MATH357 - Statistics

Based on lectures from Winter 2025 by Prof. Abbas Khalili.
Notes by Louis Meunier

Contents

§1 Review of Probability

↪**Definition 1.1** (Measurable Space, Probability Space): We work with a set Ω = sample space = {outcomes}, and a σ-algebra $\mathcal{F}$, which is a collection of subsets of Ω containing Ω and closed under taking complements and countable unions. The tuple $(\Omega, \mathcal{F})$ is called *measurable space*.

We call a nonnegative function $P : \mathcal{F} \to \mathbb{R}$ defined on a measurable space a *probability function* if $P(\Omega) = 1$ and if $\{E_n\} \subseteq \mathcal{F}$ a disjoint collection of subsets of Ω, then $P\left(\bigcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} P(E_n)$. We call the tuple $(\Omega, \mathcal{F}, P)$ a *probability space*.

↪**Definition 1.2** (Random Variables): Fix a probability space $(\Omega, \mathcal{F}, P)$. A Borel-measurable function $X : \Omega \to \mathbb{R}$ (namely, $X^{-1}(B) \in \mathcal{F}$ for every $B \in \mathfrak{B}(\mathbb{R})$) is called a *random variable* on $\mathcal{F}$.

- *Probability distribution*: X induces a probability distribution on $\mathfrak{B}(\mathbb{R})$ given by $P(X \in B)$
- *Cumulative distribution function (CDF)*:

$$F_X(x) := P(X \leq x).$$

Note that $F(-\infty) = 0, F(+\infty) = 1$ and F right-continuous.

We say X *discrete* if there exists a countable set $S := \{x_1, x_2, \ldots\} \subset \mathbb{R}$, called the *support* of X, such that $P(X \in S) = 1$. Putting $p_i := P(X = x_i)$, then $\{p_i : i \geq 1\}$ is called the *probability mass function* (PMF) of X, and the CDF of X is given by

$$P(X \leq x) = \sum_{i : x_i \leq x} p_i.$$

We say X *continuous* if there is a nonnegative function f, called the *probability distribution function* (PDF) of X such that $F(x) = \int_{-\infty}^{x} f(t) \, \mathrm{d}t$ for every $x \in \mathbb{R}$. Then,

- $\forall B \in \mathfrak{B}(\mathbb{R}), P(X \in B) = \int_B f(t) \, \mathrm{d}t$
- $F'(x) = f(x)$
- $\int_{-\infty}^{\infty} f(x) \, \mathrm{d}x = 1$

If $X : \Omega \to \mathbb{R}$ a random variable and $g : \mathbb{R} \to \mathbb{R}$ a Borel-measurable function, then $Y := g(X) : \Omega \to \mathbb{R}$ also a random variable.

↪**Definition 1.3** (Moments): Let X be a discrete/random random variable with pmf/pdf f and support S. Then, if $\sum_{x\in S}|x|f(x) / \int_S |x|f(x)\,\mathrm{d}x < \infty$, then we say the first moment/mean of X exists, and define

$$\mu_X = \mathbb{E}[X] = \begin{cases} \sum_{x\in S} xf(x) \\ \int_S xf(x)\,\mathrm{d}x \end{cases}.$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable function. Then, we have

$$\mathbb{E}[g(X)] = \begin{cases} \sum_{x\in S} g(x)f(x) \\ \int_S g(x)f(x) \end{cases}.$$

Taking $g(x) = |x|^k$ gives the so-called "kth absolute moments", and $g(x) = x^k$ gives the ordinary "kth moments". Notice that $\mathbb{E}[\cdot]$ linear in its argument.

For $k \geq 1$, if μ exists, define the central moments

$$\mu_k := \mathbb{E}\Big[(X - \mu)^k\Big],$$

where they exist.

↪**Definition 1.4** (Moment Generating Function (mgf)): If X a r.v., the mgf of X is given by

$$M(t) := \mathbb{E}\big[e^{tX}\big],$$

if it exists for $t \in (-h, h)$, $h > 0$. Then, $M^{(n)}(0) = \mathbb{E}[X^n]$.

↪**Definition 1.5** (Multiple Random Variable): $X = (X_1, ..., X_n) : \Omega \to \mathbb{R}^n$ a random vector if $X^{-1}(I) \in \mathcal{F}$ for every $I \in \mathfrak{B}_{\mathbb{R}^n}$. (It suffices to check for "rectangles" $I = (-\infty, a_1] \times \cdots \times (-\infty, a_n]$, as before.)

Let F be the CDF of X, and let $A \subseteq \{1, ..., n\}$, enumerating A by $\{i_1, ..., i_k\}$. Then, the CDF of the subvector $X_A = \left(X_{i_1}, ..., X_{i_k}\right)$ is given by

$$F_{X_A}\left(x_{i_1}, ..., x_{i_k}\right) = \lim_{\substack{x_{i_j} \to \infty, \\ i_j \in \mathcal{I} \setminus A}} F(x_1, ..., x_n).$$

In particular, the marginal distribution of X_i is given by

$$F_{X_i}(x) = \lim_{x_1, ..., x_{i-1}, x_{i+1}, ..., x_n \to +\infty} F(x_1, ..., x, ..., x_n).$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ measurable. Then,

$$\mathbb{E}[g(X_1, ..., X_n)] = \begin{cases} \sum_{(x_1, ..., x_n)} g(x_1, ..., x_n) f(x_1, ..., x_n) \\ \int \cdots \int g(x_1, ..., x_n) f(x_1, ..., x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \end{cases}.$$

We have the notion of a joint mgf,

$$M(t_1, ..., t_n) = \mathbb{E}\left[e^{\sum_{i=1}^n t_i X_i}\right],$$

if it exists for $0 < \left(\sum_{i=1}^n t_i^2\right)^{\frac{1}{2}} < h$ for some $h > 0$. Notice that $M(0, ..., 0, t_i, 0, ..., 0)$ is equal to the mgf of X_i.