1 Probability

1.1 Probability Space

- *Sample space* Ω is the collection of all possible outcomes.
- An *event* A is a subset of Ω.
- A σ-field, denoted by $\mathcal{F}$, is a collection of events such that: (i) $\emptyset \in \mathcal{F}$; (ii) if an event $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$; (iii) if $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$, then $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.
- $(\Omega, \mathcal{F})$ is called a *measure space*.
- A function $\mu : \mathcal{F} \mapsto [0, \infty]$ is called a *measure* if it satisfies (i) $\mu(A) \geq 0$ for all $A \in \mathcal{F}$; (ii) if $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, are mutually disjoint, then $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)$
- If $\mu(\Omega) = 1$, we call μ a *probability measure*. A probability measure is often denoted as P.
- $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

1.2 Random Variable

- A function $X : \Omega \mapsto \mathbb{R}$ is $(\Omega, \mathcal{F}) \setminus (\mathbb{R}, \mathcal{R})$ *measurable* if $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ for any $B \in \mathcal{R}$, where $\mathcal{R}$ is the Borel σ-field on the real line. *Random variable* is an alternative name for a measurable function.
- Discrete random variable: the set $\{X(\omega) : \omega \in \Omega\}$ is finite or countable.
- Continuous random variable: the set $\{X(\omega) : \omega \in \Omega\}$ is uncountable.
- $P_X : \mathcal{R} \mapsto [0, 1]$ is also a probability measure if defined as $P_X(B) = P\left(X^{-1}(B)\right)$ for any $B \in \mathcal{R}$. This P_X is called the probability measure *induced* by the measurable function X .

1.3 Distribution Function

- (Cumulative) distribution function

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

- Properties of CDF: $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, non-decreasing, and right-continuous

$$\lim_{y\to x^+} F(y) = F(x).$$

- Probability density function (PDF): if there exists a function f such that for all x,

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

then f is called the PDF of X.

- Properties: $f(x) \geq 0$. $\int_a^b f(x)\, dx = F(b) - F(a)$

1.4 Examples

- Binary, Poisson, uniform, normal, χ^2, t, F.
- Parametric distribution verses nonparametric distribution.
- Implementation in `R`: `d` for density, `p` for probability, `q` for quantile, and `r` for random variable. For instance, `dnorm`, `pnorm`, `qnorm`, and `rnorm`.

2 Expected Value

2.1 Integration

- X is called a *simple function* on a measurable space $(\Omega, \mathcal{F})$ if $X = \sum_i a_i 1\{A_i\}$ is a finite sum, where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$.
- Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $a_i \geq 0$ for all i. The integral of X with respect to μ is

$$\int X \mathrm{d}\mu = \sum_i a_i \mu(A_i).$$

- Let X be a non-negative measurable function. The integral of X with respect to μ is

$$\int X \mathrm{d}\mu = \sup\left\{ \int Y \mathrm{d}\mu : 0 \leq Y \leq X,\ Y \text{ is simple} \right\}.$$

- Let X be a measurable function. Define $X^+ = \max\{X, 0\}$ and $X^- = -\min\{X, 0\}$. Both X^+ and X^- are non-negative functions. The integral of X with respect to μ is

$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$

- If the measure μ is a probability measure P, then the integral $\int X \mathrm{d}P$ is called the *expected value,* or *expectation,* of X. We often use the popular notation $E[X]$, instead of $\int X \mathrm{d}P$, for convenience.

2.2 Properties

- Elementary calculation: $E[X] = \sum_x xP(X = x)$ or $E[X] = \int xf(x) \mathrm{d}x$.
- $E[1\{A\}] = P(A)$.
- $E[X^r]$ is call the r-moment of X. Mean $\mu = E[X]$, variance $\mathrm{var}[X] = E\left[(X-\mu)^2\right]$, skewness $E\left[(X-\mu)^3\right]$ and kurtosis $E\left[(X-\mu)^4\right]$.
- Skewness coefficient $E\left[(X-\mu)^3\right]/\sigma^3$, degree of excess $E\left[(X-\mu)^4\right]/\sigma^4 - 3$.
  - Application: The formula that killed Wall Street
- Jensen's inequality. If $\varphi(\cdot)$ is a convex function, then $\varphi(E[X]) \leq E[\varphi(X)]$.
  - Application: Kullback-Leibler distance $d(p,q) = \int \log(p/q) \mathrm{d}P = E_P[\log(p/q)]$
- Markov inequality: if $E[|X|^r]$ exists, then $P(|X| > \epsilon) \leq E[|X|^r]/\epsilon^r$ for all $r \geq 1$.
  - Application: Chebyshev inequality: $P(|X| > \epsilon) \leq E\left[X^2\right]/\epsilon^2$.

3 Multivariate Random Variable

- Bivariate random variable: $X : \Omega \mapsto \mathbb{R}^2$.
- Multivariate random variable $X : \Omega \mapsto \mathbb{R}^n$.
- Joint CDF: $F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$. Joint PDF is defined similarly.

3.1 Elementary Formulas

- conditional density $f(Y|X) = f(X,Y)/f(X)$
- marginal density $f(Y) = \int f(X,Y)\, dX$.
- conditional expectation $E[Y|X] = \int Y f(Y|X)\, dY$

- proof of law of iterated expectation

$$E\left[E\left[Y|X\right]\right] = \int E\left[Y|X\right] f\left(X\right) dX = \int \left(\int Y f\left(Y|X\right) dY\right) f\left(X\right) dX = \int \int Y f\left(Y|X\right) f\left(X\right) dY dX$$
$$= \int \int Y f\left(X, Y\right) dY dX = \int Y \left(\int f\left(X, Y\right) dX\right) dY = \int Y dY = E\left[Y\right].$$

- conditional probability, or Bayes' Theorem $P\left(A|B\right) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$.

3.2 Independence

X and Y are *independent* if $P\left(X \in A, Y \in B\right) = P\left(X \in A\right) P\left(Y \in B\right)$ for all A and B.

Application: (Chebyshev law of large numbers) If $X_1, X_2, \ldots, X_n$ are independent, and they have the same mean μ and variance $\sigma^2 < \infty$. Let $Z_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)$. Then the probability $P\left(|Z_n| > \epsilon\right) \to 0$ as $n \to \infty$,

3.3 Law of Iterated Expectations

- Given a probability space $(\Omega, \mathcal{F}, P)$, a sub σ-algebra $\mathcal{G} \subset \mathcal{F}$ and a $\mathcal{F}$-measurable function X with $E|X| < \infty$, the *conditional expectation* $E\left[X|\mathcal{G}\right]$ is defined as a $\mathcal{G}$-measurable function such that $\int_A X dP = \int_A E\left[X|\mathcal{G}\right] dP$ for all $A \in \mathcal{G}$.

- Law of iterated expectations
$$E\left[E\left[Y|X\right]\right] = E\left[Y\right]$$
is a trivial fact from the definition of the conditional expectation by taking $A = \Omega$.

- Properties of conditional expectations
  1. $E\left[E\left[Y|X_1, X_2\right]|X_1\right] = E\left[Y|X_1\right]$
  2. $E\left[E\left[Y|X_1\right]|X_1, X_2\right] = E\left[Y|X_1\right]$
  3. $E\left[h\left(X\right) Y|X\right] = h\left(X\right) E\left[Y|X\right]$

Application: Regression $Y = E\left[Y|X\right] + \epsilon$, where $\epsilon = Y - E\left[Y|X\right]$.