Probability Notes

Ty Darnell

Contents

Chapter 1

Combinatorial Analysis

1.1 Introduction

The math theory of counting is known as **combinatorial analysis**
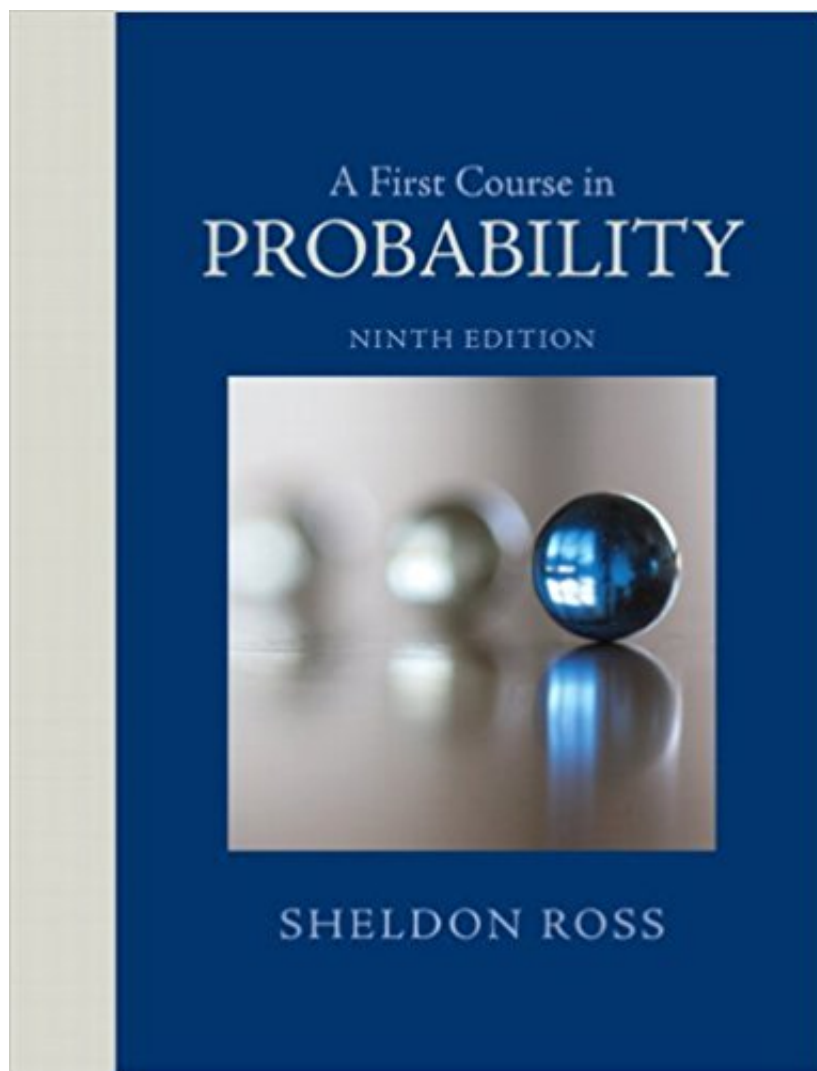
1.2 The Basic Principle of Counting

There are $n_1 \cdot n_2 \cdots n_r$ possible outcomes of the r experiments

1.3 Permutations

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

Permutation with Replacement: n^r

1.4 Combinations

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

Combination with Replacement:

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

A useful combinatorial identity:

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

1.4.1 The Binomial Theorem

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

1.5 Multinomial Coefficients

A set of n distinct items is to be divided into r distinct groups of respective sizes $n_1, n_2, \ldots, n_r$ where $\sum_{i=1}^{r} n_i = n$

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

The numbers $\binom{n}{n_1, n_2, \ldots, n_r}$ are known as **multinomial coefficients**

1.5.1 The Multinomial Theorem

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

That is, the sum is over all nonnegative integer-valued vectors $(n_1, n_2, \ldots, n_r)$ such that $n_1 + n_2 + \cdots + n_r = n$

1.6 The Number of Integer Solutions of Equations

There are r^n possible outcomes when n distinguishable balls are to be distributed into r distinguishable urns.

If the balls are indistinguishable then they outcome of the experiment can be described by a vector $(x_1, x_2, \ldots, x_r)$, where x_i denotes the number of balls in the *ith* urn.

The problem reduces to finding the number of distinct nonnegative integer-valued vectors such that

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

Proposition 6.1: There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \cdots + x_r = n \ x_i > 0, i = 1, \ldots, r$$

Proposition 6.2: There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

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

Chapter 2

Axioms of Probability

2.1 Introduction

Concept of probability as an event, show how probabilities can be computed in certain situations

2.2 Sample Space and Events

Sample Space: the set of all possible outcomes of an experiment

Event: Any subset of the sample space

For any two events E and F of a sample space S, we define the new event $E \cup F$ to consist of all outcomes that are either in E or in F or in both.

The event $E \cup F$ is called the **union** of E and F

The event EF consists of all the outcomes that are in both E and F and is called the **intersection** of E and F. Also written $E \cap F$.

