada de X, Y entonces

   $$f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2 F(x,y)}{\partial x \partial y} & \text{si esa derivada existe y es no negativa en}(x, y) \\ 0 & \text{en caso contrario} \end{cases}$$

Observación: Sea C una región en $\mathbb{R}^2$, si R es una (sub)región de C tal que $f(x, y)$ vale 0 para (x, y) fuera de R entonces:

$$\iint_C f(x, y)dxdy = \iint_R f(x, y)dxdy$$

Esperanza de una función de dos v.a.c.

Sea X, Y v.a.c. con función de densidad conjunta $f(x,y)$ y g una función de valor real y definida sobre una región que contiene a $R_x X R_y$.
Se define:

$$\mathbb{E}[g(\mathbb{X},\mathbb{Y})] = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)\cdot f(x,y)dydx$$

siempre que la integral exista.

Teorema

Sean X, Y v.a.c. con función de densidad conjunta $f(x,y)$. Entonces:

1. $\mathbb{E}[x] = \int_{\mathbb{R}}\int_{\mathbb{R}} x\cdot f(x,y)dxdy, \qquad \mathbb{E}[Y] = \int_{\mathbb{R}}\int_{\mathbb{R}} y\cdot f(x,y)dxdy$
2. Para $\alpha \in \mathbb{R}$ se tiene: $\mathbb{E}[\alpha X + Y] = \alpha\mathbb{E}[X] + \mathbb{E}[Y]$

Teorema: Función de densidad marginal

Sean X, Y v.a.c. con función de densidad conjunta f. Entonces, las **funciones de densidad marginales** de X, Y estan dadas por:

$$f_X(x)\int_{\mathbb{R}} f(x,y)dy \qquad f_Y(y)\int_{\mathbb{R}} f(x,y)dx$$

Covarianza

Sean X, Y v.a.c., se define la **covarianza** de X, Y mediante:

$$Cov(X,Y) = \mathbb{E}[X\cdot Y] - \mathbb{E}[X]\cdot\mathbb{E}[Y] = \mathbb{E}[(X-\mu_X)\cdot(Y-\mu_Y)]$$

donde $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$ siempre que la expresión converja.

Nota: La covarianza $Cov(X,Y)$ mide que tanto se relacionan dos variables X, Y, o sea "que tanto cambian juntas". Cuando $Cov(X,Y) = 0$ significa que la manera de variar de X no se relaciona tanto con la manera de variar de Y (o al menos de una forma "simple") Esto puede suceder si X, Y son independientes.

Teorema:

Sean X, Y v.a.c. y $a, b, c, d \in \mathbb{R}$ se tiene que:

1. $Cov(a\cdot X + b, c\cdot Y + d) = a\cdot c\cdot Cov(X,Y)$
2. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

Independencia

Sean X, Y v.a.c. y sea F la función de distribución acumulada de X, Y. Se dice que X, Y son **independientes** si y solo si para todos los $x, y \in \mathbb{R}$ se cumple que:

$$F(X,Y) = F_X(x)\cdot F_Y(y)$$

Equivalentemente, X, Y son independientes si y solo si para todos los $x, y \in \mathbb{R}$ se cumple que:

$$f(x,y) = f_X(x)\cdot f_Y(y)$$

donde f es la función de densidad conjunta de X, Y.

Teorema

Sean X, Y v.a.c. independientes, entonces:

1. $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

2. $Cov(X,Y) = 0$

3. $Var(X+Y) = Var(X) + Var(Y)$

4. Para T en algún intervalo que contiene 0 se cumple que:

$$M_{\underbrace{X+Y}_{Z,\text{ nueva var}}}(T) = M_X(T) \cdot M_Y(T)$$

Nota: Las v.a.c. independientes cumplen esto, pero cumplirlo no las hace v.a.c. independientes.

Observación: Condicionando funciones de densidad, esperanzas y demás Sean X, Y v.a.c., f su función de densidad conjunta y sea $\tilde{y}$ un punto de R_Y tal que $f_Y(\tilde{y}) \neq 0$.

1. La **función de densidad de X condicionada a que** $Y = \tilde{y}$ se define como la función: $f_{X|Y=\tilde{y}} : \mathbb{R} \to [0, +\infty[$ dada por

$$f_{X|Y=\tilde{y}}(x) = \frac{f(x,\tilde{y})}{f_Y(\tilde{y})}$$

2. Dado que $A \subseteq \mathbb{R}$, se define la **probabilidad condicional de que X tome un valor en A dado que** $Y = \tilde{y}$, como:

$$\begin{aligned}\mathbb{P}(\{X \in A\}|Y = \tilde{y}) &= \mathbb{P}_{\{Y=\tilde{y}\}}(\{X \in A\}) \\ &= \mathbb{P}(\{X \in A\}|\{Y = \tilde{y}\}) \\ &= \int_A f_{X|Y=\tilde{y}}(x)dx \\ &= \int_A \frac{f(x,\tilde{y})}{f_Y(\tilde{y})}dx\end{aligned}$$

3. Dada $g : R_x \to \mathbb{R}$ una función, se define la **esperanza condicional** de la variable $Z = g(x)$ **dado que** $Y = \tilde{y}$, como:

$$\mathbb{E}[Z|Y = \tilde{y}] = \mathbb{E}[g(x)|Y = \tilde{y}] = \int_R g(x) \cdot f_{X|y=\tilde{y}}(x)dx$$

