STAT 230 Notes / Definition

Thomas Liu

August 9, 2023

Contents

1 Chapter 1

1.1 Sample Space

The set of all possible distinct outcomes to a random experiemnt or process, with the property that in a single trial, one and only one of these outtimes occurs

1.2 Classical Probability

The probability of some event is

$$\frac{\text{number of ways the event can occur}}{\text{number of outcomes in S}}$$

provided all points in the sample space S are equally likely

1.3 Relative Frequency

The probability of an event is the (limiting) proposition (or fraction) of times the event occurs in a very long series of repetitions of an experiemnt or process

1.4 Subjective Probability

The probability of an event is a measure of how sure the person making the statement is that the event will happen

2 Chapter 2

2.1 Simple Event / Compound Event

Simple event if the event is indivisible so it contains only one point
Compound event if an event A made up of two or more simple events

2.2 Probability

Let $S = \{a_1, a_2, \cdots\}$ be a discrete sample space. Assign numbers (probabilities) $P(a_i)$, $i = 1, 2, \cdots$ to the a_i's such that

1. $0 \leq P(a_i) \leq 1$
2. $\sum\limits_{\text{all } i} P(a_i) = 1$

The set of probabilities $\{P(a_i), i = 1, 2, \cdots\}$ is called a probability distribution on S

2.3 Odds in Favour / Odds Against

The odds in favour of an event A is the probability the event occurs divided by the probability it does not occur, $\dfrac{P(A)}{1-P(A)}$

The odds against the event is the reciprocal, $\dfrac{1-P(A)}{P(A)}$

3 Chapter 3

3.1 Addition Rule

Suppose we can do job 1 in p ways and job 2 in q ways. Then we can do either job 1 **OR** job 2 (not both), in $p+q$ ways

3.2 Multiplication Rule

Suppose we can do job 1 in p ways and, for each of these ways, we can do job 2 in q ways. Then we can do both job 1 **AND** job 2 in $p \times q$ ways

3.3 Permutation

The sample space is a set of arrangements or sequences, called permutations

$$n^{(k)}(\text{n to k factors}) = \frac{n!}{(n-k)!}$$

3.4 Combination

The combinatorial symbol $\binom{n}{k}$ (n choose k) is used to denote the number of subsets of size k that can be selected from a set of n objects

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

Properties:

1. $n^{(k)} = \dfrac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
2. $\binom{n}{k} = \dfrac{n!}{k!(n-k)!} = \dfrac{n^{(k)}}{k!}$
3. $\binom{n}{k} = \binom{n}{n-k}$ for all $k = 0, 1, \cdots, n$
4. If we define $0! = 1$, then the formulas hold with $\binom{n}{0} = \binom{n}{n} = 1$
5. $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
6. Binomial Theorem: $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$

3.5 Number of arrangements when symbols are repeated

If we have n_i symbols of type i, $i = 1, 2, \cdots, k$ with $n_1 + n_2 + \cdots + n_k = n$, then the number of arrangements using all of the symbols is

$$\begin{aligned}&\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_k}{n_k}\\=&\frac{n!}{n_1!n_2!\cdots n_k!}\end{aligned}$$

3.6 Useful Series and Sums

3.6.1 Geometric Series

$$\sum_{i=0}^{n-1} t^i = 1 + t + t^2 + \cdots + t^{n-1} = \frac{1-t^n}{1-t} \text{ for } t \neq 1$$

If $|t| < 1$, then

$$\sum_{x=0}^{\infty} t^x = 1 + t + t^2 + \cdots = \frac{1}{1-t}$$

Differentiate:

$$\frac{d}{dt}\sum_{x=0}^{\infty} t^x = \frac{d}{dt}(\frac{1}{1-t})$$

or

$$\sum_{x=0}^{\infty} x t^{x-1} = \frac{1}{(1-t)^2} \text{ for } |t| < 1$$

3.6.2 Binomial Thoerem

$$(1+t)^n = 1 + \binom{n}{1}t + \binom{n}{2}t^2 + \cdots + \binom{n}{n}t^n = \sum_{x=0}^{n}\binom{n}{x}t^2$$