$\emptyset$ is the null event, it consists of no outcomes

If $EF = \emptyset$ then E and F are **mutually exclusive** since they have no outcomes in common.

If $E_1, E_2, \ldots$ are events, the union of these events, $\cup_{n=1}^{\infty} E_n$ is defined to be that event which consists of all outcomes that are in E_n for at least one value of $n = 1, 2, \ldots$

The intersection of the events E_n, $\cap_{n=1}^{\infty} E_n$, is the event consisting of those outcomes which are in all of the events $E_n, n = 1, 2, \ldots$

E^c is the **complement** of E. Consists of all outcomes in the sample space that are not in E

For any two events E and F, if all of the outcomes in E are also in F, then E is **contained** in F. E is a **subset** of F, written $E \subset F$ or equivalently $F \supset E$, meaning F is a **superset** of E

If $E \subset F$ and $F \supset E$, then $E = F$

Rules of Unions and Intersections:

- **Commutative Laws**: $E \cup F = F \cup E \quad EF = FE$

- **Associative Laws**: $(E \cup F) \cup G = E \cup (F \cup G) \quad (EF)G = E(FG)$

- **Distributive Laws**: $(E \cup F)G = EG \cup FG \quad EF \cup G = (E \cup G)(F \cup G)$

DeMorgan's Laws:

1.
$$\left(\bigcup_{i=1}^{n} E_i \right)^c = \bigcap_{i=1}^{n} E_i^c$$

2.
$$\left(\bigcap_{i=1}^{n} E_i \right)^c = \bigcup_{i=1}^{n} E_i^c$$

2.3 Axioms of Probability

2.3.1 Relative Frequency Definition of Probability

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

where $n(E)$ is the number of times in the first n repetitions of the experiment that the event E occurs

2.3.2 The Three Axioms of Probability

1. $0 \le P(E) \le 1$

2. $P(S) = 1$

3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$

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

Axiom 3 states that, 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.

2.4   Some Simple Propositions

Proposition 4.1 $P(E^c) = 1 - P(E)$

Proposition 4.2 If $E \subset F$ then $P(E) \leq P(F)$

Proposition 4.3 $P(E \cup F) = P(E) + P(F) - P(EF)$

Proposition 4.4, The inclusion-exclusion identity

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

$\sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots, E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \ldots, n\}$

Proposition 4.4 states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time, minus the sum of the probabilities of these events taken two at a time, plus the sum of the probabilities of these events taken three at a time, and so on.

Inclusion-exclusion identity can be written more succinctly:

$$P(\cup_{i=1}^{n} E_i) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < \cdots < i_r} P(E_{i_1} \cdots E_{i_r})$$

2.5   Sample Spaces Having Equally Likely Outcomes

$$P(E) = \frac{number\ of\ outcomes\ in\ E}{number\ of\ outcomes\ in\ S}$$

If we assume all outcomes of an experiment are equally likely to occur

Randomly selecting k items from a set of n items:
either let the outcome be the ordered selection of the k items or let it be the unordered set of items selected.

$$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

2.6   Probability as a Continuous Set Function

2.7   Probability as a Measure of Belief

Chapter 3

Conditional Probability and Independence

3.1 Introduction

3.2 Conditional Probabilities

Conditional probability that E occurs given F has occurred: $P(E|F)$

$$P(E|F) = \frac{P(EF)}{P(F)} \tag{3.2.1}$$

Multiplying both sides of (3.2.1) by $P(F)$ we get:

$$P(EF) = P(F)P(E|F) \tag{3.2.2}$$

The Multiplication Rule:

$$P(E_1 E_2 E_3 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \cdots P(E_n|E_1 \cdots E_{n-1})$$

If we consider only those experiments in which F occurs, then $P(E|F)$ will equal the long-run proportion of them in which E also occurs.

3.3 Bayes's Formula

We can write E as:

$$E = EF \cup EF^c$$

Since all outcomes in E are either in both E and F or in E but not F

Since EF and EF^c are clearly mutually exclusive we have:

$$\begin{aligned}
P(E) &= P(EF) + P(EF^c) \\
&= P(E|F)P(F) + P(E|F^c)P(F^c) \\
&= P(E|F)P(F) + P(E|F^c)[1 - P(F)]
\end{aligned} \tag{3.3.1}$$

Bayes Formula

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

Bayes' Rule with LOTP

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \tag{3.3.2}$$

3.3.1 Odds

The odds of an event A tell how much more likely it is that the event A occurs than it is that it does not occur.

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

The change in the probability of a hypothesis when new evidence is introduced can be expressed compactly in terms of the change in the odds of that hypothesis.

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \frac{P(E|H)}{P(E|H^c)} \tag{3.3.3}$$

The new value of the odds of H is the old value, multiplied by the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true.

3.3.2 Law of Total Probability

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i) \tag{3.3.4}$$

3.3.3 Bayes' Formula

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \tag{3.3.5}$$

Let $F_1, \ldots, F_n$ be a set of mutually exclusive and exhaustive events (meaning that exactly one of these events must occur)
Suppose E has occurred and we are interested in determining which one of the F_j also occurred.

