Chapter 1: Combinatorial Analysis

The Basic Principle of Counting. Suppose that two experiments are to be performed. If

- experiment 1 can result in any one of m possible outcomes; and
- experiment 2 can result in any one of n possible outcomes;

then together there are mn possible outcomes of the two experiments.

The Generalized Basic Principle of Counting. Suppose that r experiments are to be performed. If

- experiment 1 can result in any one of n_1 possible outcomes;
- experiment 2 can result in any one of n_2 possible outcomes;
- $\vdots$
- experiment r can result in any one of n_r possible outcomes;

then together there are $n_1 n_2 \cdots n_r$ possible outcomes of the r experiments.

Permutations of n Distinct Objects. Suppose there are n distinct objects, then the total number of different arrangements is

$$n(n-1)(n-2)\cdots(3)(2)(1) = n!$$

with the convention that $0! = 1$

Permutation of n Non-distinct Objects. For n object of which n_1 are alike, n_2 are alike, $\cdots$, n_r are alike, there are

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}$$

different permutations of the n objects.

n People Sitting in a Circle. Generally, for n people sitting in a circle, there are

$$\frac{n!}{n} = (n-1)!$$

possible arrangements.

Number of Arrangement for Making Necklaces. Given n different pearls string in a necklace, the number of ways of stringing the pearls is

$$\frac{(n-1)!}{2}$$

Combinations. Generally, if there are n distinct objects, of which we choose a group of r items, then the number of possible groups is given by

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Combination Remarks.

(i) For $r = 0, 1, 2, \cdots, n$,

$$\binom{n}{r} = \binom{n}{n-r}$$

(ii)

$$\binom{n}{0} = \binom{n}{n} = 1$$

(iii) When n is a non-negative integer, and $r < 0$ or $r > n$, take

$$\binom{n}{r} = 0$$

Useful Combinatorial Identities. For $1 \le r \le n$,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

* Proof: Consider the cases where the first object (i) is chosen, (ii) is not chosen.

The Binomial Theorem. Let n be a non-negative integer, then

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Corollaries to the Binomial Theorem. With appropriate substitution of x and y, the following equations can be proven.

1.
$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

2.
$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$$

3.
$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots$$

Multinomial Coefficients. The number of division to divide n objects into r distinct groups of size $n_1, n_2, \cdots, n_r$ such that $\sum_{i=1}^{r} n_i = n$ is given by

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r}$$

The above expression can be easily shown to be

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}$$

which we denote as

$$\binom{n}{n_1, n_2, \cdots, n_r}$$

The Multinomial Theorem. Let n be a nonnegative integer, then

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \cdots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

where $\binom{n}{n_1, n_2, \cdots, n_r}$ is the multinomial coefficient.

The Number of Integer Solutions of Equations. Consider the following integer equation

$$x_1 + x_2 + \cdots + x_r = n$$

The number of ***positive*** integer solutions is $\binom{n-1}{r-1}$

The number of ***non-negative*** integer solutions is $\binom{n+r-1}{r-1}$

Chapter 2: Axioms of Probability

Sample Space. The *sample space* is the set of all possible outcomes of an experiment, usually denoted by S.

Event. Any subset of the sample space is an event.

Operation of Sets.

1. Commutative laws.
   (i) $EF = FE$
   (ii) $E \cup F = F \cup E$
2. Associative laws
   (i) $(EF)G = E(FG)$
   (ii) $(E \cup F) \cup G = E \cup (F \cup G)$
3. Distributive laws
   (i) $(E \cup F)G = EG \cup FG$
   (ii) $EF \cup G = (E \cup G)(F \cup G)$
4. DeMorgan's law
   (i) $(\bigcup_{i=1}^{n} E_i)^c = \bigcap_{i=1}^{n} E_i^c$
   (ii) $(\bigcap_{i=1}^{n} E_i)^c = \bigcup_{i=1}^{n} E_i^c$

Definition of Probability

1. Classical Approach

   Assume all the sample points are equally likely to occur.

$$P(E) = \frac{|E|}{|S|}$$

   where $|E|$ is the number of sample points in event E and $|S|$ is the number of sample points in S.

2. Relative Frequency Approach

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

   where $n(E)$ is the number of time in n repetition of the experiment that E occurs.

3. Subjective Approach
   Probability is considered as a measure of belief.

Axioms of Probability. Probability, denoted by P, is a function of the collection of events satisfying

(i) For any event E,
$$0 \le P(E) \le 1$$
(ii) Let S be the sample space, then
$$P(S) = 1$$
(iii) For any sequence of *mutually exclusive* events $E_1, E_2, \cdots$ (that is, $E_i \cap E_j = \emptyset$ when $i \ne j$),
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

In other words, for any sequence of mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

Properties of Probability.

(i)
$$P(\emptyset) = 0$$

(ii) For any finite sequence of mutually exclusive events $E_1, E_2, \cdots, E_n$,
$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

(iii) Let E be an event, then
$$P(E^c) = 1 - P(E)$$

(iv) If $A \subseteq B$, then
$$P(A) \leq P(B)$$

(v) Let A and B be any two events, then
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Inclusion-Exclusion Principle. Let $E_1, E_2, \cdots, E_n$ be any events, then

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(E_{i_1} \cap E_{i_2}) + \cdots + (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} P(E_{i_1} \cap \cdots \cap E_{i_r}) + \cdots + (-1)^{n+1} P(E_1 \cap \cdots \cap E_n)$$

Example of Inclusion-Exclusion Principle. Suppose $n = 4$.