4. Se define la **varianza condicional de X dado que** $Y = \tilde{y}$, como:

$$Var(X|Y = \tilde{y}) = \mathbb{E}[X^2|Y = \tilde{y}] - (\mathbb{E}[X|Y = \tilde{y}])^2$$

Observación: Sean X, Y v.a.c. independientes con función de densidad conjunta f.
La variable $Z = X + Y$ cumple que:

1. $$\begin{aligned}F_Z(\lambda) &= F_{X+Y}(\lambda) \\ &= \mathbb{P}(\{X + Y \leq \lambda\}) \\ &= \int_{\mathbb{R}} f_X(x) \cdot F_Y(\lambda - x)dx \\ &= (f_X * F_Y)(\lambda)\end{aligned}$$

2. $$\begin{aligned}F_Z(\lambda) &= F_{X+Y}(\lambda) \\ &= \mathbb{P}(\{X + Y \leq \lambda\}) \\ &= \int_{\mathbb{R}} f_X(x) \cdot F_Y(\lambda - x)dx \\ &= (f_X * F_Y)(\lambda)\end{aligned}$$

Teorema:

Si X, Y son v.a.c. tales que $X \sim N(\mu, \sigma^2)$, $Y \sim N(\epsilon, \eta^2)$ y X, Y son independientes, entonces

$$\underbrace{X+Y}_{Z} \sim N(\mu, \epsilon, \sigma^2, \eta^2)$$

Teorema:

Sean X, Y v.a.c. tales que $X \sim Exp(\theta)$, $Y \sim Exp(\theta)$ y X, Y son independientes entonces:

1. La v.a.c. $Z = min\{X, Y\} \sim exp(2\theta)$
2. $X + Y$ tiene función de densidad:
   $f_{X+Y}(\lambda) = \begin{cases} \theta^2 \lambda e^{-\theta\lambda} & \text{si } \lambda \geq 0 \\ 0 & \text{si } \lambda < 0 \end{cases}$

5 Teoremas límite

Teorema: Desigualdad de Markov

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea $X : \Omega \to \mathbb{R}$ una v.a. que solo toma valores no negativos, tal que para todo $r \in \Omega$ se tiene $X(r) \geq 0$. Supongase que $\mathbb{E}[X]$ es finita. Entonces, para cada $\alpha \in]0, +\infty[$ se cumple la **desigualdad de Markov**:

$$\mathbb{P}\left(\{\alpha \leq X\}\right) = \frac{\mathbb{E}[X]}{\alpha}$$

Ademas, esta desigualdad también se puede rescribir para formar la **desigualdad inversa de Markov**:

$$1 - \frac{\mathbb{E}[X]}{\alpha} \leq \mathbb{P}\left(\{X < \alpha\}\right)$$

Como X se asume, tomando solo valores no negativos, entonces la desigualdad inversa también es

$$1 - \frac{\mathbb{E}[X]}{\alpha} \leq \mathbb{P}\left(\{0 \leq < \alpha\}\right)$$

Observación: Si $X : \alpha \to \mathbb{R}$ es v.a. no negativa con $\mathbb{E}[X]$ finita y $\alpha \in]0, +\infty[$, entonces:

1. Por la desigualdad de Markov:
   $\mathbb{P}\left(\{\alpha \leq X\}\right) \leq \frac{\mathbb{E}[X]}{\alpha}$,
   y una consecuencia de esto es:
   $\mathbb{P}\left(\{\alpha < X\}\right) \leq \frac{\mathbb{E}[X]}{\alpha}$.
2. Por desigualdad inversa de Markov:
   $1 - \frac{\mathbb{E}[X]}{\alpha} \leq \mathbb{P}\left(\{X \leq \alpha\}\right)$
   así que también:
   $1 - \frac{\mathbb{E}[X]}{\alpha} \leq \mathbb{P}\left(\{X < \alpha\}\right)$

La utilidad de estas probabilidades es estimar probabilidades

Teorema: Desigualdad de Chebyshev

Sea $X : \Omega \to \mathbb{R}$ una v.a.. Supongase que $\mathbb{E}[X] = \mu$ y $Var(X) = \sigma^2$ existen y son finitos con $\sigma \in]0, +\infty[$, entonces para todo $T \in]0, +\infty[$ se cumplen:

1. $\mathbb{P}\left(\{|X - \mu| \geq T\}\right) \leq \frac{\sigma^2}{T^2}$
2. $\mathbb{P}\left(\{\mu - T\sigma < X < \mu + T\sigma\}\right) = \mathbb{P}\left(\{|X - \mu| < T\sigma\}\right) \geq 1 - \frac{1}{T^2}$

Ley de grandes números

Al realizar diversos ensayos de un experimento, si los ensayos son muy pocos hay mucha posibilidad de error que hace que haya mucha variabilidad en mediciones que realizan variables aleatorias sobre los resultados del experimento
En el largo plazo, los resultados de los experimentos tienen un comportamiento mas regular tendiendo al valor promedia. Esta es la **ley de grandes números** en acción. Un principio mucho mas general, de acuerdo al cual hay mas regularidad en el comportamiento de ciertas variables aleatorias cuando se usa un numero grande de estas