3.4 Independent Events

If $P(E|F) = P(E)$, E is **independent** of F

$$P(EF) = P(E)P(F) \tag{3.4.1}$$

3.4.1 is symmetric in E and F

Proposition 3.4.1. *If E and F are independent, then so are E and F^c*

Three events are independent if:

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

An infinite set of events is independent if every finite subset of those events is independent.

3.5 $P(\cdot|F)$ is a Probability

Conditional probabilities satisfy all of the properties of ordinary probabilities

Proposition 3.5.1.

1. $0 \leq P(E|F) \leq 1$

2. $P(S|F) = 1$

3. If $E_i, i = 1, 2, \ldots$ are mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} E_i \Big| F\right) = \sum_{1}^{\infty} P(E_i|F)$$

Chapter 4

Random Variables

4.1 Random Variables

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself.

Random Variables: real-valued functions defined on a sample space of an experiment.

Since the value of an r.v. is determined by the outcome of the experiment, we may assign probabilities to the possible values of the r.v.

Cumulative Distribution Function (CDF):

$$F(x) = P(X \leq x) \qquad -\infty < x < \infty$$

4.2 Discrete Random Variables

Probability Mass Function (PMF) (of a discrete r.v.)

$$p(a) = P\{X = a\}$$

$p(a)$ is the PMF of X

$$p(x_i) \geq 0 \quad \text{for } i = 1, 2, \ldots$$
$$p(x) = 0 \quad \text{for all other values of } x$$

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

The **cumulative distribution** function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{all \ x \leq a} p(x)$$

If X is a discrete r.v. whose possible values are $x_1, x_2, x_3, \ldots$ where $x_1 < x_2 < x_3 < \ldots$ then the distribution function F of X is a step function. The value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.

The size of the step at any value is equal to the probability that X assumes that particular value.

4.3 Expected Value

Expectation or **Expected Value** (of X): weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

where $p(x)$ is the PMF of X

If X is a random variable that must take on the values $x_1, x_2, \ldots, x_n$ with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$ then we have:

$$\sum_{i=1}^{n} x_i p(x_i) = E[X]$$

I is an indicator variable for event A if:

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$$

$E[I] = P(A)$
The expected value of the indicator variable for the event A is equal to the probability that A occurs.

4.4 Expectation of a Function of a Random Variable

Proposition 4.4.1. *If X is a discrete r.v. that takes on one of the values $x_1, i \geq 1$ with respective probabilities $p(x_i)$, then for any real-valued function g:*

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

If a and b are constants, then

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

Expected value is also called the **mean** or the **first moment of X**.

The quantity $E[X^n], n \geq 1$ is called the **nth moment of X**.

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

Maximizing Expected Value: Find where the derivative equals 0

4.5   Variance

Variance: If X is a random variable with a mean μ then:

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

alternatively:

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

a useful identity for any constants a and b:

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

Standard Deviation:

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

4.6   The Bernoulli and Binomial Random Variables

$$\begin{aligned} p(0) = PX = 0 = 1 - p \\ p(1) = PX = 1 = p \end{aligned} \tag{4.6.1}$$

Bernoulli random variable: An r.v. with a PMF given by 4.6.1 Has parameters $(1, p)$

Binomial random variable: n independent Bernoulli trials, X represents number of successes and is an r.v. with parameters (n, p)

The PMF of a binomial r.v.:

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i} \qquad i = 0, 1, \ldots, n \tag{4.6.2}$$

4.6.1   Properties of Binomial Random Variables

$E[X] = np$

$Var(X) = np(1 - p)$

Proposition 4.6.1. *X is an r.v. with parameters (n, p). As k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer $\leq (n + 1)p$.*

Stirling's Approximation (for large k):

$$k! \sim k^{k+1/2} e^{-k} \sqrt{2\pi}$$

4.6.2   Computing the Binomial Distribution Function

Distribution function given X is binomial with parameters (n, p).

$$P\{X \leq i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k} \qquad i = 0, 1, \ldots, n$$

To compute the distribution function, utilize the relationship between $P\{X = k + 1\}$ and $P\{X = k\}$

$$P(X = k + 1) = \frac{p}{1-p} \frac{n-k}{k+1} P\{X = k\} \qquad (4.6.3)$$

4.7   The Poisson Random Variable

Poisson r.v.(with parameter $\lambda > 0$):

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!} \qquad i = 0, 1, 2, \ldots \qquad (4.7.1)$$

The Poisson r.v. with parameter $\lambda = np$ (the expected number of successes) may be used as an approximation for a binomial r.v. when n is large and p is small enough so that np is of moderate size.

$Var(X)$ and $E(X)$ are both equal to λ

Poisson Paradigm: Consider n events with p_i equal to the probability that event i occurs, $i = 1, \ldots, n$. If all the p_i are "small" and the trials are either independent or at most "weakly dependent," then the number of these events that occur approximately has a Poisson distribution with mean $\sum_{i=1}^{n} p_i$.

