

An overview of the Probability Theory

1 Probability rules

Sample space: the set Ω of all possible outcomes in a random experiment.

Random events: $A, B \subset \Omega$, $A \cap B = \{A \text{ and } B\}$, $A \cup B = \{A \text{ or } B \text{ or both}\}$

Division rule: $P(A) = \frac{\text{no. favorable outcomes}}{\text{total no. equally likely outcomes}}$

Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Complementary event probability: $P(\bar{A}) = 1 - P(A)$, $\bar{A} = \{A \text{ has not occurred}\}$

Conditional probability of A given B has occurred: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Multiplication rule: $P(A \cap B) = P(A|B)P(B)$

$$\boxed{\text{Independent events: } P(A \cap B) = P(A)P(B)}$$

Law of Total Probability:

if $\{B_1, \dots, B_k\}$ is a partition of Ω , then $P(A) = \sum_{i=1}^k P(B_i)P(A|B_i)$

Bayes' Probability Law

prior probabilities $P(B_i)$ and posterior probabilities $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$

2 Random variables

Discrete random variable X : probability mass function (pmf) $f(x) = P(X = x)$

Continuous random variable X : probab. density function (pdf) $f(x) \approx \frac{P(x < X < x + \Delta)}{\Delta}$

Cumulative distribution function (cdf)

$$F(x) = P(X \leq x) = \sum_{y \leq x} f(y) \text{ or } = \int_{y \leq x} f(y) dy.$$

Mean (average or expected) value of X : $\mu = E(X) = \sum_x x f(x)$ or $\mu = \int x f(x) dx$

$$E(X + Y) = E(X) + E(Y), \quad E(c \cdot X) = c \cdot E(X)$$

Variance: $\sigma_X^2 = \text{Var}(X) = E((X - \mu)^2) = E(X^2) - \mu^2$

$$E(X^2) = \sum x^2 f(x) \text{ or } E(X^2) = \int x^2 f(x) dx.$$

Standard deviation (SD): $\sigma_X = \sqrt{\text{Var}(X)}$

$$\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X), \quad \sigma_{cX} = c \cdot \sigma_X$$

3 Special distributions

Discrete uniform distribution $X \sim dU(N)$: $f(k) = \frac{1}{N}$, $1 \leq k \leq N$; $E(X) = \frac{N+1}{2}$, $\text{Var}(X) = \frac{N^2-1}{12}$

Uniform distribution $X \sim U(a, b)$: $f(x) = \frac{1}{b-a}$, $a < x < b$, $E(X) = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$

Binomial distribution $X \sim \text{Bin}(n, p)$: $f(k) = \binom{n}{k} p^k q^{n-k}$, $0 \leq k \leq n$, $E(X) = np$, $\text{Var}(X) = npq$

Hypergeometric distribution $X \sim \text{Hg}(N, n, p)$: $E(X) = np$, $\text{Var}(X) = npq(1 - \frac{n-1}{N-1})$, finite population correction $\text{FPC} = 1 - \frac{n-1}{N-1}$

$$f(k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}, \quad \max(0, n - Nq) \leq k \leq \min(n, Np)$$

Geometric distribution $X \sim \text{Geom}(p)$: $f(k) = pq^{k-1}$, $k \geq 1$, $E(X) = \frac{1}{p}$, $\text{Var}(X) = \frac{q}{p^2}$

Exponential distribution $X \sim \text{Exp}(\lambda)$: $f(x) = \lambda e^{-\lambda x}$, $x > 0$, $E(X) = \sigma_X = \frac{1}{\lambda}$

Poisson distribution $X \sim \text{Pois}(\lambda)$: $f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \geq 0$, $E(X) = \text{Var}(X) = \lambda$

$\text{Bin}(n, p) \approx \text{Pois}(np)$ if $n \geq 100$, $p \leq 0.01$, $np \leq 20$

Standard normal distribution $Z \sim N(0, 1)$: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $E(Z) = 0$, $\text{Var}(Z) = 1$,

$P(Z < z) = \Phi(z)$, $P(|Z| > z) = 2(1 - \Phi(z))$

Normal distribution $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0, 1)$, $f(x) = \frac{1}{\sigma} \phi(\frac{x-\mu}{\sigma})$, $E(X) = \mu$, $\text{Var}(X) = \sigma^2$

Central Limit Theorem (CLT):

if X_1, \dots, X_n are IID, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, then approximately $\bar{X} \stackrel{a}{\sim} N(\mu, \frac{\sigma^2}{n})$ for large n

Normal approximations

$\text{Bin}(n, p) \approx N(np, npq)$, $np \geq 5$, $nq \geq 5$

$\text{Pois}(\lambda) \approx N(\lambda, \lambda)$, $\lambda \geq 5$

$\text{Hg}(N, n, p) \approx N(np, npq \frac{N-n}{N-1})$, $np \geq 5$, $nq \geq 5$

4 Joint distributions

Joint pmf (pdf) of X and Y : $f_{X,Y}(x, y)$

Marginal distribution of X : $f_X(x) = \sum_y f_{X,Y}(x, y)$ or $\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$

Independent r.v. X and Y : $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

Conditional distribution of $(Y|X = x)$: $f_{Y|X}(y|x) = f_{X,Y}(x, y) / f_X(x)$

Conditional expectation $E(Y|X)$ and $\text{Var}(Y|X)$

$E(E(Y|X)) = E(Y)$

$\text{Var}(Y) = \text{Var}(E(Y|X)) + E(\text{Var}(Y|X))$

Addition rule for variance

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

Covariance, a measure of association between X and Y

$\sigma_{xy} = \text{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y) = E(XY) - E(\mu_X)E(\mu_Y)$

Correlation coefficient $\rho = \frac{\sigma_{xy}}{\sigma_X \sigma_Y}$, $-1 < \rho < 1$. Independent r.v. are always uncorrelated: $\rho = 0$

Multinomial distribution $(X_1, \dots, X_r) \sim \text{Mn}(n; p_1, \dots, p_r)$

$P(X_1 = k_1, \dots, X_r = k_r) = \binom{n}{k_1, \dots, k_r} p_1^{k_1} \dots p_r^{k_r}$, $\text{Cov}(X_i, X_j) = -np_i p_j$

Bivariate normal distr $(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Marginal distributions: $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$. Conditional distribution of $(Y|X = x)$ is normal with mean $\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$, and variance $\sigma_Y^2(1 - \rho^2)$