PROBABILITY

1.1 Axioms of probability theory

Throughout this book, we use Kolmogorov's axioms.

Axiom 1: $P(E) \in \mathbb{R}$, ≥ 0

Axiom 2: $P(\text{all}) = 1$

Axiom 3: σ - additivity

$$p(\cup E_i) = \sum P(E_i)$$ if E_i's are disjoint.

A probability space $= \{\Omega, F, P\}$ sample space

$\Omega \rightarrow$ non-empty set $\quad \omega \in \Omega$ outcome

$F \rightarrow$ set of subsets of Ω called events

σ - algebra satisfying following properties

A1: $\Omega \in F$

A2: If $A \in F$, $A^c \in F$

A3: If $A, B \in F$, $A \cup B \in F$

$P \rightarrow$ probability measure on F satisfying Kolmogorov's properties

Borel sets → smallest topology covering all open balls in a given space.

Borel σ algebra B^n of subsets R^n is the smallest σ algebra containing all sets of the form $[a_1, b_1] \times [a_2, b_2] \times \dots [a_n, b_n]$.

(Continuity of Probability) Suppose $B_1, B_2, \dots$ is a sequence of events

(a) If $B_1 \subset B_2 \subset \dots$ then $\lim_{j \to \infty} P(B_j) = P(\bigcup_{i=1}^{\infty} B_i)$

(b) If $B_1 \supset B_2 \supset \dots$ then $\lim_{j \to \infty} P(B_j) = P(\bigcap_{i=1}^{\infty} B_i)$

1.2 Independence and conditional probability

Events $A_1, A_2, \dots A_k$ are defined to be independent if

$$P(A_1, A_2, \dots A_k) = P(A_1) P(A_2) \dots P(A_k)$$

A_1, A_2, A_3 are independent if:

$$P(A_1, A_2) = P(A_1) P(A_2)$$
$$P(A_2, A_3) = P(A_2) P(A_3)$$
$$P(A_1, A_3) = P(A_1) P(A_3)$$
$$P(A_1, A_2, A_3) = P(A_1) P(A_2) P(A_3)$$

A_1, A_2, A_3 are pairwise independent if

$$P(A_1, A_2) = P(A_1, A_2)$$
$$\vdots$$

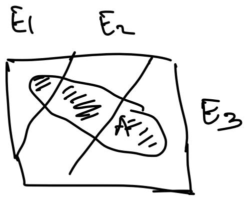
only for 2 events at a time

Conditional probability of A given B, where A,B are events and $P(B) \neq 0$.

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

Events $E_1, E_2, \ldots E_R$ are said to form a partition of Ω if the events are mutually exclusive and $E_1 \cup E_2 \cup \ldots E_R = \Omega$ and

$$P(E_1) + P(E_2) + \ldots P(E_R) = 1$$



$$P(A) = P(AE_1) + P(AE_2) + P(AE_3)$$
$$= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3)$$

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

The event $\{A_n \text{ infinitely often}\}$ is the set of $\omega \in \Omega$ such that $\omega \in A_n$, for infinitely many values of n.

Borel-Cantelli lemma: Let $(A_n : n \geq 1)$ be a sequence of events and let $p_n = P(A_n)$

(a) If $\sum_{n=1}^{\infty} p_n < \infty$, then $P\{A_n \text{ infinitely often}\} = 0$

(b) If $\sum_{n=1}^{\infty} p_n = \infty$, and $A_1, A_2, \ldots$ are mutually independent, then $P\{A_n \text{ infinitely often}\} = 1$

$$\{A_n \text{ infinitely often}\} = \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n$$

$$\{A_n \text{ eventually}\} = \bigcup_{k \geq 1} \bigcap_{n \geq k} E_n$$

1.3 Random Variables & their distributions

$$F_X(c) = P\{\omega : X \leq c\} = P\{X \leq c\}$$

$$P\{X < c\} = F_X(c - \varepsilon)$$

$$P\{X = c\} = F_X(c) - F_X(c - \varepsilon)$$

A function F is the CDF if:

A1: F is non-decreasing

A2: $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

A3: F is right continuous

Discrete random variable → probability mass function (pmf)

$$P_X(x) = P\{X = x\}$$

$$F_X(x) = \sum_{y \leq x} P_X(y)$$

Continuous random variable → cumulative density function

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

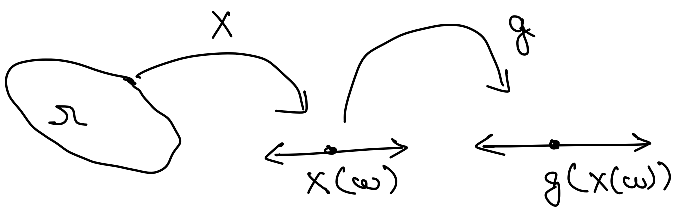
f_X → probability density function

$$f_X(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{x}^{x+\varepsilon} f_X(y)\,dy$$

probability measure P_X :

$$P_X((a,b]) = P\{X \in (a,b]\}$$

1.4 Functions of a Random Variable



① Examine possible ranges of X, Y

② Find CDF of Y : $F_Y(c) = P\{Y \le c\} = P\{g(X) \le c\}$

③ If F_Y has piecewise continuous derivative, and if pdf f_Y is desired, differentiate F_Y.

