

UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Instituto del cálculo

## **Apunte de Probabilidades**

**Ing. Jemina M. García**

Buenos Aires, 2021.



# Índice general

<b>1. Vectores aleatorios</b>	<b>1</b>
1.1. Vectores aleatorios . . . . .	1
1.1.1. Vectores aleatorios discretos . . . . .	2
1.1.2. Vectores aleatorios continuos . . . . .	2
1.1.3. Algunas distribuciones importantes de vectores aleatorios . . . . .	3
1.2. Independencia . . . . .	4
1.3. Simulación . . . . .	5
1.4. Estimación de la función de distribución . . . . .	6
1.4.1. Función empírica . . . . .	6
1.4.2. Ejemplo . . . . .	6
<b>2. Momentos</b>	<b>9</b>
2.1. Repaso . . . . .	9
2.2. Vectores aleatorios . . . . .	10
2.2.1. Algunos ejemplos importantes . . . . .	11
2.3. Covarianza . . . . .	12
2.3.1. Volvamos a la Normal Multivariada . . . . .	13
<b>3. Transformaciones</b>	<b>15</b>
3.1. Función de variable aleatoria . . . . .	15
3.1.1. Variables aleatorias . . . . .	15
3.1.2. Vectores aleatorios . . . . .	16
3.1.3. Método de transformaciones bivariadas . . . . .	16
<b>4. Predicción</b>	<b>19</b>
4.1. Variables aleatorias condicionadas . . . . .	19
4.1.1. Vectores discretos . . . . .	19
4.1.2. Vectores continuos . . . . .	20
4.2. Mezcla de variables aleatorias . . . . .	21
4.3. Función de regresión . . . . .	22
4.4. Esperanza condicional . . . . .	23

4.4.1. Predicción . . . . .	24
4.4.2. Varianza condicional . . . . .	26
<b>5. Función generadora de momentos</b>	<b>27</b>
5.1. Función generadora de momentos . . . . .	27
<b>6. Convergencia</b>	<b>31</b>
6.1. Desigualdades . . . . .	31
6.2. Convergencia . . . . .	32
6.3. Leyes de los grandes números . . . . .	33
6.3.1. Ley débil de los grandes números . . . . .	33
6.3.2. Ley fuerte de los grandes números . . . . .	34
<b>7. Teorema central del límite</b>	<b>37</b>
7.1. Convergencia en distribución . . . . .	37
7.2. Teorema Central del Límite . . . . .	38
7.2.1. El Teorema . . . . .	38
7.3. Método delta . . . . .	39
<b>8. Bonus Track: Proceso de Poisson</b>	<b>41</b>
8.1. Adelgazamiento o coloreo . . . . .	43
8.2. Superposición de procesos de Poisson . . . . .	43
8.3. Propiedades . . . . .	43

# Capítulo 1

## Vectores aleatorios

### 1.1. Vectores aleatorios

**Definición:** Sea  $(\Omega, \mathcal{A}, \mathbf{P})$  un espacio de probabilidad, se dice que  $X = (X_1, X_2, \dots, X_n)$  es un vector aleatorio de dimensión  $n$ , si para cada  $j = 1, 2, \dots, n$ ,  $X_j : \Omega \rightarrow \mathbb{R}$  es una variable aleatoria.

**Teorema:** Para todo  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  se tendrá que  $X^{-1}((-\infty, x_1) * (-\infty, x_2) * \dots * (-\infty, x_n)) \in \mathcal{A}$

Teníamos que una variable aleatoria podía ser discreta, continua o mixta. Un vector aleatorio es un vector formado por variables aleatorias, entonces en el caso de vectores podemos llegar a tener todas las combinaciones posibles. En adelante solo trabajaremos con vectores aleatorios discretos (esto es: formados solamente por variables aleatorias discretas), o continuos.

**Función de distribución de un vector aleatorio:** Sea  $X = (X_1, X_2, \dots, X_n)$  un vector aleatorio de dimensión  $n$ , definimos la función de distribución de  $X$  como:

$$F_X(x) = \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n;)$$

**Propiedades, cuando  $X = (X, Y)$ :**

1.  $\lim_{x, y \rightarrow \infty} F_X = 1, \lim_{x \rightarrow -\infty} F_X = 0, \lim_{y \rightarrow -\infty} F_X = 0$
2.  $F_X(x)$  es monótona no decreciente en cada variable
3. Es continua a derecha en cada variable
4.  $\mathbf{P}((a_1, b_1) \times (a_2, b_2)) = F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2)$

### 1.1.1. Vectores aleatorios discretos

#### **Función de probabilidad de un vector aleatorio discreto.**

Sean  $X$  e  $Y$  dos variables aleatorias discretas definidas en el espacio muestral  $\Omega$  de un experimento. La función de probabilidad conjunta se define para cada par de números  $(x, y)$  como:

$$p_{X,Y}(x, y) = \mathbf{P}(X = x, Y = y)$$

Debe cumplirse que:

1.  $p(x, y) \geq 0$
2.  $\sum_x \sum_y p(x, y) = 1$

**Ejemplo:** De una urna que contiene 3 bolillas numeradas 1, 2 y 3, se extraen sin reposición y sucesivamente 2 bolillas. Sea  $X$  el número de la primer bolilla e  $Y$  el número de la segunda.

1. Hallar la función de probabilidad conjunta de  $X$  e  $Y$
2. Calcular  $\mathbf{P}(X < Y)$
3. Hallar la función de probabilidad de  $X$  y de  $Y$ . ¿cómo se llaman?

Las funciones de probabilidad **marginales** de  $X$  e  $Y$  están dadas por:

$$p_X(x) = \sum_y p(x, y)$$

$$p_Y(y) = \sum_x p(x, y)$$

En general:

Sea  $A$  cualquier conjunto compuesto de pares de valores  $(x, y)$  entonces:

$$\mathbf{P}((X, Y) \in A) = \sum_{(x,y) \in A} p(x, y)$$

### 1.1.2. Vectores aleatorios continuos

#### **Función de densidad de un vector aleatorio continuo.**

Sean  $X$  e  $Y$  variables aleatorias continuas, una función de densidad de probabilidad conjunta  $f_{X,Y}(x, y)$  de estas dos variables es una función que satisface:

1.  $f(x, y) \geq 0$
2.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$

Entonces para cualquier conjunto  $A$

$$\mathbf{P}((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy$$

Las funciones de densidad de probabilidad marginales de  $X$  e  $Y$  están dadas por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

**Ejemplo:** Se elige al azar un punto  $(X, Y)$  en el círculo de centro  $(0, 0)$  y radio 1

1. Hallar la densidad conjunta del vector aleatorio  $(X, Y)$ .
2. Hallar la probabilidad de que la distancia del punto al centro del círculo sea menor que 0.5.
3. Hallar las funciones de densidad marginales de  $X$  e  $Y$ .

### 1.1.3. Algunas distribuciones importantes de vectores aleatorios

#### Distribución Multinomial

El vector aleatorio  $X = (X_1, X_2, \dots, X_n)$  tiene distribución multinomial de parámetros  $n, p_1, \dots, p_n$  con  $\sum_{i=1}^n p_i = 1$  si:

$$\mathbf{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{n!}{x_1! x_2! \dots x_n!} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}$$

Propiedades:

$$X_i \sim \mathcal{B}(n, p_i), \quad \forall i$$

**Ejemplo:** Se tiene una caja con 3 tipos de cartucho: 4 de tipo A, 6 de tipo T y 5 de tipo N. Se extraen con reposición 3 cartuchos. Sean  $X_1$  la cantidad de cartuchos de tipo A,  $X_2$  la cantidad de cartuchos de tipo T y  $X_3$  la cantidad de cartuchos tipo N en la muestra. Hallar la función de probabilidad conjunta del vector aleatorio  $(X_1, X_2, X_3)$

#### Distribución Normal Multivariada

Se dice que el vector aleatorio  $\underline{X} = (X_1, \dots, X_p)$  tiene distribución normal multivariada de dimensión  $p$ ,  $\underline{X} \sim \mathcal{N}_p(\mu, \Sigma)$ , de parámetros  $\mu \in \mathbb{R}^p$  y  $\Sigma \in \mathbb{R}^{p \times p}$  (simétrica y definida positiva) si su función de densidad conjunta está dada por

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-1/2(\underline{x}-\mu)^T \Sigma^{-1} (\underline{x}-\mu)}$$

El vector  $\mu \in \mathbb{R}^p$  es el vector formado por las esperanzas de las variables  $X_1, \dots, X_p$ , y la matriz  $\Sigma \in \mathbb{R}^{p \times p}$  se llama matriz de covarianzas, y el elemento  $\{i, j\}$  de la matriz consiste en la  $\mathbf{cov}(X_i, X_j)$ . Esta distribución se desarrollará con mayor detalle en el próximo capítulo.

## 1.2. Independencia

Así como definimos independencia para eventos, ahora vamos a definir independencia pero para variables aleatorias.

