Lecture 1
Introduction to Statistics

MATH 311 Statistics I: Probability Theory
January, 2023

Xuemao Zhang
East Stroudsburg University

Agenda

1. **What is Statistics?**
2. **Objectives of Statistics**
3. **Descriptive Statistics**

What is Statistics?

The science of planning studies and experiments, collecting data, and then organizing, summarizing, analyzing, interpreting, and drawing conclusions based on the **data**.

Data collection

Data are collected from a **population** of interest. We typically obtain data from two distinct sources:

- **Experiment**. Apply some treatment and then observe its effects on the experimental units (subjects in experiments).
- **Observational study**. Observe and measure specific **variables** without attempting to modify the subjects being studied.

Example 1. The Pew Research Center surveyed 2252 adults and found that 59% of them go online wirelessly. This an observational study because the adults had no treatment applied to them.

What is Statistics?

Example 2. In the largest public health experiment ever conducted, 200,745 children were given the Salk **vaccine**, while another 201,229 children were given a **placebo**. The vaccine injections constitute a treatment that modified the subjects, so this is an example of an experiment.

Variable

- A **variable** is a characteristic that changes or varies over time and/or for different individuals/objects under consideration. It is 'what' to be measured.
  Examples: Hair color, white blood cell count, time to failure of a computer component.
- Data are obtained by measuring variables of subjects.
- A data set is a set of measurements of one variable or several variables of a group of subjects.

What is Statistics?

Types of Variables

- **Categorical** (or qualitative) variable consists of names or labels (representing categories).
  1. **Example** The gender (male/female) of professional athletes.
  2. **Example** Shirt numbers on professional athletes uniforms - substitutes for names.
- **Numerical** (or quantitative) variable consists of numbers representing counts or measurements.

Categorical variables

Note. Categorical/qualitative variables cannot be used for computations or calculations. Categorical variables can be further classified using levels of measurement by looking at what is being measured.

- **Nominal** - categories only. Characterized by data that consist of names, labels, or categories only, and the data cannot be arranged in an ordering scheme (such as low to high). For example, Survey responses yes, no, undecided; Social Security Numbers.

What is Statistics?

Categorical variables - continued

- **Ordinal** - categories with some order. Involves data that can be arranged in some order, but differences between data values either cannot be determined or are meaningless. For example, course grades A, B, C, D, or F; ranks.

Numerical variables

Numerical variables can be further described by distinguishing between discrete and continuous types.

- **Discrete variable** - if the number of possible values of the variable is either finite or 'countable' (i.e. the number of possible values is 0, 1, 2, 3, . . .). For example, the number of eggs that a hen lays.
- **Continuous variable** - if the number of possible values of the variable correspond to some **continuous scale** (an interval) that covers a range of values without gaps, interruptions, or jumps. For example, the amount of milk that a cow produces; e.g. 2.343115 gallons per day.

Univariate Data

Univariate Data

- Data results from making observations on a **single variable**. A univariate data set consists of observations on a single variable.
- For example, the following sample of lifetimes (hours) of brand D batteries put to a certain use is a numerical univariate data set:
  5.6 5.1 6.2 6.0 5.8 6.5 5.8 5.5

Bivariate Data

Bivariate Data

- When two variables are measured (not always but usually on a single unit), the resulting data are called bivariate data. We denote one variable by X, and the second variable as Y.
- We consider a sample data set of size n

$$(x_1, y_1), \ldots, (x_n, y_n)$$

  which is taken from a bivariate population.

Examples

- **Example 1:** A data set might consist of a (height, weight) pair for each basketball player on a team, with the first observation as (72, 168), the second as (75, 212), and so on.
- **Example 2:** If an engineer determines the value of both x = component lifetime and y = reason for component failure, the resulting data set is bivariate with one variable numerical and the other categorical.

Multivariate Data

Multivariate Data

Multivariate data arises when observations are made on more than one variable (so bivariate is a special case of multivariate).

Example

For example, a research physician might determine the systolic blood pressure, diastolic blood pressure, and serum cholesterol level for each patient participating in a study.

Objectives of Statistics

1. Design the process of data collection.
   - Sampling (in observational study): We collect a subset/sample of a population of interest.
   - Experimental design: We apply some treatment and then observe its effects on the subjects.

Population and sample

- **Population**: The complete collection of all individuals that are being considered.
- **Sample**: Subcollection of members selected from a population of interest.

Sampling vs Census

- Sampling:
- Census: Collection of data from every member of a population. We only need **Descriptive Statistics**, procedures used to summarize and describe the set of measurements.

Objectives of Statistics

Parameter

a numerical measurement describing some variable of a population.

Example. Assume the population of interest is all ESU students in 2014-2015. It is reported that the average rent per month of all ESU students in 2014-2015 is \$525.50. This \$525.50 is a ?

Statistic

a numerical measurement describing some variable of a sample.

Example. Assume the population of interest is all ESU students in 2014-2015. A sample of 100 ESU students in 2014-2015 reported that the average rent per month of those sampled students is \$428.58. This \$428.58 is a ?

Objectives of Statistics

2. Make inferences about a population based on a sample: When we cannot enumerate the whole population, we consider **statistical inferences**, procedures used to draw conclusions or make inferences about the population from information contained in a sample.

Example. Based on a sample of 100 ESU students, there is evidence to indicate the average rent of all ESU students is significantly lower than the average rent of all MIT students. This statement is an example of **Statistical Inference**.

Example. A sample of 100 ESU students reported that the average rent per month of those sampled students is $428.58. This statement is an example of **Descriptive Statistics** since no inferences were made.

Descriptive Statistics

After collecting the numerical data we need to summarize them in order to obtain relevant information from the data? Two ways to summarize are

- Graphical
- Numerical

Univariate Categorical Data

Example: A sample of 30 persons who often consume donuts were asked what variety of donuts was their favorite. The responses from these 30 persons were as follows:
glazed filled other plain glazed other
frosted filled filled glazed other frosted
glazed plain other glazed glazed filled
frosted plain other other frosted filled
filled other frosted glazed glazed filled

Univariate Categorical Data

Frequency distributions

A **frequency distribution** of a categorical variable lists all categories and the number of elements that belong to each of the categories.