- $\sum_{1 \leq i_1 < i_2 \leq 4} P(E_{i_1} \cap E_{i_2}) = P(E_1 \cap E_2) + P(E_1 \cap E_3) + P(E_1 \cap E_4) + P(E_2 \cap E_3) + P(E_2 + E_4) + P(E_3 + E_4)$
- $\sum_{1 \leq i_1 < i_2 < i_3 \leq 4} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) = P(E_1 \cap E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_4) + P(E_1 \cap E_3 \cap E_4) + P(E_2 \cap E_3 \cap E_4)$
- $\sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq 4} P(E_{i_1} \cap E_{i_2} \cap E_{i_3} \cap E_{i_4}) = P(E_1 \cap E_2 \cap E_3 \cap E_4)$

Sample Spaces Having Equally Likely Outcomes. Let $S = \{s_1, s_2, \cdots, s_N\}$ where N denotes the number of outcomes of S. Since outcomes are equally likely to occur.

$$P(\{s_i\}) = \frac{1}{|S|}$$

Similarly, if event A has $|A|$ outcomes, then

$$P(A) = \frac{|A|}{|S|}$$

Probability as a Continuous Set Function

- A sequence of events $\{E_n\}, n \geq 1$ is said to be an <u>*increasing*</u> sequence if

$$E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$$

whereas it is said be <u>*decreasing*</u> sequence if
$$E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$$

- If $\{E_n\}, n \geq 1$ is an <u>*increasing*</u> sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ as
$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$
Similarly, if $\{E_n\}, n \geq 1$ is a <u>*decreasing*</u> sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ as
$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$
- If $\{E_n\}, n \geq 1$ is an <u>*increasing*</u> or a <u>*decreasing*</u> sequence of events, then
$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$$

Chapter 3: Conditional Probability and Independence

Conditional Probabilities. Let E and F be two events. Suppose that $P(F) > 0$, the conditional probability of E given F is defined as

$$\frac{P(EF)}{P(F)}$$

and is denoted by $P(E|F)$. It can also be read as the conditional probability that E occurs given that F has occurred.

Multiplication Rule. Suppose that $P(A) > 0$, then

$$P(AB) = P(A)P(B|A)$$

General Multiplication Rule. Let $A_1, A_2, \cdots, A_n$ be n events, then

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$$

Building Block to Bayes' Formula. Let A and B be *any* two events, then

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Partitioning. We say that $A_1, A_2, \cdots, A_n$ partition the sample space S if :
(a) They are "mutually exclusive", meaning $A_i \cap A_j = \emptyset$ for all $i \neq j$
(b) They are "collectively exhaustive", meaning $\bigcup_{i=1}^{n} A_i = S$

Bayes' First Formula. Suppose the events $A_1, A_2, \cdots, A_n$ *partition* the sample space. Assume further that $P(A_i) > 0$ for $0 \leq i \leq n$.

Let B be any event, then
$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$$

Bayes' Second Formula. Suppose the events $A_1, A_2, \cdots, A_n$ *partition* the sample space. Assume further that $P(A_i) > 0$ for $0 \leq i \leq n$.

Let B be any event, then for any $1 \leq i \leq n$,
$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)}$$

Odds. The odds of an event A is defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

Independence. Two events A and B are said to be <u>*independent*</u> if

$$P(AB) = P(A)P(B) \text{ or } P(A|B) = P(A)$$

They are said to be <u>*dependent*</u> if

$$P(AB) \neq P(A)P(B)$$

Independence and Complements. If A and B are independent, then so are

(i) A and B^c
(ii) A^c and B
(iii) A^c and B^c

* If A is independent of B, and A is also independent of C, it <u>*is not*</u> necessarily true that A is independent of *BC.*

Mutual Independence. Three events *A, B* and *C* are said to be independent if the following 4 conditions hold:

(1) $P(ABC) = P(A)\,P(B)\,P(C)$
(2) $P(AB) = P(A)\,P(B)$
(3) $P(AC) = P(A)\,P(C)$
(4) $P(BC) = P(B)\,P(C)$

* If only conditions 2, 3, 4 are met, *A, B, C* are *pairwise independent.*

Mutual Independence Effects

It should be noted that if *A, B,* and *C* are independent, then A is independent of any event formed from B and C.

(i) A is independent of $B \cup C$
(ii) A is independent of $B \cap C$

Generalised Mutual Independence. Events $A_1, A_2, \cdots, A_n$ are said to be independent if, for every sub-collection of events $A_{i_1}, A_{i_2}, \cdots, A_{i_r}$, we have

$$P(A_{i_1} A_{i_2} \cdots A_{i_r}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_r})$$

Algebra of Conditional Probability. Let A be an event with $P(A) > 0$. Then the following three conditions hold.

(i) For any event B, we have
$$0 \leq P(B|A) \leq 1$$

(ii)
$$P(S|A) = 1$$

(iii) Let $B_1, B_2, B_3, \cdots$ be a sequence of mutually exclusive events, then
$$P(\cup_{k=1}^{\infty} B_k\, |A) = \sum_{k=1}^{\infty} P(B_k|A)$$

Chapter 4: Random Variables

Random Variable. A random variable *X*, is a mapping from the sample space to real numbers.

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

Discrete Random Variable. A random variable is said to be <u>*discrete*</u> if the range of X is either finite of countably infinite.

Probability Mass Function (Discrete). Suppose a random variable X is discrete, taking values $x_1, x_2, \cdots$, then the probability mass function of X, denoted by p_X (or simply as p if the context is clear), is defined as

$$p_X(x) = \begin{cases} P(X = x), \text{if } x = x_1, x_2, \cdots; \\ 0, \qquad \text{otherwise.} \end{cases}$$

Properties of the Probability Mass Function

(i) $p_X(x_i) \geq 0$; for $i = 1, 2, \cdots$;
(ii) $p_X(x_i) = 0$, for other values of x,
(iii) Since X must take on one of the values of x_i,

$$\sum_{i=1}^{\infty} p_X(x_i) = 1$$

Cumulative Distribution Function (Discrete). The cumulative distribution function of X, abbreviated to <u>*distribution function*</u> of X, (denoted as F_X or F if the context is clear) is defined as

$$F_X : \mathbb{R} \to \mathbb{R}$$

where

$$F_X(x) = P(X \leq x), \quad \text{for } x \in \mathbb{R}$$

Expected Value. If X is a discrete random variable having the probability mass function p_X, the expectation or the expected value of X, denoted by $E(X)$ or μ_X is defined by

$$E(X) = \sum_x x\, p_X(x)$$

Bernoulli Random Variable. Suppose X takes only two values 0 and 1 with

$$P(X = 0) = 1 - p \quad \text{and} \quad P(X = 1) = p$$

X is a Bernoulli random variable of parameter p, denoted it by $X \sim Be(p)$.