**Definición:** Sea  $(X, Y)$  un vector aleatorio, las variables aleatorias  $X$  e  $Y$  son independientes sii

$$\mathbf{P}((X \in A) \cap (Y \in B)) = \mathbf{P}(X \in A) * \mathbf{P}(Y \in B), \forall A, B \in \mathcal{B}$$

**Propiedad 1:** Se dice que las variables aleatorias  $X_1, \dots, X_n$  son independientes sii

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) * \dots * F_{X_n}(x_n)$$

**Propiedad 2:** Se dice que las variables aleatorias discretas  $X_1, \dots, X_n$  son independientes sii

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) * \dots * p_{X_n}(x_n)$$

**Propiedad 3:** Se dice que las variables aleatorias continuas  $X_1, \dots, X_n$  son independientes sii

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) * \dots * f_{X_n}(x_n)$$

vale para casi todo  $x_1, \dots, x_n$

Para ir todos los días al trabajo, Dana se dirige en auto hasta la estación de tren y luego sigue su camino en tren. Dana sale de su casa en un intervalo distribuido uniformemente entre las 7:30 y las 7:50. El tiempo de viaje hasta la estación es también uniforme entre 20 y 40 minutos, e independiente de la hora en que sale de su casa. Hay un tren que sale a las 8:12 y otro que sale a las 8:26.

1. ¿Cuál es la probabilidad de que Dana pierda ambos trenes?
2. ¿Cuál es la probabilidad de que tenga que esperar más de 8 minutos en la estación hasta que sale el tren?



### 1.3. Simulación

¿Qué ocurre si dada una función de distribución  $F$  quiero conocer la variable aleatoria  $X$  cuya función de distribución coincide con  $F$ ?

En el video se explican los ejemplos de forma gráfica para poder comprender luego la definición.

#### Ejemplos:

1. Simular el tiro de una moneda usando un número al azar entre 0 y 1
2. Simular el tiro de un dado usando un número al azar entre 0 y 1
3. Simular una muestra de una población con distribución exponencial usando una uniforme entre 0 y 1
4. Simular una muestra de 5 valores de una población con la distribución mixta del ejercicio 3 de este capítulo

Sea  $F : \mathbb{R} \rightarrow [0, 1]$  una función de distribución, existe una variable aleatoria  $X$  tal que  $F_X(x) = \mathbf{P}(X \leq x)$

Definimos **inversa generalizada** como:

$$F_X^{-1}(u) = \min\{x \in \mathbb{R} : F_X(x) \geq u\}, u \in (0, 1)$$

Hallando la inversa generalizada de  $F$  podemos entonces comenzar a simular variables aleatorias

**Teorema:** Si  $F$  es una función que cumple:

- Ser no decreciente
- $\lim_{x \rightarrow \infty} F = 1, \lim_{x \rightarrow -\infty} F = 0$
- Continua a derecha

Entonces si defino  $X = F^{-1}(U)$  con  $U \sim \mathcal{U}(0, 1)$ , se tiene que  $X$  es una variable aleatoria cuya función de distribución es la función  $F$  dada.

## 1.4. Estimación de la función de distribución

### 1.4.1. Función empírica

La Función empírica es una estimación de la función de distribución acumulada de una variable aleatoria. Lo mínimo que puede pedirse a esta función es que cumpla con las condiciones que debe tener una función de distribución:

1.  $\hat{F}_X(x) \in [0, 1], \forall x \in \mathbb{R}$
2.  $\hat{F}_X(x)$  es monótona no decreciente
3.  $\hat{F}_X(x)$  es continua a derecha
4.  $\lim_{x \rightarrow -\infty} \hat{F}_X(x) = 0$  y  $\lim_{x \rightarrow \infty} \hat{F}_X(x) = 1$

Se define la Función empírica como

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$$

Donde  $x_1, x_2, \dots, x_n$  se asumen realizaciones de las variables aleatorias  $X_1, X_2, \dots, X_n$ , todas con distribución  $F_X(x)$  e independientes

### 1.4.2. Ejemplo

Este ejemplo podemos hacerlo a mano y luego ver en R la empírica del ejercicio 1.

De una variable aleatoria X se ha obtenido la muestra:  $\{2,5; 2,2; 2,4; 2,2; 2,1\}$ . Hallar y graficar la función de distribución empírica asociada a la muestra y usarla para estimar la probabilidad  $\mathbf{P}(X \leq 2,3)$

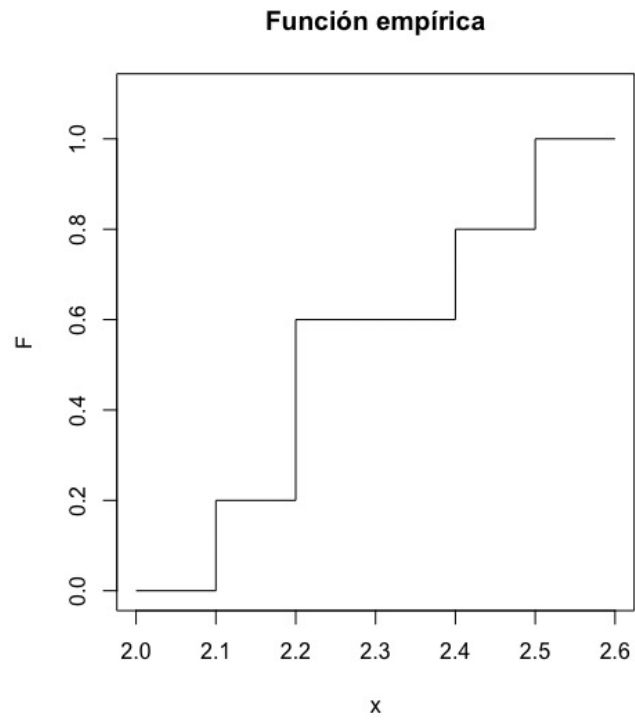
Según la definición:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$$

En este caso,  $n=5$ , por lo tanto:

$$\begin{aligned} \hat{F}(x) &= \frac{1}{5} \sum_{i=1}^n \mathbf{1}(x_i \leq x) \\ &= \frac{1}{5} \mathbf{1}(2,1 \leq x) + \frac{1}{5} \mathbf{1}(2,2 \leq x) + \frac{1}{5} \mathbf{1}(2,2 \leq x) + \frac{1}{5} \mathbf{1}(2,4 \leq x) + \frac{1}{5} \mathbf{1}(2,5 \leq x) \\ &= \frac{1}{5} \mathbf{1}(2,1 \leq x) + \frac{2}{5} \mathbf{1}(2,2 \leq x) + \frac{1}{5} \mathbf{1}(2,4 \leq x) + \frac{1}{5} \mathbf{1}(2,5 \leq x) \\ &= \frac{1}{5} \mathbf{1}(2,1 \leq x < 2,2) + \frac{3}{5} \mathbf{1}(2,2 \leq x < 2,4) + \frac{4}{5} \mathbf{1}(2,4 \leq x < 2,5) + \mathbf{1}(2,5 \leq x) \end{aligned}$$

El gráfico de la Función empírica tendrá forma de escalera:



Para estimar la probabilidad:

$$\mathbf{P}(X \leq 2,3) = F_X(2,3) \approx \hat{F}(2,3) = \frac{3}{5}$$



## Capítulo 2

# Momentos

### 2.1. Repaso

Ya hemos visto valores que representen a nuestras variables aleatorias, cuando trabajamos en el caso univariado.

**Definición:** Sea  $X$  una VAD con función de probabilidad  $p_X(x)$ , el valor esperado (o media) de  $X$  es:

$$\mathbf{E}(X) = \sum_{x \in R_X} x * p_X(x)$$

A veces, usamos como notación  $\mu_X$  para referirnos a la esperanza de la variable aleatoria  $X$ .

**Propiedad:** El valor de la esperanza de cualquier función  $h(X)$  se calcula como:

$$\mathbf{E}(h(X)) = \sum_{x \in R_X} h(x) * p_X(x)$$

**¿Qué pasa con las variables continuas?**

Sea  $X$  una variable aleatoria continua con función de densidad  $f_X(x)$ , el valor esperado o media de  $X$  es:

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

El valor de la esperanza de cualquier función  $h(X)$  se calcula como:

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

Repasemos la propiedad de la esperanza que más vamos a usar.

*Propiedad:* Si  $h(x) = ax + b$ , entonces  $\mathbf{E}(h(X)) = a\mathbf{E}(X) + b$

*Propiedad útil:*

$$\mathbf{E}(X) = \int_0^\infty (1 - F_X(x))dx + \int_{-\infty}^0 F_X(x)dx$$

**Varianza** La varianza de una variable aleatoria mide la dispersión media de los valores de una variable aleatoria.

Sea  $X$  una V.A. y  $\mu$  su valor esperado, la varianza de  $X$  se define como:

$$\mathbf{var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2)$$

Propiedad:

$$\mathbf{var}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$$

El desvío estándar se define como la raíz cuadrada de la varianza:  $\sigma = \sqrt{\mathbf{var}(X)}$ . Se expresa en las mismas unidades que la variable aleatoria.

La **mediana** es el valor de  $X$  que acumula a izquierda (o derecha) una probabilidad igual a 0.5, o sea  $F_X(x) = 0,5$ .

La **moda** es el valor de la variable con valor de probabilidad máximo.

## 2.2. Vectores aleatorios

¿Que pasa con todas estas medidas cuando estamos trabajando con vectores aleatorios? La esperanza y la varianza se calculan para una variable aleatoria. Por lo tanto, cuando estamos en presencia de un vector aleatorio, podemos querer calcular la esperanza y varianza de cada variable que compone el vector ( y eso ya aprendimos a hacerlo), o también podemos querer calcular la esperanza y varianza de una función de ese vector aleatorio, la cual devuelva una variable.

Por ejemplo, si tengo un vector aleatorio  $(X, Y)$ , una función que me devuelve una variable aleatoria puede ser  $h(X, Y) = X+Y$ , o  $h(X, Y) = X/Y$ , o  $h(X, Y) = \cos(X)^2/\ln(Y)$ ... bueno, ya se entendió la idea.