Example

Frequency Distribution of Favorite Donut Variety:

Donut Variety	Tally	Frequency (f)
Glazed	~~||||~~ |||	8
Filled	~~||||~~ ||	7
Frosted	~~||||~~	5
Plain	|||	3
Other	~~||||~~ ||	7
		Sum = 30

Univariate Categorical Data

Relative Frequency

- The **relative frequency** of a category = $\dfrac{\text{Frequency of that category}}{\text{sum of all frequencies}}$
- **Percentage** = (Relative frequency) $\times$ 100%

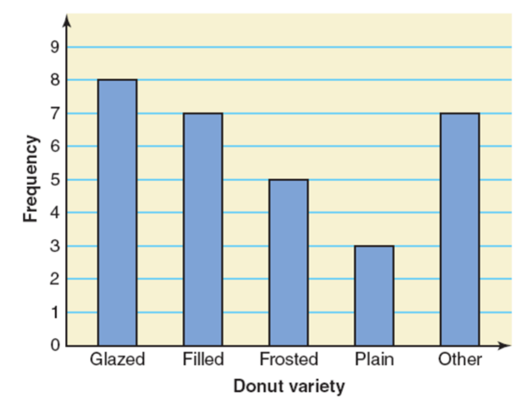
Relative Frequency and Percentage Distributions of Favorite Donut Variety:

Donut Variety	Relative Frequency	Percentage
Glazed	8/30 = .267	.267(100) = 26.7
Filled	7/30 = .233	.233(100) = 23.3
Frosted	5/30 = .167	.167(100) = 16.7
Plain	3/30 = .100	.100(100) = 10.0
Other	7/30 = .233	.233(100) = 23.3
	Sum = 1.000	Sum = 100%

Univariate Categorical Data

Bar chart/graph

- A graph made of bars whose heights represent the frequencies of respective categories is called a **frequency bar chart**.



- If the heights represent relative frequencies, the graph then is called a **relative frequency bar chart**.

Univariate Numerical Data

Notations:

- The number of observations in a single sample, that is, the sample size, will often be denoted by n.
- Given a data set consisting of n observations on some variable x, the individual observations will be denoted by $x_1, x_2, \ldots, x_n$. The subscript bears no relation to the magnitude of a particular observation.

Frequency distributions

A **frequency distribution** for numerical data lists all the classes partitioning the data and the number of values that belong to each class.

Univariate Numerical Data

Example

The following data give the total number of iPods sold by a mail order company on each of 30 days. Construct a **frequency distribution table**.

8 25 11 15 29 22 10 5 17 21
22 13 26 16 18 12 9 26 20 16
23 14 19 23 20 16 27 16 21 14

iPods Sold	Tally	f
5–9	|||	3
10–14	𝍸 |	6
15–19	𝍸 |||	8
20–24	𝍸 |||	8
25–29	𝍸	5
		$\Sigma f = 30$

Note: It is suggested to use 5-20 classes depending on the sample size.

Univariate Numerical Data

Relative Frequency

- The **relative frequency** of a class = $\dfrac{\text{Frequency of that class}}{\text{sample size}}$
- **Percentage** = (Relative frequency) $\times$ 100%

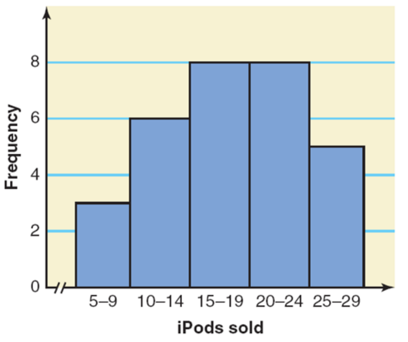
Relative Frequency and Percentage Distributions for the Data on iPods Sold:

iPods Sold	Class Boundaries	Relative Frequency	Percentage
5–9	4.5 to less than 9.5	3/30 = .100	10.0
10–14	9.5 to less than 14.5	6/30 = .200	20.0
15–19	14.5 to less than 19.5	8/30 = .267	26.7
20–24	19.5 to less than 24.5	8/30 = .267	26.7
25–29	24.5 to less than 29.5	5/30 = .167	16.7
		Sum = 1.001	Sum = 100.1

Univariate Numerical Data

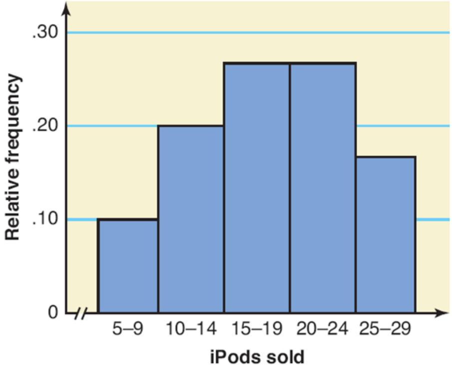
Histogram

- A **histogram** is a graph in which classes are marked on the horizontal axis and the frequencies, relative frequencies, or percentages are marked on the vertical axis.
- The frequencies, relative frequencies, or percentages are represented by the heights of the bars.
- In a histogram, the bars are drawn adjacent to each other.



Frequency histogram

Univariate Numerical Data

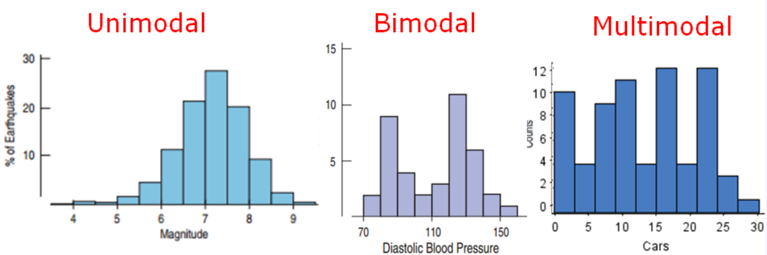


Relative frequency histogram

Univariate Numerical Data

Shapes of Histograms

- Numerically, a **mode** of a data set is the value that occurs with the greatest frequency.
- A Mode of a histogram is a hump or high-frequency bin (local maxima).
  - Unimodal: One mode
  - Bimodal: Two modes
  - Multimodal: 3 or more modes



Skewness for unimodal data

- A histogram is **skewed right** if the longer tail is on the right side of the mode.
- A histogram is **skewed left** if the longer tail is on the left side of the mode.

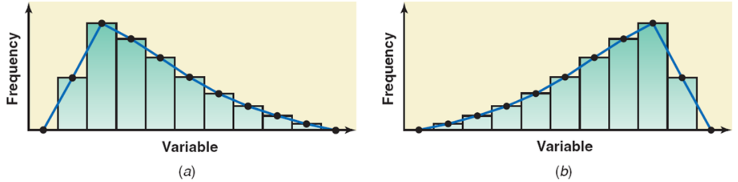


Figure (a) A histogram skewed to the right. (b) A histogram skewed to the left.

Univariate Numerical Data- Measure of Location/Cener

Median

- The **absolute center** of a data set that has been **ranked** in increasing order.
- Half of the data values are to the left of the median and half are to the right of the median:
  - If sample size n is an odd number, median is the value at position $(n+1)/2$
  - If sample size n is an even number, median is the average of the two values at position $n/2$ and $\frac{n}{2}+1$.