Basically it is not essential that all the events have the same probability of occurrence, but only that all of these probabilities be small.

Another use of the Poisson probability distribution arises in situations where "events" occur at certain points in time

Assume the following:

1. The Probability that exactly 1 event occurs in a given interval of length h is equal to $\lambda h + o(h)$
   $o(h)$ stands for any function $f(h)$ for which $\lim_{h \to 0} f(h)/h = 0$

2. The probability that 2 or more events occur in an interval of length h is equal to $o(h)$

3. For any integers $n, j_1, j_2, \ldots, j_n$ and any set of n nonoverlapping intervals, if we define E_i to be the even that exactly j_i of the events under consideration occur in the ith of these intervals, then events $E_1, E_2, \ldots, E_n$ are independent.

Assumptions 1 and 2 state that, for small values of h, the probability that exactly 1 event occurs in an interval of size h equals λh plus something that is small compared with h, whereas the probability that 2 or more events occur is small compared with h.

Assumption 3 states that whatever occurs in one interval has no (probability) effect on what will occur in other, nonoverlapping intervals.

(4.7.2) is obtained by defining the number of events occurring in any interval of length t as $N(t)$ and breaking the interval into n nonoverlapping subintervals, each of length t/n.

$P\{N(t) = k\} = P\{$k of the n subintervals contain exactly 1 event and other n-k contain 0 events$\}$

$+ P\{N(t) = k$ and at least 1 subinterval contains 2 or more events$\}$ (4.7.2)

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0, 1, \ldots \qquad (4.7.5)$$

4.7.1   Computing the Poisson Distribution Function

If X is Poisson with parameter λ, then:

$$\frac{P\{X = i + 1\}}{P\{X = i\}} = \frac{e^{-\lambda} \lambda^{i+1}/(i+1)!}{e^{-\lambda} \lambda^i/i!} = \frac{\lambda}{i+1} \qquad (4.7.6)$$

We can use (4.7.6) to compute successively:

$$P\{X = 1\} = \lambda P\{X = 0\}$$

$$P\{X = 2\} = \frac{\lambda}{2} P\{X = 1\}$$

$$\vdots$$

$$P\{X = i + 1\} = \frac{\lambda}{i+1} P\{X = i\}$$

4.8   Other Discrete Probability Distributions

4.8.1   The Geometric Random Variable

Suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. If X equals the number of trials required, then:

$$P\{X = n\} = (1 - p)^{n-1} p \qquad n = 1, 2, \ldots \qquad (4.8.1)$$

In order for $X = n$ the first $n - 1$ trials are failures and the nth trial is a success.

Any r.v. X whose PMF is given by (4.8.1) is a **geometric** r.v. with parameter p.

The probability that at least k trials are necessary to obtain a success is equal to the probability that the first $k-1$ trials are all failures:

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

Expected value of a geometric r.v. $E[X] = \dfrac{1}{p}$

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

4.8.2   The Negative Binomial Random Variable

Suppose that independent trials, each having probability p, $0 < p < 1$, of being a success are performed until a total of r successes is accumulated. X equals the number of required trials, then:

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \qquad n = r, r+1, \ldots \tag{4.8.2}$$

Any r.v. X whose PMF is given by (4.8.2) is a negative binomial r.v. with parameters (r, p)

The geometric r.v. is just a negative binomial with parameter $(1, p)$

The probability of r successes occuring before m failures:

$$\sum_{n=r}^{r+m-1} \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Expected value and variance of negative binomial:

$$E[X] = \frac{r}{p}$$

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

4.8.3   The Hypergeometric Random Variable

Suppose that a sample of size n is to be chosen randomly (without replacement) from an urn containing N balls, of which m are white and $N - m$ are black. X is the number of white balls selected.

$$PX = i = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad i = 0, 1, \ldots, n \tag{4.8.4}$$

An r.v. X whose PMF is given by (4.8.4) for some values of n, N, M is a hypergeometric r.v.

Maximum Likelihood Estimate: Parameter values are found such that they maximize the likelihood that the process described by the model produced the data that was actually observed.

$P_i(N)$ is first increasing and then decreasing, and reaches its maximum value at the largest integral value not exceeding mn/i. This value is the *maximum likelihood estimate* of N.

Expected Value and Variance

$$E[X] = \frac{nm}{N}$$

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

Let $p = m/N$ and rewrite as:

$$Var(X) = np(1-p)(1 - \frac{n-1}{N-1})$$

4.8.4   The Zeta (or Zipf) Distribution

$$P\{X = k\} = \frac{C}{k^{\alpha+1}} \qquad k = 1, 2, \ldots$$

for some $\alpha > 0$

$$C = \left[\sum_{k=1}^{\infty}(\frac{1}{k})^{\alpha+1}\right]^{-1}$$

4.9   Expected Value of Sums of Random Variables

The expected value of a sum of random variables is equal to the sum of their expectations.