El valor esperado de una función  $h(X, Y)$  está dado por:

$$\mathbf{E}(h(X, Y)) = \sum_x \sum_y h(x, y) p_{X,Y}(x, y)$$

Si  $(X, Y)$  es un vector aleatorio discreto.

$$\mathbf{E}(h(X)) = \iint_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y) dx dy$$

Si  $(X, Y)$  es un vector aleatorio continuo.

Veamos algunas *Propiedades de orden*:

Sea  $X = (X_1, \dots, X_n)$  un vector aleatorio,  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  una función, tenemos que

1. Si  $X > 0$  entonces  $\mathbf{E}(X) > 0$
2. Si  $g(x) > 0$  entonces  $\mathbf{E}(g(X)) > 0$
3. Sea  $h(X) > g(X)$  entonces  $\mathbf{E}(h(X)) > \mathbf{E}(g(X))$
4.  $\mathbf{E}(|X|) \geq \mathbf{E}(X)$
5.  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$
6. Sea  $h(X)$  una función cóncava,  $\mathbf{E}(h(X)) \leq h(\mathbf{E}(X))$
7.  $\mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$  desigualdad de cauchy shuartz
8.  $\sqrt{\mathbf{E}(X+Y)^2} \geq \sqrt{\mathbf{E}(X^2)} + \sqrt{\mathbf{E}(Y^2)}$  nombre??

**Más propiedades importantes:** (que vas a usar un montón)

1.  $\mathbf{E}[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i \mathbf{E}[X_i]$  Fíjate que es la misma idea que cuando probamos linealidad para una variable, solo que trabajando con vectores. ¿Te animas a demostrarla?
2. Si  $X_1, \dots, X_n$  son independientes entonces  $\mathbf{E}[\prod_{i=1}^n X_i] = \prod_{i=1}^n \mathbf{E}[X_i]$  Esta propiedad la demuestro en el video, pero solamente para  $(X, Y)$

### 2.2.1. Algunos ejemplos importantes

1. Suma de variables aleatorias *i.i.d.* con distribución  $Ber(p)$
2. Suma de variables aleatorias *i.i.d.* con distribución  $\mathcal{E}(\lambda)$

## 2.3. Covarianza

Lo que queremos encontrar ahora, es una medida de la relación entre las variables que componen un vector aleatorio. Buscamos tener una idea de que le pasa a la variable  $Y$  cuando la variable  $X$  varía, y viceversa. Veamos primero la definición y después busquemos entenderla.

La **covarianza** entre dos VA  $X$  e  $Y$  está dada por:

$$\mathbf{cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))]$$

Primera observación: la covarianza se calcula entre DOS variables aleatorias. ¿Qué me está diciendo esta cuenta? Observemos que mide la distancia entre  $X$  y su esperanza, la distancia entre  $Y$  y su esperanza, y las multiplica. Cuando hago esa cuenta, lo que tengo que pensar es que el significado del resultado lo voy a encontrar en el signo, no en la magnitud. A ese resultado le calcula la esperanza, esto es: el promedio ponderado por las probabilidades.

### Propiedades

1.  $\mathbf{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}[X]\mathbf{E}[Y]$
2. Si  $X$  e  $Y$  son independientes entonces  $\mathbf{E}(XY) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$ , y por lo tanto  $\mathbf{cov}(X, Y) = 0$
3.  $\mathbf{cov}(a + bX, c + dY) = b \cdot d \cdot \mathbf{cov}(X, Y)$  (propiedad de bilinealidad)
4.  $\mathbf{cov}(X + Y, Z) = \mathbf{cov}(X, Z) + \mathbf{cov}(Y, Z)$
5. Sean  $X$  e  $Y$  dos variables aleatorias,  $\mathbf{var}(X + Y) = \mathbf{var}(X) + \mathbf{var}(Y) + 2 \cdot \mathbf{cov}(X, Y)$

En general,  $\mathbf{var}(\sum X_i) = \sum \mathbf{var}(X_i) + 2 \sum \sum_{i < j} \mathbf{cov}(X_i, Y_j)$

Definamos una última medida, que es la que siempre se usa para describir grado de correlación entre dos variables aleatorias

El **coeficiente de correlación** entre  $X$  e  $Y$  está dado por:

$$\rho_{XY} = \frac{\mathbf{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Lo que estamos haciendo con esta cuenta es llevando la covarianza a que su valor absoluto esté entre 0 y 1, por lo tanto el coeficiente de correlación será un número que siempre esté entre -1 y 1, sin importar la magnitud de las variables involucradas, por lo que puede entenderse, interpretarse, y usarse para comparar. ¿Qué significan los extremos?

Propiedad:  $|\rho_{XY}| = 1$  *sii*  $\mathbf{P}(aX + b = Y) = 1$ , para algún  $a$  y  $b$



Y todos los valores del medio indicarán algún grado de relación lineal entre las variables. CUIDADO! si  $\rho$  da cero, no te está diciendo que no hay dependencia entre las variables, te está diciendo que no hay correlación lineal, eso puede significar que no hay dependencia o que hay correlación de otro tipo.

### 2.3.1. Volvamos a la Normal Multivariada

**Caso particular: normal bivariada.**

Donde  $\mu = (\mu_X, \mu_Y)$ ,  $\Sigma = \begin{bmatrix} \sigma_X^2 & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \sigma_Y^2 \end{bmatrix}$

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right]}$$

Propiedades:  $X \sim \mathcal{N}(\mu_X, \sigma_X)$ ,  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$

Veamos algunas propiedades que serán de muchísima utilidad.

1. Si  $X \sim \mathcal{N}_p(0, \text{diag}(\lambda_1, \dots, \lambda_p))$ , entonces  $X_1, \dots, X_p$  son independientes y tienen distribución  $\mathcal{N}(0, \lambda_i)$ .
2. Si  $X \sim \mathcal{N}_p(\mu, \Sigma)$  y  $A \in \mathbf{R}^{p \times p}$  es no singular, entonces  $AX + b \sim \mathcal{N}_p(A\mu + b, A\Sigma A^T)$ .  
Otra propiedad increíble: tenemos combinaciones lineales de normales las cuales van a tener distribución normal, ya que sabemos que las marginales también serán normales.

**Def:** Se dice que  $X$  es normal multivariada si y solo si  $\forall t \in \mathbb{R}^p$  se tiene que  $t^T X$  es normal univariada.

Es decir:

$$X \sim \mathcal{N}_p(\mu, \Sigma) \iff a^T X \sim \mathcal{N}(a^T \mu, a^T \Sigma a), \forall a \in \mathbb{R}^p$$

Además,  $X_1, \dots, X_p$  son independientes si y solo si  $\Sigma$  es una matriz diagonal.

A partir de las propiedades, es fácil ver que ocurre lo siguiente:

**Teorema:** Sean  $X_1, \dots, X_n$  variables aleatorias independientes con distribución  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ , sea  $Y = \sum_{i=1}^n a_i X_i$ ,  $a_i \in \mathbb{R}$  constantes, entonces  $Y$  tendrá distribución normal de parámetros  $\mu_Y = \sum_{i=1}^n a_i \mu_i$ ,  $\sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$ .



## Capítulo 3

# Transformaciones

### 3.1. Función de variable aleatoria

Sea  $X$  una variable aleatoria, sea  $Y = g(X)$  una función de la variable aleatoria  $X$ , lo que buscamos es encontrar la distribución de la variable aleatoria  $Y$ . En general podremos hallar la función de distribución de la nueva variable aleatoria con lo aprendido en el curso introductorio, con un método llamado el *método de eventos equivalentes*: como  $F_Y(y) = \mathbf{P}(Y \leq y)$  y además  $Y = g(X)$ , desarrollando el cálculo de probabilidades utilizando la relación entre las variables podemos encontrar la distribución de  $Y$ . Salvo en algunos pocos casos en el que se trate de transformaciones complicadas con vectores aleatorios, voy a usar siempre el método de eventos equivalentes. Lo importante, graficar, pintar, y calcular probabilidades. Todos los ejemplos se encuentran desarrollados en los videos.

#### 3.1.1. Variables aleatorias

---

Si  $X$  es una variable aleatoria discreta, entonces  $Y = g(X)$  también lo será.

---

$$p_Y(y) = \mathbf{P}(Y = y) = \sum_{x \in A} p_X(x), A = \{x \in \mathbb{R} : g(x) = y\}$$

**Ejemplo:** En un concurso de pesca cada pescador paga 100\$ por participar. La cantidad de peces que cada pescador puede obtener durante el desarrollo del concurso es una variable aleatoria con distribución de Poisson de parámetro 4.5. Cada pescador tiene permitido cobrar a lo sumo 8 piezas. Hay un premio de 50\$ por cada pieza. Calcular la

función de probabilidad de la ganancia.

---

$Y = g(X)$  será una variable aleatoria continua cuando la derivada de  $g(x)$  sea distinta de cero en casi todo punto

---

**Ejemplo:** Sea  $X \sim \mathcal{E}(1)$ , probar que  $Y = \frac{X}{\lambda} \sim \mathcal{E}(\lambda)$

Con esto ya agregamos una propiedad más a la exponencial!!!

