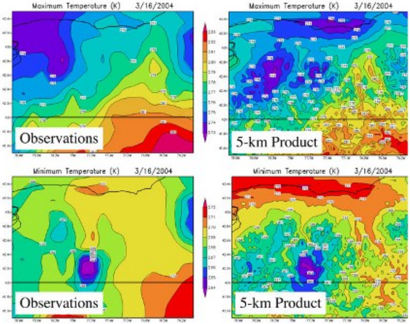
Repaso probabilidades y estadística

A. M. Alvarez-Meza, Ph.D.
amalvarezme@unal.edu.co

Departamento de ingeniería eléctrica, electrónica y computación
Universidad Nacional de Colombia-sede Manizales

Contenido

Datos financieros



El clima

Descripción y caracterización de la incertidumbre en variables aleatorias

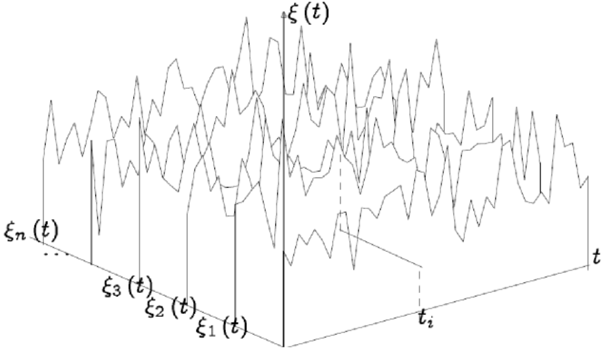
Representación de variables aleatorias II

- Cualquier modelo de un experimento analizado a priori se caracteriza por un conjunto de posibles resultados u **observaciones**.
- El subconjunto formado por un solo elemento se denomina **evento** y todo el conjunto se denomina **espacio de eventos**.
- Un espacio probabilístico está conformado entonces por tres componentes:
  - Eventos elementales.
  - Algebra de los eventos (reglas de relación).
  - Estabilidad de las frecuencias de aparición (**medida de probabilidad**).

Representación de señales aleatorias I

- La **señal aleatoria** $\xi(s)$ corresponde a un proceso que se desarrolla sobre la variable s.
- Cuando s discurre **continuamente**, i.e. el tiempo, se habla de **funciones aleatorias**.
- Cuando s corresponde a una malla de valores **discretos**, $\{s_1, s_2, \ldots, s_n\}$, se habla de **series aleatorias**.

Representación de señales aleatorias II



Las probabilidades permiten caracterizar y describir estructuras de aleatoriedad (incertidumbre) a partir de datos

- Sean dos variables aleatorias $X = \{x_1, \ldots, x_M\}$ y $Y = \{y_1, \ldots, y_L\}$.
- Se tienen N realizaciones de X y Y. Además X y Y son independientes.
- Se define la **probabilidad conjunta** como

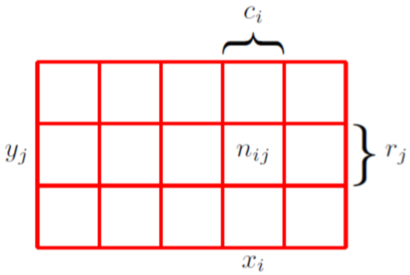
$$p\left(X = x_i, Y = y_j\right) = \frac{n_{ij}}{N},$$

donde $n_{ij} \in \mathbb{N}$ se define como el número de realizaciones en las que $X = x_i$ y $Y = y_j$ $(\forall i \in [1, M], j \in [1, L])$.

Nociones básicas II

- Sea c_i el número de realizaciones en las que X toma el valor x_i (ind. del valor de Y).
- Se define la **probabilidad marginal** de $X = x_i$ como

$$p\left(X = x_i\right) = \frac{c_i}{N} = \frac{1}{N} \sum_j n_{ij} = \sum_{j=1}^{L} p\left(X = x_i, Y = y_j\right).$$



Nociones básicas III

- Se define la **probabilidad condicional** de $Y = y_j$ dado $X = x_i$ como

$$p\left(Y = y_j | X = x_i\right) = \frac{n_{ij}}{c_i},$$

además,

$$\begin{aligned} p\left(X = x_i, Y = y_j\right) =& \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i}\frac{c_i}{N} \\ =& p\left(Y = y_j | X = x_i\right) p\left(X = x_i\right). \end{aligned}$$

- **Regla de la suma:** $p\left(X\right) = \sum_Y p\left(X, Y\right)$.
- **Regla del producto:**

$$\begin{aligned} p\left(X, Y\right) &= p\left(Y|X\right) p\left(X\right) \\ &= p\left(X|Y\right) p(Y). \end{aligned}$$

Nociones básicas IV

- **Teorema de Bayes**: dado que

$$p(X, Y) = p(Y, X)$$
$$p(Y|X)\,p(X) = p(X|Y)\,p(Y)$$

entonces:

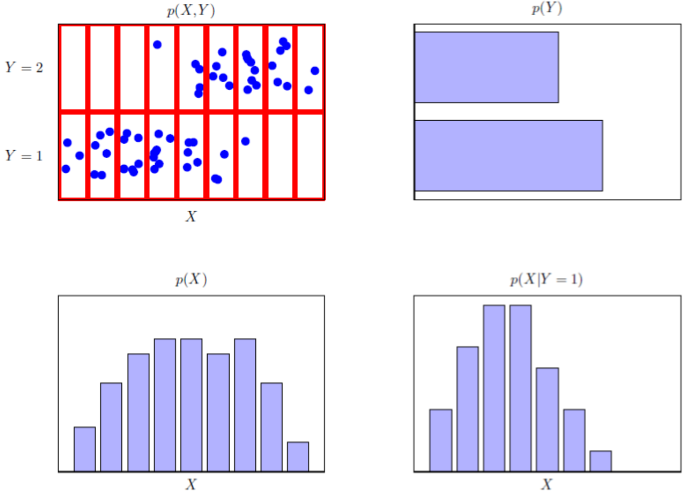
$$p(Y|X) = \frac{p(X|Y)\,p(Y)}{p(X)}.$$

- **Independencia**:

$$p(Y|X) = p(Y)\,, p(X, Y) = p(X)\,p(Y)\,.$$

Ejemplo

Suponga que X toma nueve valores y Y toma dos valores. Se tienen $N = 60$ realizaciones.



Densidad de probabilidad I

- La **función de densidad de probabilidad** $p(x)$ debe cumplir que

$$p(x) \geq 0, \int_{-\infty}^{\infty} p(x)dx = 1.$$

- Para un intervalo

$$p\left(x \in (a, b)\right) = \int_{a}^{b} p(x)dx.$$

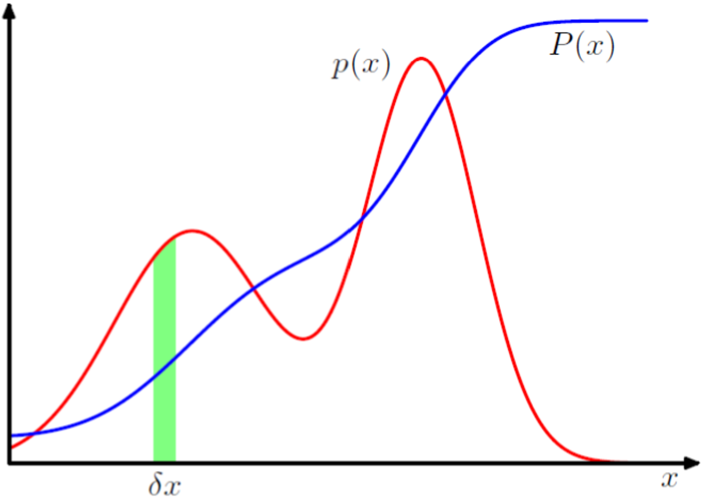
- La **función de distribución de probabilidad** (ó distribución acumulativa) se define como

$$P(x) = \int_{-\infty}^{x} p(z)dz,$$

igualmente

$$p(x) = \frac{dP(x)}{dx}.$$

Densidad de probabilidad II



Vectores aleatorios

- Sea $\Psi = \{X_1, X_2, \ldots, X_D\}$ un conjunto de D variables aleatorias.
- Estas variables aleatorias pueden representarse como un vector columna de dimensiones $D \times 1$,

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix}$$

- Un valor específico de $\boldsymbol{X}$ se denota como $\boldsymbol{x} = (x_1, x_2, \ldots, x_D)^\top$

Densidad de probabilidad conjunta

- La densidad de probabilidad conjunta para $\boldsymbol{X}$, $p(\boldsymbol{x}) = p(x_1, \ldots, x_D)$, debe satisfacer

$$p(\boldsymbol{x}) \geq 0$$
$$\int_{-\infty}^{\infty} p(\boldsymbol{x}) d\boldsymbol{x} = 1 \tag{1}$$

- **Regla de la suma:**

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y}.$$

- **Regla del producto:**

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{y}) p(\boldsymbol{y}).$$

- **Teorema de Bayes:**

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{y}) p(\boldsymbol{y})}{p(\boldsymbol{x})}.$$

Valor esperado y covarianza I

- El **valor esperado** o la **esperanza** de una función $f(x)$ se define como:

$$\mathbb{E}\left(f\right) = \sum_{x} p(x)f(x), \quad , \mathbb{E}\left(f\right) = \int p(x)f(x)dx.$$

- La **esperanza muestral** de una función $f(x)$ se define como:

$$\mathbb{E}\left(f\right) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i).$$