Let F be a function satisfying the properties of a CDF, and let U be uniformly distributed over the interval $[0,1]$. The problem is to find a function

g such that F is CDF of $g(U)$.

$$F_X(x) = P\{g(U) \le x\} = P\{U \le g^{-1}(x)\}$$
$$= g^{-1}(x)$$

Thus, $g(x) = F^{-1}(x)$

1.5 Expectation of a Random Variable

A random variable is called simple if there is a finite set $\{x_1, \dots, x_m\}$ such that $X(\omega) \in \{x_1, \dots, x_m\}$ for all ω. The expectation

$$E[X] = \sum_{i=1}^{m} x_i P\{X = x_i\}$$

For continuous random variable,

$$E[X] = \int_{\Omega} X(\omega) P(d\omega)$$

obviously, by definition : $E[X+Y] = E[X] + E[Y]$

Expectation has the following properties :

1. Linearity $\rightarrow$ $E[X+Y] = E[X] + E[Y]$

2. Preservation of order $\rightarrow$ If $P\{X \ge Y\} = 1$; then $E[X] \ge E[Y]$

3. If X has pdf f_X then
$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

4. If X has pmf P_X then
$$E[X] = \sum_{x>0} x P_X(x) + \sum_{x<0} x P_X(x)$$

5. Law of the unconscious statistician (LOTUS) $\rightarrow$ If g is Borel measurable,
$$E[g(X)] = \int_{\Omega} g(X(\omega)) P(d\omega)$$

If X is a continuous random variable

6. Integration by parts formula $\rightarrow$
$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$
Integration by parts
$$\int uv\,dx = u \int v\,dx - \int u' \left(\int v\,dx\right) dx$$
$$= x - \int_{-\infty}^{\infty} F(x) dx$$

$$\text{Var}[X] = E\left[(X - E[X])^2\right] = E[X^2] - E[X]^2$$

- Markov Inequality : $Y \geq 0$, $c > 0$
$$P\{Y \geq c\} \leq \frac{E[Y]}{c}$$
$\hookrightarrow \int_{c}^{\infty} f(x) dx$ $\quad \hookrightarrow \int_{0}^{\infty} y f(y) dy$

- Chebychev Inequality : If X is a random variable with finite mean μ and variance σ^2, then for any $d > 0$,
$$P\{|X - \mu| \geq d\} \leq \frac{\sigma^2}{d^2}$$

The characteristic function ϕ_X of a random variable X is defined by

$$\phi_X(u) = E[e^{juX}]$$

$u \rightarrow$ real
$j \rightarrow$ imaginary $\sqrt{-1}$

Two random variables have the same probability distribution if and only if they have the same characteristic function.

If $E[X^k]$ exists and finite, then

$$\underbrace{\phi_X^{(k)}}_{\text{derivative}}(0) = j^k E[X^k]$$

upto order k

z-transform of the pmf

$$\psi_X(z) = E[z^X] = \sum_{k=0}^{\infty} z^k p_X(k)$$

analogous to characteristic function

1.6 Frequently Used Distributions

- Bernoulli :

$$\begin{cases} p & , i=1 \\ 1-p & , i=0 \\ 0 & , \text{otherwise} \end{cases}$$

mean $= p$
var $= p(1-p)$

z-transform $= 1-p+pz$

- Poisson: $p(i) = \dfrac{\lambda^i e^{-\lambda}}{i!}$ mean $= \lambda$, var $= \lambda$

  z-transform $= \exp(\lambda(z-1))$

- Binomial: $\binom{n}{i} p^i (1-p)^{n-i}$ mean $= np$, var $= np(1-p)$

  z-transform $= (1-p+pz)^n$

- Geometric: $p(i) = (1-p)^{i-1} p$ mean $= 1/p$, var $= \dfrac{1-p}{p^2}$

  z-transform $= \dfrac{pz}{1-z+pz}$

  memoryless property
  $\rightarrow p(x > i+j \mid x > i) = p(x > j)$

- Exponential: $f(x) = \lambda e^{-\lambda x}$ mean $= 1/\lambda$, var $= 1/\lambda^2$

  $\phi = \dfrac{\lambda}{\lambda - j\mu}$

  memoryless property

- Uniform: $f(x) = \dfrac{1}{b-a}$ mean $= a+b/2$, var $= (b-a)^2/12$

  $\phi = \dfrac{e^{j\mu b} - e^{j\mu a}}{j\mu(b-a)}$

- Gamma: $f(x) = \dfrac{\alpha^n x^{n-1} e^{-\alpha x}}{\Gamma(n)}$, $x \geq 0$ mean $= n/\alpha$, var $= n/\alpha^2$

  where $\Gamma(n) = \int_0^\infty s^{n-1} e^{-s}\, ds$

$$\phi = \left(\frac{\alpha}{\alpha - j\mu}\right)^{n}$$

- Rayleigh : $f(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right)$ mean $= \sigma\sqrt{\pi/2}$

  $CDF = 1 - \exp\left(\frac{-r^2}{2\sigma^2}\right)$ var $= \sigma^2\left(2 - \frac{\pi}{2}\right)$

1.7 Failure Rate Functions

Eventually a system or component of a system would fail.