- Si  $X \sim \mathcal{E}(1)$  entonces  $Y = \frac{X}{\lambda} \sim \mathcal{E}(\lambda)$

**Ejemplo más difícil:** Sea  $X$  una variable aleatoria con distribución normal estandar, hallar la distribución de  $Y = X^2$

### 3.1.2. Vectores aleatorios

Buscaremos nuevamente funciones de variables aleatorias, solo que en este caso la información dada será acerca de vectores aleatorios, pero la idea es siempre la misma. Tenemos un vector aleatorio del cual conocemos toda la información, y ahora tenemos una nueva variable, o mas de una variable, que queremos conocer. La forma de encontrar su distribución será encontrar si tienen alguna relación con las variables conocidas, y planteando eventos equivalentes encontraremos la nueva distribución.

**Ejemplo:**

1. Sean  $X$  e  $Y$  variables aleatorias independientes con  $X \sim P(\lambda)$  e  $Y \sim P(\mu)$ , hallar la distribución de  $X + Y$
2. Sean  $X_1, \dots, X_n$  variables aleatorias independientes con distribución uniforme en el intervalo  $(0, 1)$ , hallar la distribución de  $U = \min\{X_1, \dots, X_n\}$

### 3.1.3. Método de transformaciones bivariadas

Suponga que  $Y_1$  e  $Y_2$  son variables aleatorias continuas con función de densidad conjunta  $f_{Y_1, Y_2}(y_1, y_2)$  y que para todo  $(y_1, y_2)$  tal que  $f_{Y_1, Y_2}(y_1, y_2) > 0$ ,  $u_1 = h_1(y_1, y_2)$  y  $u_2 = h_2(y_1, y_2)$  es una transformación uno a uno de  $(y_1, y_2)$  y  $(u_1, u_2)$

con inversa  $y_1 = h_1^{-1}(u_1, u_2)$  y  $y_2 = h_2^{-1}(u_1, u_2)$ .

Si las inversas tienen derivadas parciales continuas respecto a  $u_1$  y  $u_2$  y jacobiano  $J$ , entonces la densidad conjunta de  $U_1, U_2$  será

$$f_{U_1, U_2}(u_1, u_2) = \frac{f_{Y_1, Y_2}(y_1, y_2)}{|J|} |h_1^{-1}, h_2^{-1}|$$

**Ejemplo:**  $X, Y$  iid  $\mathcal{N}(0, 1)$ , hallar la densidad conjunta de  $(W, Z) = (X + Y, X - Y)$

**Método del Jacobiano generalizado**

Si  $\underline{X}$  un vector aleatorio e  $\underline{Y} = g(\underline{X})$  con  $g|_{A_i} = g_i : A_i \rightarrow B$  biyectiva, continua, con derivada continua, donde  $A_1, \dots, A_k$  es una partición del  $\text{supp}(X)$ , entonces

$$f_{\underline{Y}}(\underline{y}) = \sum_{i=1}^k \frac{f_{\underline{X}}(\underline{x}) \mathbf{1}\{\underline{x} \in A_i\}}{|J_{g_i}(\underline{x})|} | \underline{x} = g_i^{-1}(\underline{y})$$



## Capítulo 4

# Predicción

### 4.1. Variables aleatorias condicionadas

Buscamos entender como se comporta la variable  $Y$  si conocemos el valor de  $X$ , y encontrar cual es su distribución de probabilidades. Ya aprendimos como se comporta cada variable por separado, como se comportan en conjunto, ahora queremos saber como se comporta una variable si conocemos el valor de la otra.

#### 4.1.1. Vectores discretos

**Definición:** Sean  $X$  e  $Y$  variables aleatorias discretas con  $p_X(x) > 0$ , la función de probabilidad condicional de  $Y$  dado que  $X = x$  es

$$p_{Y|X=x}(y) = \mathbf{P}(Y = y|X = x) = \frac{\mathbf{P}(Y = y, X = x)}{\mathbf{P}(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

Se define como 0 a  $p_{Y|X=x}(y)$  cuando  $p_X(x) = 0$ .

(Para cada valor  $x$  que tome la variable aleatoria  $X$ , tendremos una variable aleatoria  $Y|X = x$  diferente, con su correspondiente función de probabilidad)

Observemos que lo que estoy usando es la definición de probabilidad condicional.

Así como usamos la definición de probabilidad condicional para encontrar una fórmula para el cálculo de la probabilidad de la intersección de dos eventos, de la misma manera podemos ver que

$$p_{X,Y}(x, y) = p_{Y|X=x}(y)p_X(x)$$

Vamos a usar mucho esta propiedad, porque muchas veces, al observar la función de probabilidad conjunta vamos a poder factorizar en estas dos funciones y a partir de ello conocer la distribución de las variables involucradas (la marginal y la condicionada).

A partir de ella también deducimos otra *Propiedad útil*: Sean  $X$  e  $Y$  vectores aleatorios discretos tal que

$p_{Y|X=x}(y) = p_Y(y)$  para todo  $x \in \mathbb{R}$ , entonces  $X$  e  $Y$  son independientes.

**Ejemplo 1:** Un almacén tiene en su depósito 35 productos de cierto tipo, 15 de los cuales fueron proporcionados por el proveedor 1, 7 por el proveedor 2 y 13 por el proveedor 3. Se van a seleccionar al azar y sin reposición 2 de los productos del depósito. Sean las variables aleatorias

$X$ : Cantidad de productos seleccionados que provienen del proveedor 1

$Y$ : Cantidad de productos seleccionados que provienen del proveedor 2

1. ¿Son  $X$  e  $Y$  independientes?
2. Calcular la función de probabilidad condicional  $p_{Y|X=1}(y)$ . ¿Qué distribución conocida tiene?
3. Calcular  $\mathbf{P}(Y > 0 | X = 1)$

Veamos ahora que pasa cuando trabajamos con vectores aleatorios continuos.

#### 4.1.2. Vectores continuos

**Definición:** Sea  $(X, Y)$  un vector aleatorio con densidad conjunta  $f_{X,Y}(x, y)$  y densidad marginal  $f_X(x)$ , entonces para cualquier valor de  $X$  con el cual  $f_X(x) > 0$ , la función de densidad condicional de  $Y$  dado  $X = x$  es

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Si  $f_X(x) = 0$  se define como 0.

Mucho más con vectores aleatorios, vamos a usar la propiedad de que

$$f_{X,Y}(x, y) = f_{Y|X=x}(y)f_X(x)$$

Veamos un ejemplo

**Ejemplo 2:** Sea  $(X, Y)$  un vector aleatorio con densidad conjunta

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-y/x} & \text{si } \{0 < y, 0 < x < 1\} \\ 0 & \text{en otro caso} \end{cases}$$



1. ¿Son  $X$  e  $Y$  independientes?
2. Hallar  $f_{Y|X=1/3}(y)$
3. Calcular  $\mathbf{P}(Y < 1/5|X = 1/3)$

Un ejemplo de distribuciones con variables condicionadas conocidas es el de la **Distribución Normal bivariada**. Dado el vector aleatorio  $(X, Y) \sim \mathcal{N}_2(\mu, \Sigma)$ , se tiene que:

$$Y|X = x \sim \mathcal{N}\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right)$$

$$X|Y = y \sim \mathcal{N}\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right)$$

## 4.2. Mezcla de variables aleatorias

Este tema vamos a tratar de entenderlo primero con un ejemplo. Supongamos que saliendo de su casa, Juan tiene 3 posibles medios de transporte para ir al centro: Tren, colectivo y moto. El 70 % de las veces viaja en moto, el 20 % de las veces en tren y el 10 % de las veces en colectivo. El tiempo que tarda en viajar (en horas) depende del medio de transporte que elija, siendo una variable aleatoria  $X_1, X_2$  o  $X_3$  si viana en tren, colectivo o moto respectivamente. Las distribuciones de cada una de esas variables las conozco. Ahora, me interesa calcular por ejemplo, la probabilidad de que un día cualquiera Juan tarde más de una hora en llegar al centro. ¿Cómo calculo esa probabilidad?

Si defino  $X$ : "tiempo que tarda Juan en viajar al centro (en horas)", entonces, usando probabilidad total puedo calcular la probabilidad en cuestión. Generalicemos.

Sean  $A_1, A_2, \dots, A_n$  una partición de  $\Omega$  y  $X$  una variable aleatoria, de manera que se conocen las distribuciones de  $X|A_i$  para  $i = 1, 2, \dots, n$

Entonces la distribución de  $X$ , aplicando la fórmula de probabilidad total, será

$$F_X(x) = \sum_{i=1}^n F_{X|A_i}(x) \mathbf{P}(A_i)$$

Si redefino la partición, puedo crear una variable aleatoria  $M$  de manera tal que  $\mathbf{P}(A_i) = \mathbf{P}(M = i), i = 1, \dots, n$

Entonces lo que conozco son las distribuciones de  $X|M = m$ . De esta manera puedo definir la distribución de  $X$ :

$$F_X(x) = \sum_{i=1}^n F_{X|M=i}(x) \mathbf{P}(M = i)$$

Si la V.A.X es discreta, entonces la función de probabilidad de X será:

$$p_X(x) = \sum_{i=1}^n p_{X|M=i}(x) \mathbf{P}(M = i)$$

Si la V.A.X es continua, entonces la función de densidad de X será:

$$f_X(x) = \sum_{i=1}^n f_{X|M=i}(x) \mathbf{P}(M = i)$$

A partir de esta forma para encontrar la distribución de una variable aleatoria, podemos calcular la esperanza de una variable, si lo que conocemos es su distribución pero solo a partir de una partición del espacio muestral. Podemos calcular la  $\mathbf{E}[X]$  como

$$\mathbf{E}[X] = \mathbf{E}[X|X \in A] \mathbf{P}(X \in A) + \mathbf{E}[X|X \in A^c] \mathbf{P}(X \in A^c)$$

**Ejemplo 3:** Una rata está atrapada en un laberinto. Inicialmente puede elegir una de tres sendas. Si elige la primera se perderá en el laberinto y luego de 4 minutos volverá a su posición inicial; si elige la segunda volverá a su posición inicial luego de 7 minutos; si elige la tercera saldrá del laberinto luego de 3 minutos. En cada intento, la rata elige con igual probabilidad cualquiera de las tres sendas. Calcular la esperanza del tiempo que demora en salir del laberinto.

