Chapters 1-6

Part 1

Probability

Ω = sample space: set of all possible outcomes of an experiment

$\mathcal{F}$ = σ-algebra: set of subsets of Ω, the events

P = probability measure: a function from subsets of Ω to $\mathbb{R}$

Axioms
- $P(\Omega) = 1$
- If $A \subset \Omega \Rightarrow P(A) \geq 0$ (nonnegative)
- If A_1 and A_2 are disjoint ($A_1 \cap A_2 = \emptyset$) $\Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$

more generally, if $A_1, A_2, \ldots, A_n$, are mutually disjoint

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Properties
- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- If $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability: Let A and B be two events with $P(B) \neq 0$.
The conditional probability of A given B is defined as

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

Multiplication law: Let A and B be two events with $P(B) \neq 0$
Then $P(A \cap B) = P(A|B)\,P(B)$

Law of total probability: Let $B_1, B_2, \ldots, B_n$ be n events such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$ with $P(B_i) > 0$ for all i (a partition). Then for any event A

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\,P(B_i)$$

Baye's rule: Let A and $B_1, \dots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^n B_i = \Omega$ and $P(B_i) > 0$ for all i. Then

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i) P(B_i)}$$

Independence: A and B are said to be independent events if $P(A \cap B) = P(A)P(B)$ (not the only way to characterize independence)

Random variable: is a function from Ω to $\mathbb{R}$

Discrete: can take only a finite or at most a countably infinite number of values

Continuous: can take values on a continuum

Probability mass function: ~~[illegible]~~ If X is a discrete random variable that can take the values $x_1, x_2, \dots$, then the pmf is a function p such that $p(x_i) = P(X = x_i)$ and $\sum_i p(x_i) = 1$

Probability density function: the continuous version of the pmf. Is a function $f(x)$ such that $f(x) \geq 0$, is ~~piecew~~ piecewise continuous and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. If X is a continuous random variable, then ~~[illegible]~~ $P(a < X < b) = \int_a^b f(x)\,dx$

Common discrete random variables

- Bernoulli: takes only the values 0 and 1, result of an experiment (succes or failure

$$p(x) = \begin{cases} p^x (1-p)^{1-x} & \text{if } x=0 \text{ or } x=1 \\ 0 & \text{otherwise} \end{cases}$$

- Binomial: total number of "successes" when n independent trial are performed

takes values $0, 1, 2, \dots, n$

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

- Geometric: Number of trials (independent) up to (inclusive) the first success, thus X takes values $1, 2, 3, \dots$

$$p(x) = \begin{cases} (1-p)^{x-1} p & , x = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

- Negative binomial: Number of trials (independent) until there are r successes, then X takes values $1, 2, 3, \dots$

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

- Hypergeometric: Number of successes in a sequence of n draws from a finite population without replacement (binomial draws ~~...~~ with replacement).
~~n: size of population, m: size of the sample~~
~~$p(x) = \binom{r}{x}$~~ Suppose an urn with n balls of which r are black and $n-r$ are white. If X denotes the number of black balls when taking m ball without replacement then

$$p(x) = \frac{\binom{r}{x}\binom{n-r}{m-x}}{\binom{n}{m}}$$

- Poisson: number of events occurring in a fixed period of time when they occur with a know average rate λ and independently of the of the time since the last event. Is the limit of a binomial distribution when the number of trials n approaches to infinity and the probability of success p approaches to zero in such a way that $np = \lambda$.

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{setting } np = \lambda \longrightarrow p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Common continuous random variables

- Uniform $[a,b]$: $U[a,b]$ assigns equal probability to any number in $[a,b]$

$$f(x) = \begin{cases} 1/b-a & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$Exp(\lambda)$

- Exponential: models lifetimes or waiting times

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$Gamma(\alpha, \lambda)$ α: shape, λ: scale $\quad Gamma(1,\lambda) = Exp(\lambda)$

- Gamma: a more flexible density to model waiting times

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad \Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du \quad \alpha > 0$$

$N(\mu, \sigma^2)$

- Normal: distribution of a sum of a large number of independent random variables