Promediante

Sea $X_1 : \Omega \to \mathbb{R}$, $X_2 : \Omega \to \mathbb{R}$, $X_3 : \Omega \to \mathbb{R}$, ... una sucesión (infinita) de variables aleatorias todas continuas o todas discretas. Para cualquier entero positivo n, se define una nueva variable aleatoria $\overline{X}_n : \Omega\mathbb{R}$, llamada **promediante** de $x_1, ..., x_n$ o **n-esimo promediante**, como la variable aleatoria:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

Es decir, $\overline{X}_n : \Omega \to \mathbb{R}$ es la v.a. que. al tomar un $r \in \Omega$, se evalua como

$$\overline{X}_n(r) = \frac{X_1(r) + X_2(r) + ... + X_n(r)}{n}$$

de modo que $\overline{X}_n(r)$ es simplemente el promedio de los números reales $X_1(r), X_2(r), ..., X_n(r)$

Sucesión de una v.a. i.i.d.

Sea $X_1 : \Omega \to \mathbb{R}$, $X_2 : \Omega \to \mathbb{R}$, $X_3 : \Omega \to \mathbb{R}$, ... una sucesión (infinita) v.a.

1. $X_1, X_2, X_3, ...$ son **mutuamente independientes** si y solo si para cualquier cantidad finita n de v.a. distintas $X_{i1}, X_{i2}, ..., X_{in}$ tomados de esta sucesión se tiene que para cada $\lambda_1 \in \mathbb{R}_{xi1}$, $\lambda_2 \in \mathbb{R}_{xi2}$, ..., $\lambda_n \in \mathbb{R}_{xin}$ se cumple la **función de distribución acumulada de** $X_{i1}, ...X_{in}$:

$$\mathbb{P}\left(\{X_{i1} \leq \lambda_1\} \cap \{X_{i2} \leq \lambda_2\} \cap ... \cap \{X_{in} \leq \lambda_n\}\right) = \underbrace{\mathbb{P}\left(\{X_{i1} \leq \lambda_1\}\right)}_{F_{X_i1}(\lambda_i)} \cdot \underbrace{\mathbb{P}\left(\{X_{i2} \leq \lambda_2\}\right)}_{F_{X_i2}(\lambda_i)} \cdot ... \cdot \underbrace{\mathbb{P}\left(\{X_{in} \leq \lambda_n\}\right)}_{=F_{X_in}(\lambda_n)}$$

2. $X_1, X_2, X_3, ...$ se dicen estar **idénticamente distribuidas** si y solo si para cualesquiera indices i, j en x_i y y_j tienen una misma función de distribución acumulada de modo que para cada $\lambda \in \mathbb{R}$:

$$\underbrace{\mathbb{P}\left(\{X_i \leq \lambda\}\right)}_{F_{xi}(\lambda)} = \underbrace{\mathbb{P}\left(\{X_j \leq \lambda\}\right)}_{F_{xj}(\lambda)}$$

3. $X_1, X_2, X_3, ...$ se dicen ser **i.i.d.** si y solo si son mutuamente independientes e igualmente distribuidas

Observación Si $X_1, X_2, X_3, ...$ son v.a. igualmente distribuidas, entonces cualesquiera dos de estas X_i, X_j cumplen que:

1. Para $g : \mathbb{R} \to \mathbb{R}$ una función se cumple: $\mathbb{E}[g(x_i)] = \mathbb{E}[g(x_j)]$ siempre que existan
2. $Var(X_i) = Var(X_j)$

Teorema: Ley fuerte de grandes números

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea $X_1 : \Omega \to \mathbb{R}$, $X_2 : \Omega \to \mathbb{R}$, $X_3 : \Omega \to \mathbb{R}$, ... una sucesión (infinita) de v.a. i.i.d. tales que sus esperanzas son finitas. Por la observación anterior, todas las esperanzas de $X_1, X_2, X_3, ...$ son iguales, así que sea mu el valor de estas, o sea, que para indice i en $\{1, 2, 3, ...\}$: $\mathbb{E}[X_i] = \mu$, entonces:

1. **Versión 1:** $\mathbb{P}(\{\lim_{n\to+\infty} \overline{X}_n = \mu\}) = 1$
   Esto es la probabilidad de que un resultado $r \in \Omega$ del experimento dado al azar cumpla que: $\lim_{n\to+\infty} \overline{X}_n(r) = \mu$ es 1. La sucesión numérica $\overline{X}_1(r), \overline{X}_2(r), \overline{X}_3(r), ...$ converge a μ. Esto no quiere decir que todos los resultados $r \in \Omega$ del experimento cumplen $\lim_{n\to+\infty} \overline{X}_n(r) = \mu$, si no que el conjunto de resultados $r \in \Omega$ que no la cumplen es "muy pequeño" respecto a la probabilidad $\mathbb{P}$