### 4.3. Función de regresión

Hasta ahora aprendimos a encontrar la distribución de las variables condicionadas  $Y|X = x$  y  $X|Y = y$ . Ahora, si tenemos nuevas variables aleatorias, de las cuales ya conocemos su distribución, tranquilamente podemos calcularles la esperanza y la varianza, ¿no? ¿Cómo haríamos en cada caso?

Sea  $Y|X = x$  una variable aleatoria discreta, entonces

$$\mathbf{E}[Y|X = x] = \sum_{y \in R_Y} y \cdot p_{Y|X=x}(y)$$

Sea  $Y|X = x$  una variable aleatoria continua, entonces

$$\mathbf{E}[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy$$

A estas funciones, que son funciones de  $x$ , se las llama **funciones de regresión**, y se las denota  $\varphi(x)$

**Ejemplos:** Busquemos la función de regresión para los dos primeros ejemplos de este capítulo

## 4.4. Esperanza condicional

Si llamamos  $\varphi(x) = \mathbf{E}[Y|X = x]$  a la esperanza de la variable condicionada  $Y|X = x$ , luego  $\varphi : \text{sop}(X) \rightarrow \mathbb{R}$ .

Vamos a definir ahora una variable aleatoria, llamada **Esperanza condicional de Y dado X**, denotado por  $\mathbf{E}[Y|X]$ , como  $\varphi(X) = \mathbf{E}[Y|X]$ . (es decir, estamos evaluando la función de regresión en la variable aleatoria X. La función de regresión era un número, pero al evaluarla de esta manera tenemos una variable aleatoria nueva, que es una función de la variable aleatoria X)

Veamos una propiedad importante que tiene esta nueva variable aleatoria que llamamos **esperanza condicional**.

**Propiedad:**  $\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|X]]$

**Demostración:**(para variables aleatorias continuas)

$$\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx$$

Pero

$$\varphi(x) = \mathbf{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy$$

Entonces

$$\mathbf{E}[\mathbf{E}[Y|X]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X=x}(y) f_X(x) dx dy = \mathbf{E}[Y]$$

**Ejemplo:** Hallar la esperanza condicional para los dos primeros ejemplos de este capítulo

**Mas propiedades:**

1. X e Y vectores aleatorios, s y r funciones medibles tales que las variables aleatorias  $r(X) * s(Y)$ ,  $r(X)$  y  $s(Y)$  tienen esperanza finita, entonces  $\mathbf{E}(r(X) * s(Y)|X) = r(X) * \mathbf{E}(s(Y)|X)$
2.  $Y_1, Y_2$ , V.A. con esperanza finita, X vector aleatorio,  $\mathbf{E}(aY_1 + bY_2 | X) = a\mathbf{E}(Y_1 | X) + b\mathbf{E}(Y_2 | X)$

3.  $\mathbf{E}(Y|X) = \mathbf{E}(Y)$  si  $X$  e  $Y$  son independientes
4.  $\mathbf{E}(r(X)|X) = r(X)$

Demostremos sólo la primera...

**Ejemplo copado** Sean  $X_1, \dots, X_n$  variables i.i.d, y  $S = \sum X_i$ ,

$$\mathbf{E}(\sum X_i|S) = S = \sum \mathbf{E}(X_i|S) = n\mathbf{E}(X_i|S)$$

Por lo tanto,  $\mathbf{E}(X_i|S) = S/n$

**Definición:** La variable aleatoria esperanza condicional de  $Y$  dada  $X$  se define como  $\varphi(X) = \mathbf{E}[Y|X]$  con  $\varphi$  una función medible tal que  $\mathbf{E}((Y - \varphi(X)) * t(X)) = 0$  para toda función  $t$  medible  $t : R_X \rightarrow \mathbb{R}$  tal que  $Y * t(X)$  tiene esperanza finita. Definiremos  $\varphi(x) = \mathbf{E}[Y|X = x]$ . (podemos ver que si  $t$  es cualquier función en el plano,  $\varphi(X)$  será el mejor predictor de  $Y$  basado en  $X$ , es la proyección ortogonal).

*Obs: La esperanza condicional siempre existe y a demás es única con probabilidad 1*

¿Y que estamos queriendo decir con esta definición? Genial, tenemos una nueva variable aleatoria, que la definiste de una forma un poco rebuscada, pero ya entendí como encontrarla, ¿Es alguien importante? ¿porque quiero aprenderla?

Uno de los objetivos principales de la estadística es tratar de predecir el futuro. Estudiamos experimentos aleatorios para poder adivinar, con algún grado de certeza, que es lo que va a pasar después. Es como la forma más cercana que tenemos a una bola de cristal. Lo que vamos a intentar es analizar si una variable aleatoria  $Y$  tiene alguna relación con otras variables  $X$ , de manera que si yo pudiera medir fácilmente las  $X$ , utilizando esa función podría conocer el valor de  $Y$ . Sabemos que no vamos a poder encontrar una relación exacta, pero podríamos pensar en encontrar una función de las variables  $X$  que estén “lo más cerca posible” de  $Y$ .

#### 4.4.1. Predicción

Sea  $Y$  una V.A.,  $X = (X_1, X_2, \dots, X_n)$  un vector aleatorio, existirá alguna función  $g(X)$  que nos sirva para predecir a  $Y$ . Para encontrar dicha función se calcula el error cuadrático medio:

$$ECM = \mathbf{E}[(Y - g(X))^2]$$

Esta va a ser nuestra medida de distancia, porque como te comenté antes, estamos buscando la función de  $X$  que esté lo más cerca posible de  $Y$ , y necesitamos definir qué

significa **más cerca**. Bueno, el error cuadrático medio nos da una idea de cercanía entre  $Y$  y  $g(X)$ .

Empecemos de a poco. Si buscamos el mejor predictor constante, ¿qué haríamos?. El mejor predictor constante es buscar  $g$  tal que  $g(X) = c \in \mathbb{R}$ . ¿Se te ocurre quien puede ser ese valor constante que minimiza el error cuadrático medio?

Lo que vamos a buscar es la función  $g$  que minimiza el error, con  $g$  una función constante.

Busco  $c$  tal que  $ECM = \mathbf{E}[(Y - c)^2]$  sea mínimo. Como  $c \in \mathbb{R}$ , para encontrar el mínimo derivamos e igualamos a cero. Primero desarrollamos el cuadrado y aplicamos propiedades de la esperanza. Luego derivamos, igualamos a cero y despejamos  $c$ . Conclusión: El mejor predictor constante es la esperanza de  $Y$ . Y en este caso, el ECM coincide con la varianza.

Busquemos ahora el mejor predictor lineal, siguiendo la misma idea: El mejor predictor lineal es la ecuación que resulta de minimizar  $ECM = \mathbf{E}[(Y - (aX + b))^2]$ .

Aplico propiedades de la esperanza, luego derivo parcialmente en función de  $a$  y de  $b$  y despejo, resultando:

$$g(X) = \hat{Y} = \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - \mathbf{E}[X]) + E[Y]$$

A esta función, que es una recta, la llamamos **Recta de regresión**.

Finalmente, queremos encontrar el predictor óptimo.

Sea  $\mathcal{P}$  un conjunto de predictores para la variable aleatoria  $Y$ , que forman un espacio vectorial. Cada elemento de  $\mathcal{P}$  es una variables aleatoria observable. Supongamos que se quiere predecir a  $Y$  a través de  $\hat{Y} \in \mathcal{P}$ . Si usamos como criterio de bondad de un predictor el error cuadrático medio, diremos que  $\hat{Y}_0 \in \mathcal{P}$  es es un predictor óptimo de  $Y \in \mathcal{P}$ , si dado otro  $\hat{Y} \in \mathcal{P}$  se tiene

$$ECM(\hat{Y}_0, Y) \leq ECM(\hat{Y}, Y) \quad (4.1)$$

A continuación damos un criterio suficiente para obtener un predictor óptimo usando el criterio del error cuadrático medio:

Una condición suficiente para que  $\hat{Y}_0 \in \mathcal{P}$  sea un predictor óptimo usando el criterio del error cuadrático medio es que

$$\mathbf{E}((Y - \hat{Y}_0)\hat{Y}) = 0$$

para todo  $\hat{Y} \in \mathcal{P}$ . Además, si  $\hat{Y}_0$  satisface (5.1), es esencialmente el único predictor óptimo. Es decir si  $\hat{Y} \in \mathcal{P}$  satisface  $ECM(\hat{Y}_0, Y) = ECM(\hat{Y}, Y)$ , entonces  $\mathbf{P}(\hat{Y} = \hat{Y}_0) = 1$ .

Entonces, habiendo visto la definición de esperanza condicional, la esperanza condicional de  $Y$  dada  $X$  cumple que:

$$\mathbf{E}[(Y - g(X))^2] \geq \mathbf{E}[(Y - \mathbf{E}(Y|X))^2]$$

Para cualquier función  $g(X)$

**Predicción:** “La esperanza condicional de  $Y$  dado  $X$  es la función de la variable aleatoria  $X$  que mejor predice o se aproxima a  $Y$ ”

¿Entonces? Lo que estoy diciendo, de forma muy rebuscada quizás (pero matemáticamente correcta), es que la esperanza condicional es el mejor predictor de  $Y$  en términos del error cuadrático medio.