Tail Sum Formula for Expectation. For non-negative integer-valued random variable X (that is, X takes values 0, 1, 2, ⋯),

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X \geq k)$$

Expectation of a Function of a Random Variable. If X is a discrete random variable that takes values x_i, $i \geq 1$, with respective probabilities $p_X(x_i)$, then for any real value function g.

$$\begin{aligned} E[g(X)] &= \sum_i g(x_i)\, p_X(x_i) \quad \text{or equivalently} \\ &= \sum_x g(x)\, p_X(x) \end{aligned}$$

Corollary: Expectation of a Linear Function. Let a and b be constants, then

$$E[aX + b] = aE(X) + b$$

k^{th} Moment of X. For $k \geq 1$, $E(X^k)$ is called the k^{th} moment of X.

k^{th} Central Moment of X. Let $\mu = E(X)$, and take $g(x) = (x - \mu)^k$, then

$$E[(X - \mu)^k]$$

is called the k^{th} central moment.

Variance. If X is a random variable with mean μ, then the variance of X, denoted by $Var(X)$, is defined by

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

This is also known the 2^{nd} central moment of X.

Standard Deviation. The standard deviation of X, denoted by σ_X or $SD(X)$, is defined as

$$\sigma_X = \sqrt{Var(X)}$$

Remarks Regarding Variance.

(1) Note that $Var(X) \geq 0$
(2) $Var(X) = 0$ if and only if X is a degenerate random variance (that is, the random variable takes only one value)
(3) It follows from the formula that $E(X^2) \geq [E(X)]^2 \geq 0$.

Scaling and Shifting Property of Variance and Standard Deviation.

(i) $Var(aX + b) = a^2 Var(X)$
(ii) $SD(aX + b) = |a|\, SD(X)$

Bernoulli Random Variable. A Bernoulli random variable, denoted by $Be(p)$, is defined by

$$X = \begin{cases} 1, & \text{if it is a success;} \\ 0, & \text{if it is a failure;} \end{cases}$$

It is used to model a trial in which a particular event occurs or does not. Occurrence of this event is called success and non-occurrence is called failure. Each trial has a probability of success of p and a probability of failure of $q = 1 - p$

$$E(X) = p \quad Var(X) = p(1 - p)$$

Binomial Random Variable. A binomial random variable, denoted by $Bin(n, p)$, is defined by

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

where X represent the number of success in n independent Bernoulli(p) trials. Hence X takes values $0, 1, 2, \cdots, n$

$$E(X) = np \quad Var(X) = np(1 - p)$$

Note: Let $X_i, i = 1, \cdots, n$ be n independent $Bernoulli(p)$ random variables. Then

$$X = X_1 + X_2 + \cdots + X_n$$

where $X \sim Bin(n, p)$

Geometric Random Variable. A geometric random variable, denoted by $Geom(p)$, is defined as

$$P(X = k) = pq^{k-1}$$

where X represents the number of $Bernoulli(p)$ trials required to obtain the first success. Therefore, X takes values $1, 2, 3, \cdots$, and so on.

$$E(X) = \frac{1}{p} \quad Var(X) = \frac{1 - p}{p^2}$$

Similarly, we could let X' represent the number of failures of $Bernoulli(p)$ trial to obtain the first success.

In that case,

$$X' = X - 1$$

$$P(X' = k) = pq^k$$

$$E(X') = \frac{1 - p}{p} \quad Var(X') = \frac{1 - p}{p^2}$$

Negative Binomial Random Variable. A negative binomial random variable, denoted bye $NB(r, p)$, is defined as

$$P(X = k) = \binom{k - 1}{r - 1} p^r q^{k-r}$$

where X represents the number of $Bernoulli(p)$ trials required to obtain r successes. Therefore, X takes values $r, r + 1, \cdots$, and so on.

$$E(X) = \frac{r}{p} \quad Var(X) = \frac{r(1 - p)}{p^2}$$

Remark: Note that $Geom(p) = NB(1, p)$

Poisson Random Variable. A Poisson random variable, denoted by $P(\lambda)$, defined as

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

where X represents the number of occurrences of an event within a given time interval where λ is the average number of occurrences of that event within the same time interval. Therefore, X takes values $0, 1, 2, \cdots$, and so on.

$$E(X) = \lambda \quad Var(X) = \lambda$$

Estimating a Binomial Random Variable with Poisson. The Poisson random variable can be used as an approximation for a binomial random variable with parameter (n, p) when n is large and p is small such that np is of moderate size.

Hypergeometric Random Variable. A hypergeometric random variable, denoted by $H(n, N, m)$, defined as

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

Suppose that we have a set of N balls, of which m are red and $N - m$ are blue. We choose n of these balls, ***without replacement***. X represents the number of red balls in our sample.

$$E(X) = \frac{nm}{N} \quad Var(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$

Properties of Distribution Function.

(i) F_X is a non-decreasing function, i.e., if $a < b$, then $F_X(a) \leq F_X(b)$

(ii)
$$\lim_{b\to\infty} F_X(b) = 1 \quad \lim_{b\to-\infty} F_X(b) = 0$$

(iii) F_X has left limits. i.e.

$$\lim_{x\to b^-} F_X(x) \text{ exists for all } b \in \mathbb{R}$$

(iv) F_X is right continuous. That is, for any $b \in \mathbb{R}$
$$\lim_{x\to b^+} F_X(x) = F_X(b)$$

Useful Calculation with Distribution Function.

