1 Combinatorial Analysis

The *basic principle of counting*: If an experiment consists of two phases, with n possible outcomes in the first phase, and m possible outcomes in the second phase for each of those outcomes, then the total possible number of outcomes is nm.

There are $n!$ possible linear orderings of n items.

The number of distinct ways to choose k items from n without regard to order is given by

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

For non-negative integers $n_1, n_2, \ldots, n_r$ with $\sum n_i = n$, the number distinct of ways to divide n items into non-overlapping groups of sizes $n_1, n_2, \ldots, n_r$ is given by

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

Stars and Bars

For any positive integers n and k, the number of k-tuples of positive integers that sum to n is $\binom{n-1}{k-1}$

For any positive integers n and k, the number of k-tuples of non-negative integers that sum to n is $\binom{n+k-1}{k-1}$

2 Axioms of Probability

Some notation and terminology:

- S is the set of all possible outcomes of an experiment, a.k.a. the *sample space*

- An *event* is a subset of S

- For any event A we define A^c to be the complement of A in S

- $S^c = \emptyset$ is the null set

- $A \cap B$ or sometimes AB is the intersection of the sets A and B. If $A \cap B = \emptyset$ then we say that A and B are mutually exclusive

Axioms

If E is an event in a sample space S then the probability that event E occurs as the outcome of a single experiment is denoted by $P(E)$ and satisfies the following axioms

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

2. $P(S) = 1$

3. If $\{E_i\}_{i=1}^{\infty}$ is a set of pairwise mutually exclusive events then $P(\cup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} P(E_i)$

Useful Results

$$P(A^c) = 1 - P(A)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Informally, the generalization of this formula: add up all of the individual event probabilities, then subtract all of the double counted intersections, then add back all of the twice removed triple intersections, then remove all of the twice restored double counted quadruple intersections, etc.

If S is finite and every point in S is assumed to have equal probability of occurring then

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

3 Conditional Probability and Independence

Conditional Probability

For events E and F, the conditional probability of E occurs, given that F has occurred is denoted by $P(E|F)$ and is defined by

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(F|E)P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)}$$

where the last expression is *Bayes's Formula*

Independence

Two events are said to be *independent* if $P(E \cap F) = P(E)P(F)$. Events that are not independent are said to be *dependent*. This generalizes to sets of events if every the probability of intersection of subsets is equal to the product of the probabilities of the individual events in the subset.

4 Random Variables

Random Variables

Given an experiment with sample space S, a *random variable* is a real-valued function, X, defined on S. Because the value of a random variable is determined by the outcome of the experiment, we can assign probabilities to the possible values of the random variable, denoted by $P(X = x)$.

Discrete Random Variables

If the range of X is countable, $\{x_1, x_2, \ldots\}$, then X is said to be discrete. In this case we can define the *probability mass function* $p : \mathbb{R} \to \mathbb{R}$ of X to be $p(a) = P(X = a)$ which has the following properties:

1. $p(x_i) \geq 0$ for $i = 1, 2, \ldots$
2. $p(x) = 0$ for any other values of x
3. $\sum_{i=1}^{\infty} p(x_i) = 1$

Expected Value

If X is a discrete random variable with probability mass function $p(x)$ then the *expected value* of X, denoted as $E[X]$ is defined by

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

Some useful properties, for a random variable X with non-zero outputs $\{x_1, x_2, \ldots\}$:

- If g is any real value function then $E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p(x_i)$
- If a and b are constants, then $E[aX + b] = aE[X] + b$

Variance, Standard Deviation, Covariance, and Correlation

If X is a random variable with *mean* $\mu = E[X]$ then the variance of X, denoted by $\text{Var}(X)$, is defined to be $\text{Var}(X) = E[(X - \mu)^2]$ Some useful formulas that follow from this definition:

- $\text{Var}(X) = E[X^2] - (E[X])^2$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$

The *standard deviation* of X is the square root of the variance: $\sigma = \sqrt{\text{Var}(X)}$
Given random variables, X and Y:

- $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ and
  - If X and Y are independent, then $E[XY] = E[X]E[Y]$, so $\text{Cov}(X, Y) = 0$
  - $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
  - $\text{Cov}(aX + bY, cZ) = ac\text{Cov}(X, Z) + bc\text{Cov}(Y, Z)$
- $\text{Cor}(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$

Units of covariance are the products of the units of X and Y, while correlation is unit-less.