#### 4.4.2. Varianza condicional

Sean  $X$  e  $Y$  variables aleatorias con  $\mathbf{var}(Y)$  finita, la varianza de la variable  $Y|X = x$  será

$$\mathbf{var}(Y|X = x) = \mathbf{E}((Y - \mathbf{E}(Y|X = x))^2|X = x)$$

Ahora, si llamamos  $\tau(x) = \mathbf{var}(Y|X = x)$ , tenemos que  $\tau : \text{sup}(X) \rightarrow \mathbb{R}$ . Llamaremos varianza condicional de  $Y$  dada  $X$  a la variable aleatoria  $\tau(X) = \mathbf{var}(Y|X) = \mathbf{E}((Y - \mathbf{E}(Y|X))^2|X)$ . Desarrollando el cuadrado y utilizando las propiedades de esperanza condicional, se tiene que  $\mathbf{var}(Y|X) = \mathbf{E}(Y^2|X) - \mathbf{E}(Y|X)^2$ .

#### Propiedad (Pitágoras)

$$\mathbf{var}(Y) = \mathbf{E}(\mathbf{var}(Y|X)) + \mathbf{var}(\mathbf{E}(Y|X))$$

#### Demostración:

$$\begin{aligned} \mathbf{var}(Y|X) &= \mathbf{E}(Y^2|X) - \mathbf{E}(Y|X)^2 \\ \mathbf{E}(\mathbf{var}(Y|X)) &= \mathbf{E}[\mathbf{E}(Y^2|X)] - \mathbf{E}[\mathbf{E}(Y|X)^2] = \mathbf{E}(Y^2) - \mathbf{E}[\mathbf{E}(Y|X)^2] \\ \mathbf{var}(\mathbf{E}(Y|X)) &= \mathbf{E}[\mathbf{E}(Y|X)^2] - \mathbf{E}[\mathbf{E}(Y|X)]^2 = \mathbf{E}[\mathbf{E}(Y|X)^2] - \mathbf{E}[Y]^2 \\ \mathbf{var}(Y) &= \mathbf{E}(\mathbf{var}(Y|X)) + \mathbf{var}(\mathbf{E}(Y|X)) \end{aligned}$$

**Ejemplo:** Usar pitágoras para hallar la varianza de  $Y$  en el segundo ejemplo de este capítulo.

## Capítulo 5

# Función generadora de momentos

### 5.1. Función generadora de momentos

Esta función (como lo dice el nombre) se usa para encontrar momentos de variables aleatorias, pero también para encontrar la distribución de suma de variables aleatorias y se usa para demostrar muchos teoremas más adelante.

**Definición:** La función generadora de momentos de  $X$  se define como

$$\psi_X(t) = \mathbf{E}(e^{tX}), \quad t \in \mathbb{R}$$

Cuando la función generadora de momentos (*gfm*) está bien definida, se pueden intercambiar las operaciones de diferenciación y esperanza. Esto nos lleva a la siguiente expresión:

$$\begin{aligned} \psi'(0) &= \frac{d}{dt} \mathbf{E}(e^{tX})|_{t=0} \\ &= \mathbf{E} \left( \frac{d}{dt} e^{tX} \right) |_{t=0} \\ &= \mathbf{E} (X e^{tX}) |_{t=0} = E(X) \end{aligned}$$

Derivando  $k$  veces podemos concluir que

$$\psi^{(k)}(0) = E(X^k)$$

y de esta manera encontramos un método para encontrar todos los momentos de una distribución.

**Comentario:** La función característica, definida como  $\mathbf{E}(e^{itX})$  está siempre bien definida para todo  $t$ , pero no la veremos en este curso)

**Ejemplo:** Sea  $X \sim \mathcal{E}(1)$ , encontrar la *fgm* de  $X$ , y a partir de ella su esperanza y varianza.

Tenemos que, para cualquier  $t > 1$ ,

$$\begin{aligned}\psi_X(t) &= \mathbf{E}(e^{tX}) = \int_0^\infty e^{tx} e^{-x} dx \\ &= \int_0^\infty e^{x(t-1)} dx \\ &= \frac{1}{1-t}\end{aligned}$$

Por lo tanto

$$\psi'_X(t) = \frac{1}{(1-t)^2} \rightarrow E(X) = \psi'_X(0) = 1$$

Y para la varianza:

$$\psi''_X(t) = \frac{2}{(1-t)^3} \rightarrow \psi''_X(0) = 2, \text{var}(X) = 1$$

### Propiedades:

1. Si  $Y = aX + b$  entonces  $\psi_Y(t) = e^{bt}\psi_X(at)$

Prueba:

$$\begin{aligned}\psi_Y(t) &= \mathbf{E}(e^{tY}) = \mathbf{E}(e^{t(aX+b)}) \\ &= \mathbf{E}(e^{taX} e^{tb}) \\ &= e^{tb} \mathbf{E}(e^{taX}) = e^{tb} \psi_X(at)\end{aligned}$$

2. Si  $X_1, \dots, X_n$  son independientes, y sea  $Y = \sum_{i=1}^n X_i$ , entonces

$$\psi_Y(t) = \prod_{i=1}^n \psi_{X_i}(t)$$

Usaremos la propiedad 2 para encontrar la función generadora de momentos de una distribución binomial.

**Ejemplo:** Sea  $X \sim \mathcal{B}(n, p)$ . Sabemos que  $X = \sum_{i=1}^n X_i$  con  $X_i \sim \text{Ber}(p)$  independientes. por lo tanto

$$\psi_{X_1}(t) = \mathbf{E}(e^{tX_1}) = (1-p) + pe^t$$

Esto vale para todas las  $X_i$  porque son *iid*, resultando

$$\psi_X(t) = \prod_{i=1}^n \psi_{X_i}(t) = \prod_{i=1}^n [(1-p) + pe^t] = [(1-p) + pe^t]^n$$



Diremos que  $X$  e  $Y$  son iguales en distribución (o idénticamente distribuidas) si tienen la misma función de distribución, y lo escribiremos como  $X \stackrel{D}{=} Y$ .

**Teorema:** Sean  $X$  e  $Y$  variables aleatorias, si  $\psi_X(t) = \psi_Y(t), \forall t$  en un intervalo abierto al rededor del cero, entonces  $X \stackrel{D}{=} Y$ .

### Ejemplo: sigamos con las binomiales

Definamos  $X \sim \mathcal{B}(n, p)$ ,  $Y \sim \mathcal{B}(m, p)$ , ambas variables independientes. Ahora definimos  $W = X + Y$ , ¿que distribución tiene  $W$ ?, busquemos su *fgm* usando propiedades resulta

$$\begin{aligned}\psi_W(t) &= \mathbf{E}(e^{tW}) = \mathbf{E}(e^{tX})\mathbf{E}(e^{tY}) \\ &= [(1-p) + pe^t]^n [(1-p) + pe^t]^m \\ &= [(1-p) + pe^t]^{n+m}\end{aligned}$$

Que es la *fgm* de una distribución binomial de parámetros  $(n+m)$  y  $p$ . Por lo tanto  $W \sim \mathcal{B}(n+m, p)$ .

**Se recomienda como práctica realizar el mismo trabajo cuando tenemos dos variables con distribución de Poisson independientes.**

**Función generadora de momentos para una distribución normal**  $X \sim \mathcal{N}(0, 1)$

$$\begin{aligned}\psi_X(t) &= \mathbf{E}(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx} (2\pi)^{-1/2} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} e^{t^2/2} (2\pi)^{-1/2} e^{(-x^2+2tx-t^2)/2} dx \\ &= e^{t^2/2} \int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-(x-t)^2/2} dx = e^{t^2/2}\end{aligned}$$

Resultando

$$\psi_X(t) = e^{t^2/2}$$

Para encontrar la función para cualquier normal basta con plantear que si  $X \sim \mathcal{N}(0, 1)$  entonces  $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$ , y usar propiedades.



## Capítulo 6

# Convergencia

### 6.1. Desigualdades

Se busca estimar probabilidades conociendo, no la distribución de nuestra variable, si no solamente las medidas resumen ya vistas, tales como la esperanza y la varianza. Servirán para demostrar muchos de los teoremas que veremos más adelante, y para encontrar cotas para probabilidades de eventos.

#### Desigualdad de Markov

Sea  $h : \mathbb{R} \rightarrow \mathbb{R}^+$  tal que  $h$  es par y restringida a  $\mathbb{R}^+$  es creciente, y sea  $X$  una V.A. tal que  $\mathbf{E}(h(X))$  existe, entonces para todo  $t$  en  $\mathbb{R}$

$$\mathbf{P}(|X| \geq t) \leq \frac{\mathbf{E}[h(X)]}{h(t)}$$

Si además  $X$  es no negativa, para todo  $a > 0$

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$$

Ejemplo: Sea  $X$  una variable aleatoria positiva de media 15. Demostrar que  $\mathbf{P}(X > 60) < 0,25$ .

#### Desigualdad de Tchevycheff

Sea  $X$  una V.A. con varianza finita, para todo  $k > 0$

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq k) \leq \frac{\mathbf{var}[X]}{k^2}$$

Observemos que la desigualdad de Chevychev se puede ver como un caso particular de la desigualdad de markov.

