Week 1

Basic probability

1. A and B **mutually exclusive** or disjoint iff

$$A \cap B = \emptyset$$

Remark. If disjoint then cannot be independent iff neither A nor B has 0 probability. Also,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

Therefore

$$B = (A \cap B) \cup (A^c \cap B)$$

and

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

2. A and B are independent if occurence of one event does not affect the other

Remark. If A and B are independent then

$$P(A|B) = P(A) \iff P(A \cap B) = P(A) \cdot P(B)$$

which can be used to check if A and B are independent. More generally

$$P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

3. Probability as relative frequency

$$P(A) = \frac{n(A)}{n(\Omega)}$$

works only if each outcome is equally likely. Therefore can counting method is important in detemrining probability in this context

Week 2

Counting Methods

1. **Fundamental Principle of Counting** *With replacement and order matters* Let $n_1, n_2, \ldots, n_m$ be possible outcome in m independent stages, then total number of possible outcome is

$$\prod_{i=1}^{m} n_i$$

2. **Permutation** *without replacement and order matters* The number of ways to order k distinct items from a group of n distinct items is

$$_nP_k = \frac{n!}{(n-k)!}$$

To adopt for cases where n items are not distinct just divide by number of repeats for al repeats $r!$

3. **Combination** *without replacement where order does not matter* The number of unique combinations of k items from a group of n distinct items is

$$\binom{n}{k} =_n C_k = \frac{n!}{(n-k)! \cdot k!}$$

Compare to permutation, combination divide by an additioal $k!$ to account for different orderings of k items

Remark. Selecting 4 face cards + 1 numeric card

$$\binom{12}{4}\binom{36}{1}$$

Selecting triple of one rank and a double of another

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}$$

Select 3 products with at least 1 defective from a bin containing 4 defectives and 8 non-defectives

$$\binom{4}{1}\binom{8}{2} + \binom{4}{2}\binom{8}{1} + \binom{4}{3}$$

4. **Combination with Replacement** *with replacement and order does not matter* To count the number of cases of selecting/arranging k indistinguishable objects amongst n distinguishable bins where replacement is allowed (i.e. select a bin multiple times) Use stick and stone method by ordering k stones out of $(n + k - 1)$ sticks and stones.

$$_{n+k-1}C_k = \binom{n+k-1}{k}$$

Definition 0.1. Conditional Probability Probability of event A occurring given the condition that event B has occurred

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

Theorem 0.1. *Law of Total Probability If $B_1, B_2, \ldots, B_k$ is a collection of mutually exclusive and exhaustive events, then for any A,*

$$P(A) = \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)$$

Theorem 0.2. *Bayes' Rule Together with Conditional probability and Law of total probability.. Let B_i be partition of sample space and let A be an event in Σ. Then,*

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{i=1}^{k} P(A|B_i) \cdot P(B_i)}$$

Week 3

Intro to Discrete Random Variable

Definition 0.2. A **Random Variable** is a real-valued function that assigns a numerical value to each event in sample space Ω

$$X : \Omega \to \mathbb{R}$$

Definition 0.3. Discrete Random Variable is a random variable X that take on a finite number or a countably infinite number of possible values x. with **Probability Mass Function (PMF)** satisfying

1. $0 \le P(X = x) \le 1$

2. $\displaystyle\sum_{x \in X} P(X = x) = 1$

and the **Cumulative Distribution Function (CDF)** $F(x)$ of a discrete RV having probability mass function $P(x)$ is the cumulative probability up to and including $X = x$

$$F(b) = P(X \le b) = \sum_{x \in \{x \le b\}} P(x)$$

which satisfies,

1. The graph of CDF is a non-decreasing step-function (i.e. $\forall a < b : F(a) \le F(b)$)

2. The graph of CDF is right continuous (i.e. $\displaystyle\lim_{x \to c^+} F(x) = F(c)$)

3. $\displaystyle\lim_{x \to \infty} F(x) = 1$ and $\displaystyle\lim_{x \to -\infty} F(x) = 0$

Definition 0.4. Expected Value of a discrete random variable with PMF $f(x)$ is the long-run value

$$E(X) = \sum_{x \in X} x \cdot f(x)$$

The expected value of any real-valued function of random variable $g(X)$ is given by

$$E[g(X)] = \sum_{x \in X} g(x) \cdot f(x)$$

The expected value has properties

1. $E[c] = c$

2. $E[aX + b] = a \cdot E[X] + b$

Definition 0.5. The **Variance** of X is the averaged squared variability of X from the expected value

$$V(X) = E[(X - \mu)^2] = \sum_{x \in X} (x - \mu)^2 \cdot f(x) = E[X^2] - \mu^2$$

which has properties

1. $V(c) = 0$

2. $V(aX + b) = a^2 \cdot V(X)$

The **Standard Deviation** is

$$SD(X) = \sigma = \sqrt{Var(X)}$$

Theorem 0.3. *Chebyshev's Inequality* Let X *be a random variable with mean* μ *and variance* μ^2. *Then for any positive* k,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Bernoulli trial is a random experiment consisting of exactly one trail with 2 possible outcomes, specifically $X = 1$ if success and $X = 0$ if failure.

$$p(X = x) = p^x \cdot (1 - p)^{1-x}$$

with

$$E[X] = p \text{ and } V(X) = p(1 - p)$$

Binomial Distribution Let X be the random variable representing number of successes arising from n trials. Then X has binomial distribution with n trials and probability of success p, or $X \sim Bin(n, p)$ with PMF,

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

with

$$E(X) = np \text{ and } V(X) = np(1 - p)$$

Remark. A binomial distribution must have fixed n of identical trials where each is Bernoulli. Probability of success p is constant from trial to trial.

