Math 461: Probability Theory

Lanxiao Hermite Bai

October 20, 2016

Contents

1 Combinatorial Analysis

1.1 Introduction

Combinatorial Analysis The mathematical theory of counting is formally known as **combinatorial analysis**.

1.2 The Basic Principle of Counting

Proposition 1.2.1 (Product Rule) *Suppose a procedure can be broken down into a sequence of k parts for each the number of possible results denoted as n_k, the number of the possible outcomes of the procedure*

$$N = \prod_{i=1}^{k} n_i$$

1.3 Permutations

Proposition 1.3.1 (Permutations)

$$P(n) = n! = n \cdot (n-1) \cdots 1 \tag{1}$$

1.4 Combinations

Definition 1.4.1 *We define $\left(\begin{array}{c} n \\ r \end{array} \right)$, for $r \leq n$, by*

$$\left(\begin{array}{c} n \\ r \end{array} \right) = \frac{n!}{(n-r!)r!} \tag{2}$$

and say that $\left(\begin{array}{c} n \\ r \end{array} \right)$ represents the number of possible combinations of n objects taken r at a time.

Proposition 1.4.1 (The Binomial Theorem)

$$(x+y)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^k y^{n-k} \tag{3}$$

1.5 Mutinomial Coefficient

Notation If $n = \sum_{i=1}^{r} n_i$, we define $\left(\begin{array}{c} n \\ n_1, n_2, \cdots, n_r \end{array} \right)$ by

$$\left(\begin{array}{c} n \\ n_1, n_2, \cdots, n_r \end{array} \right) = \frac{n!}{n_1! n_2! \cdots n_r!} \tag{4}$$

Thus, $\left(\begin{array}{c} n \\ n_1, n_2, \cdots, n_r \end{array} \right)$ represents the number of possible divisions of n distinct objects into r distinct groups of respective sizes $n_1, n_2, \cdots, n_r$.

Proposition 1.5.1 (The Multinomial Theorem)

$$(\sum_{i=1}^{r} x_r)^n = \sum_{(n_1, \cdots, n_r) \ : \ n = \sum_{i=1}^{r} n_i} \left(\begin{array}{c} n \\ n_1, n_2, \cdots, n_r \end{array} \right) \prod_{j=1}^{r} x_j^{n_j} \tag{5}$$

1.6 The number of Interger Solutions of Equations

Proposition 1.6.1 *There are* $\left(\begin{array}{c} n-1 \\ r-1 \end{array} \right)$ *distinct nonnegative integer-value vectors* $(x_1, x_2, \cdots, x_r)$ *satisfying the equation*

$$n = \sum_{i=1}^{r} n_i, x_i > 0, i = 1, \cdots, r \tag{6}$$

Proposition 1.6.2 *There are* $\left(\begin{array}{c} n+r-1 \\ r-1 \end{array} \right)$ *distinct nonnegative integer-value vectors* $(x_1, x_2, \cdots, x_r)$ *satisfying the equation*

$$n = \sum_{i=1}^{r} n_i \tag{7}$$

2 Axioms of Probability Theory

2.1 Sample Space and Events

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

Event Any subset $E \subseteq S$ is known as an event.

Union Event $E \cap F$ is called the **union** of E and F. Union of events of $E_1, E_2, ...$ is denoted by

$$\bigcup_{n=1}^{\infty} E_n$$

Intersection Event EF is called **Intersection** of E and F. Intersection of events of $E_1, E_2, ...$ is denoted by

$$\bigcap_{n=1}^{\infty} E_n$$

Complement Complement of E is denoted by E^C.

Contain Contained relationship is denoted by $E \subset F$. If $E \subset F$ and $F \subset E$, then $E = F$.

Rules

1. Commutative Laws $E \cup F = F \cup E$ and $EF = FE$
2. Associative Laws $(E \cap F) \cap G = F \cap (F \cap G)$ and $(EF)G = E(FG)$
3. Distributive Laws $(E \cap F)G = EG \cap FG$ and $EF \cap G = (E \cap G)(F \cap G)$
4. De Morgan's Laws

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

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

2.2 Axioms of Probability

Definition 2.2.1 *For each event E of the sample space S, we define $n(E)$ to be the number of times in the first n reptitions of the experiment that event E occurs. Then probability is defined as*