**Ejemplo:** Una máquina produce rieles cuya longitud (en metros) es una v.a. con distribución  $\mathcal{U}(0,8,1,2)$ . Se eligen al azar  $n$  rieles en forma independientes. Sea  $\bar{X}$  el promedio de sus longitudes. Hallar el mínimo valor de  $n$  tal que:

$$\mathbf{P}(0,99 < \bar{X} < 1,01) > 0,9$$

## 6.2. Convergencia

Cuando estudiamos convergencia en esta parte de probabilidades lo que nos interesa es entender el comportamiento límite de una sucesión de variables aleatorias  $X_1, X_2, \dots$ . Cada vez nos encontramos con más datos y estamos interesados en saber que pasa si podemos conocer aún más.

Definiremos 3 tipos de convergencia, pero solo trabajaremos sobre los dos que mas utilizaremos y nos acompañarán a lo largo de la carrera.

Consideremos un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbf{P})$  y sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias definidas sobre ese espacio, y  $X$  otra variable aleatoria definida sobre el mismo espacio.

**Definición:** Diremos que  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias converge casi seguramente a otra variable aleatoria  $X$  ( $X_n \xrightarrow{c.s.} X$ ) si y solo si

$$\mathbf{P}(\{w \in \Omega : X_n(w) \rightarrow X(w)\}) = 1$$

es decir

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \bigcup_{m=n}^{\infty} \{|X_m(w) - X(w)| \geq \epsilon\} \right) = 0$$

Para probar este tipo de convergencia necesitamos aprender conceptos mucho más complicados que escapan el alcance de este curso.

**Definición:** Diremos que  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias converge en probabilidad a otra variable aleatoria  $X$  ( $X_n \xrightarrow{p} X$ ) si y solo si para todo  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X_n - X| \geq \epsilon) = 0$$

En general, para probar este tipo de convergencia utilizaremos la desigualdad de Markov.

Propiedad: Si  $X_n \xrightarrow{c.s.} X$  entonces  $X_n \xrightarrow{p} X$ .

### Ejemplos:

1. Sean  $\{X_n\}_{n \geq 2}$  variables aleatorias *i.i.d* con distribución  $\mathcal{U}(0, 1)$ , y sea  $Y_n = n^{-X_n}$ , probar que  $Y_n \xrightarrow{p} 0$

2.

### Propiedades:

Sean  $\{X_n\}_{n \geq 1}$  e  $\{Y_n\}_{n \geq 1}$  sucesiones de variables aleatorias,  $X$  e  $Y$  variables aleatorias.

1. Si  $X_n \xrightarrow{p} X$  y  $Y_n \xrightarrow{p} Y$ , entonces
  - a)  $c \cdot X_n \xrightarrow{p} c \cdot X$
  - b)  $g(X_n) \xrightarrow{p} g(X)$ , para toda función  $g$  continua
  - c)  $X_n + Y_n \xrightarrow{p} X + Y$
  - d)  $X_n \cdot Y_n \xrightarrow{p} X \cdot Y$
  - e)  $g(X_n, Y_n) \xrightarrow{p} g(X, Y)$  para toda  $g$  uniformemente continua. Si  $(X, Y) = (a, b)$  basta con que  $g$  sea continua.
2. Si  $|X_n| \leq c, \forall n$  entonces

$$X_n \xrightarrow{p} 0 \iff \lim_{n \rightarrow \infty} \mathbf{E}(|X_n|) = 0$$

3. Si  $X_n \xrightarrow{p} c$  y  $f$  es acotada y continua en  $c$ , entonces

$$\lim_{n \rightarrow \infty} \mathbf{E}(f(X_n)) = f(c)$$

## 6.3. Leyes de los grandes números

### 6.3.1. Ley débil de los grandes números

Sean  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias no correlacionadas tales que  $\mathbf{E}(X_i) = \mu_i$ ,  $\mathbf{var}(X_i) = \sigma_i^2, \forall i$ . Consideremos la sucesión de variables  $\{\bar{X}_n\}_{n \geq 1}$  con  $\bar{X}_n$  el promedio de las primeras  $n$  variables, y con

$$\mathbf{E}(\bar{X}_n) = \bar{\mu}_n = \sum_{i=1}^n \mu_i$$

entonces si

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \xrightarrow{n \rightarrow \infty} 0$$

se tiene que

$$\bar{X}_n - \bar{\mu} \xrightarrow{p} 0$$

**Observación:** Si todas las variables tienen igual media  $\mu_i = \mu, \forall i$  entonces  $\bar{\mu} = \mu$  resultando

$$\bar{X}_n \xrightarrow{p} \mu$$

Gracias a la desigualdad de Markov, podemos probar la ley débil de los grandes números, que lo que dice en términos de convergencia es que a medida que  $n$  aumenta y promediamos más variables, la probabilidad de que  $\bar{X}$  se aleje de  $\mu$  en más de un  $\epsilon$  es casi cero, y tiende a cero cuando  $n$  tiende a infinito.

Esta ley permite fundamentar el concepto de probabilidad de un evento

Si definimos  $X_i = \mathbf{I}\{\text{en el experimento } i \text{ ocurre el evento } A\}$ , tenemos que

$$\mathbf{E}(X_i) = \mathbf{P}(A)$$

$$\mathbf{var}(X_i) = \mathbf{P}(A)(1 - \mathbf{P}(A)) < \infty$$

Como además las  $X_i$  son independientes, por la ley de los grandes números

$$\bar{X}_n \xrightarrow{p} \mathbf{E}(X_1) = \mathbf{P}(A)$$

Observemos que  $\bar{X}_n$  es la frecuencia relativa de ocurrencia del evento  $\mathbf{A}$  en  $n$  realizaciones independientes del experimento. Por lo tanto tenemos que la frecuencia relativa converge a la probabilidad del evento.

### 6.3.2. Ley fuerte de los grandes números

Sean  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias independientes tales que  $\mathbf{E}(X_i) = \mu_i$ ,  $\mathbf{var}(X_i) = \sigma_i^2, \forall i$ . Si

$$\sum_{i=1}^n \frac{\sigma_i^2}{i^2} < \infty$$

(basta con que exista  $\mathbf{E}(|X_i|)$ )

se tiene que

$$\bar{X}_n - \bar{\mu} \xrightarrow{c.s.} 0$$

En particular, si todas las variables son idénticamente distribuidas con  $\mathbf{E}(X_i) < \infty$ ,  $\mathbf{var}(X_i) < \infty$  entonces

$$\bar{X}_n \xrightarrow{c.s.} \mu$$

**Ejemplo:** Sean  $\{X_n\}_{n \geq 1}$  variables aleatorias *i.i.d* con distribución  $\mathcal{E}(\lambda)$ , y sea  $Y_n = \sum_{i=1}^n X_i^3/n$ , hallar el límite en probabilidad de  $Y_n$ .

Ver anexo: Archivo `LGN.R` con simulaciones para entender la ley de los grandes números de forma empírica.





## Capítulo 7

# Teorema central del límite

### 7.1. Convergencia en distribución

Es un tipo de convergencia que se basa en la proximidad entre las funciones de distribución de las variables aleatorias.

**Definición:** Sean  $X, X_1, X_2, \dots$  variables aleatorias con función de distribución  $F, F_1, F_2, \dots$  respectivamente. Diremos que  $X_n$  converge en distribución a  $X$  ( $X_n \xrightarrow{D} X$ ) si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

para todo punto de continuidad  $x$  de  $F$ .

**Teorema 1:** Si  $X_n \xrightarrow{P} X$  entonces  $X_n \xrightarrow{D} X$ .

Si  $X_n \xrightarrow{D} X$  y  $\mathbf{P}(X = c) = 1$  entonces  $X_n \xrightarrow{P} X$ .

**Teorema 2:**  $X_n \xrightarrow{D} X \iff \varphi_{X_n}(t) \rightarrow \varphi_X(t)$ .

**Corolario:** Si  $\varphi_{X_n}(t) \rightarrow e^{-t^2/2}$  entonces  $X_n \xrightarrow{D} Z$  con  $Z \sim \mathcal{N}(0, 1)$ .

**Proposición:** Sean  $X, X_1, X_2, \dots$  variables aleatorias y  $g : \mathbb{R} \rightarrow \mathbb{R}$  una función continua, entonces

1.  $X_n \xrightarrow{c.s.} X$  entonces  $g(X_n) \xrightarrow{c.s.} g(X)$
2.  $X_n \xrightarrow{P} X$  entonces  $g(X_n) \xrightarrow{P} g(X)$
3.  $X_n \xrightarrow{D} X$  entonces  $g(X_n) \xrightarrow{D} g(X)$

**Teorema de Slutsky**

Sean  $\{X_n\}_{n \geq 1}$  e  $\{Y_n\}_{n \geq 1}$  sucesiones de variables aleatorias, tales que  $X_n \xrightarrow{D} X$  y  $Y_n \xrightarrow{p} c$ , entonces

1.  $X_n + Y_n \xrightarrow{D} X + c$
2.  $X_n - Y_n \xrightarrow{D} X - c$
3.  $X_n \cdot Y_n \xrightarrow{D} X \cdot c$
4. Si  $c \neq 0$  y  $\mathbf{P}(Y_n \neq 0) = 1$  entonces  $X_n/Y_n \xrightarrow{D} X/c$

**7.2. Teorema Central del Límite**

En el capítulo anterior analizamos que ocurre en el límite cuando promediamos variables aleatorias *i.i.d*. Lo que buscamos ahora es entender el comportamiento de la suma de variables aleatorias *i.i.d*, y cuando hablamos del comportamiento nos referimos a la distribución de probabilidades que puede llegar a tener esa nueva variable *suma*. Algunos casos que ya estudiamos son:

1. Si  $X_1, \dots, X_n \sim \mathcal{B}(p)$  i.i.d, entonces  $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$
2. Si  $X_1, \dots, X_n \sim \mathcal{G}(p)$  i.i.d, entonces  $\sum_{i=1}^n X_i \sim \text{Pas}(n, p)$
3. Si  $X_1, \dots, X_n \sim \mathcal{E}(\lambda)$  i.i.d, entonces  $\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$

Ahora, no se si ya simulaste el experimento de sumar variables y observar como se comporta esa suma. Es muy interesante *observar*. Veamos que ocurre en el anexo **SUMA.R**.