Let T be the random variable that denotes the lifetime of this item with pdf f_T.

The failure rate function h:

$$h(t) \triangleq \lim_{\varepsilon \to 0} \frac{P(t < T \le t+\varepsilon \mid T > t)}{\varepsilon}$$

$$= \lim_{\varepsilon \to 0} \frac{F(t+\varepsilon) - F(t)}{[1 - F(t)]\,\varepsilon}$$

$$= \frac{f_T(t)}{(1 - F(t))}$$

$f_T(t) \to$ pdf

$F(t) \to$ cdf

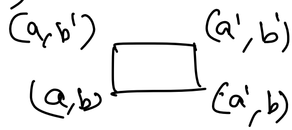
Here, $F = 1 - e^{\int_0^t h(s)\,ds}$

1.8 Jointly Distributed random variables

Let $X_1, X_2, \ldots X_m$ be random variables on a single probability space $(\Omega, \mathcal{F}, P)$. The joint cumulative distribution function (CDF) is the function on R^m defined by:

$$F_{X_1, X_2, \ldots X_m}(x_1, \ldots, x_m) = P(X_1 \le x_1, X_2 \le x_2, \ldots)$$

If R is the rectangular region R $(a, b] \times (a', b']$ in the plane, then



$$P\{(X_1, X_2) \in R\} = F(a', b') - F(a, b') - F(a', b) + F(a, b)$$

By countable additivity axiom of probability,

$$F_{X_1, X_2}(x_1, \infty) = F_{X_1}(x_1)$$

$$f_{X_1}(u_1) = \int_{-\infty}^{\infty} f_{X_1 X_2}(u_1, u_2)\, du_2$$

↳ marginal pdf

The joint characteristic function of $X_1, \ldots, X_m$ is the function on R^m defined by

$$\Phi_{X_1, X_2 \ldots}(u_1, u_2, \ldots u_m) = E\left[e^{j(X_1 u_1 + X_2 u_2 + \cdots)}\right]$$

Random Variables are said to be independent if for any borel subsets, the events are independent.

$$F_{X_1, X_2 \dots}(x_1, \dots x_m) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_m}(x_m)$$

1.9 Conditional Densities

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

If $f_Y(y)$ is fixed, conditional pdf is a function of pdf.

conditional expectation / mean :

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$$

1.10 Correlation and Covariance

$$\text{correlation} = E[XY]$$

$$\text{covariance} = E[(X-E[X])(Y-E[Y])] = \mathrm{Cov}(X,Y)$$

$$\text{correlation coefficient} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

A fundamental inequality is Schwarz's inequality

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$$

The Schwarz's inequality is equivalent to the L^2 triangle inequality for random variables:

$$E[(X+Y)^2]^{1/2} \leq E[X^2]^{1/2} + E[Y^2]^{1/2}$$

$$Cov(X,Y) \leq \sqrt{Var(X)\,Var(Y)}$$

↳ Applying Schwarz's inequality on $(X-E[X])$ and $(Y-E[Y])$

$$Cov(X,Y) = E[X(Y-E[Y])] = E[(X-E[X])Y]$$
$$= E[XY] - E[X]E[Y]$$

If X or Y has mean 0, $Cov(X,Y)=0$

If X, Y are independent, they are uncorrelated.

$$Cov(X+Y, U+V) = Cov(X,U) + Cov(X,V) + Cov(Y,U) + Cov(Y,V)$$

$$Cov(aX+b, cY+d) = ac\,Cov(X,Y)$$

Suppose $S_m = X_1 + X_2 + \dots X_m$

$$Var(S_m) = Cov(S_m, S_m)$$
$$= \sum_i Var(X_i) + \sum_{i,j;\, i\neq j} Cov(X_i, X_j)$$

1.11 Transformation of random vectors

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$$

Let g be a one-to-one mapping from R^m to R^m.

$$y = g(x)$$

Suppose that the Jacobian matrix of derivatives $\frac{\partial y}{\partial x}(x)$ is continuous in x and non singular for all x. By the inverse function theorem of vector calculus, it follows that the Jacobian matrix of the inverse mapping

$$\frac{\partial x}{\partial y}(y) = \left(\frac{\partial y}{\partial x}(x) \right)^{-1}$$

$$f_Y(y) = \frac{f_X(x)}{\left| \frac{\partial y}{\partial x}(x) \right|}$$