3.6.3 Multinomial Theorem

$$(t_1 + t_2 + \cdots + t_k)^n = \sum \frac{n!}{x_1!x_2!\cdots x_k!} t_1^{x_1} t_2^{x_2} \cdots t_k^{x_k}$$

3.6.4 Hypergeometric Identity

$$\sum_{x=0}^{\infty}\binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

3.6.5 Exponential Series

$$e^t = \frac{t^0}{0!} + \frac{t^1}{1!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!} \text{ for all } t \in \mathbb{R}$$

$$e^t = \lim_{n\to\infty} (1 + \frac{t}{n})^n \text{ for all } t \in \mathbb{R}$$

3.6.6 Special series

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = [\frac{n(n+1)}{2}]^2$$

4 Chapter 4

4.1 De Morgan's Laws

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$

4.2 Addition Law of Probability of the Union of n Events

4.2.1 Two Events

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

4.2.2 Three Events

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

4.2.3 n events

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) = \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) \\ - \sum_{i<j<k<l} P(A_iA_jA_kA_l) + \cdots \end{aligned}$$

4.3 Mutually Exclusive

Events A and B are mutually exclusive if $A \cap B = \emptyset$

4.4 Probability of the Union of Mutually Exclusive Events

4.4.1 Two Events

$$P(A \cup B) = P(A) + P(B)$$

4.4.2 n Events

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

4.5 Complement

Probability of the complement of an event

$$P(A) = 1 - P(\overline{A})$$

4.6 Independent / Dependent Events

Events A and B are **independent events** if and only if $P(A \cap B) = P(A)P(B)$
If the events are not independent, the events are **dependent**
The events $A_1, A_2, \cdots, A_n$ are mutually independent if and only if

$$P(A_{i1} \cap A_{i2} \cap \cdots \cap A_{ik}) = P(A_{i1})P(A_{i2}) \cdots P(A_{ik})$$

4.7 Conditional Probability

The **conditional probability** of event A, given event B, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ provided } P(B) > 0$$

A and B are independent events if and only if

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

4.8 Product Rule

Let $A, B, C, D, \cdots$ be arbitrary events in a sample space. Assume that $P(A) > 0$, $P(A \cap B) > 0$, and $P(A \cap B \cap C) > 0$. Then

$$\begin{aligned} P(AB) &= P(A)P(B|A) \\ P(ABC) &= P(A)P(B|A)P(C|AB) \\ P(ABCD) &= P(A)P(B|A)P(C|AB)P(D|ABC) \end{aligned}$$

and so on

4.9 Law of Total Probability

Let $A_1, A_2, \cdots, A_k$ be a partition of the sample space S into disjoint (mutually exclusive) events

$$A_1 \cup A_2 \cup \cdots \cup A_k = S \text{ and } A_i \cap A_j = \emptyset \text{ if } i \neq j$$

Let B be an arbitrary event in S

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \cdots + P(BA_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned}$$

4.10 Baye's Theorem

Suppose A and B are events defined on a sample space S. Suppose $P(B) > 0$

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

5 Chapter 5

5.1 Range

The set of possible values for the variable

5.2 Random Variable

A **random variable** is a function that assigns a real number to each point in a sample space S

5.3 Discrete Random Variables

Take integer values or, values in a countable set

5.4 Continuous Random Variables

Tae values in some interval of real numbers like $(0, 1)$ or $(0, \infty)$ or $(-\infty, \infty)$

5.5 Probability Function

Let X be a discrete random variable with $range(X) = A$. The **probability function** of X is the function

$$f(x) = P(X = x), \text{ defined for all } x \in A$$

The set of pairs $\{(x, f(x)) : x \in A\}$ is called the **probability distribution** of X

1. $f(x) \geq 0$ for all $x \in A$
2. $\sum_{\text{all } x \in A} f(x) = 1$

5.6 Cumulative Distribution Function (CDF)

The **cumulative distribution function (CDF)** of X is the function denoted by $F(x)$

$$F(x) = P(X \leq x) \text{ defined for all } x \in \mathbb{R}$$

5.7 Discrete Uniform Distribution

5.7.1 Physical Setup

Suppose the range of X is $\{a, a+1, \cdots, b\}$ where a and b are integers and suppose all values are equally probable. Then X has a Discrete Uniform distribution on the set $\{a, a+1, \cdots, b\}$. The variables a and b are called the parameters of the distribution