(i) $P(a < X \leq b) = F_X(b) - F_X(a)$

(ii) $P(X = a) = F_X(a) - F_X(a^-)$, where $F_X(a^-) = \lim_{x\to a^-} F_X(x)$

(iii) From (i) and (ii), we can compute $P(a \leq X \leq b)$; $P(a \leq X < b)$; and $P(a < X < b)$. For example,
$$\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X \leq b) \\ &= F_X(b) - F_X(a^-) \end{aligned}$$

(iv) Calculating probabilities from probability mass function

$$P(A) = \sum_{x\in A} p_X(x)$$

(v) Calculating probability mass function from distribution function.

$$p_X(x) = F_X(x) - F_X(x^-), \quad x \in \mathbb{R}$$

(vi) Calculating distribution function from probability mass function.

$$F_X(x) = \sum_{y\leq x} p_X(y), \quad x \in \mathbb{R}$$

Chapter 5: Continuous Random Variables

Probability Density Function. We say that X is a continuous random variable if there exists a non-negative function f_X. defined for all real $x \in \mathbb{R}$, such that

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx, \quad \text{for } -\infty < a < b < +\infty,$$

The function f_X is called the probability density function (p.d.f.) of the random variable X.

Distribution Function. We define the distribution function of X by

$$F_X(x) = P(X \leq x), \quad \text{for } x \in \mathbb{R}$$

Note: The definition for distribution function is the same for discrete and continuous random variables. In the continuous case,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad x \in \mathbb{R}$$

and

$$f_X(x) = \frac{\partial}{\partial x} F_X(x), \quad x \in \mathbb{R}$$

Determining the Constant in the Probability Density Function. We can determine the constant by using the fact that

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 0$$

Expectation of Continuous Random Variable. Let X be a continunous random variable with probability density function f_X, then

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\,dx$$

$$Var(X) = \int_{-\infty}^{\infty} \left(x - E(X)\right)^2 f_X(x)\,dx$$

Expectation of a Function of a Continuous Random Variable. If X is a continuous random variable with probability density function f_X, then for any real value function $g(X)$.

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

(ii) Linearity Property

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2 Var(X)$$

$$SD(aX + b) = |a|\, SD(X)$$

(iii) Formula for variance.

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

Tail Sum Formula (Continuous). Suppose X is a *non-negative* continuous random variable, then

$$E(X) = \int_0^{\infty} P(X > x)\,dx = \int_0^{\infty} P(X \geq x)\,dx$$

Uniform Distribution. A random variable X is said to be uniformly distributed over the interval (a, b), denoted by $X \sim U(a, b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & \text{otherwise;} \end{cases}$$

and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy = \begin{cases} 0, & \text{if } x < a; \\ \frac{x-a}{b-a}, & \text{if } a \leq x < b \\ 1, & \text{if } b \leq x \end{cases}$$

It can be shown that

$$E(X) = \frac{a+b}{2} \quad Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution. A random variable X is said to be normally distributed with parameters μ and σ^2, denoted by $X \sim N(\mu, \sigma^2)$ if its probability density function is given by

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

Complete this another time

Exponential Distribution. A random variable X is said to be exponentially distributed with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

and its distribution function is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x \leq 0; \\ 1 - e^{-\lambda x}, & \text{if } x > 0; \end{cases}$$

It can be shown that

$$E(X) = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2}$$

The exponential distribution also has a memoryless property.

$$P(X > s + t | X > s) = P(X > t), \quad \text{for } s, t > 0$$

Exponential Distribution is usually used to model the time between events, where events occur continuously and independently at a constant average rate

Gamma Distribution. A random variable X is said to have a gamma distribution with parameters (α, λ), denoted by $Gamma(\alpha, \lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

where $\lambda > 0, \alpha > 0$ and $\Gamma(\alpha)$, called the gamma function, is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy$$

If $X \sim Gamma(\alpha, \lambda)$, then

$$E(X) = \frac{\alpha}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2}$$

Remarks Regarding the Gamma Distribution

(a) $\Gamma(1) = \int_0^\infty e^{-y}\, dy = 1$
(b) It can be shown, via integration by parts, that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
(c) for integral values of α, say $\alpha = n$,

$$\Gamma(n) = (n-1)!$$

(d) $Gamma(1, \lambda) = Exp(\lambda)$
(e) If $X_i \sim Exp(\lambda)$ independently, then $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$
(f) If $X \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, then $X \sim \chi^2(n)$
(g) $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-y} y^{-\frac{1}{2}}\, dy = \sqrt{\pi}$

Interpretation of Gamma Distribution when $\alpha = n$. If events are occurring randomly in time then the amount of time one has to wait until a total of n events has occured is a random variable which follows a Gamma distribution with parameters (n, λ)

Weibull Distribution. A random variable X is said to have a Weibull Distribution with parameters (ν, α, β), denoted by $W(\nu, \alpha, \beta)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{\beta}{\alpha}\left(\dfrac{x-\nu}{\alpha}\right)^{\beta-1} \exp\left(-\left(\dfrac{x-\nu}{\alpha}\right)^\beta\right), & \text{if } x > \nu; \\ 0, & \text{if } x \leq \nu; \end{cases}$$

If $X \sim W(\mu, \alpha, \beta)$, then

$$E(X) = \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \quad Var(X) = \alpha^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2\right]$$

Remark: The $Exp(\lambda)$ is a special case of a Weibull distribution with $\alpha = 1, \beta = \lambda$ and $\nu = 0$.

Cauchy Distribution. A random variable X is said to follow a Cauchy distribution with parameter θ and α, where $-\infty < \theta < \infty$ and $\alpha > 0$ if its density is given by

$$f_X(x) = \frac{1}{\pi\alpha\left[1 + \left(\frac{x-\theta}{\alpha}\right)^2\right]}, \text{ for } -\infty < x < \infty$$

Both $E(X)$ and $Var(X)$ do not exist.

