M 362K Midterm Note

Xiaohui Chen

March 11, 2015

1 Combinatorial Probability

The multiplication principle: Suppose an experiment can be broken down into a first stage A consisting of $N(A)$ outcomes and that for each of these outcomes. Then the total number of outcomes for the two states combined is equal to $N(A) \times N(B)$

Permutations: Given a set of n distinguishable object, an ordered selection of r different elements of the set is called a permutation of n objects chosen r at a time

Factorials: Let n be a whole number. The $n!$ is defined by $n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$. By convention, we define $0! = 1$

${}_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$

Combinations: Given a set of n distinguishable objects, an ordered selection of r different elements of the set is called a combination of n objects chosen r at a time and is denoted by ${}_nC_r$ and read as n choose r

${}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}$. The form ${}_nC_r = \begin{pmatrix} n \\ r \end{pmatrix}$ is especially common and is referred as the binomial coefficient

Partitions: Let A be a set of n distinguishable objects. Let whole numbers $\{r_1, r_2, \cdots, r_k\}$ be given such that $r_1 + r_2 + \cdots r_k = n$. A partition of A into subsets of sizes $\{r_1, r_2, \cdots, r_k\}$

is a particular distribution of the n objects into disjoint subsets $A_1, A_2, \cdots, A_k$ of sizes $r_1, r_2, \cdots, r_k$ respectively

Multinomial Coefficients: The number of partitions of n distinct objects into k subsets of sizes $r_1, r_2, \cdots, r_k$, where $r_1 + r_2 + \cdots r_k = n$ is called multinomial coefficient, denoted by
$\begin{pmatrix} & n & \\ r_1 & \cdots & r_k \end{pmatrix} = \frac{n!}{r_1! r_2! \cdots r_k!}$

The number of unordered samples of r objects, with replacement, from n distinguishable objects is ${}_{n+r-1}C_r = \begin{pmatrix} n+r-1 \\ r \end{pmatrix} = \begin{pmatrix} n+r-1 \\ n-1 \end{pmatrix}$. This is equivalent to the number of ways to distribute r indistinguishable balls into n distinguishable urns without exclusion

Samples of size r from n distinguishable objects	Without replacement	With replacement	
Order matter	${}_nP_r$	n^r	Distinguishable balls
Order doesn't matter	$\begin{pmatrix} n \\ r \end{pmatrix}$	$\begin{pmatrix} n+r-1 \\ r \end{pmatrix}$	Indistinguishable balls
	Exclusive	Non-exclusive	Distributions of r balls into n distinguishable urns

The Binomial Theorem: For every non-negative integer n and real numbers x and y, we have $(x+y)^n = \sum_{r=0}^{n} {}_nC_r \cdot x^r \cdot y^{n-r} = \sum_{r=0}^{n} {}_nC_r \cdot x^{n-r} \cdot y^r$

The Multinomial Theorem: $(x_1 + x_2 + \cdots + x_r)^n = \sum_{n_1 + \cdots + n_r = n} \begin{pmatrix} & n & \\ n_1, & \cdots, & n_r \end{pmatrix} \cdot x_1^{n_1} \cdot x_2^{n_2} \cdots x_r^{n_r}$

The odds against the event A are quoted as ratio $Pr(A\ does\ not\ occur) : Pr(A\ does\ occur) =$

$Pr(A^C) : Pr(A) = (1-p) : p$

If the odds against the event A are quoted as $b : a$, then $Pr(A) = \frac{a}{a+b}$

2 General Rules of Probability

The **sample space** is the set (collection) of all possible outcomes of a probability experiment.

An **event** is a subset of the sample space

2.1 Axioms of Probability Theory

(1) $0 \leq Pr(E) \leq 1$ for any event E

(2) $Pr(U) = 1$, where U denotes the entire sample space

(3) The probability of the union of mutually exclusive events is the sum of the individual probabilities of the disjoint sets: $Pr\left(\bigcup_{mutually\ exclusive}\right) = \sum_i Pr(E_i)$

2.2 Two Important Probability Rules

(1) **Negation Rule**: $Pr(E') = 1 - Pr(E)$

(2) **Inclusion-Exclusion Rule**: $Pr(E) + Pr(F) = Pr(E \cup F) + Pr(E \cap F)$

2.3 De Morgan's Laws

For any two sets A and B

(1) $(A \cap B)' = A' \cup B'$

(2) $(A \cup B)' = A' \cap B'$

2.4 The Venn Box Diagram

	A	A’	
B	$Pr(A \cap B)$	$Pr(A' \cap B)$	$Pr(B)$
B’	$Pr(A \cap B')$	$Pr(A' \cap B')$	$Pr(B')$
	$Pr(A)$	$Pr(A')$	1

2.5 Conditional Probability

The conditional probability that event A occurs given that event B occurred is

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

If the sample space consists of equally likely outcomes, then

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

Independence: Let A and B be events with non-zero probabilities. We say A and B are independent if any (and hence all) of the following hold:

(1) $Pr(A|B) = Pr(A)$

(2) $Pr(B|A) = Pr(B)$ or

(3) $Pr(A \cap B) = Pr(A) \cdot Pr(B)$. This is called the **multiplicative rule**

Otherwise the events are said to be **dependent**

2.6 Bayes' Theorem

Suppose that the sample space S is partitioned into disjoint subsets $B_1, B_2, \cdots, B_n$, That is, $S = B_1 \cup B_2 \cup \cdots \cup B_n$, $Pr(B_i) > 0$ for all $i = 1, 2, \ldots, n$, and $B_i \cap B_j = \varnothing$ for all $i \neq j$. Then for an event A,

$$Pr(B_j|A) = \frac{Pr(B_j) \cdot Pr(A|B_j)}{\sum\limits_{i=1}^{n} Pr(B_i) \cdot Pr(A|B_i)}$$

3 Discrete Random Variables

3.1 Discrete Random Variable

Discrete Random Variable: We say X is a **discrete random variable** if X is a numerically valued function whose domain is the sample space of a probability experiment with a finite or countably infinite number of outcomes

Every random variable has a **probability distribution** associated with it

The tabulation of the probabilities for each possible value x of a discrete random variable X is called its **probability distribution**. The probabilities must be positive and sum to one

The function $p(x_i) = Pr(X = x_i)$ on the values of the random variable X is called the **probability function** of X

3.2 Cumulative Probability Distribution

Let X be a discrete random variable. For each real number x, let $F(x) = Pr(X \leq x)$. The function $F(x)$ is called the **cumulative distribution function** (CDF) for the random variable X and satisfies

(1) $0 \leq F(x) = Pr(X \leq x)$ for all X

(2) If $x_{i-1} < x_i$ are consecutive values in the probability distribution table of X, then $Pr(X = x_i) = F(x_i) - F(x_{i-1}) = Pr(X \leq x_i) - Pr(X \leq x_{i-1}) = p(x_i)$

(3) We define $F(\infty) = Pr(X < \infty) = 1$

If X is a discrete random variable with probability function $Pr(X = x_i) = p(x_i)$, then the **expected value (mean)** of the random variable X is given by $\mu_X = E[X] = \sum_i x_i \cdot p(x_i)$

If X is a discrete random variable with probability function $Pr(X = x_i) = p(x_i)$, and $Y = g(X)$ is a transformation of X, then $\mu_Y = E[Y] = E[g(X)] = \sum_i g(x_i) \cdot p(x_i)$

3.3 Median

If $x_1, x_2, \cdots, x_n$ is a collection of n data points listed from smallest to largest, then the **median** of the data equals

(a) $x_{\frac{n+1}{2}}$ if n is odd. This is just the middle term in the sequence

(b) $\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$ if n is even. This is the mean of the two middle terms

3.4 Midrange

If $\{x_1, x_2, \cdots, x_n\}$ is a collection of n data points listed from smallest to largest, then the **midrange** od the data is defined to be

$$\frac{x_1 + x_n}{2} = \frac{minimum + maximum}{2}$$

3.5 Mode

If $x_1, x_2, \ldots, x_n$ is a collection of n data points, then the **mode** of the data is defined as,

(a) The value x_i that occurs most frequently

(b) The two values x_i and x_j of they occur the same number of times, and more frequently than the remaining points. In this case we say the data is **bi-modal**

(c) Otherwise, the mode does not exist

3.6 Percentiles

If $x_1, x_2, \ldots, x_n$ are n data points arranged in ascending order, then x_i corresponds to the

$$\left(100 \cdot \frac{i}{n+1}\right)^{th} percentile$$

3.7 Quartiles

The **first quartile** corresponds to the 25^{th} percentile and is denoted: Q_1

The **second quartile** corresponds to the 50^{th} percentile and is denoted: Q_2

The **third quartile** corresponds to the 75^{th} percentile and is denoted: Q_3

The **inter-quartile range(IQR)** is $IQR = Q_3 - Q_1$, where Q_3 is the third quartile and Q_1 is the first quartile

3.8 Variance

$$Var[X] = \sigma_X^2 = \sum_{x_i}(x_i - \mu_X)^2 \cdot p(x_i) = E[X^2] - E[X]^2 = \sum_{x_i} x_i^2 p(x_i) - \left(\sum_{x_i} x_i p(x_i)\right)^2 = \sum_{x_i} x_i^2 p(x_i) - (\mu_X)^2$$

Let X be a discrete random variable and let $Y = a \cdot X + b$, where a and b are real numbers.