- La **esperanza condicional** de una función $f(x)$ dado $Y = y$ se define como:

$$\mathbb{E}_x\left(f|y\right) = \int p(x|y)f(x)dx.$$

- La **varianza** de una función $f(x)$ está definida como:

$$\mathrm{var}\left(f\right) = \mathbb{E}\left(\left[f(x) - \mathbb{E}\left(f(x)\right)\right]^2\right) = \mathbb{E}\left(f(x)^2\right) - \mathbb{E}\left(f(x)\right)^2.$$

Valor esperado y covarianza II

- La **covarianza** de dos variables aleatorias X y Y se define como:

$$\begin{aligned}\operatorname{cov}(x,y) =& \mathbb{E}_{x,y}\left([x-\mathbb{E}(x)][y-\mathbb{E}(y)]\right)\\ =& \mathbb{E}_{x,y}(xy)-\mathbb{E}(x)\,\mathbb{E}(y)\,.\end{aligned}$$

- La esperanza de un vector de variables aleatorias $\boldsymbol{X}$, se define como:

$$\mathbb{E}(\boldsymbol{x}) = \begin{bmatrix}\mathbb{E}(x_1)\\ \mathbb{E}(x_2)\\ \vdots\\ \mathbb{E}(x_D)\end{bmatrix}$$

- La covarianza para el caso de vectores de variables aleatorias $\boldsymbol{X}$ y $\boldsymbol{Y}$ se define como:

$$\begin{aligned}\operatorname{cov}(\boldsymbol{x},\boldsymbol{y}) =& \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}\left([\boldsymbol{x}-\mathbb{E}(\boldsymbol{x})]\left[\boldsymbol{y}^\top-\left(\mathbb{E}(\boldsymbol{y})\right)^\top\right]\right)\\ =& \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}\left(\boldsymbol{x}\boldsymbol{y}^\top\right)-\mathbb{E}(\boldsymbol{x})\left(\mathbb{E}(\boldsymbol{y})\right)^\top.\end{aligned}$$

Distribución Gaussiana I

- La distribución Gaussiana en el caso univariado se define como:

$$\mathcal{N}\left(x|\mu,\sigma^2\right)=\frac{1}{\left(2\pi\sigma^2\right)^{\frac{1}{2}}}\exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right).$$

- Esperanza

$$\mathbb{E}\left(x\right)=\int x\mathcal{N}\left(x|\mu,\sigma^2\right)dx=\mu.$$

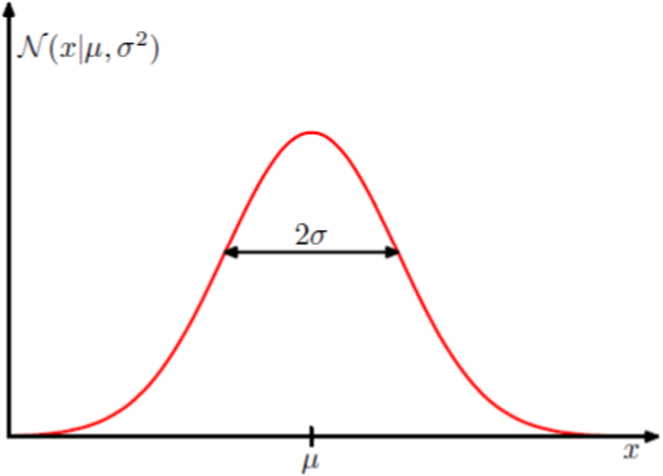
- Varianza

$$\operatorname{var}\left(x\right)=\mathbb{E}\left(x^2\right)-\mathbb{E}\left(x\right)^2=\sigma^2.$$

- Para el caso multivariado:

$$\mathcal{N}\left(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\frac{1}{\left(2\pi\right)^{\frac{D}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\mu\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\mu\right)\right).$$

Distribución Gaussiana II



Ejercicios Laboratorio

En un cuaderno (notebook) de jupyter responda a las siguientes preguntas con ejemplos concretos de implementación sobre Python 3. Envie su notebook al correo `amalvarezme@unal.edu.co`.

- Consultar en qué consisten las siguientes distribuciones de probabilidad (incluir modelo de pdf): i) Bernoulli, ii) Binomial, iii) beta, iv) Dirichlet, y v) Gaussiana

Referencias I

Castellanos Domínguez, C. G., Shinakov, Y. S., et al. (2007).
Análisis de aleatoriedad en senales y sistemas.
Universidad Nacional de Colombia-Sede Manizales.

Meyer, P. L., Campos, C. P., and Cuéllar, G. A. (1973).
Probabilidad y aplicaciones estadísticas.
Number QA273. 25. M49 1973. Fondo educativo interamericano.

Alvarez, M. (2014).
Material teoría estimación y análisis de aleatoriedad - Cursos de maestría y doctorado en ingeniería eléctrica.
Universidad Tecnológica de Pereira.
https://sites.google.com/site/maalvarezl/teaching-in-spanish