Beta Distribution. A random variable X is said to have a Beta distribution with parameters (a, b), denoted by $Beta(a, b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1}, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise}; \end{cases}$$

where $-\infty < a, b < \infty$ and $\mathrm{B}(a, b)$, called the Beta function, is defined by

$$\mathrm{B}(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$$

If $X \sim Beta(a, b)$, then

$$E(X) = \frac{a}{a+b} \quad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Remarks:

1. $U(0,1)$ is a special Beta distribution. $Beta(1,1) \equiv U(0,1)$
2. It can be shown that

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Approximation of Binomial Random Variables. Let $X \sim Bin(n, p)$. Here, n is assumed to be large. There are two commonly used approximation of the binomial distribution.

(a) Normal approximation
(b) Poisson approximation

Normal Approximation of Binomial Random Variable. Suppose that $X \sim Bin(n, p)$. Then for any $a < b$,

$$P\left(a < \frac{X - np}{\sqrt{npq}} \leq b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$, where $q = 1 - p$ and $\Phi(z) = P(Z \leq z)$ with $Z \sim N(0,1)$

That is,

$$Bin(n, p) \approx N(np, npq)$$

Equivalently,

$$\frac{X - np}{\sqrt{npq}} \approx Z$$

where $Z \sim N(0, 1)$

Remark: The normal approximation will be generally good for values of n satisfying $np(1-p) \geq 10$

The approximation can be further improved with continuity correction.

Continuity-correction. If $X \sim Bin(n, p)$, then

$$P(X = k) = P\left(k - \frac{1}{2} < x < k + \frac{1}{2}\right)$$

$$P(X \geq k) = P\left(X \geq k - \frac{1}{2}\right)$$

$$P(X \leq k) = P\left(X \leq k + \frac{1}{2}\right)$$

Poisson Approximation of Binomial Random Variable. The Poisson distribution is used as an approximation to the binomial distribution when the parameter n and p are large and small, respectively and that np is moderate.

As a working rule, use the Poisson approximation if $p < 0.1$ and put $\lambda = np$. If $p > 0.9$, put $\lambda = n(1-p)$ and work in terms of "failure".

Distribution of a Function of a Random Variable. Let X be a continuous random variable having a probability density function f_X. Suppose that $g(x)$ is a strictly monotonic, differentiable function of X. Then the random variable Y defined by $Y = g(X)$ has probability density function given by

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\left|\dfrac{d}{dy}g^{-1}(y)\right|\right), & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \neq g(x) \text{ for all } x; \end{cases}$$

where $g^{-1}(y)$ is defined to be equal that value of X such that $g(x) = y$.

Chapter 6: Jointly Distributed Random Variables.

Joint Distribution Function. For any two random variables X and Y defined on the same sample space, we define the joint distribution function of X and Y, denoted by $F_{X,Y}(x, y)$, by

$$F_{X,Y}(x, y) \coloneqq P(X \leq x, Y \leq y) \quad for\ x, y \in \mathbb{R}$$

Obtaining Individual Distribution Function. The distribution function of X can be obtained from the joint density function of X and Y in the following way.

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

We call $F_X(x)$ the marginal distribution function of X

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

and $F_Y(y)$ the marginal distribution function of Y.

Joint Probability Mass Function of Discrete Random Variables. In the when both X and Y are discrete random variables, we define the joint probability mass function of X and Y, denoted by $p_{X,Y}(x, y)$, as

$$p_{X,Y}(x, y) \coloneqq P(X = x, Y = y)$$

We can recover the probability mass function of X and Y in the following manner.

$$p_X(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{X,Y}(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_{x \in \mathbb{R}} p_{X,Y}(x, y)$$

We call $p_X(x)$ the marginal probability mass function of X and $p_Y(y)$ the marginal probability mass function of Y.

Useful Formulas of Discrete P.M.F.

(i)

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \sum_{a_1 < x \leq a_2} \sum_{b_1 < y < b_2} p_{X,Y}(x, y)$$

(ii)

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = \sum_{x \leq a} \sum_{y < b} p_{X,Y}(x, y),$$

(iii)

$$P(X > a, Y > b) = \sum_{x > a} \sum_{y > b} p_{X,Y}(x, y)$$

Jointly Probability Mass Function Continuous Random Variables. We say that X and Y are jointly continuous random variable if there exist a function (which is denoted by $f_{X,Y}(x, y)$, called the joint probability density function of X and Y) defined for all real x and y, having the property that for every set C of pairs of real numbers, we have

$$P\big((X, Y) \in C\big) \coloneqq \iint_{(x,y) \in C} f_{X,Y}(x, y)\ dx\ dy$$

Some Useful Formula

(i) Let $A, B \subset \mathbb{R}$, take $C = A \times B$ above

$$P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)\ dy\ dx$$

(ii) In particular, let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ where $a_1 < a_2$ and $b_1 < b_2$, we jabe

$$\mathrm{P}(\mathrm{a}_1 < X < a_2, b_1 < Y < b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)\ dy\ dx$$

(iii) Let $a, b \in \mathbb{R}$, we have

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a_2} \int_{-\infty}^{b_2} f_{X,Y}(x, y)\ dy\ dx$$

As a result of this,

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

We can recover the probability density function of X and Y in the following manner.

$$f_X(x) \coloneqq \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy$$

$$f_Y(y) \coloneqq \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx$$

Independent Random Variables. Two random variables X and Y are said to be independent if

$$P(X \in A, Y \in B) \coloneqq P(X \in A)\ P(Y \in B) \quad \text{for any } A, B \subset \mathbb{R}$$

In general, X and Y are independent if and only if there exist function $g, h : \mathbb{R} \to \mathbb{R}$ such that for all $x, y \in \mathbb{R}$, we have

$$f_{X,Y}(x, y) = h(x)\ g(y)$$

Random variable that are not independent are said to be dependent.