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n} \tag{8}$$

Axiom 2.2.1

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

Axiom 2.2.2

$$P(S) = 1$$

Axiom 2.2.3 *For any sequence of mutually exclusive events*

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

Proposition 2.2.1

$$P(E^C) = 1 - P(E)$$

Proposition 2.2.2 *If $E \subset F$, then $P(E) \leq P(F)$.*

Proposition 2.2.3

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

Proposition 2.2.4

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

2.3 Probability As A Continuous Set Function

Proposition 2.3.1 *If $E_n, n \geq 1$ is either an increasing or a decreasing sequence of events, then*

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

3 Conditional Probability and Independence

Definition 3.0.1 (Conditional Probability) *If $P(F) > 0$, then*

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

Proposition 3.0.1

$$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_2E_1)\cdots P(E_n|E_1\cdots E_{n-1})$$

3.1 Bayes's Formula

Proposition 3.1.1

$$P(E) = P(E|F)P(F) + P(E|F^C)[1 - P(F)]$$

Proposition 3.1.2 (Bayes's Theorem)

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

Definition 3.1.1 (Odd) *The odds of an event A are defined by*

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

That is, 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.

3.2 Independent Events

Definition 3.2.1 *Two events E and F are said to be independent if Equation*

$$P(EF) = P(E)P(F)$$

holds. Two events E and F that are not independent are said to be dependent.

Proposition 3.2.1 *If E and F are independent, then so are E and F^C.*

4 Random Variables

4.1 Random Variable

Definition 4.1.1 (Random Variable) *Real-valued functions defined on the sample space, are known as random variables.*

Proposition 4.1.1

$$1 = P(\bigcup_{i=0}^{n}\{Y = i\}) = \sum_{i=0}^{n} P\{Y = i\}$$

4.2 Discrete Random Variables

Definition 4.2.1 (Probability mass function)

$$p(a) = P(X = a)$$

Proposition 4.2.1

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

Definition 4.2.2 *Commutative distribute function*

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

4.3 Expected Value

Definition 4.3.1 (Expected Value)

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

4.4 Exceptation of A Function of A Random Variable

Proposition 4.4.1

$$E[g(X)] = \sum_{i} g(x_i)p(x_i)$$

Corollary 4.4.1

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

4.5 Variance

Definition 4.5.1 (Variance) *If X is a random variable with mean μ, then the **variance** of X is*

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

Proposition 4.5.1

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

Definition 4.5.2 (Standard Deviation)

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

4.6 The Bernoulli And Binomial Random Variables

Definition 4.6.1 (Bernoulli Random Variable) *A random variable X is said to be a Bernoulli random variable if its probability mass function is given by Equations*

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

for some $p \in (0,1)$.

Definition 4.6.2 (Binomial Random Variables) *Suppose now that n independent trials, each of which results in a success with probability p and in a failure with probability 1 p, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p).*

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \cdots, n.$$

4.6.1 Properties of Binomial Random Variables

Proposition 4.6.1 *For the Binomial Random Variables,*

1. $E[X^k] = npE[(Y+1)^{k-1}]$, $Y = \binom{n-1}{i} p^i (1-p)^{n-1-i}$.

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

Proposition 4.6.2 *If X is a binomial random variable with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $PX = k$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n+1)p$.*

4.6.2 Computing the Binomial Distribution Function

Proposition 4.6.3

$$P\{x \leq i\} = \sum_{k=0}^{i} \begin{pmatrix} n \\ i \end{pmatrix} p^k (1-p)^{n-k}$$

Corollary 4.6.1

$$P\{X = k+1\} = \frac{p}{1-p} \frac{n-k}{k+1} P\{X = k\}$$

4.7 The Poisson Random Variable

Definition 4.7.1 (Poisson Random Variable)

$$p(i) = e^{-\lambda} \frac{\lambda^i}{i!}, \lambda \geq 0 (\lambda = np)$$

Proposition 4.7.1

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

Proposition 4.7.2

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

Proposition 4.7.3 *Under the following conditions:*

- *In a time interval of arbitrary length h, $o(h) = \lim_{h \to 0} /h = 0, p = \lambda h + o(h)$*
- *$p(x = 2) = o(h)$ in a time interval of arbitrary length.*

- *For any integers n, $j_1, j_2, \cdots, j_n$ and any set of n nonoverlapping intervals, if we define E_i to be the event that exactly j_i of the events under consideration occur in the ith of these intervals, then events $E1, E2, \cdots, En$ are independent.*

The the number of events occurring in any interval of length t is a Poisson random variable with parameter λt.