In sample space S, for an r.v. X, lets $X(s)$ be the value of X when $s \in S$ is the outcome of the experiment. If X and Y are both r.v.s, then so is their sum. That is,
$Z = X + Y$ is an r.v. and $Z(s) = X(s) + Y(s)$

Let $p(s) = P(\{s\})$ be the probability that s is the outcome of an experiment. Because we can write any event A as the finite or countably infinite union of mutually exclusive events $\{s\}, s \in A$ we have:

$$P(A) = \sum_{s \in A} p(s)$$

Proposition 4.9.1.

$$E[X] = \sum_{s \in S} X(s)p(s)$$

For r.v.s $X_1, X_2, \ldots, X_n$

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

4.10   Properties of the Cumulative Distribution Function

For the distribution function F of X, $F(b)$ denotes the probability that the r.v. X takes on a value that is less than or equal to b.

Properties of the CDF F:

1. F is a nondecreasing function, if $a < b$, then $F(a) \leq F(b)$

2. $\lim_{b \to \infty} = 1$

3. $\lim_{b \to -\infty} F(b) = 0$

4. F is right continuous. For any b and any decreasing sequence $b_n, n \geq 1$ that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$

$$P\{x < X \leq b\} = F(b) - F(a) \qquad \text{for all } a < b$$

$$P\{X < b\} = \lim_{n \to \infty} F\left(b - \frac{1}{n}\right)$$

Chapter 5

Continuous Random Variables

5.1 Introduction

X is a **continuous random variable** if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers:

$$P\{X \in B\} = \int_B f(x) \ dx \tag{5.1.1}$$

The function f is called the **probability density function** of X

(5.1.1) states that the probability that X will be in B can be obtained by integrating the pdf over the set B. Since X must assume some value, f must satisfy:

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) \ dx$$

Letting $B = [a, b]$ we obtain:

$$P\{a \leq X \leq b\} = \int_a^b f(x) \ dx \tag{5.1.2}$$

The probability that a c.r.v. will assume any fixed value is 0.

$$P\{X < a\} = P\{X \leq a\} = F(a) = \int_{-\infty}^a f(x) \ dx$$

The relationship between the cumulative distribution F and the probability distribution f is:

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) \ dx$$

Differentiate both sides yields:

$$\frac{d}{da}F(a) = f(a)$$

The density is the derivative of the cdf.

$f(a)$ is a measure of how likely it is that the r.v. will be near a.

5.2 Expectation and Variance of Continuous Random Variables

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

Proposition 5.2.1. *If X is a continuous r.v. with pdf $f(x)$, then for any real-valued function g:*

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

Lemma 2.1
For a nonnegative r.v. Y:

$$E[Y] = \int_{0}^{\infty} P\{Y > y\} \ dy$$

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

Variance

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

5.3 The Uniform Random Variable

An r.v. is said to be **uniformly** distributed over the interval $(0, 1)$ if its pdf is given by:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.3.1}$$

The probability that X is in any particular subinterval of $(0, 1)$ equals the length of that subinterval

In general X is a uniform r.v. on the interval (α, β) if the pdf of X is given by:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases} \tag{5.3.2}$$

The distribution function of a uniform r.v. on the interval (α, β) is given by:

$$F(a) = \begin{cases} 0 & a \leq \alpha \\ \dfrac{a - \alpha}{\beta - \alpha} & \alpha < a < \beta \\ 1 & a \geq \beta \end{cases}$$

$$E[X] = \frac{\beta + \alpha}{2}$$

The expected value of an r.v. that is uniformly distributed over some interval is equal to the midpoint of that interval.

$$Var(X) = \frac{(\beta - \alpha)^2}{12}$$

The variance is the square of the length of the interval divided by 12.

5.4   Normal Random Variables

We say that X is a **normal random variable** or X is **normally distributed** with parameters μ and σ^2 if the density of X is given by:

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

The density function is a bell-shaped curve that is symmetric about μ.

If X is normally distributed with parameters μ and σ^2, then $Z = (X - \mu)/\sigma$ is normally distributed with parameters 0 and 1.
Such an r.v. is said to be a **standard,** or a **unit**, normal variable.

$$E[X] = \mu$$

$$Var(X) = \sigma^2$$

The cdf of a standard normal r.v. is denoted by $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \, dy$$

Use the table to get nonnegative values of $\Phi(x)$

TABLE 5.1: AREA $\Phi(x)$ UNDER THE STANDARD NORMAL CURVE TO THE LEFT OF X

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Negative values of x, $\Phi(x)$ can be obtained from the relationship:

$$\Phi(-x) = 1 - \Phi(x) \qquad -\infty < x < \infty \tag{5.4.1}$$

(5.4.1) states that if Z is a standard normal r.v. then:

$$P\{Z \le -x\} = P\{Z > x\} \qquad -\infty < x < \infty$$

Since $Z = (X - \mu)/\sigma$ is a standard normal r.v. whenever X is normally distributed with parameters μ and σ^2, it follows that the distribution function of X can be expressed as:

$$F_X(a) = P\{X \le a\} = P\left(\frac{X - \mu}{\sigma} \le \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

5.4.1   The Normal Approximation to the Binomial Distribution

The **DeMoivre-Laplace limit theorem** states that when n is large, a binomial r.v. with parameters n and p will have approximately the same distribution as a normal r.v. with the same mean and variance as the binomial.