5.7.2 Probability Function

There are $b - a + 1$ values in the set $\{a, a+1, \cdots, b\}$ so the probability of each value must be $\frac{1}{b-a+1}$ in order that $\sum_{x=a}^{b} f(x) = 1$

$$f(x) = P(X = x) = \begin{cases} \frac{1}{b-a+1} & \text{for } x = a, a+1, \cdots, b \\ 0 & \text{otherwise} \end{cases}$$

5.8 Hypergeometric Distribution

5.8.1 Physical Setup

We have a collection of N objects which can be classified into two distinct types. Call on type 'success' (S) and the other type 'failure' (F). There are r success and $N - r$ failures. Pick n objects at random without replacement. Let X be the number of successes obtained. Then X has a Hypergeometric distribution. The parameters of the distribution are N, r and n

5.8.2 Probability Function

Using counting techniques we note there are $\binom{N}{n}$ points in the sample space S if we don't consider order of selection. There are $\binom{r}{x}$ ways to choose the x success objects from the r available and $\binom{N-r}{n-x}$ ways to choose the remaining $(n - x)$ objects from the $(N - r)$ failures.

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

The range of values for x is somewhat complicated. Of course, $x \geq 0$. However if the number, n, picked exceeds the number $N - r$, of failures, the difference, $n - (N - r)$ must be successes. So $x \geq max(0, n - N + r)$. Also $x \leq r$ since we can't get more successes than the number available. But $x \leq n$, since we can't get more successes than the number of objects chosen. Therefore $x \leq min(r, n)$

5.9 Binomial Distribution

5.9.1 Physical Setup

Suppose an experiemnt has two types of distinct outcomes. Call these types 'success' (S) and 'failur' (F). Let $P(S) = p$ and $P(F) = 1 - p$. Repeat the experiemnt n independent times. Let X be the number of successes obtained. Then X has what is called a Binomial distribution. The parameters of the distribution are n and p. We write $X \sim binomial(n, p)$ as a shorthand for 'X is distributed according to a Binomial distribution with n repetitions and probability p of success' or 'X has a Binomial distribution with parameters n and p'. The n individual experiemnts in the process just described are often called 'trials' or 'Bernoulli trials' and the process is called a Bernoulli process or a Binomial process

5.9.2 Probability Function

There are $\binom{\frac{n!}{x!(n-x)!}=n}{x}$ different arrangements of x S's and $(n - x)$ F's over the n trials. The probability for each of these arrangements has p multiplied together x times and $(1-p)$ multiplied $(n-x)$ times, in some order, since the trials are independent. So each arrangement has probability $p^x(1-p)^{n-x}$

$$f(x) = P(X = x) = \binom{n}{x} p^x(1-p)^{n-x} \quad \text{for } x = 0, 1, \cdots, n \text{ and } 0 < p < 1$$

5.10 Negative Binomial Distribution

5.10.1 Physical Setup

The setup for this distribution is almost the same for Binomial; that is, an experiemnt (trial) has two dinstinct types of outcome, S and F, and is repeated independently with $P(S) = p$ on each trial. Continue doing the experiemnt until a specified number, k, of successes have been obtained. Let X be the number of failures obtained before the kth success. Then X has a Negative Binomial distribution. Write $X \sim Negative\ Binomial(k, p)$ to denote this. The parameters of the distribution are k and p

5.10.2 Probability Function

In all there will be $x + k$ trials (x F's and k S's) and the last trial must be a success. In the first $x + k - 1$ trials we therefore need x failures and $(k - 1)$ successes, in any order. There are $\frac{(x+k-1)!}{x!(k-1)!} = \binom{x+k-1}{x}$ different orders. Each order will have probability $p^k(1-p)^x$ since there must be x trials which are failures and k which are success

$$f(x) = P(X = x) = \binom{x + k - 1}{x} p^k(1-p)^x \text{ for } x = 0, 1, \cdots \text{ and } 0 < p < 1$$

