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 $(\Omega, \mathcal{F}) \backslash (\mathbb{R}, \mathcal{R})$ *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`. Execute online `http://www.tutorialspoint.com/execute_r_online.php`.

2   Expected Value

- *Expected value,* or *expectation,* is an average of a random variable. Expectation is nothing but an integral. We write $E[X]$, instead of $\int X dP$, just for a concise notation when the underlying probability measure is clear.

- Elementary calculation: $E[X] = \sum_x x P(X = x)$ or $E[X] = \int x f(x)\, \mathrm{d}x$.

- $P(X \in A) = E[1\{X \in A\}]$, where $1\{\cdot\}$ is the indicator function, and $\{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\}$.

- $E[X^r]$ 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 $E\left[(X - \mu)^4\right]/\sigma^4 - 3$.

- Jensen's inequality. If $\varphi(\cdot)$ is a convex function, then $\varphi(E[X]) \leq E[\varphi(X)]$.

- Chebyshev inequality: $P(|X| > \epsilon) \leq E[X^2]/\epsilon^2$.

3   Multivariate Random Variable

- Bivariate random variable: $X : \Omega \mapsto \mathbb{R}^2$.

- Multivariate random variable $X : \Omega \mapsto \mathbb{R}^d$.

- Joint CDF: $F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$. Joint PDF is defined similarly.

3.1 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* $E[X|\mathcal{G}]$ is defined as a $\mathcal{G}$-measurable function such that $\int_A X dP = \int_A E[X|\mathcal{G}] dP$ for all $A \in \mathcal{G}$.

- Law of iterated expectation

$$E[E[Y|X]] = E[Y]$$

  is a trivial fact from the definition of the conditional expectation by taking $A = \Omega$.

- Properties of conditional expectations

  1. $E[E[Y|X_1, X_2]|X_1] = E[Y|X_1]$

  2. $E[E[Y|X_1]|X_1, X_2] = E[Y|X_1]$

  3. $E[h(X)Y|X] = h(X)E[Y|X]$

3.2 Elementary Formulation

- conditional distribution

- conditional probability

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}.$$

- 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

4 Independence

X and Y are *independent* if $P[X \in A, Y \in B] = P(X \in A) P(Y \in B)$ 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 $Z_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)$. Then the probability $P(|Z_n| > \epsilon) \to 0$ as $n \to \infty$,