If we "standardize" the binomial by first subtracting its mean np and then dividing the result by its standard deviation $\sqrt{np(1-p)}$, then the distribution function of this standardized r.v. (which has mean 0 and variance 1) will converge to the standard normal distribution function as $n \to \infty$

The DeMoivre-Laplace limit theorem
If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are preformed, then, for any $a < b$:

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} \to \Phi(b) - \Phi(a)$$

as $n \to \infty$

The normal approximation will, in general, be good for values of n satisfying $np(1-p) \ge 10$

Continuity Correction
Since the binomial is discrete and the normal is continuous, in order to use the normal approximation you must:

$$\text{write } P\{X = i\} \text{ as } P\{i - 1/2 < X < i + 1/2\}$$

5.5   Exponential Random Variables

Exponential random variable: with parameter λ whose probability density function for some $\lambda > 0$ is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The cumulative distribution function $F(a)$ of an exponential r.v. is:

$$F(a) = P\{X \leq a\}$$
$$= \int_0^a \lambda e^{-\lambda x} \, dx$$
$$= -e^{-\lambda x}\big|_0^a$$
$$= 1 - e^{-\lambda a} \quad a \geq 0$$

$$E[X] = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda}$$

We say that a nonnegative r.v. X is **memoryless** if:

$$P\{X > s + t | X > t\} = P\{X > s\} \quad \text{for all } s, t \geq 0 \tag{5.5.1}$$

Think of X as the lifetime of some instrument, (5.5.1) states that the probability that the instrument survives for at least $s + t$ hours, given that it has survived t hours, is the same as the initial probability that it survives for at least s hours. Basically if the instrument is alive at age t, the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution. The instrument does not "remember" that it has already been in use for a time t.

(5.5.1) is equivalent to:

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

$$P\{X > s + t\} = P\{X > s\}P\{X > t\} \tag{5.5.2}$$

Exponentially distributed r.v.s are memoryless.

$$g(s + t) = g(s)g(t)$$

The only right continuous solution of this functional equation is:

$$g(x) = e^{-\lambda x} \tag{5.5.3}$$

Since a distribution function is always right continuous, we have:

$$\overline{F}(x) = e^{-\lambda x} \quad \text{or} \quad F(X) = P\{X \leq x\} = 1 - e^{-\lambda x}$$

Laplace distribution: A variation of the exponential distribution in which an r.v. that is equally likely to be positive or negative and whose absolute value is exponentially distributed with parameter $\lambda, \lambda \geq 0$.

Its density is given by:

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \qquad -\infty < x < \infty$$

Its distribution function is given by:

$$F(x) = \begin{cases} \frac{1}{2} \int_{-\infty}^{x} \lambda e^{\lambda x}\, dx & x < 0 \\[2mm] \frac{1}{2} \int_{-\infty}^{0} \lambda e^{\lambda x}\, dx + \frac{1}{2} \int_{0}^{x} \lambda e^{-\lambda x}\, dx & x > 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{\lambda x} & x < 0 \\[2mm] 1 - \frac{1}{2} e^{-\lambda x} & x > 0 \end{cases}$$

5.5.1   Hazard Rate Functions

X is a positive continuous r.v. that is the lifetime of some item. Let X have distribution function F and density f.

The **hazard rate**(also called the *failure rate*) function $\lambda(t)$ of F is defined by:

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}, \quad \text{where } \overline{F} = 1 - F$$

To interpret $\lambda(t)$, suppose that the item has survived for a time t and we desire the probability that it will not survive for an additional time dt. That is, $P\{X \in (t, t+dt)|X > t\}$

which $\approx \dfrac{f(t)}{\overline{F}(t)} dt$

Thus $\lambda(t)$ represents the conditional probability intensity that a t-unit-old item will fail.

If the lifetime distribution is exponential, then the distribution of the remaining life for a t-year-old item is the same as that for a new item. Which means $\lambda(t) = \lambda$

Thus the failure rate function for the exponential distribution is constant

λ is the rate of the distribution

A distribution function of a positive c.r.v. can be specified by giving its hazard rate function:

$$F(t) = 1 - exp\left\{ -\int_{0}^{t} \lambda(t)\, dt \right\} \tag{5.5.4}$$

If an r.v. has a linear hazard rate function, that is if:

$$\lambda(t) = a + bt$$

Then:

$$F(t) = 1 - e^{-at - bt^2/2}$$

and differentiation yields its density:

$$f(t) = (a + bt)e^{-(at + bt^2/2)} \quad t \geq 0$$

When $a = 0$, the preceding equation is known as the *Rayleigh density function.*

5.6 Other Continuous Distributions

5.6.1 The Gamma Distribution

An r.v. has a **gamma distribution** with parameters (α, λ), $\lambda < 0, \alpha > 0$, if its density function is given by:

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha - 1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$\Gamma(\alpha)$ is the **gamma function**:

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

Integration by parts yields:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \tag{5.6.1}$$

For integral values of α say $\alpha = n$, we obtain by applying (5.6.1) repeatedly:

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

When α is a positive integer, $\alpha = n$, the gamma distribution with parameters (α, λ) often arises as the distribution of the amount of time one has to wait until a total of n evens has occurred.