Example 1: Quiz Scores 2, 4, 5, 6, 7, 9, 9.
Example 2: Quiz Scores 2, 4, 5, 6, 7, 9, 9, 10.

Univariate Numerical Data- Measure of Location/Cener

Mean

- The Mean is what most people think of as the average. It takes every data value into account.
- Add up all the numbers and divide by the number of values (sample size). Denote the sample mean by $\overline{x}$, where x is the symbol of the variable.

Mean vs. Median:

- For symmetric data, the mean and the median are equal.
- The tail “pulls” the mean towards it more than it does to the median.
- The mean is more sensitive to outliers than the median. One extreme value can affect it dramatically

Univariate Numerical Data- Measure of Spread/Dispersion

Range

- Range = Largest value - Smallest Value
- The range is sensitive to outliers. A single high or low value will affect the range significantly. Therefore, it is not as useful as other measures of variation.

Variance and Standard Deviation

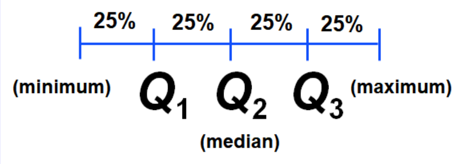
The variance of a set of values is a measure of how much data values deviate away from the mean.

- Sample variance $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$.
- Sample standard deviation $s = \sqrt{s^2}$.
- If data values are close to the mean (less spread out), then s^2 or s will be small; If data values are far from the mean (more spread out), then s^2 or s will be large.

Univariate Numerical Data- Measure of Spread/Dispersion

Percentiles and Quartiles

- Percentiles divide the data in one hundred groups. The *nth* percentile is the data value such that *n* percent of the data lies below that value.
- The median is the 50th percentile(Q_2).
- The median of the lower half of the data is the 25th percentile and is called the first quartile (Q_1).
- The median of the upper half of the data is the 75th percentile and is called the third quartile (Q_3).
- There is no universal agreement on the method of calculating percentiles and quartiles Q_1 and Q_3, and different algorithms often yield different results.



Univariate Numerical Data- Measure of Spread/Dispersion

Interquartile Range

- The Interquartile Range (IQR) is the difference between the upper quartile and the lower quartile: IQR = $Q_3 - Q_1$.
- The IQR measures the range of the middle half of the data.

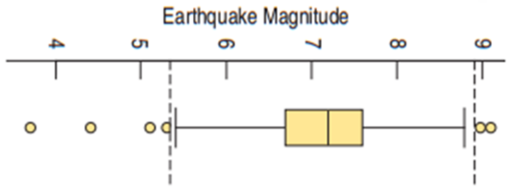
Box Plot

- A Boxplot is a chart that displays the 5-Point Summary (Minimum, Q_1, Q_2, Q_3, and Maximum) and the outliers.
- Any data values smaller than the lower fence or larger than the upper fence are defined as outliers in a boxplot.
  - Lower Fence = $Q_1 - 1.5 \times IQR$
  - Upper Fence = $Q_3 + 1.5 \times IQR$

Univariate Numerical Data- Measure of Spread/Dispersion

Box Plot

- The Box shows the Q_1, Q_2 and Q_3. The line inside the box shows the median.
- The dashed lines are called fences, outside the fences lie the outliers.
- Above and below the box are the whiskers that display the most extreme data values within the fences. The whiskers extend only as far as the minimum data value that is not an outlier and the maximum data value that is not an outlier.



Lecture 2
Probability

MATH 311 Statistics I: Probability Theory
January, 2023

Xuemao Zhang
East Stroudsburg University

Agenda

1. A Review of Set Notation
2. A Probabilistic Model for Experiments (Discrete Case)
3. Counting techniques
4. Conditional Probability and Bayes' Rule
5. Independence and Two Laws of Probability

A Review of Set Notation

We will use capital letters, A, B, C, ... to denote sets. If the elements in the set A are $a_1, a_2, \ldots a_k$, $k \geq 1$, we will write

$$A = \{a_1, a_2, \ldots a_k\}.$$

- Let S denote the set of all elements under consideration; that is, S is the universal set.
- Let $\emptyset$ denote the null, or empty set. It is the set consisting of no points.

Set Relation - subset

We will say that A is a subset of B, or A is contained in B (denoted by $A \subseteq B$), if every point in A is also in B. That is,

$$A \subseteq B \Leftrightarrow \forall x \in A, x \in B.$$

A Review of Set Notation

Set Relation - Union

The union of A and B, denoted by $A \cup B$, is the set of all points in A or B (or both). That is,

$$A \cup B = \{x \in S | x \in A \text{ or } x \in B\}.$$

Set Relation - Intersection

The intersection of A and B, denoted by $A \cap B$, is the set of all points in both A and B. That is,

$$A \cap B = \{x \in S | x \in A \text{ and } x \in B\}.$$

A Review of Set Notation

Set Relation - Complement

$\overline{A}$ or A^c denotes the complement of A. $\overline{A}$ or $A^c$is the set of all points that are in S but not in A. Therefore,

- $A \cup \overline{A} = S$
- $A \cap \overline{A} = \emptyset$

Set Relation - Disjoint

A and B are said to be disjoint or mutually exclusive if $A \cap B = \emptyset$.

Example. A and $\overline{A}$ are disjoint.

A Review of Set Notation

Distributive Law

For any 3 sets, A, B, and C,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

DeMorgan's Law

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$
$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

A Review of Set Notation

Table: Summary of Symbols in Set Relations

Symbol	Meaning	Example
$\{\}$	a set notation	$A = \{1, 2, 3, 4, 5, 6\}$
$\in$	element of, belongs to	$x \in A$: x is an element of set A
$\forall$	For all (any),	$\forall\ x \in A$: For all x in set A
$\cup$	union	$A \cup B$: union of set A and B
$\cap$	intersection	$A \cap B$: intersection of set A and B
$\emptyset$	empty set	$\{0, 1\} \cap \{2, 3\} = \emptyset$
$\subseteq$	subset	$A \subseteq B$: A is a subset of B
A^C or $\overline{A}$	complement	A^C or $\overline{A}$: the complement of set A
$\Leftrightarrow$	equivalent, if and only if	$x - 2 > 0 \Leftrightarrow x > 2$: $x - 2 > 0$ if and only if $x > 2$

Basic Concepts of Probability

Basic Concepts

- An **experiment** (statistical observation) is the process by which an observation is made. Some examples: (1) Record an age (2) Toss a die (3) Toss two coins
- A **simple event**, or **sample point** is an event that cannot be decomposed to simpler components. We use letter E with a subscript to denote a simple event.
  - The basic element to which probability is applied.
  - One and only one simple event can occur when the experiment is performed.
- The set of all simple events or sample points of an experiment is called the **sample space**, denoted by S or Ω.
- A **discrete sample space** is one that contains either a finite or a countable number of distinct sample points.
- An (compound) **event** is a collection of one or more simple events.
- An event **occurs** if one of its simple events occurs.