Week 4

Geometric Distribution Let X be random variable representing number of failures before the first success with probability of success p fixed for each trial. Then

$$P(X = x) = (1 - p)^x \cdot p$$

with

$$E[X] = \frac{1 - p}{p} \text{ and } V(x) = \frac{1 - p}{p^2}$$

Remark. Geometric Distribution is memoryless, that is probability of observing another k failure given j consecutive failures is the probabiilty of observing k failures,

$$P(X \geq j + k | X \geq j) = P(X \geq k)$$

Negative Binomial Distribution Let X be the number of failures before r^{th} success and p the probability of success is fixed for each trial

$$P(X = x) = \binom{x + r - 1}{r - 1} \cdot p^r \cdot (1 - p)^x$$

with

$$E[X] = \frac{r(1 - p)}{p} \text{ and } V(X) = \frac{r(1 - p)}{p^2}$$

Hypergeometric Distribution A pool of N objects can be partitioned into 2 groups with k objects of type A and $N - k$ objects of type B. Then a random sample of size n without replacement form this pool. Let X denote random variable for number of objects of type A selected

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

with

$$E[X] = n \cdot \frac{k}{N} \text{ and } V(X) = n \cdot \frac{k}{N}(1 - \frac{k}{N})(\frac{N-n}{N-1})$$

Week 6

Poisson Distribution A discrete random variable X denoting the nmber of events of interest in an interval, with λ the mean number of occurrences per unit interval, can be modeled by a Poisson distribution with probability mass function

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

with

$$E[X] = \lambda \text{ and } V(X) = \lambda$$

Remark. A poisson distribution is valid if no more than one event can occur simultaneously, events occuring in different intervals are independent and expected number of events in each time interval is constant

Also Poisson can be used to approximate binomial distribution

Continuous Random Variable

The **Probability Density Function (PDF)** of a continous random variable X is a function $f(x)$ with following properties

1. $f(x) \geq 0$ for all x

2. $\int_{-\infty}^{\infty} f(x)dx = 1$

3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

The **Cumulative Distribution Function (CDF)** of a continuous random variable X is the function $F(x)$ such that

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

Note that

1. $F'(x) = f(x)$

2. $P(a \leq X \leq b) = F(b) - F(a)$

3. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$

The kth **percentile** x_k with random variable X having PDF $f(x)$ is

$$\frac{k}{100} = P(X \le x_k) = \int_{-\infty}^{x_k} f(x)dx = F(x_k)$$

The **Expected Value** of continuous random variable X with PDF $f(x)$ is given by

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

and for any real-valued function $g(x)$ of X

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

The **Variance** is given by

$$V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (X - \mu)^2 f(x)dx = E[X^2] - \mu^2$$

having properties

1. $E(aX + bY) = aE(X) + bE(Y)$

2. $E(XY) = E(X) \cdot E(Y)$ only if X and Y independent

3. $V(aX + bY) = a^2 \cdot V(X) + b^2 \cdot V(Y)$ only if X and Y are independent

A continuous random variable following **Uniform Distribution** on the interval $a \le X \le b$ if it has probability density function

$$f(x) = \begin{cases} \dfrac{1}{b - a}, & a \le x \le b \\ 0, & elsewhere \end{cases}$$

with

$$E(X) = \frac{b + a}{2} \text{ and } V(X) = \frac{(b - a)^2}{12}$$

An **Exponential Distribution** with mean $\theta > 0$ for a continous random variable X has probability density function

$$f(x) = \begin{cases} \dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x \ge 0 \\ 0, & elsewhere \end{cases}$$

with

$$E(X) = \theta \text{ and } V(X) = \theta^2$$

7

Remark. Note exponential distribution is memoryless. the probability that $X \sim Exp(\theta)$ is greater than $b + a$ if known X is greater than a is the same as probability that X is greater than b

$$P(X > a + b | X > a) = P(X > b)$$

Also Assume $X \sim Poisson(\lambda t)$ where $E(X) = \lambda t$. The waiting time from occurrence of any one event till occurrence of next is modeled by $Exp(\theta = \dfrac{1}{\lambda t})$

An **Gamma Distribution** with shape α and scale β for a continuous random variable X has probability density function

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x \geq 0 \\ 0, & elsewhere \end{cases}$$

Week 8

Moment Generating Functions The moments of a random variable are the expected values of powers itself, i.e. $E(X^k)$ is called the kth moment of X. The moment generating function (MGF) is defined as

$$M_X(t) = E[e^{tX}] = \begin{cases} \displaystyle\int_{x \in X} e^{tx} f(x)dx, & X \text{ continuous} \\ \displaystyle\sum_{x \in X} e^{tx} f(x), & X \text{ discrete} \end{cases}$$

So then if $M(t)$ exists and differentiable in neighbourhood of $t = 0$ then

$$E[X^k] = M^{(k)}(0)$$

Remark. Note MGF are unique. That is two random variable with same MGF also have the same probability distribution Also

1. $M_Y(t) = e^{tb} M_X(at)$ when $Y = aX + b$

2. $M_{X+Y}(t) = M_X(t)M_Y(t)$ only if X, Y are independent

Multivariate Distribution

Joint Probability Mass Function (PMF) of two discrete random variable X, Y is defined,

$$f(x, y) = P(X = x, Y = y)$$

satisfying

1. $0 \leq f(x, y) = P(X = x, Y = y) \leq 1$

2. $\displaystyle\sum_{x \in X} \sum_{y \in Y} f(x, y) = 1$