If events are occurring randomly, the amount of time one has to wait until a total of n events has occurred will be a gamma r.v. with parameters (n, λ).

Density function of T_n (the time its at which the nth even occurs):

$$f(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{n - 1}}{(n - 1)!}$$

Expected Value and Variance:

$$E[X] = \frac{\alpha}{\lambda}$$

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

5.6.2   The Weibull Distribution

The Weilbull distribution provides a close approximation to the distribution of the lifetime of an item, when the item dies6 when any of its parts fail.

The Weibull distribution Function:

$$F(x) = \begin{cases} 0 & x \leq v \\ 1 - exp\left\{-\left(\dfrac{x - v}{\alpha}\right)^{\beta}\right\} & x > v \end{cases} \tag{5.6.2}$$

An r.v. whose cdf is given by (5.6.2) is a *Weibull r.v.* with parameters v, α, β

Weibull Density:

$$f(x) = \begin{cases} 0 & x \leq v \\ \dfrac{\beta}{\alpha}\left(\dfrac{x - v}{\alpha}\right)^{\beta - 1} exp\left\{-\left(\dfrac{x - v}{\alpha}\right)^{\beta}\right\} & x > v \end{cases}$$

5.6.3   The Cauchy Distribution

An r.v. has a **Cauchy distribution** with parameter $\theta, \infty < \theta < \infty$, if its density is given by:

$$f(x) = \frac{1}{\pi}\frac{1}{1 + (x - \theta)^2} \quad -\infty < x < \infty$$

5.6.4   The Beta Distribution

Density of a **beta distribution**:

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

where

$$B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1} \ dx$$

The beta distribution can be used to model a random phenomenon whose set of possible values is some finite interval $[c, d]$, which by letting c denote the origin and taking $d - c$ as a unit measurement can be transformed into the interval $[0, 1]$

When $a = b$ the beta density is symmetric about $\frac{1}{2}$, giving more and more weight to regions about $\frac{1}{2}$ as the common value a increases.

When $b > a$, the density is skewed to the left, skewed to right when $a > b$.

The relationship

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{5.6.3}$$

can be shown to exist between

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\ dx$$

and the gamma function

Expected Value and Variance of the Beta Distribution:

$$E[X] = \frac{a}{a+b}$$

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

5.7 The Distribution of a Function of a Random Variable

Suppose we know the distribution of X and want to find the distribution of $g(X)$.

We need to express the event $g(X) \leq y$ in terms of X being in some set.

Theorem 5.7.1: Let X be a c.r.v. having probability density function f_X. Suppose that $g(x)$ is a strictly monotonic differentiable (and thus continuous) function of x. Then the r.v. Y defined by $Y = g(X)$ has a probability density function given by:

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

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

Chapter 6

Jointly Distributed Random Variables

6.1   Joint Distribution Functions

Joint Cumulative Probability Distribution of X and Y:

$$F(a,b) = P\{X \leq a, Y \leq b\} \quad -\infty < a, b < \infty$$

Marginal Distributions of X and Y:

$F_X = F(a, \infty)$
$F_Y = F(\infty, b)$

Joint Probability Mass Function of X and Y:

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

To obtain marginal pmf of X or Y:

$$p_X(x) = P\{X = x\}$$
$$= \sum_{y:p(x,y)>0} p(x,y)$$

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y)$$

X and Y are **jointly continuous** if there exists a function $f(x,y)$ defined for all real x and y having the property that for every set C of pairs of real numbers:

$$P\{(X,Y) \in C\} = \iint\limits_{(x,y) \in C} f(x,y) \; dx \; dy$$

Joint Probability Density Function: $f(x,y)$

To find the joint cumulative distribution function from $f(x,y)$:
If A and B are any sets of real numbers, then by defining
$C = \{(x,y) : x \in A, y \in B\}$, we have:

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

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\}$$
$$= \int_A \int_{-\infty}^{\infty} f(x,y)\ dy\ dx$$
$$= \int_A f_X(x)\ dx$$

where

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

is thus the probability density function of X. Similarly, the probability density function of Y is given by

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

6.2   Independent Random Variables

The r.v.s X and Y are **independent**, if for any two sets of real numbers A and B:

$$P\{X \in A, Y \in B\} = P\{X \in A\}\,P\{Y \in B\}$$

The above equation follows if and only if:

$$P\{X \le a, Y \le b\} = P\{X \le a\}\,P\{Y \le b\}$$

In terms of the joint distribution function, X and Y are independent if:

$$F(a,b) = F_X(a)F_Y(b) \quad \text{for all } a, b$$

If X and Y are discreet r.v.s they are independent if:

$$p(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$

In the jointly continuous case, the condition of independence is equivalent to:

$$f(x,y) = f_X(x)f_Y(y) \quad \text{for all } x, y$$

X and Y are independent if knowing the value of one does not change the distribution of the other. If they are not independent, they are **dependent**.