Equivalent Statements for Independence (Discrete). The following three statements are equivalent:

(i) Random variables X and Y are independent,

(ii) For all $x, y \in \mathbb{R}$, we have

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

(iii) For all $x, y \in \mathbb{R}$, we have

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

Equivalent Statements for Independence (Continuous). The following three statements are equivalent:

(iv) Random variables X and Y are independent,

(v) For all $x, y \in \mathbb{R}$, we have

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

(vi) For all $x, y \in \mathbb{R}$, we have

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

Sums of Independent Random Variables. Under the assumption of independence of X and Y, we have

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad \text{for } x, y \in \mathbb{R}$$

Then it follows that

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a - y) f_Y(y)\ dy = \int_{-\infty}^{\infty} F_Y(a - x) f_X(x)\ dx$$

We can also show that

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y)\ dy = \int_{-\infty}^{\infty} f_Y(a - x) f_X(x)\ dx$$

Sum of 2 Independent Gamma Random Variables. Assume that $X \sim Gamma(\alpha, \lambda)$ and $Y \sim Gamma(\beta, \lambda)$, are X and Y are mutually independent,

Then,

$$X + Y \sim Gamma(\alpha + \beta, \lambda)$$

Note that both X and Y must have the same second parameter.

Sum of Independent Normal Random Variables. If $X_i, i = 1, \cdots, n$ are independent random variables that are normally distributed with respective parameters $\mu_i, \sigma_i^2, i = 1, \cdots, n$, then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Sum of Independent Poisson Random Variables. Let $X \sim Bin(n, p)$ and $Y \sim Bin(m, p)$. The probability mass function of $X + Y$ is given as

$$X + Y \sim Bin(n + m, p)$$

Note that both X and Y must have the same success probability, p.

Conditional Probability Mass Function (Discrete). The conditional probability mass function of X given $Y = y$ is defined by

$$p_{X|Y}(x|y) \coloneqq P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Conditional Distribution Function (Discrete). The conditional probability mass function of X given $Y = y$ is defined by

$$F_{X|Y}(x|y) = P(X \leq x | Y = y)$$

$F_{X\|Y}(x\|y)$	$= P(X \leq x\|Y = y)$ $= \frac{\sum_{a \leq x} p_{X,Y}(a, y)}{p_Y(y)}$ $= \sum_{a \leq x} \frac{p_{X,Y}(x, y)}{p_Y(y)}$ $= \sum_{a \leq x} p_{X\|Y}(a\|y)$

Conditional Probability Mass Function and Independence. If X is independent of Y, then the conditional probability mass function of X give $Y = y$ is the same as the marginal probability mass function of X for every Y such that $p_Y(y) > 0$

$$p_{X|Y}(x|y) = p_x(x)$$

Conditional Probability Density Function. Suppose that X and Y are jointly distributed continuous random variables . We define the conditional probability density function of X given that $Y = y$ as

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

We can find the conditional probability of events associated with one random variable when we are given the value of the second random variable.

That is, for $A \subset \mathbb{R}$ and Y such that $f_Y(y) > 0$,

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)\ dx$$

Conditional Distribution Function (Discrete). The conditional distribution function of X given $Y = y$ is defined by

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) \int_{-\infty}^{x} f_{X|Y}(t|y)\ dt$$

Conditional Probability Mass Function and Independence. If X is independent of Y, then the conditional probability mass function of X give $Y = y$ is the same as the marginal probability density function of X for every Y such that $f_Y(y) > 0$

$$f_{X|Y}(x|y) = f_x(x)$$

The Bivariate Normal Distribution. We say that the random variables X, Y have a bivariate normal distribution if, for constant $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$f_{X,Y}(x,y) := \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

Finish this up

Joint Probability Distribution Function of Functions of Random Variables. Let X and Y be jointly distributed random variables with joint probability density function $f_{X,Y}(x,y)$. It is sometime necessary to obtain the joint distribution of the random variables U and V, which arise as functions of X and Y.

Specifically, suppose that

$$U = g(X,Y) \quad V = h(X,Y)$$

for some functions g and h.

We want to find the joint probability function of U and V in terms of the joint probability density function $f_{X,Y}(x,y)$, g and h.

Assume the following conditons are satisfied.

1. Let X and Y be jointly continuous distributed random variables with known joint probability density function.
2. Let U and V be given functions of X and Y in the form:
   $$U = g(X,Y) \quad V = h(X,Y)$$
   And we can uniquely solve X and Y in terms of U and V, say $x = a(u,v)$ and $y = b(u,v)$
3. The functions g and h have continuous partial derivatives at all points (x,y) and
   $$J(x,y) := \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\frac{\partial h}{\partial y} - \frac{\partial g}{\partial y}\frac{\partial h}{\partial x} \neq 0$$
   at all points (x,y)

The joint probability density function of U and V is given by

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\ |J(x,y)|^{-1}$$

where $x = a(u,v)$ and $y = b(u,v)$

Generalised Joint Probability Distribution Function of Functions of Random Variables. When the joint density function of n random variable $X_1, X_2, \cdots, X_n$ is given and we want to compute the joint density function of $Y_1, Y_2, \cdots, Y_n$, where

$$Y_1 = g_1(X_1, \cdots, X_n), Y_2 = g_2(X_1, \cdots, X_n), \cdots, Y_n = g_n(X_1, \cdots, X_n)$$

Assume that the function g_j have continuous partial derivatives and that the Jacobian determinant $J(x_1, x_2, \cdots, x_n) \neq 0$ at all points $(x_1, x_2, \cdots, x_n)$, where

$$J(x_1, x_2, \cdots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

Furthermore, we suppose the equations

$$y_1 = g_1(x_1, x_2, \cdots, x_n)$$
$$y_2 = g_2(x_1, x_2, \cdots, x_n)$$
$$\vdots$$
$$y_n = g_n(x_1, x_2, \cdots, x_n)$$

have a unique solution, say

$$x_1 = h_1(y_1, y_2, \cdots, y_n)$$
$$x_2 = h_2(y_1, y_2, \cdots, y_n)$$
$$\vdots$$
$$x_n = h_n(y_1, y_2, \cdots, y_n)$$

Under these assumptions, the joint density function of the random variables $(Y_1, Y_2, \cdots, Y_n)$ is given by

$$f_{Y_1,Y_2,\cdots,Y_n}(y_1, y_2, \cdots, y_n) = f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n)\ |J(x_1, x_2, \cdots, x_n)|^{-1}$$

where $x_i = h_i(y_1, y_2, \cdots, y_n), i = 1, 2, \cdots, n$.