Then,

(1) $E[X] = E[a \cdot X + b] = a \cdot E[X] + b$

(2) $Var[Y] = Var[a \cdot X + b] = a^2 \cdot Var[X]$

Standard deviation $\sigma_X = \sqrt{Var[X]}$

3.9 Standardized Random Variable

Let X be a discrete random variable and let $Z = \frac{X-\mu}{\sigma}$. Then Z is called the **standardization of X**. The random variable Z always has mean equal to 0 and standard deviation equal to 1

z-score: $z = \frac{X-\mu}{\sigma}$

Markov Inequality: $Pr[Y > a] \leq \frac{\mu_Y}{a}$ for any $a > 0$

Chebychev's Theorem:

$Pr(X < \mu_X - k \cdot \sigma_X \; or \; X > \mu_X + k \cdot \sigma_X) = Pr(|X - \mu| > k \cdot \sigma_X) \leq \frac{1}{k^2}$

$Pr(\mu_X - k \cdot \sigma_X \leq X \geq \mu_X + k \cdot \sigma_X) \geq 1 - \frac{1}{k^2}$

Outliers: We define an outlier to be any data ppint with a z-score less than $z = -3$ or greater than $z = 3$

Coefficient of Variation: $\frac{100 \cdot \sigma}{\mu}\%$

3.10 Joint Distributed Random Variables

Let X and Y be random variables arising from the same discrete probability experiment. The **joint distribution** of X and Y is given by $p(x, y) = Pr[\{X = x\} \cap \{Y = y\}]$

We say X and Y are **independent** if dor all x and y the events $\{X = x\}$ and $\{Y = y\}$

are independent. That is, $p(x,y) = Pr[\{X = x\} \cap \{Y = y\}] = Pr[X = x] \cdot Pr[Y = y] = p_x(x) \cdot p_Y(y)$

Let X and Y be random variables arising from the same probability experiment. Then,

(a) $E[X + Y] = E[X] + E[Y]$. This formula extends to sums of any length

Further, if X and Y are **independent**, then

(b) $E[X \cdot Y] = E[X] \cdot E[Y]$, and

(c) $Var[X + Y] = Var[X] + Var[Y]$

This formula extends to sums of any length provided the summands are pair-wise independent

4 Some Discrete Distributions

4.1 Discrete Uniform Distribution

arithmetic series: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

sums of squares: $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

finite geometric series: $\sum_{n=0}^{N} ax^n = \frac{a(1-x^{N+1})}{1-x}$, for any $x \neq 1$

infinite geometric series: $\sum_{n=0}^{N} ax^n = \frac{a}{1-x}$, for any $|x| < 1$

A random variable X is said to have a **discrete uniform distribution** if its probability function is $Pr(X = x) = p(x) = \frac{1}{n}$ for $x = 1, 2, \ldots, n$

$E[X] = \frac{n+1}{2}$

$Var[X] = \frac{n^2-1}{12}$

4.2 Bernoulli Trials

Suppose that the random variable X has property function given by $Pr[X = 1] = p$ and $Pr[X = 0] = q = 1 - p$. Then X is called a **Bernoulli random variable** with probability of success P

$E[X] = p$ and $Var[X] = pq = p(1 - p)$

4.3 Binomial Distribution

Suppose that the random variable Y has probability function given by $Pr(Y = y) = p(y) = {}_nC_y p^y q^{n-y}$ for $y = 0, 1, 2, \ldots, n$ and $0 \leq p \leq 1$. Then the random variable Y is called a **binomial random variable** with **parameters** n and p

Properties:

(a) There are n identical trials

(b) For each (Bernoulli) trial, there are two outcomes called success and failure

(c) The probability of success is p and the probability of failure is $q = 1 - p$

(d) Each trial is independent of the other trials

$\mu_Y = E[Y] = np$

$\sigma_Y^2 = Var[Y] = npq = np(1 - p)$

4.4 Geometric Distribution

Suppose that the random variable X has probability function given by $Pr(X = k) = p(1 - p)^k = pq^k$ for $k = 0, 1, 2 \cdots$, $q = 1 - p$ and $0 < p < 1$. Then X is called the **geometric random variable with parameter** p

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

$Var[X] = \frac{q}{p^2} = \frac{1-p}{p^2}$

4.5 Negative Binomial Distribution

Requirements:

(a) The trials are identical

(b) Each trial is independent of te other trials

(c) The random variable M denotes the number of failures prior to the r^{th} sucess

(d) The probability of success is p and the probability of failure is $q = 1 - p$

$p_k = Pr(M = k) = {}_{r+k-1}C_k p^r q^k = {}_{r+k-1}C_{r-1} p^r (1-p)^k$

$E[M] = \frac{rq}{p}$

$Var[M] = \frac{rq}{p^2}$

4.6 Hyper-geometric Random Variable

$Pr(X = k) = p_k = \frac{{}_GC_k \cdot {}_BC_{n-k}}{{}_{B+G}C_n}$

$\mu_X = E[X] = n\left(\frac{G}{B+G}\right)$

$\sigma_X^2 = Var[X] = n\left(\frac{G}{B+G}\right)\left(\frac{B}{B+G}\right)\left(\frac{B+G-n}{B+G-1}\right)$

4.7 Poisson Distribution

Suppose that the random variable Z has probability function given by $Pr(Z = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$ and $\lambda > 0$. Then Z is called a **Poison random variable with parameter λ**

$E[Z] = \lambda$ and $Var[Z] = \lambda$

Suppose that Z_i are independent **Poisson random variables with mean** λ_i for $i = 1, 2$.

Then $Z = Z_1 + Z_2$ is a Poisson random variable with mean (parameter) $E[Z] = \lambda_1 + \lambda_2$