2. **Versión 2**
   Conforme el entero positivo n se hace cada vez mas grande, el n-esimo promediante $\overline{X}_n$ toma valores cada vez mas cercanos al valor esperado μ en casi todos los resultados $r \in \Omega$.
   O sea, para casi todos los posibles resultados $r \in \Omega$ del experimento, el valor del n-esimo promediante $\overline{X}_n$ en r, o sea, $\overline{X}_n(r)$, se acerca cada vez mas al valor esperado μ conforme n se hace cada vez mas grande.
   Es decir, el valor $\overline{X}_n$ se acerca al valor esperado μ casi siempre cuando n es un numero muy grande

El teorema del limite central, en una versión simplificada, establece que al realizar un experimento y tomar distintas muestras a los resultados del experimento, si se toma un numero suficientemente grande de muestras, entonces sigue una distribución acumulada aproximadamente normal.
En la practica, no siempre es posible conocer la distribución acumulada precisa de las v.a. con las que se trabaja, o bien estas no siguen una distribución normal. Sin embargo, bajo ciertas condiciones, por **teorema del limite central**, se sabe que la estandarización del promedio de estas sigue una distribución acumulada aproximadamente normal.
Esto permite poder estimar ciertas probabilidades relacionadas con las variables originales, a pesar de que no se conozcan las distribuciones exactas de esas variables. De este modo, las distribuciones normales aparecen en muchos fenómenos del mundo real.

Estandarización de una variable aleatoria

Sea $X : \Omega \to \mathbb{R}$ una v.a.. Supongase que $\mathbb{E}[X] = \mu$, $Var(X) = \sigma^2$ son finitos y que $Var(X) = \sigma^2 > 0$, donde $\sigma \in]0, +\infty[$ es la desviación estándar. Entonces la v.a.:

$$\widehat{X} = \frac{X - \mathbb{E}[X]}{\sqrt{Var(X)}} = \frac{X - \mu}{\sigma}$$

es una nueva v.a. $\widehat{X} : \Omega \to \mathbb{R}$ denominada como **estandarización** de X. Esta cumple que:

$$\mathbb{E}[\widehat{X}] = 0 \qquad\qquad Var(\widehat{X}) = 1$$

Teorema: Limite central

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea $X_1 : \Omega \to \mathbb{R}$, $X_2 : \Omega \to \mathbb{R}$, $X_3 : \Omega \to \mathbb{R}$, ... una sucesión (infinita) de v.a. i.i.d. Se sabe que $X_1, X_2, X_3, ...$ tienen misma esperanza y varianza. Supongase que existen la esperanza$\mu = \mathbb{E}[X_i]$ y la varianza $\sigma^2 = Var(X_i)$ y que $\sigma^2 = Var(X_i) > 0$, donde $\sigma \in]0, +\infty[$.
Para cada $n \in \{1, 2, 3, ..\}$ considérese al n-esimo promediante:

$$\overline{X}_n(r) = \frac{X_1(r) + X_2(r) + ... + X_n(r)}{n}$$

Entonces, para todo $\lambda \in \mathbb{R}$:

$$\lim_{n \to +\infty} \left(F_{\widehat{\overline{X}}_n}(\lambda) \right) = F_{N(0,1)}(\lambda)$$

o sea:

$$\lim_{n \to +\infty} \left(\mathbb{P}\left(\{ \widehat{\overline{X}}_n \le \lambda \} \right) \right) = \mathbb{P}\left(\{ N(0,1) \le \lambda \} \right)$$

En palabras: Cuando n va creciendo, la función de distribución acumulada $F_{\widehat{\overline{X}}_n}$ de la estandarización $\widehat{\overline{X}}_n$ del n-esimo promediante se va aproximando cada vez mas a la función de distribución acumulada $F_{N(0,1)}(\lambda)$ de una variable normal estándar $N(0, 1)$. Así, cuando n es grande, el comportamiento de la estandarización:

$$\widehat{\overline{X}}_n = \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}\left(\overline{X}_n - \mu \right)}{\sigma} \sim N(0, 1)$$

6 Cadenas de Markov

Una cadena de Markov es un tipo particular de proceso estocástico en el que el modo en el que se produce un estado futuro depende unicamente del estado inmediatamente anterior tomado.
Intuitivamente, es un proceso cuyo estado futuro depende unicamente del estado presente y no de los estados pasados
Estas sirven para modelar la manera en que ciertos fenómenos y procesos aleatorios evolucionan en función del tiempo, tales como tasas de interés, precios de activos financieros, entre otros. Permiten también hacer predicciones sobre el futuro. ¿A que se refiere un proceso estocástico? Es una colección de variables aleatorias que "van evolucionando en el tiempo."

Proceso estocástico (real)

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea $\mathbb{T}$ un conjunto no vacío de indices. (Representa un conjunto de instantes temporales de interés).
Para cada instante T en $\mathbb{T}$ sea $X_T : \Omega \to \mathbb{R}$ una variable aleatoria. A la colección:

$$\underbrace{(X_T : T \in \mathbb{T})}_{\text{Los } X_T \text{ tal que } T \text{ es un instante de } \mathbb{T}}$$

de todas esas v.a. se le denomina como **proceso estocástico** (real).
Cuando el conjunto de tiempos $\mathbb{T}$ es finito o infinito numerable se dice que $(X_T : T \in \mathbb{T})$ es un proceso estocástico **discreto** (y usualmente $\mathbb{T} \subseteq \{0, 1, 2, 3, ...\}$.
Cuando el conjunto de tiempos $\mathbb{T}$ es infinito no numerable se dice que $(X_T : T \in \mathbb{T})$ es un proceso estocástico **continuo** (y usualmente $\mathbb{T} \subseteq [0, +\infty[$).
Asimismo, al conjunto S formado por todos los valores reales alcanzados por variables aleatorias del proceso estocástico $(X_T : T \in \mathbb{T})$ se le denomina como **conjunto de estados** o **espacio de estados** del proceso estocástico, o sea:

$$S = \{X_T(r) : T \in \mathbb{T}, r \in \Omega\}$$

Entonces, S es el conjunto de todos los posibles valores reales alcanzados por las v.a. X_T del proceso estocástico en instantes T y elementos $r \in \Omega$ del espacio muestral