Example. Toss a die. Define the sample space, simple events and two events: $A = \{\text{an odd number}\}$ and $B = \{\text{a number} > 2\}$.

Basic Concepts of Probability

Axioms of probability

Let $P(A)$ be the probability of event A occurs, $A \subseteq S$.

1. Nonnegativity: $P(A) \geq 0$, for every event $A \subseteq S$.
2. Additivity: If A and B are two disjoint or mutually exclusive events, then
$$P(A \cup B) = P(A) + P(B).$$
3. Normalization: $P(S) = 1$.

Remarks about additivity.

- **Finite additivity**. If $E_1, E_2, \cdots, E_k$ are simple events, then
$$P(\{E_1, \cdots, E_k\}) = P(E_1) + \cdots + P(E_k).$$
Therefore, $P(A)$ is found by adding the probabilities of all simple events contained in A.
- **Countable additivity**. If $A_1, A_2, A_3, \ldots$ form a sequence of pairwise mutually exclusive events in S (that is, $A_i \cap A_j = \emptyset$ for $i \neq j$), then
$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Basic Rules for Computing Probability

We measure “how often" using Relative frequency f/n, where n is the sample size. As n gets larger,

$$\boxed{\begin{array}{ccc} \textit{Sample} & \longrightarrow & \textit{Population} \\ \textit{Relativefrequency} & \longrightarrow & \textit{Probability} \end{array}}$$

Relative Frequency Approximation of Probability

Conduct (or observe) an experiment, and count the number of times event A actually occurs. Based on these actual results, $P(A)$ is approximated as follows:

$$P(A) = \frac{\text{number of times A occurred}}{\text{number of times the experiment was repeated}}.$$

Basic Rules for Computing Probability

Classical Approach to Probability (Requires Equally Likely Outcomes)

Assume that a given experiment has N different simple events and that each of those simple events has an equal chance of occurring. If event A can occur in n_A of these N ways (A contains n_A simple events), then

$$\boxed{P(A) = \frac{\text{number of simple events in A}}{\text{Total number of simple events}}.}$$

Example. A bag contains 6 red marbles, 3 blue marbles, and 7 green marbles. If a marble is randomly selected from the bag, what is the probability that it is blue?

Basic Rules for Computing Probability

Example 2.1.
A manufacturer has five seemingly identical computer terminals available for shipping. Unknown to her, two of the five are defective. A particular order calls for two of the terminals and is filled by randomly selecting two of the five that are available.

a. List the sample space for this experiment.
b. Let A denote the event that the order is filled with two nondefective terminals. List the sample points in A.
c. Construct a Venn diagram for the experiment that illustrates event A.
d. Assign probabilities to the simple events in such a way that the information about the experiment is used and the axioms in Definition 2.6 are met.
e. Find the probability of event A.

Basic Rules for Computing Probability

Example. A balanced coin is tossed 3 times. Calculate the probability that exactly two of the three tosses result in heads.(Hint: use tree diagram to count the simple events in the sample space.)

Counting techniques

To calculate probabilities using the Classical Approach, we need counting techniques. **Outline.**

1. $m \times n$ rule
2. Permutations
3. Combinations
4. Partitions: Permutations of different sets of objects

$m \times n$ rule

- With m distinct elements $a_1, \ldots, a_m$ and n distinct elements $b_1, \ldots, b_n$, it is possible to form $m \times n$ pairs containing one element from each group.
- If an experiment is performed in two stages, with m distinct ways to accomplish the first stage and n distinct ways to accomplish the second stage, then there are $m \times n$ ways to accomplish the experiment.

Note. This rule is easily extended to $k > 2$ stages with n_i distinct ways to accomplish the ith stage, $i = 1, \ldots, k$. then the number of ways to accomplish the experiment equals to $n_1 \cdot n_2 \cdots n_k$.

Counting techniques

Example 2.5. An experiment involves tossing a pair of dice and observing the numbers on the upper faces. Find the number of sample points in S, the sample space for the experiment.

Example 2.6. Refer to the coin-tossing experiment in Example 2.3. We found for this example that the total number of sample points was eight. Use the extension of the mn rule to confirm this result.

Example 2.7. Consider an experiment that consists of recording the birthday for each of 20 randomly selected persons. Ignoring leap years and assuming that there are only 365 possible distinct birthdays, find the number of points in the sample space S for this experiment. If we assume that each of the possible sets of birthdays is equiprobable, what is the probability that each person in the 20 has a different birthday?

Counting techniques

Definition. An ordered arrangement of r distinct objects is called a permutation.

Notation. The factorial symbol ! denotes the product of decreasing positive whole numbers. $n! = n \times (n-1) \times (n-2) \cdots \times 2 \times 1$. For example,

$$4! = 4 \times 3 \times 2 \times 1.$$

By special definition, $0! = 1$.

Permutations

The number of permutations (or sequences) of r ($r \leq n$) items selected from n distinct items, denoted by P_r^n is

$$P_r^n = \frac{n!}{(n-r)!}$$

Proof.

Counting techniques

Example 2.8. The names of 3 employees are to be randomly drawn, without replacement, from a bowl containing the names of 30 employees of a small company. The person whose name is drawn first receives \$100, and the individuals whose names are drawn second and third receive \$50 and \$25, respectively. How many sample points are associated with this experiment?

Example 2.9. Suppose that an assembly operation in a manufacturing plant involves four steps, which can be performed in any sequence. If the manufacturer wishes to compare the assembly time for each of the sequences, how many different sequences will be involved in the experiment?

Counting techniques

Definition. If the **order does not matter** in a permutation then it is called a combination. That is, sequences consisting the same elements with different order are regarded as a single arrangement.

Combinations

The number of distinct combinations(subsets) of n distinct objects that can be formed, taking them r at a time, denoted by C_r^n or $\binom{n}{r}$ is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{P_r^n}{r!}.$$

Proof.

Note. $\binom{n}{r}$ are called *binomial coefficients*. $\sum_{r=0}^{n} \binom{n}{r} = ?$

Proof.

Counting techniques

Example 2.11. Find the number of ways of selecting two applicants out of five and hence the total number of sample points in S for Example 2.2.

Example 2.12. Let A denote the event that exactly one of the two best applicants appears in a selection of two out of five. Find the number of sample points in A and $\mathrm{P}(A)$.

Example 2.13. A company orders supplies from M distributors and wishes to place n orders ($n < M$). Assume that the company places the orders in a manner that allows every distributor an equal chance of obtaining any one order and there is no restriction on the number of orders that can be placed with any distributor. Find the probability that a particular distributor -say, distributor I- gets exactly k orders ($k \leq n$).