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

- Beta: a more flexible density to model variables on the interval $[0,1]$

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-u)^{b-1} \qquad 0 \le u \le 1$$

$Beta(a,b)$ $\quad Beta(1,1) = U[0,1]$

$\boxed{\text{Part 2}}$

Cumulative distribution function: $F(x) := P(X \le x) = \int_{-\infty}^{\infty} f(x)\,dx$

Expected value: If X is a discrete r.v. with p.m.f. $p(x)$, the expected value of X, denoted by $E[X]$ is

$E[X] = \sum_i x_i p(x_i)$ / $E[g(X)] = \sum_i g(x_i) p(x_i)$

If X is continuous r.v. with p.d.f. $f(x)$, then

$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$ / $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$

$E[a+bX] = a + bE[X]$

~~Standard~~ Variance: If X is a r.v. with expected value $E[X]$, the variance, $Var(X)$ is

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

That can be computed as $Var(X) = E(X^2) - [E(X)]^2$

Standard deviation: $std(X) = \sqrt{var(X)}$

Moment generating function: The mgf of a random variable X is $M(t) = E(e^{tX})$. In the discrete case

$M(t) = \sum_x e^{tx} p(x)$ and in the continuous case

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

Properties: • If $M(t)$ exists in an open interval containing ~~an~~ zero

then $M^{(r)}(0) = E(X^r)$

• If " " " " " " " "

then it uniquely determines the probability distribution.

(important property to prove the Central Limit Theorem)

- Chi-square distribution: If $Z \sim N(0,1)$ then $U = Z^2 \sim \chi^2_{(1)}$

$$f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2}$$

If $U_1, U_2, \ldots \sim \chi^2_{(1)}$ i.i.v $\Rightarrow U = \sum_{i=1}^{n} U_i \sim \chi^2_{(n)}$

$\text{Gamma}(\frac{1}{2}, \frac{1}{2}) = \chi^2_{(1)}$

- t-student: If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ and they are independent then $\frac{Z}{\sqrt{U/n}} \sim t_{(n)}$

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}$$

- F: Let $U \sim \chi^2_{(m)}$ and $V \sim \chi^2_{(n)}$ be independent, then

$$W = \frac{U/m}{V/n} \sim F_{(m,n)}$$

$$F(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n} w\right)^{-\frac{(m+n)}{2}}$$

Joint distribution: Determines the joint behaviour of two or more random variables. We define the ~~joint distribution~~ joint cumulative distribution function of the two random variables X and Y as $F(x,y) = P(X \leq x, Y \leq y)$, regardless if they are continuous or discrete.

Discrete r.v. Suppose X and Y are discrete r.v. on the same sample space that take values $x_1, x_2, \dots$ and $y_1, y_2, \dots$ respectively. Their joint frequency function is $p(x_i, y_j) = P(X = x_i, Y = y_j)$

The marginal frequency function is defined, for the r.v. X as $p_X(x) = \sum_i p(x, y_i)$ and for Y as $p_Y(y) = \sum_j p(x_j, y)$

Continuous r.v. Suppose X and Y are continuous r.v. with a joint cdf $F(x,y)$. Their joint density function is a piecewise continuous function of two variables, $f(x,y)$, is nonnegative and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy\, dx = 1$.

For any reasonable two-dimensional set A

$$P((X,Y) \in A) = \iint_A f(x,y)\, dy\, dx$$

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$$

The marginal cdf of X is $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,y)\, dy\, du$

The marginal density of X is $f_X(x) = F_X'(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

Independency. The r.v.'s $X_1, X_2, \dots X_n$ are independent if their joint cdf has the property that $F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n)$
$\forall\, x_1, x_2, \dots, x_n$

Covariance: If X and Y are jointly distributed r.v. with expectations μ_X and μ_Y respectively, the covariance of X and Y is

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

It can be shown that $Cov(X,Y) = E(XY) - E(X)E(Y)$

If X and Y are independent $Cov(X,Y) = 0$ (the converse is not true)

Properties: $Cov(aW + bX, cY + dZ) = ac\,Cov(W,Y) + bc\,Cov(X,Y) + ad\,Cov(W,Z) + bd\,Cov(X,Z)$