Cadena de Markov (discreta)

Sea $(\Omega, A, \mathbb{P})$ un espacio de probabilidad. Sea $(X_T : T \in \mathbb{T})$ un proceso estocástico con $T = \{1, 2, 3, ..., m\}$ (puede iniciar en 0 también) finito o infinito. $(X_T : T \in \mathbb{T})$ se dice ser una **cadena de Markov** si y solamente si para cualquier cantidad finita de estados $s_0, s_1, s_2, ..., s_n, s_{n+1}$ en el conjunto de estados S de esta cadena y para $X_0, X_1, ..., X_n, X_{n+1}$ las primeras variables se cumple que la probabilidad de pasar al siguiente estado depende solo del estado X_n actual y no de los anteriores, tal que:

$$\mathbb{P}\left(\{X_{n+1} = S_{n+1}\} | \{X_1 = S_1, X_2 = S_2, ..., X_n = S_n\}\right) = \mathbb{P}\left(\{X_{n+1} = S_{n+1}\} | \{X_n = S_n\}\right)$$

En este curso se va a suponer que las cadenas de Markov tienen la siguiente **propiedad estacionaria**:
Para cada n y cualesquiera estados x, y en S se cumple que:

$$\mathbb{P}\left(\{X_{n+1} = y\} | \{X_n = x\}\right) = \mathbb{P}\left(\{X_2 = y\} | \{x_1 = x\}\right)$$

Esto nos dice que, la probabilidad de que X_{n+1} pase al estado y dado que el estado anterior era $X_n = x$ es igual a la probabilidad de que X_2 haya pasado a y dado que el estado inicial era $X_1 = x$. En esencia, que las probabilidades al cambiar de estados se dan del mismo modo sin importar en cual instante se encuentre en la cadena.
Gracias a esta propiedad estacionaria, tiene sentido definir una **función de probabilidad de transición de un paso** $p : S^2 \to [0, 1]$ que toma un par de estados x, y en S y devuelve la probabilidad

$$p(x, y) = \mathbb{P}\left(\{X_{n+1} = y\} | \{X_n = x\}\right)$$

de que el valor de la siguiente variable siguiente X_{n+1} sea y en caso de que el valor de la variable anterior haya sido $X_n = x$
O sea, $p(x, y)$ describe la probabilidad de cambios entre estados x y y

Matriz de transición de un paso

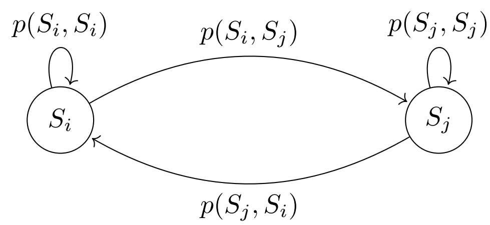
Sea $X_T : T \in \mathbb{T}$ una cadena de Markov con $\mathbb{T} = \{1, 2, 3, ...\}$ y con un conjunto de estados $S = \{s_1, s_2.s_3, ...\}$. Sea $p(x, y)$ la función de transición de la cadena de Markov $X_T : T \in \mathbb{T}$. La matriz:

$$M = \begin{pmatrix} p(s_1, s_1) & p(s_1, s_2) & p(s_1, s_3) & \dots \\ p(s_2, s_1) & p(s_2, s_2) & p(s_2, s_3) & \dots \\ p(s_3, s_1) & p(s_3, s_2) & p(s_3, s_3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

se denomina como **matriz de transición** de la cadena de Markov $X_T : T \in \mathbb{T}$.
Nota: Cada fila de la matriz de transición suma 1.

Grafo de una cadena de Markov

Un grafo (dirigido) es un conjunto de vértices junto con aristas (flechas) que conectan vértices. En el contexto de una cadena de Markov $X_T : T \in \mathbb{T}$ con conjunto de estados $S = \{s_1, s_2.s_3, ...\}$ los vértices se usan para representar los estados y las aristas (flechas) se usan para expresar la función de transición $p(x, y)$:



Si $p(s_m, s_n) = 0$, se puede omitir su flecha.
Nota: En términos de grafo, para s_m un estado, la suma de los valores de todas las flechas que salen de s_m es 1.