Bernoulli Random Variable

A Bernoulli trial is an experiment that has two outcomes: *success* or *failure*. The sample space is $S = \{\text{success}, \text{failure}\}$ and a *Bernoulli random variable*, X, is defined by $X(\text{success}) = 1$ and $X(\text{failure}) = 0$, along with the probability mass function is given by

$$p(0) = P(X = 0) = 1 - p \qquad\qquad p(0) = P(X = 1) = p$$

where p is the probability of success in a given trial.

Binomial Random Variable

A *binomial random variable* is a random variable X that represents the number of successes that occur in n Bernoulli trials, each with probability of success p.
Properties of Binomial Random Variables:

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $E[X] = np$
- $\text{Var}(X) = np(1-p)$

Poisson Random Variable

A *Poisson Random Variable* is a random variable X that takes on one of the values $0, 1, 2, \ldots$, or the number of successes that occur in a given interval, together with a parameter $\lambda > 0$ that describes the expected number of success in that interval.
Properties of Poisson Random Variables:

- $P(X = k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$
- $E[X] = \lambda$
- $\text{Var}(X) = \lambda$

Geometric Random Variable

A *Geometric Random Variable* is a random variable X that takes on one of the values $0, 1, 2, \ldots$, or the number of trials required until a success occurs, when repeatedly performing a Bernoulli trial with probability of success p.
Properties of Geometric Random Variables:

- $P(X = k) = p(1-p)^{k-1}$
- $E[X] = \frac{1}{p}$
- $\text{Var}(X) = \frac{1-p}{p^2}$

Negative Binomial Random Variable

Suppose that independent Bernoulli trials, with proability of success p, are perfomed until r successes occur. A *negative binomial random variable* is a random variable X that takes on one of the values $0, 1, 2, \ldots$, or the number of trials required. Properties of Negative Binomial Random Variables:

- $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
- $E[X] = \frac{r}{p}$
- $\text{Var}(X) = \frac{r(1-p)}{p^2}$

Hypergeometric Random Variables

A *Hypergeometric Random Variable* is a random variable X that takes on one of the values $0, 1, 2, \ldots$, describing the number of successes in n draws, without replacement, from a finite population of size N, where exactly K members are successes. (In contrast, a binomial random variable can be thought of as describing the probability of k successes with replacement.)
Properties of Hypergeometric Random Variables:

- $P(X = k) = \dfrac{\binom{K}{k}\binom{N-k}{n-k}}{\binom{N}{n}}$
- $E[X] = \frac{nK}{N}$
- $\text{Var}(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{n-1}$

Note: $\text{Var}(X) \approx n(p)(1-p)$ when N is large relative to n, where $p = K/N$.

Expected Value and Sums of Random Variables

If $X_1, X_2, \ldots, X_n$ are random variables then

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

5 Continuous Random Variables

We say that X is a *continuous random variable* if there exists a non-negative function f, defined on $\mathbb{R}$, with the property that, for any measurable set B of real numbers, $P\{X \in B\} = \int_B f(x)\,dx$, and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. The function f is called the *probability density function* of the random variable X.

The *cumulative distribution function*, or just *distribution function* of a continuous random variable is $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$. Notice that the Fundamental Theorem of Calculus then gives us that $f(x) = \frac{d}{dx}F(x)$.

Expectation and Variance

If X is a continuous random variable with probability density function $f(x)$ then the *expected value* of X, denoted as $E[X]$ is defined by

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

A useful property that follows from this definition is that, for any function g, $E[g(X)] = E[X] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$

If X is a random variable with *mean* $\mu = E[X]$ then the variance of X, denoted by $\mathrm{Var}(X)$, is defined to be $\mathrm{Var}(X) = E[(X - \mu)^2]$ This definition is identical to the discrete case, and the same properties hold:

- $\mathrm{Var}(X) = E[X^2] - (E[X])^2$
- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(x)$

Uniform Random Variable

In general, we say that X is a *Uniform Random Variable* on the interval (α, β) if the probability of outcomes is even distributed among all outcomes in that interval. Properties of a Uniform Random Variable:

- $f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$
- $E(x) = \dfrac{\alpha + \beta}{2}$
- $\mathrm{Var}(X) = \dfrac{(\beta - \alpha)^2}{12}$

Normal Random Variable

In general, we say that X is a *Normal Random Variable* if the probability of outcomes is normally distributed with parameters μ and σ^2. Properties of a Normal Random Variabls:

- $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
- $E(x) = \mu$
- $\mathrm{Var}(X) = \sigma^2$

If X is normal, with mean μ and variance σ^2 then $Z = \frac{X-\mu}{\sigma}$ is a *standard* normal random variable with mean 0 and variance 1.

