2 Probability Theory

Definitions

- Sample space Ω
- Outcome (point or element) $\omega \in \Omega$
- Event $A \subseteq \Omega$
- σ-algebra $\mathcal{A}$
  1. $\varnothing \in \mathcal{A}$
  2. $A_1, A_2, \ldots, \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
  3. $A \in \mathcal{A} \implies \neg A \in \mathcal{A}$
- Probability Distribution $\mathbb{P}$
  1. $\mathbb{P}[A] \geq 0 \quad \forall A$
  2. $\mathbb{P}[\Omega] = 1$
  3. $\mathbb{P}\left[\bigsqcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$
- Probability space $(\Omega, \mathcal{A}, \mathbb{P})$

Properties

- $\mathbb{P}[\varnothing] = 0$
- $B = \Omega \cap B = (A \cup \neg A) \cap B = (A \cap B) \cup (\neg A \cap B)$
- $\mathbb{P}[\neg A] = 1 - \mathbb{P}[A]$
- $\mathbb{P}[B] = \mathbb{P}[A \cap B] + \mathbb{P}[\neg A \cap B]$
- $\mathbb{P}[\Omega] = 1 \qquad \mathbb{P}[\varnothing] = 0$
- $\neg(\bigcup_n A_n) = \bigcap_n \neg A_n \quad \neg(\bigcap_n A_n) = \bigcup_n \neg A_n$ DeMorgan
- $\mathbb{P}[\bigcup_n A_n] = 1 - \mathbb{P}[\bigcap_n \neg A_n]$
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$

  $\implies \mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$
- $\mathbb{P}[A \cup B] = \mathbb{P}[A \cap \neg B] + \mathbb{P}[\neg A \cap B] + \mathbb{P}[A \cap B]$
- $\mathbb{P}[A \cap \neg B] = \mathbb{P}[A] - \mathbb{P}[A \cap B]$

Continuity of Probabilities

- $A_1 \subset A_2 \subset \ldots \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A] \quad$ where $A = \bigcup_{i=1}^{\infty} A_i$
- $A_1 \supset A_2 \supset \ldots \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A] \quad$ where $A = \bigcap_{i=1}^{\infty} A_i$

Independence $\perp\!\!\!\perp$

$$A \perp\!\!\!\perp B \iff \mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$$

Conditional Probability

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \qquad \mathbb{P}[B] > 0$$

Law of Total Probability

$$\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i] \qquad \Omega = \bigsqcup_{i=1}^{n} A_i$$

Bayes' Theorem

$$\mathbb{P}[A_i \mid B] = \frac{\mathbb{P}[B \mid A_i]\,\mathbb{P}[A_i]}{\sum_{j=1}^{n} \mathbb{P}[B \mid A_j]\,\mathbb{P}[A_j]} \qquad \Omega = \bigsqcup_{i=1}^{n} A_i$$

Inclusion-Exclusion Principle

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{r=1}^{n} (-1)^{r-1} \sum_{i \leq i_1 < \cdots < i_r \leq n} \left|\bigcap_{j=1}^{r} A_{i_j}\right|$$

3 Random Variables

Random Variable (RV)

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

Probability Mass Function (PMF)

$$f_X(x) = \mathbb{P}[X = x] = \mathbb{P}[\{\omega \in \Omega : X(\omega) = x\}]$$

Probability Density Function (PDF)

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)\,dx$$

Cumulative Distribution Function (CDF)

$$F_X : \mathbb{R} \to [0,1] \qquad F_X(x) = \mathbb{P}[X \leq x]$$

1. Nondecreasing: $x_1 < x_2 \implies F(x_1) \leq F(x_2)$
2. Normalized: $\lim_{x\to-\infty} = 0$ and $\lim_{x\to\infty} = 1$
3. Right-Continuous: $\lim_{y \downarrow x} F(y) = F(x)$

$$\mathbb{P}[a \leq Y \leq b \mid X = x] = \int_a^b f_{Y|X}(y \mid x) dy \qquad a \leq b$$

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

Independence

1. $\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x]\,\mathbb{P}[Y \leq y]$
2. $f_{X,Y}(x,y) = f_X(x) f_Y(y)$