Teorema: Algunas propiedades de cadenas de Markov Sea $X_T : T \in \mathbb{T}$ una cadena de Markov con $\mathbb{T} = \{1, 2, 3, ...\}$ y estados $S = \{s_1, s_2.s_3, ...\}$

1. Para $s_1, s_2, ..., s_n$ es S se cumple que:

$$\mathbb{P}(X_1 = s_1, X_2 = s_2, \dots, X_n = s_n) = \underbrace{\mathbb{P}(X_1 = s_1)}_{\text{Esto pone } X_1 = s_1} \cdot \underbrace{p(s_1, s_2)}_{\text{paso } X_2 = s_2} \cdot \underbrace{p(s_2, s_3)}_{\text{paso } X_3 = s_3} \cdot ... \cdot \underbrace{p(s_{n-1}, s_n)}_{\text{paso } X_n = s_n}$$

2. Para $s_1, s_2, ..., s_n$ es S:

$$\mathbb{P}\left(\{X_{n+m} = s\} | \{X_1 = s_1, X_2 = s_2...., X_n = s_n\}\right) = \mathbb{P}\left(\{X_{n+m} = s\} | \{X_n = s_n\}\right)$$

   Esto se refiere a que, cuando se avanzan m pasos o unidades temporales en el proceso, la probabilidad de tomar el estado $X_{n+m} = s$ dado que el estado en el instante n era $X_n = s_n$ no depende de los estados previos a este.
   O sea, la probabilidad sigue dependiendo unicamente en el ultimo estado conocido, sin importar que ahora se avancen m pasos

Función de transición en m pasos

Sea $(X_T : T \in \mathbb{T})$ una cadena de Markov con un conjunto de estados $S = \{s_1, s_2, ...\}$ y $\mathbb{T} = \{1, 2, 3, 4, ..\}$. Sea m un entero positivo adecuado. Se define la **función de probabilidad de transición en m pasos** $p^m : S^2 \to \mathbb{R}$ como la función que toma estados x, y en S y devuelve:
$p^m(x, y) = \mathbb{P}\left(\{X_{n+m} = Y\} | \{X_n = x\}\right)$ la probabilidad de que, dado que en un instante n se tiene que el estado es $X_n = x$, en m pasos futuros al estado pase a ser $X_{n+m} = y$.

Nota: si $m = 0$ se toma:

$$p^0(x,y) = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{si } x \neq y \end{cases}$$

Teorema: Ecuaciones de Chapman-Kolmogorov

Sea $(X_T : T \in \mathbb{T})$ una cadena con conjuntos de estados S. Si x, y son estados en S entonces:

$$p^{n+m}(x,y) = \sum_{z \in S} (p^n(x,z) \cdot p^m(z,y))$$

Así, si $S = \{s_1, s_2, s_3, ...\}$ entonces:

$$p^{n+m}(x,y) = p^n(x,s_1) \cdot p^m(s_1,y) + p^n(x,s_2) \cdot p^m(s_2,y) + p^n(x,s_3) \cdot p^m(s_3,y) + ...$$

Matriz de transición en m pasos

Sea $(X_T : T \in \mathbb{T})$ una cadena de Markov con un conjunto de estados $S = \{s_1, s_2, ...\}$. La **matriz de transición en m pasos** de esta cadena de Markov es una matriz que codifica la información de la función de transición en m pasos p^m, como sigue:

$$M_m = \begin{pmatrix} p^m(s_1,s_1) & p^m(s_1,s_2) & p^m(s_1,s_3) & \dots \\ p^m(s_2,s_1) & p^m(s_2,s_2) & p^m(s_2,s_3) & \dots \\ p^m(s_3,s_1) & p^m(s_3,s_2) & p^m(s_3,s_3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Teorema: Sea $(X_T : T \in \mathbb{T})$ es una cadena de Markov. Si M es una matriz de transición de un paso, entonces el producto matricial:

$$M^m = \underbrace{MM \dots M}_{m \text{ veces}}$$

es igual a la matriz de transición M_m de m pasos de la cadena de Markov. Como consecuencia:

$$M_{m+n} = M_m \cdot M_n$$

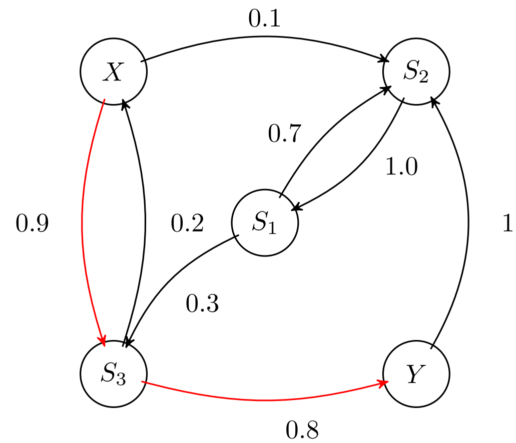
Relación de comunicación entre estados

Sea $(X_T : \in \mathbb{T})$ una cadena de Markov con $T = \{1, 2, 3, ..\}$ (infinito o infinito) y S el conjunto de sus estados.Para dos estados x, y en S se tiene:

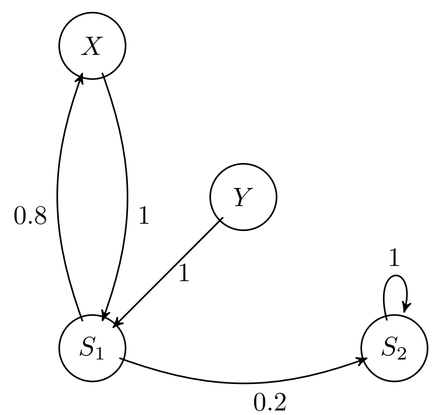
1. Se dice que x se **comunica** (en ese orden) con y si y solo si existe un instante n en $\mathbb{T}$, tal que

   $$\mathbb{P}(\{X_N = y\} | \{X_1 = x\}) > 0$$

   (No es imposible que el sistema alcance e ventualmente el estado y dado que comenzo en el estado x). En tal caso se denota $x \to y$

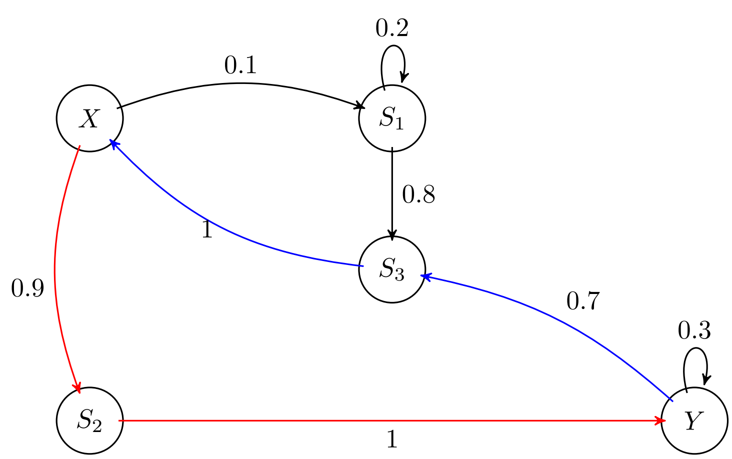


Esto significa que hay un camino de flechas de probabilidad no nula que va de x hasta y.
En caso contrario, cuando x no se comunica con y, se denota $x \not\rightarrow y$.

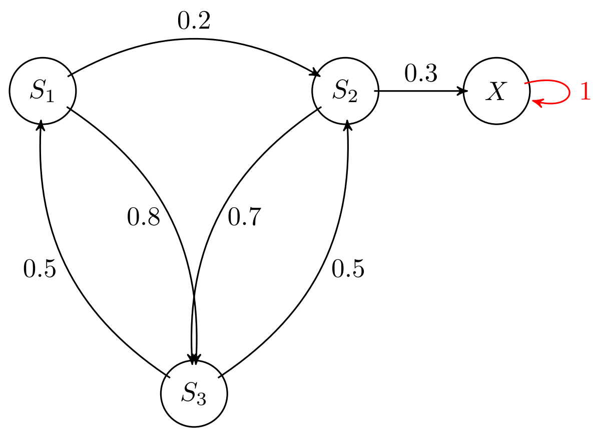


Aqui x no se comunica con y, pues no hay flechas que lleguen a y. Sin embargo, aqui partiendo de y si se puede llegar a x. De este modo $x \not\rightarrow y$, pero $y \rightarrow x$. Esto nos dice que, puede suceder que un estado no se comunique con otro, pero al reves si.

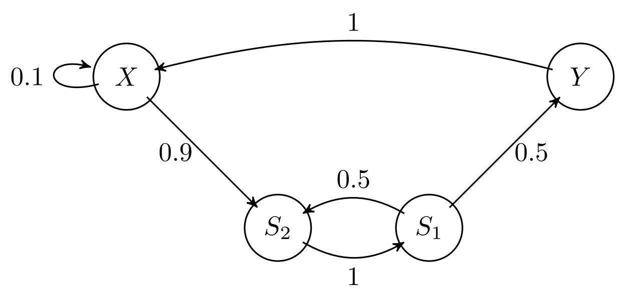
2. x, y se **intercomunican**, escrito como $x \leftrightarrow y$, si y solo si se cumplen simultaneamente que $x \rightarrow y$, $y \rightarrow x$.



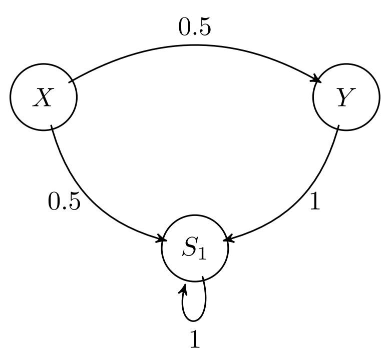
3. El estado x se dice **absorbente** si y solo si $p(x,x) = 1$. Se dice estado absorbente, pues al llegar al estado de x es imposible salir de ahi



4. La cadena de Markov $(X_T : T \in \mathbb{T})$ se dice **irreducible** si y solo si para cualesquiera dos estados x, y en S se cumple que $x \leftrightarrow y$. Esto nos dice que no es imposible (proba $\neq 0$) ir de un estado a cualquier otro.