A necessary and sufficient condition for the r.v.s X and Y to be independent is for their joint pdf $f(x,y)$ to factor into two terms, one depending only on x and the other only on y.

$$f_{X,Y}(x,y) = h(x)g(y) \quad -\infty < x < \infty, -\infty < y < \infty$$

6.3 Sums of Independent Random Variables

The cumulative distribution function F_{X+Y} is called the **convolution** of the distributions F_X and F_Y, where X and Y are independent.

$$
\begin{aligned}
F_{X+Y}(a) &= P\{X + Y \leq a\} \\
&= \iint\limits_{x+y \leq a} f_X(x) f_Y(y) \, dx \, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) \, dx \, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) \, dx f_Y(y) \, dy \\
&= \int_{-\infty}^{\infty} F_X(a - y) f_Y(y) \, dy
\end{aligned}
$$

Differentiate to get the pdf f_{X+Y} of $X + Y$:

$$
\begin{aligned}
f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy \qquad (3.2) \\
&= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy
\end{aligned}
$$

6.3.1 Identically Distributed Uniform Random Variables

$X_1, X_2, \ldots, X_n$ are independent uniform $(0,1)$ r.v.s, and:

$$F_n(x) = P\{X_1 + \cdots + X_n \leq x\}$$

Then:

$$F_n(x) = x^n / n! \, , \quad 0 \leq x \leq 1$$

The mean number of uniform $(0,1)$ r.v.s that must be summed for the sum to exceed 1 is equal to e

6.3.2 Gamma Random Variables

A gamma r.v. has a density of the form:

$$f(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)} \quad 0 < y < \infty$$

Proposition 3.1 For a fixed value of λ, the gamma distribution is closed under convolutions.

If X and Y are independent gamma r.v.s with respective parameters (s, λ) and (t, λ), then $X + Y$ is a gamma r.v. with parameters $(s + t, \lambda)$.

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a}(\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

6.3.3 Normal Random Variables

Proposition 3.2 If $X_i, i = 1, \ldots, n$ are independent r.v.s that are normally distributed with respective parameters $\mu_i, \sigma_i^2, i = 1, \ldots, n$ then $\sum_{i=1}^{n} X_i$ is normally distributed with parameters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$

Y is a **lognormal** r.v. with parameters μ and σ if $\log(Y)$ is a normal r.v. with mean μ and variance σ^2. Y is lognormal if it can be expressed as:

$$Y = e^X$$

where X is a normal r.v.

6.3.4 Poisson and Binomial Random Variables

6.3.5 Geometric Random Variables

Proposition 3.3 Let $X_1, \ldots, X_n$ be independent geometric r.v.s, with X_i having a parameter p_i for $i = 1, \ldots, n$. If all the p_i are distinct, then, for $k \geq n$:

$$P\{S_n = k\} = \sum_{i=1}^{n} p_i q_i^{k-1} \prod_{j \neq i} \frac{p_j}{p_j - p_i}$$

6.4 Conditional Distributions: Discrete Case

Conditional PMF of X given $Y = y$:

$$\begin{aligned}
p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\
&= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\
&= \frac{p(x, y)}{p_Y(y)}
\end{aligned}$$

Conditional PDF of X given that $Y = y$, for all y such that $p_Y(y) > 0$:

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

These are the same as the unconditional case, except that everything is now conditional on the event that $Y = y$. If X is independent of Y, then the conditional mass function and the distribution function are the same as the unconditional ones.

6.5   Conditional Distributions: Continuous Case

Conditional probability density function of X given $Y = y$, $f_Y(y) > 0$:

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

If X and Y are jointly continuous, then for any set A:

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

Let $A = (-\infty, a]$, then the conditional cdf of X given that $Y = y$ is:

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

If X and Y are independent, continuous r.v.s, the conditional density of X given that $Y = y$ is just the unconditional density of X.

Conditional distributions when the r.v.s are neither jointly continuous nor jointly discrete
If X is a continuous r.v. having a pdf f and N is a discrete r.v., the conditional distribution of X given that $N = n$ is:

$$f_{X|N}(x|n) = \frac{P\{N = n|X = x\}}{P\{N = n\}} f(x)$$

6.6   Order Statistics

Let $X_1, X_2, \ldots, X_n$ be n independent and identically distributed continuous r.v.s having a common density f and distribution function F.
Define $X_{(1)}$ as the smallest, increasing with $X_{(n)}$ as the largest.

The ordered values $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are known as the **order statistics** corresponding to the r.v.s $X_1, X_2, \ldots, X_n$.

Joint density function of the order statistics:

$$f_{X(1), \ldots, X(n)}(x_1, x_2, \ldots, x_n) = n! f(x_1) \cdots f(x_n) \quad x_1 < \cdots < x_n$$