Counting techniques

Example. A batch of 18 items contains 4 defectives. If three items are sampled at random, find the probability of the events: (1) A = None of the defectives appears (2) B = Exactly 2 of the defectives appears.

Solution.

Counting techniques

Partitions: Permutations of different sets of objects

Suppose there are n distinct items available, we select all of the n items **without replacement**). The number of ways of **partitioning** these n distinct objects into k distinct groups containing the $n_1, n_2, \ldots, n_k$ objects is

$$\frac{n!}{n_1! \times n_2! \times \cdots n_k!}.$$

Note: Order does not matter for each group.

Proof.

Example. A class consisting of 4 graduate and 12 undergraduate students is randomly divided into four groups of 4. What is the probability that each group includes a graduate student?

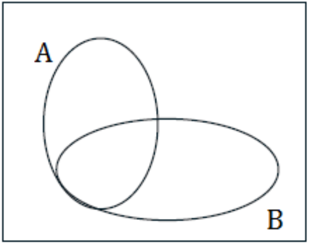
Solution.

Counting techniques

Example 2.10. A labor dispute has arisen concerning the distribution of 20 laborers to four different construction jobs. The first job (considered to be very undesirable) required 6 laborers; the second, third, and fourth utilized 4, 5, and 5 laborers, respectively. The dispute arose over an alleged random distribution of the laborers to the jobs that placed all 4 members of a particular ethnic group on job 1. In considering whether the assignment represented injustice, a mediation panel desired the probability of the observed event. Determine the number of sample points in the sample space S for this experiment. That is, determine the number of ways the 20 laborers can be divided into groups of the appropriate sizes to fill all of the jobs. Find the probability of the observed event if it is assumed that the laborers are randomly assigned to jobs.

Conditional Probability

Conditional probability



Conditional Probability

Conditional Probability

The conditional probability of an event A, given that an event B has occurred (B is our new universe), is equal to

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

provided $P(B) > 0$. Similarly we can define $P(B|A)$ if $P(A) > 0$.

Multiplication rule

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Conditional Probability

Example. Refer to the table to find the probability that a subject actually uses drugs, given that he or she had a positive test result.

	Positive Drug Test	Negative Drug Test
Subject Uses Drugs	44 (True Positive)	6 (False Negative)
Subject Does Not Use Drugs	90 (False Positive)	860 (True Negative)

Solution.

Example 2.14. Suppose that a balanced die is tossed once. Use Definition 2.9 to find the probability of a 1, given that an odd number was obtained.

Bayes' Rule

Total probability theorem

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

Proof.

Partition of a Sample Space

For some positive integer k, let the sets $S_1, S_2, \cdots, S_k$ be such that

1. $S = S_1 \cup S_2 \cup \cdots \cup S_k$
2. $S_i \cap S_j = \emptyset$, for $i \neq j$

Then the collection of sets $\{S_1, S_2, \cdots, S_k\}$ is said to be a **partition** of S.

Note. If A is any subset of S and $\{S_1, S_2, \cdots, S_k\}$ is a partition of S, A can be decomposed as follows:

$$A = (A \cap S_1) \cup (A \cap S_2) \cup \cdots \cup (A \cap S_k).$$

Bayes' Rule

Total probability theorem

Let $S_1, S_2, \cdots, S_k$ be a partition of the sample space S with **prior probabilities** $P(S_1), P(S_2), \cdots, P(S_k)$. Let $A \subseteq S$. Then

$$P(A) = \sum_{i=1}^{k} P(A|S_i)P(S_i).$$

Example. A study of residents of a region showed that 20% were smokers. The probability of death due to lung cancer, given that a person smoked, was ten times the probability of death due to lung cancer, given that the person did not smoke. If the probability of death due to lung cancer in the region is 0.006, what is the probability of death due to lung cancer given that the person is a smoker?

Solution.

Bayes' Rule

Bayes' Rule

Let $S_1, S_2, \cdots, S_k$ be a partition of the sample space S with **prior probabilities** $P(S_1), P(S_2), \cdots, P(S_k)$. Suppose an event A occurs and $P(A|S_i)$ is known for each $i = 1, \ldots, k$. Then the **posterior probability** of S_i, given that A occurred is

$$P(S_i|A) = \frac{P(A \cap S_i)}{P(A)} = \frac{P(S_i)P(A|S_i)}{\sum_{j=1}^{k} P(S_j)P(A|S_j)}, i = 1, \ldots, k.$$

Bayes' Rule

Example. In Orange County, 51% are males and 49% are females. One adult is selected at random for a survey involving credit card usage.

(a) Find the prior probability that the selected person is male.

(b) It is later learned the survey subject was smoking a cigar, and 9.5% of males smoke cigars (only 1.7% of females do). Now find the probability the selected subject is male.

Solution.

Independence

Independence Definition 1

Two events, A and B, are said to be independent if and only if the probability that event A occurs does not change, depending on whether or not event B has occurred. That is, $P(A|B) = P(A)$ (or $P(B|A) = P(B)$ by symmetry of A and B.)

Note. If the events A and B are independent, then the probability that both A and B occur is

$$P(A \cap B) = P(A)P(B).$$

Independence Definition 2

Two events, A and B, are said to be independent if and only if

$$P(A \cap B) = P(A)P(B).$$

Q. If A and B are disjoint, and $P(A) > 0$ and $P(B) > 0$, are they independent?

Two Laws of Probability

Multiplication rule

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Addition rule

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

probability of the complement

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

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

Example. Topford supplies X-Data DVDs in lots of 50, and they have a reported defect rate of 0.5% so the probability of a disk being defective is 0.005. It follows that the probability of a disk being good is 0.995. What is the probability of getting at least one defective disk in a lot of 50?

Solution.

Two Laws of Probability

Example (Ex. 2.95). Two events A and B are such that
$P(A) = 0.2$, $P(B) = 0.3$ and $P(A \cup B) = 0.4$.
(1) Find $P(A \cap B)$
(2) Find $P(\overline{A} \cup \overline{B})$
(3) Find $P(\overline{A} \cap \overline{B})$
(4) Find $P(\overline{A}|B)$
(5) Are A and B independent?

Solution.

Lecture 3
Discrete Random Variables

MATH 311 Statistics I: Probability Theory
February, 2023

Xuemao Zhang
East Stroudsburg University

Agenda

1. Random Variables
2. Discrete Random Variables
3. The Binomial Distribution
4. The Poisson Distribution
5. The Geometric Distributions
6. The Negative Binomial Distribution
7. The Hypergeometric Distribution
8. Moments and Moment Generating Functions
9. Probability-Generating Functions
10. Tchebysheff's Theorem

Random Variables

Random variable

A random variable is numerical variable (typically represented by uppercase letters such as X, Y, ...) that has a single numerical value, **determined by chance**, for each outcome of an experiment.
Mathematically a **random variable** is a function from a sample space S into the real numbers though we usually ignore this fact and consider the values only of the function.