4.8 Other Distributions

4.8.1 Geometric Random Variable

Definition 4.8.1

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

Proposition 4.8.1

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

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

4.8.2 Negative Binomial Random Variable

Definition 4.8.2

$$P\{X = n\} = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} p^r(1-p)^{n-r}$$

4.8.3 Hypergeometric Random Variable

Definition 4.8.3 *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. If we let X denote the number of white balls selected, then*

$$P\{X = i\} = \frac{\begin{pmatrix} m \\ 1 \end{pmatrix}\begin{pmatrix} N-m \\ n-i \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

4.8.4 The Zeta (or Zipf) Distribution

Definition 4.8.4

$$P\{X = k\} = \frac{C}{k^{\alpha+1}}$$

while,

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

and

$$\zeta(s) = \sum_{k=1}^{\infty}(\frac{1}{k})^{s}$$

5 Continuous Random Variables

5.1 Introduction

Definition 5.1.1 (Continuous Random Variable) *We say that 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 number,*

$$P\{X \in B\} = \int_B f(x)dx^1$$

Proposition 5.1.1

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

5.2 Expectation and Variance of Continuous Random Variables

Definition 5.2.1 (Expectation)

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

[1] $f(x)$ is called the **probability density function**.

Proposition 5.2.1

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

Corollary 5.2.1

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

Definition 5.2.2 (Variance)

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

5.3 The Uniform Random Variable

Definition 5.3.1 (Uniform Distribution) *A random variable is said to be uniformly distributed over the interval (0, 1) if its probability density function is given by,*

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

Similarly, a random variable is said to be uniformly distributed over the interval (α, β) if its probability density function is given by,

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & otherwise \end{cases}$$

Proposition 5.3.1

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

5.4 Normal Random Variables

Definition 5.4.1 (Normal Random Variable) *We say that X is a normal random variable, or simply that X is normally distributed, with parameters and ^{2} if the density of X is given by*

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

Proposition 5.4.1 *For a random variable with parameters μ and σ^2,*

1. *Expectation:* $E[X] = \mu$
2. *Variance:* $Var[X] = \sigma^2$
3. $P\{Z \leq -x\} = P\{Z > x\}$, $-\infty < x < \infty$

If we denote the ***cumulative distribution function*** *of a standard normal random variable by*

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

We have $\Phi(-x) = 1 - \Phi(x)$, $-\infty < x < \infty$.

And

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

5.4.1 The Normal Approximation to the Binomial Distribution

Theorem 5.4.1 (The DeMoivre-Laplace Limit Theorem) *When n is large, a binomial random variable with parameters n and p will have approximately the same distribution as a normal random variable with the same mean and variance as the binomial.*

If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed, then, for any $a < b$,

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

as $n \to \infty$.

5.5 Exponetial Random Variable

Definition 5.5.1 *A continuous random variable whose probability density function is given, for some $\lambda > 0$, by*

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

is said to be an ***exponential random variable*** *(or, more simply, is said to be exponentially distributed) with parameter λ.*

Proposition 5.5.1 *Let X be an exponential random variable with parameter λ.*

1.
$$E[X^n] = \frac{n}{\lambda} E[x^{n-1}]$$

2.
$$E[X] = \frac{1}{\lambda}$$

3.
$$Var(X) = \frac{1}{\lambda^2}$$

5.6 Other Continuous Distributions

5.6.1 The Gamma Distribution

Definition 5.6.1 (Gamma Distribution) *A random variable is said to have a* ***gamma distribution*** *with parameters* $(\alpha, \lambda), \lambda > 0, \alpha > 0$, *if its density function is given by*

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

where

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy.^2$$

Proposition 5.6.1 *Let X a* ***gamma distribution*** *with parameters* $(\alpha, \lambda), \lambda > 0, \alpha > 0$.

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

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

5.6.2 The Weibull Distribution

Definition 5.6.2 (The Weibull Distribution) *A random variable whose cumulative distribution function is said to be a* ***Weibull random variable*** *with parameters* ν, α, *and* β *if*

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

[2] $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) = (n-1)!$.

5.6.3 The Cauchy Distribution

Definition 5.6.3 (The Cauchy Distribution) *A random variable is said to have a Cauchy distribution with parameter θ, $-\infty < \theta < \infty$, if its density is given by*

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

5.6.4 The Beta Distribution

Definition 5.6.4 (The Beta Distribution) *A random variable is said to have a Beta distribution if its density is given by*

$$f(x) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, 0 < x < 1$$

where

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

Proposition 5.6.2

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

Proposition 5.6.3 *Let X a **beta distribution***

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

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