3.1 Transformations

Transformation function

$$Z = \varphi(X)$$

Discrete

$$f_Z(z) = \mathbb{P}[\varphi(X) = z] = \mathbb{P}[\{x : \varphi(x) = z\}] = \mathbb{P}\left[X \in \varphi^{-1}(z)\right] = \sum_{x \in \varphi^{-1}(z)} f_X(x)$$

Continuous

$$F_Z(z) = \mathbb{P}[\varphi(X) \le z] = \int_{A_z} f(x)\,dx \quad \text{with } A_z = \{x : \varphi(x) \le z\}$$

Special case if φ strictly monotone

$$f_Z(z) = f_X(\varphi^{-1}(z)) \left|\frac{d}{dz}\varphi^{-1}(z)\right| = f_X(x)\left|\frac{dx}{dz}\right| = f_X(x)\frac{1}{|J|}$$

The Rule of the Lazy Statistician

$$\mathbb{E}[Z] = \int \varphi(x)\,dF_X(x)$$

$$\mathbb{E}[I_A(x)] = \int I_A(x)\,dF_X(x) = \int_A dF_X(x) = \mathbb{P}[X \in A]$$

Convolution

- $Z := X + Y \qquad f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx \overset{X,Y\ge 0}{=} \int_0^z f_{X,Y}(x, z-x)\,dx$
- $Z := |X - Y| \qquad f_Z(z) = 2\int_0^{\infty} f_{X,Y}(x, z+x)\,dx$
- $Z := \dfrac{X}{Y} \qquad f_Z(z) = \int_{-\infty}^{\infty} |x| f_{X,Y}(x, xz)\,dx \overset{\perp\!\!\!\perp}{=} \int_{-\infty}^{\infty} x f_x(x) f_X(x) f_Y(xz)\,dx$

4 Expectation

Definition and properties

- $\mathbb{E}[X] = \mu_X = \int x\,dF_X(x) = \begin{cases} \sum_x x f_X(x) & \text{X discrete} \\ \int x f_X(x)\,dx & \text{X continuous} \end{cases}$
- $\mathbb{P}[X = c] = 1 \implies \mathbb{E}[X] = c$
- $\mathbb{E}[cX] = c\,\mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[XY] = \int_{X,Y} xy f_{X,Y}(x,y)\,dF_X(x)\,dF_Y(y)$
- $\mathbb{E}[\varphi(Y)] \ne \varphi(\mathbb{E}[X]) \qquad$ (cf. Jensen inequality)
- $\mathbb{P}[X \ge Y] = 1 \implies \mathbb{E}[X] \ge \mathbb{E}[Y]$
- $\mathbb{P}[X = Y] = 1 \iff \mathbb{E}[X] = \mathbb{E}[Y]$
- $\mathbb{E}[X] = \sum_{x=1}^{\infty} \mathbb{P}[X \ge x]$

Sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

Conditional expectation

- $\mathbb{E}[Y \mid X = x] = \int y f(y \mid x)\,dy$
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$
- $\mathbb{E}_{\varphi(X,Y) \mid X=x}[=] \int_{-\infty}^{\infty} \varphi(x,y) f_{Y|X}(y \mid x)\,dx$
- $\mathbb{E}[\varphi(Y,Z) \mid X = x] = \int_{-\infty}^{\infty} \varphi(y,z) f_{(Y,Z)|X}(y,z \mid x)\,dy\,dz$
- $\mathbb{E}[Y + Z \mid X] = \mathbb{E}[Y \mid X] + \mathbb{E}[Z \mid X]$
- $\mathbb{E}[\varphi(X)Y \mid X] = \varphi(X)\mathbb{E}[Y \mid X]$
- $\mathbb{E}_{Y \mid X}[=] c \implies \text{Cov}[X,Y] = 0$

5 Variance

Definition and properties

- $\mathbb{V}[X] = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
- $\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i] + 2\sum_{i \ne j} \text{Cov}[X_i, Y_j]$
- $\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i] \quad \text{if } X_i \perp\!\!\!\perp X_j$

Standard deviation

$$\mathsf{sd}[X] = \sqrt{\mathbb{V}[X]} = \sigma_X$$

Covariance

- $\text{Cov}[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$
- $\text{Cov}[X,a] = 0$
- $\text{Cov}[X,X] = \mathbb{V}[X]$
- $\text{Cov}[X,Y] = \text{Cov}[Y,X]$