A normal distribution can give a good approximation for a binomial random variable when $np(1-p) \geq 10$.

Exponential Random Variable

An *Exponential Random Variable* is a random variable whose probability distribution models the distance between evens in a Poisson point process. Properties of an Exponential Random Variable:

- $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
- $E(x) = \dfrac{1}{\lambda}$
- $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$

A key property of exponential random variables is that they are *memoryless*, so $P\{X > s + t \mid X > t\} = P\{X > s\}$

Applications and Other Common Distributions

Failure Rate

Let X be a non-negative random variable that represents rate of survivors to time t during the next instant in time, with probability density function $f(t)$ and cumulative density function $F(t)$, then the *failure rate* (or *hazard rate*) is defined as $h(t) = \dfrac{f(t)}{1 - F(t)}$. For small values of dt, $h(t)\,dt$ is approximately the probability that a t-unit-old item will die within the next dt-time.

In the special case that F is the exponential distribution with parameter λ, then the failure rate becomes constant with $h(t) = \gamma$.

Gamma Distribution

A *gamma distribution* with shape parameter α and scale parameter λ is defined in terms of the gamma function: $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$.
Gamma random variables are commonly used in modeling waiting times, or the distance between events in a Poisson point process.
Properties of a Gamma Random Variable:

- $f(x) = \dfrac{\lambda e^{\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}$
- $E(x) = \dfrac{\alpha}{\lambda}$
- $\text{Var}(X) = \frac{\alpha}{\lambda^2}$

Beta Distribution

A *beta distribution* with parameters a and b is defined in terms of the beta function: $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx$.

Beta random variables are commonly used in modeling the behavior of random variables limited to intervals of finite length, or the randome behavior of percentages or proportions.
Properties of a Beta Random Variable for $0 \leq x \leq 1$:

- $f(x) = \dfrac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$
- $E(x) = \dfrac{a}{a+b}$
- $\text{Var}(X) = \dfrac{ab}{(a+b)^2(a+b+1)}$

When $a = b$ the beta density is symmetric about $\frac{1}{2}$, and when $b > a$ the density is skewed to the left (meaning smaller values are more likely).

One connection between gamma and beta distributions is that $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

6   Limit Theorems

Chebyshev's Inequality and the Weak Law of Large Numbers

Theorem (Markov's Inequality). *If X is a random variable that takes on only non-negative values, then, for any $a > 0$,*

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

Theorem (Chebyshev's Inequality). *If X is a random variable with finite mean μ and variance σ^2, then, for any value $k > 0$*

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Note: If $\bar{X}$ is the sample mean of n independent and identically distributed observations from a populations with finite mean μ and variance σ^2, then we can apply Chebyshev's Inequality with $E[\bar{X}] = \mu$ and $\text{Var}(\bar{X}) = \sigma^2/n$ to conclude that

$$P\{|\bar{X} - \mu| \geq k\} \leq \frac{\sigma^2}{nk^2}$$

As n becomes large, the upper bound on this probability reaches zero. This leads to the proof of the Weak Law of Large Numbers.

Theorem (The Weak Law of Large Numbers). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$ and finite variance σ^2. Then, for any $\epsilon > 0$,*

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

The Central Limit Theorem

Theorem (The Central Limit Theorem). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$ and finite 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 $n \to \infty$.

7 Additional Topics

The Poisson Process

A function f is said to be $o(h)$ if $\lim\limits_{h \to 0} \frac{f(h)}{h} = 0$. Intuitively, this means that for small values of h, $f(h)$ is small relative to h.

Definition. Consider events that are occurring at random points in time, and let $N(t)$ be the number of events that occurs in the interval $[0, t]$. The collection of random variables $\{N(t), t \geq 0\}$ is a *Poisson process having rate* $\lambda > 0$ if the folowing conditions are satisfied:

(i) $N(0) = 0$. The process starts at time 0.

(ii) The number of events that occur in disjoint time intervals are independent.

(iii) The distribution of the number of events that occur in a given interval depends only on the length of that interval (and not on its location)

(iv) $P\{N(h) = 1\} = \lambda h + o(h)$

(v) $P\{N(h) \geq 2\} = o(h)$