**7.2.1. El Teorema**

Sean  $X_1, \dots, X_n$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con  $\mathbf{E}[X_i] = \mu$  y  $\mathbf{var}(X_i) = \sigma^2$ ,  $i = 1, \dots, n$ , entonces bajo ciertas condiciones generales se tiene que:

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0, 1)$$

o dicho de otra forma

$$\lim_{n \rightarrow \infty} \mathbf{P}(Z_n < z) = \phi(z)$$

Podrás usar el teorema central del límite para encontrar aproximaciones cuando tengas que calcular probabilidades de variables aleatorias que se pueden expresar como sumas de

variables *i.i.d.*, en casos en que no conozcas las distribuciones, o que las conozcas pero la cuenta sea muy difícil. La aproximación será mejor cuanto mayor sea el valor de  $n$ , pero no significa que tenga que usarla solo si  $n$  toma cierto valor, dependerá de la dificultad con respecto al cálculo exacto.

### 7.3. Método delta

Sea  $X_1, X_2, X_3, \dots$  una sucesión de variables aleatorias con tales que:

$$\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

Si  $g$  es una función continua y derivable en  $\mu$  tal que  $g'(\mu) \neq 0$ , entonces:

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} \mathcal{N}(0, [g'(\mu)]^2 \sigma^2)$$



## Capítulo 8

# Bonus Track: Proceso de Poisson

### Proceso puntual de Poisson

Un proceso puntual aleatorio es un conjunto enumerable de puntos aleatorios ubicados sobre la recta real. En la mayoría de las aplicaciones un punto de un proceso puntual es el instante en que ocurre algún evento, motivo por el cual los puntos también se llaman eventos o arribos.

Llamamos  $N(t)$  al número de eventos durante un intervalo específico  $[0, t]$

### Propiedades:

1. El número de eventos durante intervalos de tiempo no sobrepuestos son variables aleatorias independientes
2. La probabilidad de cada evento en particular es la misma para todos los intervalos de longitud  $t$ , independientemente de la ubicación de cada intervalo y de la historia pasada del sistema. Las variables  $N(0, t)$  y  $N(t_1, t_1 + t)$  tienen la misma distribución de probabilidades.
3. Si el intervalo es suficientemente pequeño, la probabilidad de obtener exactamente una emisión durante ese intervalo es directamente proporcional a la longitud de ese intervalo
4. La probabilidad de obtener 2 o más eventos en un intervalo suficientemente pequeño es despreciable.

La variable  $N(t)$  toma los valores posibles  $\{0, 1, 2, 3, \dots\} = \mathbb{N}_0$ , y su función de probabilidad está dada por:

$$p_N(n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Donde  $\lambda > 0$ . Entonces  $N(t)$  tiene distribución de Poisson de parámetro  $\mu = \lambda \cdot t$ .

**Propiedad:** La variable aleatoria  $N(t)$ : cantidad de eventos en un intervalo de longitud  $t$ , tiene distribución de Poisson *si y solo si* la variable  $T$ : tiempo entre 2 eventos consecutivos de Poisson, tiene distribución exponencial de parámetro  $\lambda$

**Demostración:**

$$\mathbf{P}(T > t) = \mathbf{P}(N(t) = 0) = e^{-\lambda t}, t > 0$$

Por lo tanto  $F_T(t) = P(T \leq t) = (1 - e^{-\lambda t})\mathbf{1}\{t > 0\}$

Es decir  $T \sim \mathcal{E}(\lambda)$

La variable  $G$ : "tiempo hasta el  $k$ -ésimo evento de Poisson" tiene distribución Gamma de parámetros  $k$  y  $\lambda$

Se puede probar que

$$G = \sum_{i=1}^k T_i$$

donde  $T_1, \dots, T_k$  son variables iid exponenciales de parámetro  $\lambda$

Ejemplos:

1. Los tiempos (en minutos) entre la llegada de personas a la cola de un cajero automático son independientes y con distribución exponencial de parámetro igual a 0.5.
  - a) Hallar la probabilidad de que en 5 minutos llegue al menos una persona
  - b) Hallar la probabilidad de que pasen mas de 2 minutos hasta que llegue la primera persona
  - c) Hallar la probabilidad de que pasen más de 10 minutos para que lleguen 3 personas
2. La cantidad de insectos que arriban a la mesa de un asado responde a una distribución Poisson de media 60. Cada insecto puede ser una mosca con probabilidad  $2/3$ . La cantidad de insectos y la naturaleza de cada uno de ellos son independientes.
  - a) Dado que arribaron 60 insectos, ¿cuál es la distribución de la cantidad de moscas en la mesa?

- b) Hallar la probabilidad de que arriben 50 insectos y 35 de ellos sean moscas.  
 c) ¿Cuál es la distribución de la cantidad de moscas que arriban a la mesa?

### 8.1. Adelgazamiento o coloreo

Sea  $\{N(t), t > 0\}$  un proceso de Poisson de tasa  $\lambda$ . Cada vez que ocurre un arribo se lo clasifica como de tipo I o como de tipo II. Más aún, cada arribo se clasifica como de tipo I con probabilidad  $p$  o como de tipo II con probabilidad  $(1 - p)$ , independientemente de todos los demás arribos. Sean  $N_1(t)$  y  $N_2(t)$  la cantidad de arribos de tipo I y de tipo II que ocurren en  $[0, t]$ , respectivamente. Es claro que  $N(t) = N_1(t) + N_2(t)$ .

**Teorema:**  $\{N_1(t), t \geq 0\}$  y  $\{N_2(t), t \geq 0\}$  son procesos de Poisson independientes de tasas  $\lambda p$  y  $\lambda(1 - p)$ , respectivamente.

### 8.2. Superposición de procesos de Poisson

Sean  $N_1(t), t \geq 0$  y  $N_2(t), t \geq 0$  procesos de Poisson independientes de tasas  $\lambda_1$  y  $\lambda_2$ , respectivamente. Entonces  $N(t) := N_1(t) + N_2(t)$  define un proceso de Poisson de tasa  $\lambda = \lambda_1 + \lambda_2$ .

### 8.3. Propiedades

Como el proceso de Poisson es temporalmente homogéneo y tiene incrementos independientes es razonable pensar que los intervalos de igual longitud contenidos en el intervalo  $[0, t]$  deben tener la misma probabilidad de contener al arribo. En otras palabras, si ocurre un arribo, el tiempo en que ocurrió el arribo debe estar distribuido uniformemente sobre el intervalo  $[0, t]$ .

**Teorema:** Bajo la condición de que ocurrieran exactamente  $n$  arribos en el intervalo  $[0, t]$ , los tiempos de los  $n$  arribos  $S_1, S_2, \dots, S_n$ , consideradas como variables aleatorias desordenadas, son independientes y están distribuidos uniformemente sobre el intervalo  $[0, t]$

*Prueba para  $s < t$ :*

$$\mathbf{P}(S_1 < s | N(t) = 1) = \frac{s}{t}$$

Si defino la variable aleatoria  $S_i$ : momento en el que ocurre el  $i$ -ésimo arribo, entonces

$(S_1, S_2, \dots, S_n) | N(t) = n$  son variables aleatorias i.i.d. desordenadas y con distribución  $\mathcal{U}(0, t)$

Teniendo en cuenta estas distribuciones, podemos contar la cantidad de arribos en un sub-intervalo del  $(0, t)$  de manera que se definen las siguientes variables aleatorias:

$N(a, b) | N(0, t) = n$  : "cantidad de arribos en el intervalo  $(a, b)$  de  $n$  que ocurrieron en el intervalo  $(0, t)$ " donde  $0 < a < b < t$

Esta variable aleatoria tendrá distribución binomial de parámetros  $n$  y  $\frac{b-a}{t}$

Ejemplo: A un banco llegan clientes de acuerdo con un proceso de Poisson de intensidad 10 por hora. Suponiendo que dos clientes llegaron durante la primer hora, ¿cuál es la probabilidad de que:

1. ambos lleguen durante los primeros 20 minutos?
2. al menos uno llegue durante los primeros 20 minutos?

Tambien podemos definir un vector aleatorio, si  $0 < a < b < c < t$ ,  
 $(N(0, a), N(a, b), N(b, c), N(c, t)) | N(0, t) = n$  tendrá distribución multinomial de parámetros  $(n, (a)/t, (b-a)/t, (c-b)/t, (t-c)/t)$

Ejemplo:

1. Cierta alambre tiene fallas distribuidas según un proceso de Poisson de intensidad 2 por metro. Las fallas pueden ser de tipo I o II con probabilidad  $2/3$  y  $1/3$  respectivamente. Si en los primeros 10 metros de alambre se encontraron exactamente 16 fallas de tipo I, hallar la probabilidad de que halla a lo sumo una falla (de cualquier tipo) en el primero de esos 10 metros.