In general, if $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$ then

$$Cov(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \, Cov(X_i, Y_j)$$

And, since $Var(X) = Cov(X,X)$ we get that

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$. More generally

$$Var\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_j \, Cov(X_i, X_j)$$

If the X_i are independent $var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$

Correlation coefficient. If X and Y are jointly distributed r.v.'s, the variances and covariances of both X and Y exist and the variances are not zero, then the correlation of X and Y, denoted by ρ is

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

Properties: $-1 \le \rho \le 1$, is dimensionless and $\rho = \pm 1 \Leftrightarrow P(Y = a + bX) = 1$ for some constants a, b.

Law of Large Numbers. Let $X_1, X_2, \dots X_i, \dots$ be a sequence of independent r.v. with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ $\forall i$. Let $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

Then for any $\varepsilon > 0$ $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ as $n \to \infty$

If a sequence of random variables $\{X_n\}$ is such that $P(|Z_n - \alpha| > \varepsilon) \underset{n \to \infty}{\longrightarrow} 0$ $\forall \varepsilon > 0$ then Z_n is said to converge in probability to α.

~~In~~

Let $X_1, X_2, \dots$ be a sequence of r.v. with cdf's $F_1, F_2,$ and let X be a r.v. with cdf F. We say that $X_n \xrightarrow{d} X$ if $\lim_{n \to \infty} F_n(x) = F(x)$

Notation: $X \overset{d}{\approx} F$

② Central Limit Theorem. Let $X_1, X_2, \dots$ be a sequence of independent r.v. with mean μ and variance σ^2 and distribution function F. Let $S_n = \sum_{i=1}^{n} X_i$, then $\lim_{n \to \infty} \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right) = \Phi(x)$ $-\infty < x < \infty$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

① Continuity theorem. Let F_n be a sequence of cumulative distribution functions with corresponding moment-generating function M_n and let F be a cumulative distribution function with moment generating function M. If $M_n(t) \to M(t)$ $\forall t$ in an open interval containing zero then $F_n(x) \to F(x)$ at all continuity points of F.

Normal distribution was proposed by Gauss as a model for measurement errors. Also used to model peoples' height, weight, IQ score, velocity of gas molecules, noise. Standard Normal $:= N(0,1)$ If $X \sim N(\mu, \sigma^2) \Rightarrow (X-\mu)/\sigma \sim N(0,1)$

Normal aproximations to other distributions

- If $X \sim Bin(n,p) \Rightarrow X = \sum_{i=1}^{n} Y_n$ where $Y_i \sim Ber(p)$ $\quad \mu = p, \ \sigma^2 = p(1-p)$

by the CLT $\dfrac{X - np}{\sqrt{(np)(1-p)}} \overset{d}{\approx} N(0,1)$

X plays the role of S_n
np " " " " μ
$\sqrt{np(1-p)}$ " " " " $\sigma\sqrt{n}$

$\Rightarrow \quad X \overset{d}{\approx} N(np, np(1-p))$

- If $X \sim Poisson(\lambda) \Rightarrow X = \sum_{i=1}^{n} Y_i$ where $Y_i \sim Poisson(\frac{\lambda}{n})$ $\quad \mu = \lambda/n, \ \sigma^2 = \lambda/n$

by the CLT $\dfrac{X - n\lambda/n}{\sqrt{n\lambda/n}} \overset{d}{\approx} N(0,1)$

$\Rightarrow \quad X \overset{d}{\approx} N(\lambda, \lambda)$

- If $X \sim Gamma(n, \lambda) \Rightarrow X = \sum_{i=1}^{n} Y_k$ where $Y_i \sim Gamma(1, \lambda)$ $\quad \mu = 1 \cdot \lambda, \ \sigma^2 = 1\lambda^2$

by the CLT $\dfrac{X - n\lambda}{\sqrt{n\lambda^2}} \overset{d}{\approx} N(0,1)$

$\Rightarrow \quad X \overset{d}{\approx} N(n\lambda, n\lambda^2)$