Statistique pour Mathématiciens - Introduction

1 Probability Space and Probability measures

- A set $\Omega \neq \emptyset$
    - $\omega \in \Omega$: elementary events
    - subsets $F \subseteq \Omega$: events
    - $\mathcal{F} \subseteq \mathcal{P}(\Omega)$: sample space
- A probability measure is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that
    (i) $\mathbb{P}(\Omega) = 1$,
    (ii) For every disjoint collection of events $(F_n)_{n\geq 1}$ in $\mathcal{F}$

$$\mathbb{P}\left(\bigcup_{n\geq 1} F_n\right) = \sum_n \mathbb{P}(F_n).$$

- $\Omega = \{\omega_1, \dots, \omega_n\}$ finite space, $G = \{\omega_{i_1}, \dots, \omega_{i_k}\} \in \mathcal{P}(\Omega)$ event,

$$\mathbb{P}(G) = \sum_{j:\omega_j \in G} \mathbb{P}(\{\omega_j\}).$$

1.1 Conditional Probability, Independence

- Let $G, H \in \mathcal{F}$ with $\mathbb{P}(H) > 0$
    - Conditional probability $\mathbb{P}(G|H) = \frac{\mathbb{P}(G\cap H)}{\mathbb{P}(H)}$
- Let $G, H_1, H_2, \ldots \in \mathcal{G}$, $H_1, H_2, \ldots$ disjoint, $\mathbb{P}(H_i) > 0$, $\cup_i H_i = \Omega$,
    - law of total probability: $\mathbb{P}(G) = \sum_{i=1}^{\infty} \mathbb{P}(G|H_i)\mathbb{P}(H_i)$

- Bayes theorem: $\mathbb{P}(H_j|G) = \mathbb{P}(H_j \cap G)/\mathbb{P}(G) = \frac{\mathbb{P}(G|H_j)\mathbb{P}(H_j)}{\sum_{i=1}^{\infty}\mathbb{P}(G|H_i)\mathbb{P}(H_i)}$

- The events $G_1, G_2, \dots \in \mathcal{G}$ are *independent* iff for any finite sub-collection $G_{i_1}, \dots, G_{i_k}$:

$$\mathbb{P}(G_{i_1} \cap \dots \cap G_{i_k}) = \mathbb{P}(G_{i_1}) \times \mathbb{P}(G_{i_2}) \times \ldots \times \mathbb{P}(G_{i_k})$$

2 Random Variables

- A random variable is a function $X : \Omega \to \mathbb{R}$ such that $\forall x \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

- The distribution function of a random variable X is $F : \mathbb{R} \to \mathbb{R}$ defined as

$$F(x) = \mathbb{P}(X \leq x).$$

- Properties of the distribution function:

  (i) $x \leq y \;\Rightarrow\; F(x) \leq F(y)$

  (ii) $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$

  (iii) $\lim_{y\downarrow x, y\neq x} F(y) = F(x)$, that is, F is right-continuous.

- Consider the set $D_F = \{x \in \mathbb{R} : F(x) - F(x^-) > 0\}$

  - If $\mathbb{P}(\{X \in D_F\}) = 1$ then X is a *discrete* random variable.
  - If $D_F = \emptyset$ then X is a *continuous* random variable.
  - Note that r.v. could be neither continuous or discrete (mixed r.v., ...).

2.1 Density

- Discrete case:
  The density function of a discrete r.v. X is $f : \mathbb{R} \to \mathbb{R}$ defined via

$$f(x) = \mathbb{P}(X = x).$$

  - $\mathbb{P}_X(G) = \mathbb{P}(X \in G) = \sum_{j:x_j\in G} f(x_j)$

- $F(x) = \sum_{j:x_j \leq x} f(x_j)$
- $F_X(x)$ piecewise constant with possible jumps at $x_1, x_2, \dots$
- $f_X(x) = \mathbb{P}(X = x)\ \forall x$
- $f_X(x) \leq 1\ \forall x$

- Continuous case:
  The distribution function F of a continuous random variable admits density function if there exists $f : \mathbb{R} \to \mathbb{R}_+$ such that

$$F(b) - F(a) = \int_a^b f(x)\mathrm{d}x.$$

  - $\mathbb{P}(G) = P(X \in G) = \int_G f(x)\mathrm{d}x$
  - $F(x) = \int_{-\infty}^{x} f(x)\mathrm{d}x$
  - $F(x)$ continuous
  - $f(x) = \frac{d}{\mathrm{d}x} F(x)$
  - $f(x) \neq \mathbb{P}(X = x) = 0$!!!
  - Can be $f(x) > 1$ for some x. In fact, f can be unbounded!