5.11 Geometric Distribution

5.11.1 Physical Setup

Conside the Negative Binomial distribution with $k = 1$. In this case we repeat independent Bernoulli trials with two types of outcome, S and F, and $P(S) = p$ each time until we obtain the first success. Let X be the number of failures obtained before the first success. We write $X \sim Geometric(p)$. The parameter of the distribution is p

5.11.2 Probability Function

There is only the one arrangement with x failures followed by 1 success. This arrangement has probability

$$f(x) = P(X = x) = (1-p)^x p \text{ for } x = 0, 1, \cdots \text{ and } 0 < p < 1$$

5.12 Piosson Distribution

- The random variable X represents the number of events of some type
- The events occur according to some rate, denoted by μ, $\mu > 0$
- Write $X \sim Poisson(\mu)$

5.12.1 Probability Function

$$f(x) = \frac{e^{-\mu}\mu^x}{x!} \text{ for } x = 0, 1, 2, \cdots$$

$$\mu = np$$

Let this be the rate of success. That is the number of trials n-however many there are-multiplied by the chance of success p for each of those trials

5.13 Two ways to interpret Poisson distribution

1. limiting case of binomial distribution, when you fix $\lambda = np$, and let $n \to \infty$ and $p \to 0$
2. Poisson Process

5.14 Poisson Process

Suppose the events you are counting satisfy the following assumption:

1. Independence: the number of occurences in non-overlapping intervals are independent
2. Individuality: for sufficiently short time periods of length Δt, the probability of 2 or more events occuring in the intervals is close to zero

$$\frac{P(2 \text{ } or \text{ } more \text{ } events \text{ } in \text{ } (t, t+\Delta_t))}{\Delta_t} \to 0, \ \Delta_t \to 0$$

3. Homogeneity or Uniformity: events occur at a uniform or homogeneous rate λ and propositional to time interval Δ_t,

$$\frac{P(one\ event\ in\ (t, t+\Delta_t)) - \lambda\Delta_t}{\Delta_t} \to 0$$

If X = occurences in a time period of length t, then

$$X \sim Poi(\lambda t)$$

5.14.1 Definition

A process that satisfies the prior conditions on the occurrence of events is often called Poisson Process. More precisely, if X_t, for $t \geq 0$, (a random variable for each t) denotes the number of events that have occurred up to time t, then X_t is called a Poisson process

6 Chapter 7

6.1 Sample Mean

Let $x_1, x_2, \cdots, x_n$ be n outcomes for a random variable X (such a set is called sample). Then its sample mean is defined as

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

6.2 Median & Mode

6.2.1 Median

A value such that half of the results are below it and the other half above it, when the sample is arranged in numerical order

6.2.2 Mode

The most frequency-occuring value in a sample. (can have more than one mode in a sample)

6.3 Expected Value

Suppose X is a discrete random variable with probability function $f_X(x)$. Then $E(X)$ is called the expected value of X, defined by

$$E(X) = \sum_{x \in X(S)} x \cdot f_X(x)$$

The expected value of X is sometimes referred to as the mean of X or the first moment of X

6.4 Linearity Properties of Expectation

For constants a and b

$$E[ag(X) + b] = aE[g(X)] + b$$

6.5 Variance

The variance of a random variable X, denoted by $Var(X)$ or by σ^2, is

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

The variance is the average square of the distance from the mean

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

$$Var(X) = E[X(X - 1)] + E(X) - [E(X)]^2 = E[X(X - 1)] + \mu - \mu^2$$

6.6 Standard Deviation

The standard deviation of a random variable X is

$$\sigma = sd(X) = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}$$

7 Chapter 8

7.1 Probability Density Function

$f(x)$ for a continuous random variable X is the derivative

$$f(x) = \frac{dF(x)}{dx}$$

where $F(x)$ is the cumulative distribution function for X

7.2 Properties of a probability density function

1. $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$
2. $f(x) \geq 0$
3. $\int_{-\infty}^{\infty} f(x)dx = \int_{allx} f(x)dx = 1$ $(P(-\infty \leq X \leq \infty) = 1)$
4. $F(x) = \int_{-\infty}^{x} f(u)du$