Notation. Random variables will always be denoted with uppercase letters (e.g. X) and the realized values (e.g. x) of the variable (or its range) will be denoted by the corresponding lowercase letters.

Random Variables

Examples. Random variables in some experiments.

Experiment	Random variable
1 coin toss	$X(H) = 1, X(T) = 0$
Toss two dice	X = sum of the numbers
Toss a coin 25 times	X = number of heads in 25 tosses

Discrete Random Variable

A random variable Y is said to be discrete if it can assume only a finite or countably infinite number of distinct values.

Continuous Random Variable

A random variable Y is said to be continuous if it can assume infinitely many values corresponding to the points on a real line interval.

Probability Distribution

Probability mass function

The probability mass function (pmf) of a discrete random variable Y is given by $p(y) = P(Y = y)$ for all y.

Calculation of a pmf:

(1) Collect all possible outcomes for which Y is equal to y;

(2) Add their probabilities;

(3) Repeat for all y.

Example. Toss a fair coin three times and define X = number of heads. Find the pmf $p(x)$.

Solution.

Discrete Probability Distribution Requirements

- $0 \leq p(y) \leq 1$ for all y.
- $\sum_y p(y) = 1$, where the summation is over all values of y with nonzero probability.

Expected Value

Expected value

Let Y be a discrete random variable with the probability mass function $p(y)$. Then the expected value of Y, denoted by $E(Y)$ or μ, is defined to be

$$\mu = E(Y) = \sum_y yp(y).$$

Remark. The expected value of a discrete random variable is said to exist if the sum is absolutely convergent - That is, if $\sum_y |y|p(y) < \infty$. This absolute convergence will hold for all examples in this text and will not be mentioned each time an expected value is defined.

Expected Value of a Transformation

If X is a random variable, then any function of X, say $g(X)$, is also a random variable. Often $g(X)$ is of interest itself and we write $Y = g(X)$ to denote the new random variable $g(X)$. Since Y is a function of X, we can describe the probability distribution of Y in terms of that of X (to be discussed). But we can calculate the expected value of Y without deriving its probability distribution

If X is a discrete random variable and g is a function, then

$$E[g(X)] = \sum_x g(x)p(x).$$

Remark. In general, $E[g(X)] \neq g[E(X)]$.

Proof.

Variance and Standard Deviation

Variance

$\sigma^2 = Var(Y) = E[(y-\mu)^2] = \sum_y (y-\mu)^2 p(y) = E(Y^2) - \mu^2$

Proof.

Standard Deviation

$\sigma = \sqrt{\sigma^2}$

Example. The following table describes the probability distribution of X. Find the mean, variance, and standard deviation.

$X = x$	$P(X = x)$	$xp(x)$	$x^2 p(x)$
0	0.25		
1	0.50		
2	0.25		

EAST STROUDSBURG UNIVERSITY
FOUNDED 1893

Expected Value

Properties

Let Y be a discrete random variable with probability mass function $p(y)$, mean μ and variance σ^2.

1. $E(c) = c$ for any constant c.
2. $E[cg(Y)] = cE[g(Y)]$ for any function g of Y and constant c.
3. $E\left(\sum_{i=1}^{k} g_i(Y)\right) = \sum_{i=1}^{k} E[g_i(y)]$.
4. $E(c_1 Y + c_2) = c_1 E(Y) + c_2$ for any constants c_1 and c_2.
5. $Var(c) = 0$ for any constant c.
6. $Var(cY + b) = c^2 Var(Y)$ for any constant c and b.
7. If X and Y are two **independent** random variables, and g and h are two functions, then
$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Remark. Let X and Y be two random variables. Then X and Y are independent if, for any subsets A and B of the real numbers ($A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$),

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

There will be more discussions of independence of two random variables in Chapter 4.

Binomial random variables

Binomial experiment

(1) The experiment consists of a fixed number, n, of **identical** Bernoulli(binary) trials.

(2) Each trial results in one of two outcomes: success, S, or failure, F.

(3) The trials must be independent. (The outcome of any individual trial does not affect the probabilities in the other trials.)

(4) The probability of a success p remains the same in all trials.

(5) We are interested in the number of successes in n trials.

Definition. The binomial random variable is defined as Y = number of successes out of n Bernoulli trials.

Example. The coin-tossing experiment is a simple example of a binomial random variable. Toss a fair coin $n = 3$ times and record X = number of heads.

Binomial Probability Distribution

Binomial distribution

The probability mass function of the binomial random variable Y is given by

$$p(y) = P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0, 1, 2, \ldots, n, 0 \leq p \leq 1.$$

Note. $\binom{n}{y}$ is the number of outcomes with exactly y successes among n trials.

Proof.

Binomial Probability Distribution

Mean and variance

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = Var(Y) = np(1-p).$$

Proof.

Binomial Probability Distribution

Example. The probability that a patient recovers from a stomach disease is 0.8. Suppose that 20 people are known to have contracted this disease. Let Y be the number of recoveries out of 20.

(1) What is the probability that exactly 14 recover?

(2) What is the probability that at least 12 recover?

(3) What is the probability that at least 14 but not more than 18 recover?

(4) Find the expected number of recoveries and give a standard deviation.

Solution.

Poisson Distribution

A Poisson random variable describes the number of events, that occur over a specified interval (a period of time or, space, distance, area, volume or some similar unit) during which an average of λ such events can be expected to occur.

pmf

A random variable Y is said to have a Poisson probability distribution if and only if its probability mass function is given by

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}, y = 0, 1, 2, \ldots, \lambda > 0,$$

where $e \approx 2.71828$ is a constant.

Poisson experiment

(1) Consists of an **infinite** number of identical trials.

(2) Each trial results in one of two outcomes: success, S, or failure, F.

(3) The trials are independent.

(4) The probability of a success (p) is the same for all trials.

(5) The random variable of interest is the number of successes (Y) observed.

Poisson Distribution

Example 3.18. Show that the probabilities assigned by the Poisson probability distribution satisfy the requirements that $0 \leq p(y) \leq 1$ for all y and $\sum_y p(y) = 1$.

Proof.

Poisson Distribution

Example 3.19. Suppose that a random system of police patrol is devised so that a patrol officer may visit a given beat location $Y = 0, 1, 2, 3, \ldots$ times per half-hour period, with each location being visited an average of once per time period. Assume that Y possesses, approximately, a Poisson probability distribution. Calculate the probability that the patrol officer will miss a given location during a half-hour period. What is the probability that it will be visited once? Twice? At least once?