3 Random Vectors

- Equivalent definitions of a random vector $\mathbf{X} = (X_1, \dots, X_d)^T$
  - a vector of random variables defined on the same space Ω
  - a random variable with values in $\mathbb{R}^d$.

- Joint distribution function:

$$F_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

- Joint density:
  - discrete:

$$f_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}(X_1 = x_1, \dots, X_d = x_d)$$

- continuous:

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{\mathbf{X}}(u_1, \ldots, u_d) \mathrm{d}u_1 \ldots \mathrm{d}u_d$$

$$f_{\mathbf{X}}(x_1, \ldots, x_d) = \frac{\partial^d}{\partial x_1 \ldots \partial x_d} F_{\mathbf{X}}(x_1, \ldots, x_d)$$

3.1 Marginal Distributions

- $f_{X_i} : \mathbb{R} \to \mathbb{R}_+$ defined via
  - Discrete case:

$$f_{X_i}(x_i) = \mathbb{P}(X_i = x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_d} f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d)$$

  - Continuous case:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_d) \mathrm{d}y_1 \ldots \mathrm{d}y_{i-1} \mathrm{d}y_{i+1} \mathrm{d}y_d.$$

- Marginals DO NOT determine the joint distribution

3.2 Independence

- $X_1, \ldots, X_n$ (discrete or continuous) are independent iff $\forall$ collection of intervals $G_1 \subset \mathbb{R}, \ldots, G_n \subset \mathbb{R}$,

$$\mathbb{P}\{X_1 \in G_1, \ldots, X_n \in G_n\} = \prod_i \mathbb{P}\{X_i \in G_i\}.$$

  - $X_1, \ldots, X_d$ are independent iff for all $x_1, \ldots, x_d \in \mathbb{R}$

$$F_{(X_1, \ldots, X_d)}(x_1, \ldots, x_d) = F_{X_1}(x_1) \times \ldots \times F_{X_d}(x_d)$$

  - (if densities exist) $X_1, \ldots, X_d$ are independent iff

$$f_{(X_1, \ldots, X_d)}(x_1, \ldots, x_d) = f_{X_1}(x_1) \times \ldots \times f_{X_d}(x_d)$$

4 Expectation, Variance, Covariance

- Expectation:
  - For continuous variables:

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, f(x)\mathrm{d}x.$$

  - For discrete variables:

$$\mathbb{E}[X] = \sum_j x_j\, f(x_j) = \sum_j x_j \mathbb{P}(X = x_j).$$

  - Linearity: $\mathbb{E}[X_1 + \alpha X_2] = \mathbb{E}[X_1] + \alpha \mathbb{E}[X_2]$.
  - $\mathbb{E}[h(x)] = \sum_j h(x_j)\mathbb{P}(X = x_j)$ (discrete)
    or
    $\mathbb{E}[h(x)] = \int_{\mathbb{R}} h(x) f(x)\mathrm{d}x$ (continuous).

- Variance

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \qquad (\text{if } \mathbb{E}[X^2] < \infty)$$

- Covariance between X_1 and X_2

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right] \qquad (\text{if } \mathbb{E}[X_i^2] < \infty).$$

- Correlation between X_1 and X_2

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \in [-1, 1].$$

  Expresses the degrees of linear dependency.

- Useful formulae:
  - $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathrm{Cov}(X, X)$
  - $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$
  - $\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$
  - $\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$

- $\text{Cov}(aX_1 + bX_2, Y) = a\text{Cov}(X_1, Y) + b\text{Cov}(X_2, Y)$

- if $\mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] < \infty$, then the following are equivalent:
  - $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$
  - $\text{Cov}(X_1, X_2) = 0$
  - $\text{Var}(X_1 \pm X_2) = \text{Var}(X_1) + \text{Var}(X_2)$

  Independence implies this, but the converse does not hold!

5 Moments

- The moment of order k of a random variable X is $\mathbb{E}[X^k]$ (if $\mathbb{E}[|X^k|] < \infty$).
- Moment generating function: $M_X : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ defined by

$$M_X(t) = \mathbb{E}[e^{tX}].$$

  - If $M_X(t)$ exists and is finite for $|t| < b$ for some $b > 0$ then $\mathbb{E}[X^k] = M_X^{(k)}(0)$
  - The moment generating function identifies the distribution, i.e. $M_X(t) = M_Y(t) < \infty$ around a neighbourhood of 0 $\implies \mathbb{P}_X = \mathbb{P}_Y$
  - If X and Y are independent then $M_{X+Y}(y) = M_X(y)M_Y(y)$.