7.3 Expectation, Mean, and Variance for Continuous Random Variable

When X is a continuous random variable we define

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

For $\mu = E(x)$

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

7.4 Continuous Uniform Distribution

Write $X \sim Unifrom(a, b)$
Probability density function:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{cases}$$

Cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < a \\ \displaystyle\int_a^x \frac{1}{b-a}dx & a \leq x \leq b \\ 1 & x > b \end{cases}$$

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

7.5 Exponential Distribution

In a Poisson process for events in time, let X be the length of time until the first event occur, show X has exponential distribution
Probability density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Cumulative distribution function:

$$F(x) = \begin{cases} 1 - \dfrac{(\lambda x)^0 e^{-\lambda x}}{0!} = 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Or let $\theta = \dfrac{1}{\lambda}$, $\theta = E(X)$
Probability density function:

$$f(x) = \begin{cases} \dfrac{1}{\theta} e^{-\dfrac{x}{\theta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Cumulative distribution function:

$$F(x) = \begin{cases} 1 - e^{-\dfrac{x}{\theta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$E(X) = \theta$ and $Var(X) = \theta^2$

7.6 Gamma Function

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

7.7 Memoryless Property of the Exponential Distribution

$$P(X > c + b | X > b) = P(X > c)$$

7.8 Normal Distribution

7.8.1 Physical Setup

A A random variable X has a Normal distribution if it has probability density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ are parameters of the distribution. It turns out that $E(X) = \mu$ and $Var(X) = \sigma^2$ for this distributionl that is why its probability density function is written using the symbols μ and σ^2
We write $X \sim N(\mu, \sigma^2)$ to denote that X has a Normal distribution with mean μ and variance σ^2

7.8.2 Cumulative distribution function

The cumulative distribution function of the NOrmal distribution $N(\mu, \sigma^2)$ is

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dy \quad x \in \mathbb{R}$$

7.9 Theorem

Let $X \sim N(\mu, \sigma^2)$ and define $Z = \dfrac{X-\mu}{\sigma}$. Then $Z \sim N(0, 1)$ and

$$P(X \leq x) = P(Z \leq \frac{x-\mu}{\sigma})$$

8 Chapter 9

8.1 Joint Probability Function

Suppose there are two discrete random variables X and Y, define function

$$f(x,y) = P(X = x, Y = y)$$

Call $f(x,y)$ the joint probability function of (X,Y)
$f(x,y) \geq 0$, $\sum_{all(x,y)} f(x,y) = 1$

8.2 Independent Random Variables

X and Y are independent random variables if $f(x,y) = f_1(x)f_2(y)$ for all values (x,y)

8.3 Conditional Probability Functions

The conditional probability function of X given $Y = y$ is

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)},\ f_2(y) > 0$$

Y given $X = x$ is

$$f_2(y|x) = \frac{f(x,y)}{f_1(x)},\ f_1(x) > 0$$

8.4 Poisson

If $X \sim Poisson(\mu_1)$ and $Y \sim Poisson(\mu_2)$ independently then

$$T = X + Y \sim Poisson(\mu_1 + \mu_2)$$

8.5 Binomial

If $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ independently then

$$T = X + Y \sim Binomial(n + m, p)$$

8.6 Joint Probability Function

The joint probability function of $X_1, X_2, \cdots, X_k$ is given by extending the argument in the sprinters example from $k = 3$ to general k. There are $\frac{n!}{x_1!x_2!\cdots x_k!}$ different outcomes of the n trials in which x_1 are of the 1^{st} type, x_2 are of 2^{nd} type, etc. Each of these arrangements has probability $p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$ since p_1 is multiplied x_1 times in some order, etc.

$$f(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$$

8.7 Expectation for Multivariate Distributions: Covariance and Correlation

8.7.1 Expected Value

$$E[g(X,Y)] = \sum_{all(x,y)} g(x,y)f(x,y)$$

and

$$E[g(X_1, X_2, \cdots, X_n)] = \sum_{all(x_1,x_2,\cdots,x_n)} g(x_1, x_2, \cdots, x_n)f(x_1, x_2, \cdots, x_n)$$