Si la cadena es no irreducible, es imposible ir de un estado cualquiera a otro, por ejemplo: $y \not\to x$



Observacion En una cadena de Markov $(X_T :\in \mathbb{T})$ dos estados x, y diferentes $x \neq y$) se comunican si y solo si existe un entero positivo m tal que la funcion de transicion en m pasos p^m es estrictamente positiva en x, y o sea:

$$p^m(x, y) > 0$$

En resumen: $x \to y$ si y solo si existe un entero positivo tal que $p^m(x, y) > 0$

Definicion

Sea $(X_T :\in \mathbb{T})$ una cadena de Markov. Para n un entero positivo y x, y dos estados en S se denota mediante $f^n(x,y)$ a la probabilidad a partir del estado x llegar al estado por primera vez exactamente en el instante n, o sea:

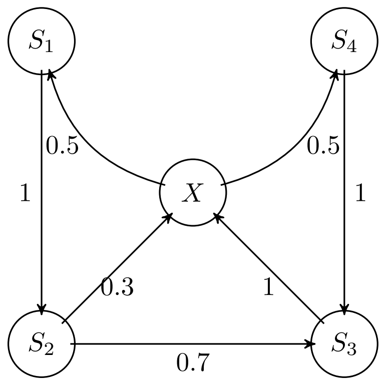
$$f^n(x,y) = \mathbb{P}\left(\underbrace{\{X_n = y, X_{n-1} \neq y, ..., x_2 \neq y\}}_{\text{Llegar a y en n sin llegar antes}} | \underbrace{\{X_1 = X\}}_{\text{Partiendo de X}} \right)$$

Ademas X un estado en S:

1. X es **recurrente** si y solo si:

   $$\sum_{n=1}^{+\infty} (f^n(x,x) = 1)$$

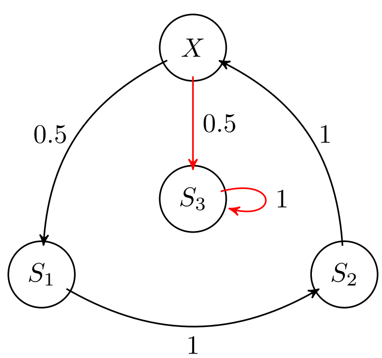
   Es imposible, al comenzar en el estado X, no regresar eventualmente al estado X en un tiempo finito.
   Teorema: Si x es un estado absorbente entonces x es recurrente, pero un estado recurrente no es absorbente.



2. X es **transitorio** si y solo si:

   $$\sum_{n=1}^{+\infty} (f^n(x,x) < 1)$$

   Existe una posibilidad de partir de x y nunca regresar a un estado x en el futuro.



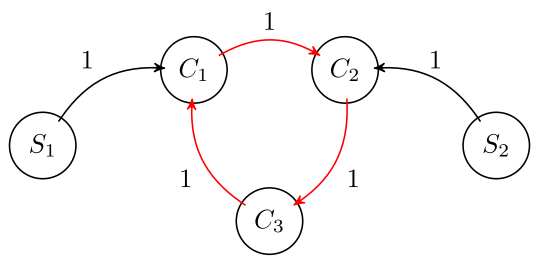
Teorema: Sea $(X_T :\in \mathbb{T})$ una cadena de Markov. Sean x, y estados de S.

1. Si x es un estado recurrente y $x \to y$, entonces y es tambien recurrente y $y \to x$
2. Si $x \leftrightarrow y$ entonces:
   - x es transitorio si y solo si y es transitorio
   - x es recurrente si y solo si y es recurrente

Definicion

Sea $(X_T :\in \mathbb{T})$ una cadena de Markov. Sea C un conjunto formado por algunos estados S $(C \leq S)$.
C se dice estar **cerrado** si y solo si cuando el sistema llega a algun estado en C entonces en e; futuro sigue tomando estados en C, o sea:

$$\underbrace{\mathbb{P}\left(\{X_2 \notin C \text{ o } X_3 \notin C \text{ o } ...\}|\{X_1 \in C\}\right) = 0}_{\text{Es imposible salir de } C \text{ y llegar a un estado fuera de } C}$$



Dado $C = \{C_1, C_2, C_3\}$, es imposible salir de C y llegar a un estado fuera de C

Teorema: Sea $(X_T :\in \mathbb{T})$ una cadena de Markov y S su conjunto de estados.

1. Si $C \subseteq S$ es un subconjunto cerrado de estados tal que C es irreducible y tambien finito, entonces todos los estados x de C son recurrentes.

2. Si la cadena de Markov dada cumple que su conjunto de estados S es finito e irreducible, entonces todos los estados x de la cadena de Markov son recurrentes

Definicion

Sea $(X_T :\in \mathbb{T})$ una cadena de Markov con estados $S = \{s_1, s_2, s_3, ...\}$. Una **distribucion de probabilidad** r es un vector de probabilidades $r = (r(s_1), r(s_2), ...)$. Esta distribucion r se dice ser **estacionaria** o **invariante** si y solo si cumple para cada estado y en S se cumple:

$$r(y) = \sum_{s \in S} (r(s) \cdot p(s, y))$$

Teorema: Sea $(X_T :\in \mathbb{T})$ una cadena de Markov con estados $S = \{s_1, s_2, s_3, ...\}$. Sea M su matriz de transicion de un paso. Una distribucion de probabilidad $r = (r(s_1), r(s_2), ...)$ es estacionaria si y solo si:

$$r = r \times M$$

Siendo este un producto matricial con r vector fila
Ademas, si $r = (r(s_1), r(s_2), ...)$ es una distribucion de probabilidad estacionaria de la cadena de Markov, entonces todo estado x que sea transitorio cumple que $r(x) = 0$