Jointly Distributed Random Variable ($n \geq 3$). Assume we have 3 jointly distributed random variables, called X, Y and Z such that

$$F_{X,Y,Z}(x,y,z) = P(X \leq x, Y \leq y, Z \leq z)$$

There are a number of marginal distribution functions, namely

$$F_{X,Y}(x,y) := \lim_{z\to\infty} F_{X,Y,Z}(x,y,z)$$

$$F_{X,Z}(x,z) := \lim_{y\to\infty} F_{X,Y,Z}(x,y,z)$$

$$F_{Y,Z}(y,z) := \lim_{x\to\infty} F_{X,Y,Z}(x,y,z)$$

$$F_X(x) := \lim_{y\to\infty, z\to\infty} F_{X,Y,Z}(x,y,z)$$

$$F_Y(y) := \lim_{x\to\infty, z\to\infty} F_{X,Y,Z}(x,y,z)$$

$$F_Z(z) := \lim_{x\to\infty, y\to\infty} F_{X,Y,Z}(x,y,z)$$

Joint Probability Density Function of X, Y and Z

Finish this up

Chapter 7: Properties of Expectation

Elementary Properties of Expected Values. Let X be a random variable.

$$\text{If } a \leq X \leq b, \text{ then } a \leq E(X) \leq b$$

Expectation of Functions of Random Variables.

(a) If X and Y are jointly discrete with joint probability mass function $p_{X,Y}(x,y)$, then

$$E[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y)$$

(b) If X and Y are jointly continuous with joint probability mass function $f_{X,Y}(x,y)$, then

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

Consequences of Previous Result.

1. If $g(x,y) \geq 0$ whenever $p_{X,Y}(x,y) > 0$, then $E[g(X,Y)] \geq 0$
2. $E[g(X,Y) + h(X,Y)] = E[g(X,Y)] + E[h(X,Y)]$
3. $E[g(X) + h(Y)] = E[g(X)] + E[h(Y)]$

4. If jointly distributed random variable X and Y satisfy $X \le Y$, then,

$$E(X) \le E(Y)$$

Mean of Sums Equals Sum of Means.

$$E(X+Y) = E(X) + E(Y)$$

This can be extended to

$$E(a_1X_1 + \cdots + a_nX_n) = a_1E(X_1) + \cdots + a_nE(X_n)$$

Boole's Inequality

Let $A_1, \cdots, A_n$ denote events and define the indicator variable $I_k, k = 1, \cdots, n$, by

$$I_k \begin{cases} 1 & \text{if } A_k \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

it can be shown that

$$P\left(\bigcup_{k=1}^{n} A_k\right) \le \sum_{k=1}^{n} P(A_k)$$

Covariance. The covariance of jointly distributed random variable X and Y, denoted by $Cov(X,Y)$, is defined by

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

where μ_X and μ_Y denote the means of X and Y respectively.

Remark:

- If $Cov(X,Y) = 0$, we say X and Y are correlated
- If $Cov(X,Y) \ne 0$, we say X and Y are uncorrelated

Alternate Formula for Covariance

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

Expectation and Independence. If X and Y are independent, then for any function $g, h: \mathbb{R} \to \mathbb{R}$, we have

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Covariance and Independence. If X and Y are independent, then $Cov(X,Y) = 0$. However, the inverse is not true.

Properties of Covariance

(i) $Var(X) = Cov(X,X)$
(ii) $Cov(X,Y) = Cov(Y,X)$
(iii) $Cov\left(\sum_{i=1}^{n} a_iX_i, \sum_{j=1}^{m} b_jY_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jCov(X_iY_j)$

Note that no independence is assumed.

Variance of a Sum

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k) + 2\sum_{1\le i<j\le n} Cov(X_i, X_j)$$

If $X_i, \cdots, X_n$ are independent random variables, then

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k)$$

Hence, under independence, variance of sum is equal to the sum of variances.

Correlation. The correlation of random variables X and Y, denoted by $\rho(X,Y)$, is defined by

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

where

$$-1 \le \rho(X,Y) \le 1$$

Remarks Regarding Correlation.

(1) The correlation coefficient is a measure of the degree of linearity between X and Y, where a magnitude close to one represents high linearity while a magnitude close to 0 represents no linearity
(2) $\rho(X,Y) = 1$ if and only if $Y = aX + b$ where $a = \frac{\sigma_Y}{\sigma_X} > 0$
(3) $\rho(X,Y) = -1$ if and only if $Y = aX + b$ where $a = -\frac{\sigma_Y}{\sigma_X} < 0$
(4) $\rho(X,Y)$ is dimensionless
(5) If X and Y are independent, then $\rho(X,Y) = 0$. The inverse is not true.

Conditional Expectation.

(1) If X and Y are jointly distributed discrete random variables, then

$$E[X|Y=y] = \sum_x x p_{X|Y}(x|y), \quad \text{if } p_Y(y) > 0$$

(2) If X and Y are jointly distributed continuous random variables, then

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx, \quad \text{if } f_Y(y) > 0$$

Important Formula Regarding Conditional Expectation.