Solution.

Poisson Distribution

Mean and variance

If Y is a Poisson random variable with parameter λ, then

$$\boxed{\mu = E(Y) = \lambda \text{ and } \sigma^2 = Var(Y) = \lambda.}$$

Proof.

Poisson Distribution

Poisson as an Approximation to the Binomial Distribution

Let Y be a binomial random variable with size n and probability of success p. Let $\lambda = np$. Then

$$\lim_{n\to\infty} \binom{n}{y} p^y (1-p)^{n-y} = \frac{\lambda^y}{y!} e^{-\lambda}, y = 0, 1, 2, \ldots$$

Note. The approximation is valid when n is large ($n \geq 100$) and p is small ($np \leq 10$).

Proof.

Geometric Probability Distribution

Geometric experiment

(1) Each trial results in one of two outcomes: success, S, or failure, F.

(2) The trials must be independent.

(3) The probability of a success p remains the same in all trials.

(4) We are interested in the number of trials until the first success.

pmf

A random variable Y is said to have a geometric probability distribution if and only if

$$p(y) = (1-p)^{y-1}p, y = 1, 2, 3, \dots, 0 \leq p \leq 1.$$

Note.
(1) $Y - 1$ is the number of failures before the first success.
(2) $\sum_{y=1}^{\infty} p(y) = 1$ follows from properties of the geometric series.

Geometric Probability Distribution

memoryless property

The geometric distribution “forgets" what has occurred. That is,

$$P(Y > s | Y > t) = P(Y > s - t) \text{ for integers } s > t.$$

Note. $P(Y > n) = P(\text{no successes in } n \text{ trials}) = (1 - p)^n$.

Proof.

Geometric Probability Distribution

The geometric distribution is sometimes used to model "lifetimes" or "time until failure" of components.

Example 3.11. Suppose that the probability of engine malfunction during any one-hour period is $p = 0.02$. Find the probability that a given engine will survive two hours.

Solution. Let X be the number of 1-hour intervals until 1st malfunction. $p = 0.02$.

Geometric Probability Distribution

Mean and variance

Let Y be a geometric random variable with probability of success p. Then

$$\mu = E(Y) = \frac{1}{p} \text{ and } \sigma^2 = Var(Y) = \frac{1-p}{p^2}.$$

Proof.

Example 3.12. If the probability of engine malfunction during any one-hour period is $p = 0.02$ and Y denotes the number of one-hour intervals until the first malfunction, find the mean and standard deviation of Y.

Geometric Probability Distribution

Exercise 3.85. Show that for a geometric random variable Y with probability of success p,

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

Negative Binomial Probability Distribution

The geometric distribution handles the case where we are interested in the number of trials on which the first success occurs. What if we are interested in knowing the number of the trials on which the second, third, or fourth success occurs? The distribution that applies to the random variable

$Y =$ the number of the trial on which the rth success occurs, $r = 2, 3, 4, \ldots$

Negative Binomial experiment

(1) Each trial results in one of two outcomes: success, S, or failure, F.

(2) The trials must be independent.

(3) The probability of a success p remains the same in all trials.

(4) Sample Space is considered infinite.

(5) We are interested in the number of trials to find the rth success.

Negative Binomial Probability Distribution

pmf

The probability mass function of a Negative Binomial random variable is

$$p(y) = \binom{y-1}{r-1}(1-p)^{y-r}p^r, y = r, r+1, r+2, \ldots, 0 \le p \le 1,$$

where p is the probability of success and r is the number of successes.

Mean and variance

The mean (expected value) and variance for the negative binomial random variable Y above are

$$\mu = E(Y) = \frac{r}{p} \text{ and } \sigma^2 = Var(Y) = \frac{r(1-p)}{p^2}.$$

Example 3.14. A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. Find the probability that the third oil strike comes on the fifth well drilled.

Negative Binomial Probability Distribution

The negative binomial distribution is sometimes defined in terms of the random variable X = number of failures before the rth success. This formulation is statistically equivalent to the one given above in terms of Y = trial at which the rth success occurs, since $X = Y - r$. Using the relationship between X and Y, the alternative form of the negative binomial distribution is

$$\boxed{P(X = x) = \binom{r + x - 1}{x} p^r (1 - p)^x, x = 0, 1, 2, \ldots}$$

$$\boxed{\mu = E(X) = \frac{r(1 - p)}{p} \text{ and } \sigma^2 = Var(X) = \frac{r(1 - p)}{p^2}.}$$

Proof.

Negative Binomial Probability Distribution

Example. A geological study indicates that an exploratory oil well should strike oil with probability 0.2.
(a) What is the probability that the first strike comes on the third well drilled?
(b) What is the probability that the third strike comes on the seventh well drilled.
(c) Find the mean and variance of the number of wells that must be drilled if the company wants to set up three producing wells.

Solution.

Hypergeometric Distribution

Suppose that a population contains a finite number of elements N that posses one of two characteristics, say red and white. Thus r of the elements might be red and $N - r$ is white. A sample of n elements is randomly selected from the population and define Y to be the number of red elements in the sample. This random variables Y is said to have a **hypergeometric distribution**.

Hypergeometric experiment

(1) Sample Space is finite.

(2) Each trial results in one of two outcomes: success, S, or failure, F.

(3) The trials are dependent.

(4) The probability of a success for each trial is different.

(5) We are interested in the number of successes in sample size n.

Hypergeometric Distribution

pmf

A random variable Y is said to have a hypergeometric probability distribution if and only if its probability mass function is given by

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer $0, 1, 2, \ldots, n$, subject to the restrictions $y \leq r$ and $n - y \leq N - r$.

Mean and variance

If Y is a hypergeometric random variable, then

$$\boxed{\mu = E(Y) = \frac{nr}{N} \text{ and } \sigma^2 = Var(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).}$$

Hypergeometric Distribution

Exercise 3.120. The size of animal populations are often estimated by using a capture-recapture method. In this method, k animals are captured, tagged, and then released into the population. Some time later n animals are captured, and Y, the number of tagged animals among the n, is noted. The probabilities associated with Y are a function of N, the number of animals in the population, so the observed value Y contains information on this unknown N. Suppose that $k = 4$ animals are tagged and then released. A sample of $n = 3$ animals is then selected at random from the same population. Find $P(Y = 1)$ as function of N. What value of N will maximize $P(Y = 1)$.

Solution.

Moments

The various moments of a distribution are an important class of expectations.

Moment and central moment

Let X be a random variable. For each integer k, the kth moment of X, denoted by μ_k', is

$$\mu_k' = E(X^k).$$

the kth central moment of X, denoted by μ_k, is

$$\mu_k = E[(X-\mu)^k],$$

where $\mu = E(X)$.