8.7.2 Property of Multivariate Expectation

$$E[ag_1(X,Y) + bg_2(X,Y)] = aE[g_1(X,Y)] + bE[g_2(X,Y)]$$

8.8 Covariance

The covariance of X and Y, denoted $Cov(X,Y)$ or σ_{XY}

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

If X and Y are independent, $Cov(X,Y) = 0$, $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$

8.9 Correlation Coefficient

The correlation coefficient of X and Y is

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

8.10 Results for Means

- $E(aX + bY) = aE(x) + bE(Y) = a\mu_X + b\mu_Y$
- $E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mu_i$, $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$
- let $X_1, X_2, \cdots, X_n$ be random variables with mean μ, the sample mean is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $E(\overline{X}) = \mu$

8.11 Results for Covariance

- $Cov(X,X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu)^2] = Var(X)$
- $Cov(aX + bY, cU + dV) = acCov(X,U) + adCov(X,V) + bcCov(Y,U) + bdCov(Y,V)$

8.12 Results for Variance

- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$
- let X and Y be independent, since $Cov(X, Y) = 0$, result gives

$$Var(X + Y) = \sigma_X^2 + \sigma_Y^2$$

  for independent variables, the variance of a sum is the sum of the variances

$$Var(X - Y) = \sigma_X^2 + (-1)^2\sigma_Y^2 = \sigma_X^2 + \sigma_Y^2$$

  for independent variables, the variance of a difference is the sum of the variances

- let a_i be constants and $Var(X_i) = \sigma_i^2$

$$Var(\sum_{i=1}^{n} a_iX_i) = \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_ia_jCov(X_i, X_j)$$

8.13 Linear Combinations of Independent Normal Random Variables

- let $X \sim N(\mu, \sigma^2)$, $Y = aX + b$ where a and b are constant real number

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

- let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ independently, let a and b be constants

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

- let $X_1, X_2, \cdots, X_n$ be independent $N(\mu, \sigma^2)$ random variables

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

9 Chpater 10

9.1 Central Limit Theorem

If $X_1, X_2, \cdots, X_n$ are independent random variables all having the same distribution, with mean μ and variance σ^2, then as $n \to \infty$, the cumulative distribution function of the random variable

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

approacges the $N(0, 1)$ cumulative distribution function. Similarly, the cumulative distribution function of

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the $N(0, 1)$ cumulative distribution function

9.2 Normal Approximation to Poisson

Suppose $X \sim Poisson(\mu)$. Then the cumulative distribution function of the standardized random variable

$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

approaches that of a standard Normal random variable as $\mu \to \infty$

9.3 Normal Approximation to Binomial

Suppose $X \sim Binomial(n, p)$. Then for n large, the random variable

$$W = \frac{X - np}{\sqrt{np(1-p)}}$$

has approximately a $N(0, 1)$ distribution

9.4 Moment Generating Function (m.g.f.)

Consider a discrete random variable X with probability function $f(x)$. The moment generating function (m.g.f.) of X is defined as

$$M(t) = E(e^{tX}) = \sum_{all\ x} e^{tx} f(x)$$

We will assume that the moment generating function is defined and finite for values of t in an interval around 0

9.5

Suppose the random variable has moment generating function $M(t)$ defined for all $t \in [-a, a]$ for some $a > 0$

$$E(X^k) = M^{(k)}(0)\ for\ k = 1, 2, \cdots$$

where

$$M^{(k)}(0) = \frac{d^k}{dt^k} M(t)|_{t=0}\ for\ k = 1, 2, \cdots$$

9.6 Uniqueness Theorem for Moment Generating Functions

Suppose that random variables X and Y have moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. If $M_X(t) = M_Y(t)$ for all t then X and Y have the same distribution

9.7 m.g.f. of a Continuous Random Variable

Consider a continuous random variable X with probability density function $f(x)$. The m.g.f. of X is defined as

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$