(1)

$$E[g(X)|Y=y] = \begin{cases} \sum_x g(x)p_{X|Y}(x|y), & \text{for discrete case;} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y), & \text{for continuous case;} \end{cases}$$

(2)

$$E\left[\sum_{k=1}^{n} X_k | Y=y\right] = \sum_{k=1}^{n} E[X_k|Y=y]$$

Conditional Expectation to Expectation

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

Computing Probability by Conditioning. Let $X = I_A$ where A is an event, then we have

$$E(I_A) = P(A)$$

$$E(I_A|Y=y) = P(A|Y=y)$$

and hence

$$P(A) = \begin{cases} \sum_y P(A|Y=y)p(Y=y), & \text{for } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P(A|Y=y)\,f_Y(y)\,dy, & \text{for } Y \text{ is continuous} \end{cases}$$

Law of Total Variance.

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

Moment Generating Functions. The moment generating function of random variable X, denoted by M_X, is defined as

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} f_X(x), & \text{if } X \text{ is discrete with pmf } p_X(x) \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx, & \text{if } X \text{ is continous with pdf } f_X(x) \end{cases}$$

This function generates all the moments of this random variable X

For $n \ge 0$,

$$E(X^n) = M_X^{(n)}(0)$$

where

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

Multiplicative Property (Independence). If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)\,M_Y(t)$$

Uniqueness Property. Let X and Y be random variables with their moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Suppose that there exists an $h > 0$ such that

$$M_X(t) = M_Y(t), \quad \text{for all } t \in (-h, h)$$

then X and Y have the same distribution

Moment Generating Functions of Common Distribution.

1. When $X \sim Be(p)$, $M_X(t) = 1 - p + pe^t$
2. When $X \sim Bin(n,p)$, $M_X(t) = (1 - p + pe^t)^n$

3. When $X \sim Geom(p), M_X(t) = \frac{pe^t}{1-(1-p)e^t}$
4. When $X \sim Poisson(\lambda), M_X(t) = \exp(\lambda(e^t - 1))$
5. When $X \sim U(\alpha, \beta), M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$
6. When $X \sim Exp(\lambda), M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$
7. When $X \sim N(\mu, \sigma^2), M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Joint Moment Generating Functions. For any n random variable $X_1, X_2, \cdots, X_n$, the joint moment generating function, $M_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = E[e^{t_1 X_1 + \cdots + t_n X_n}]$

The individual moment generating functions can be obtained from $M_{X_1,\cdots,X_n}(t_1, \cdots, t_n)$ by letting all but one of the t_i's be 0

That is,

$$M_{X_i}(t) = E[e^{tX_i}] = M_{X_1,\cdots,X_n}(0, \cdots, 0, t, 0, \cdots, 0)$$

where the t is in the i^{th} place.

Unique Property of Joint MGF.

$M_{X_1,\cdots,X_n}(t_1, \cdots, t_n)$ uniquely determines the joint distribution of $X_1, X_2, \cdots, X_n$

Multiplicative Property (Independence). n variables $X_1, X_2, \cdots, X_n$ are independent if and only if

$$M_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n)$$

Independence of Mean and Variance for Normal Sample. Let $X_1, X_2, \cdots, X_n$ are independent and identically distributed normal variables with mean μ and variance σ^2, then the sample mean $\bar{X}$ and the sample variance S^2 are independent. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Chapter 8: Limit Theorem

Markov's Inequality. Let X be a non-negative random variable. For $a > 0$, we have

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Chebyshev's Inequality. Let X be a random variable with mean μ, then for $a > 0$, we have

$$P(|X - \mu| \geq a) \leq \frac{Var(X)}{a^2}$$

Consequences of Chebyshev's Inequality. If $Var(X) = 0$, then the random variable X is a constant.

The Weak Law of Large Numbers. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, with common mean μ. Then, for any $\epsilon > 0$

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \quad as\ n \to \infty$$

Central Limit Theorem. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-t^2/2}\ dt$$

Lemma 8.3

Let $Z_1, Z_2, \cdots$ be a sequence of random variables having distribution function F_{Z_n} and moment generating function M_{Z_n}, for $n \geq 1$. Let Z be a random variable having distribution function F_Z and moment generating function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then

$$F_{Z_n}(x) \to F_Z(x)$$

for all x at which $F_Z(x)$ is continuous.

Normal Approximation. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables, each having mean μ and variance σ^2. Then, for large n, the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

is approximately standard normal.

In other words, for $-\infty < a < b < \infty$, we have

$$P\left(a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-t^2/2}$$

The Strong Law of Large Numbers. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then with probability 1.

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty$$

Difference between the Weak and the Strong Laws of Large Numbers. The weak law states that, for any specified large number n^*, $\frac{(X_1 + X_2 + \cdots + \backslash X_{n^*})}{n^*}$ is likely to be near μ. However, it does NOT say that $\frac{(X_1 + X_2 + \cdots + X_n)}{n}$ is bound to stay near μ for all values of n larger than n^*. Hence, it leaves open the possibility that large values of $\left|\frac{(X_1 + X_2 + \cdots + X_n)}{n} - \mu\right|$ can occur indefinitely often. The strong law shows that this cannot occurs. In particular, it implies that, with probability 1, for any positive

$$\left|\sum_{i=1}^{n} \frac{X_i}{n} - \mu\right|$$

will be greater than ϵ only a finite number of times.

One-sided Chebyshev's Inequality. If X is random variable with mean 0 and finite variance σ^2, then, for any $a > 0$

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Jensen's Inequality. If $g(x)$ is convex function, then

$$E[g(X)] \geq g(E[X])$$

provided that the expectations exist and are finite.

1. A function $g(x)$ is convex if for all $0 \leq p \leq 1$ and all $x_1, x_2 \in R_X$

   $$g(px_1 + (1-p)x_2) \leq pg(x_1) + (1-p)g(x_2)$$

2. A differentiable function of one variable is convex on interval if and only if

   $$g(x) \geq g(y) + g'(y)(x - y)$$

   for all x and y in the interval
3. A twice differentiable function of one variable is convex over interval if and only if its second derivative is non-negative there.