Note. $\sigma^2 = \mu_2$.

Moment Generating Function

Moment Generating Function

Let X be a random variable. The moment generating function (mgf) of X , denoted by $M_X(t)$, is

$$M_X(t) = E\left(e^{tX}\right)$$

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E\left(e^{tX}\right)$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

Q. Can $\frac{t}{1-t}$ be a mgf of a random variable?

Moment Generating Function

Uniqueness of mgf

If the mgf of a random variable X exists in a neighborhood of 0, then the distribution of X is uniquely determined. Also, if the MGFs of two random variables X and Y are equal for t near 0, then X and Y have the same probability distribution.

If X has mgf $M_X(t)$, then

$$E(X^n) = M_X^{(n)}(0),$$

where we define $M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$. That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$.

Proof.

Moment Generating Function

Example 3.23. Find the moment-generating function $m(t)$ for a Poisson distributed random variable with parameter λ. And find the mean and variance of the distribution using this mgf.

Solution.

Moment Generating Function

Example. Suppose that the discrete random variable X has the probability distribution of $P(X = 2) = 1/2$, $P(X = 5) = 1/3$, and $P(X = 7) = 1/6$.
(1) Find the moment generating function, $M_X(t)$, of X.
(2) Find the mean of X using the mgf method.

Solution.

Moment Generating Function

Exercise 3.147. Suppose Y has the geometric distribution with $p = P(S)$. Find mgf of Y.

Solution.

Moment Generating Function

Binomial Moment Generating Function.

$$\boxed{m(t) = [pe^t + (1-p)]^n.}$$

Geometric Moment Generating Function.

$$\boxed{m(t) = \frac{pe^t}{1-(1-p)e^t}.(\textbf{Exercise 3.145})}$$

Hypergeometric Moment Generating Function.

Hypergeometric has no moment generating function

Poisson Moment Generating Function.

$$\boxed{m(t) = \exp[\lambda(e^t - 1)].}$$

Negative Binomial Moment Generating Function.

$$\boxed{m(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r.}$$

Moment Generating Function

Linear combination

For any constants a and b, the mgf of the random variable $aX + b$ is given by

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

Proof.

Independence

Let X and Y be two random variables that are independent. Then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Proof.

Moment Generating Function

Let X and Y be two **independent** random variables. Combine the above two theorems, we have, for any two constants a and b,

$$\boxed{M_{aX+bY}(t) = M_{aX}(t)M_{bY}(t) = M_X(at)M_Y(bt).}$$

It can be used to find the distribution of $aX + bY$ if X and Y are independent.

Example. Let X and Y be two independent Poisson random variables with parameter λ_1 and λ_2, respectively. Find the distribution of $X + Y$.

Solution.

Probability-Generating Functions

An important class of discrete random variables is one in which Y represents a count and consequently takes integer values: $Y = 0, 1, 2, 3, \ldots$

Factorial moment

Let k be a positive integer. The kth factorial moment for a random variable Y is defined to be

$$\mu_{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-k+1)].$$

Probability-Generating Function

Let Y be a discrete random variable taking values in the non-negative integers $\{0, 1, \ldots\}$, then the probability generating function of Y is defined as

$$P_Y(t) = E(t^Y) = \sum_y t^y p(y).$$

Probability-Generating Functions

Probability-Generating Function is used to calculate a factorial moment of a random variable.

$$\left.\frac{d^k P_Y(t)}{dt^k}\right|_{t=1} = \mu_{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-k+1)].$$

Proof.

Probability-Generating Functions

Example. Find the probability-generating function for a geometric random variable. And find the mean and variance using the derived probability-generating function.

Solution.

Probability-Generating Functions

Example. Find the probability-generating function for a binomial random variable. And find the mean and variance using the derived probability-generating function.

Solution.

Probability-Generating Functions

Independence

Let X and Y be two random variables that are independent. Then

$$P_{X+Y}(t) = P_X(t)P_Y(t).$$

Proof.

The Empirical Rule

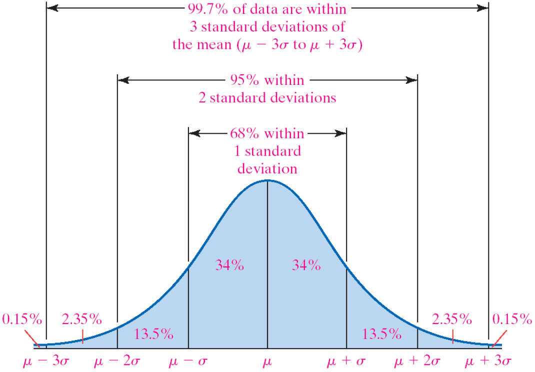
The Empirical Rule

For any symmetric bell-shaped distribution,

(1) approximately 68% of the data lie within $\mu \pm \sigma$

(2) approximately 95% of the data lie within $\mu \pm 2\sigma$

(3) approximately 99.7% of the data lie within $\mu \pm 3\sigma$



Tchebysheff's Theorem

Tchebysheff's Theorem. Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,

$$\boxed{P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}}$$

or

$$\boxed{P(|Y - \mu| \geq k\sigma) < \frac{1}{k^2}}$$

Example. 3.28 The number of customers per day at a sales counter, Y, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of Y is not known. What can be said about the probability that, tomorrow, Y will be greater than 16 but less than 24?

Solution.

Lecture 4
Continuous Random Variables

MATH 311 Statistics I: Probability Theory
March, 2022

Xuemao Zhang
East Stroudsburg University

Agenda

cumulative distribution function

Recall. A random variable Y is said to be continuous if it can assume infinitely many values corresponding to the points on a real line interval. In this section we will discuss some of the more common families of continuous distributions, those with well-known names.

cumulative distribution function

The cumulative distribution function or cdf of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P(X \leq x) \text{ for all } x.$$

cumulative distribution function of a discrete r.v.

If X is a discrete random variable X, its cdf is

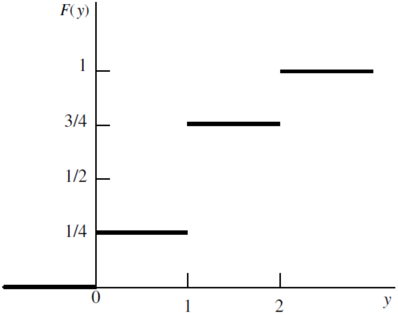
$$F_X(x) = \sum_{t \leq x} p(t).$$

cumulative distribution function

Example 4.1. Suppose that Y has a binomial distribution with $n = 2$ and $p = 1/2$. Find F(y).

Solution.

FIGURE **4.1**
Binomial distribution function,
$n = 2,\ p = 1/2$



cumulative distribution function

Properties of cdf

The function $F(x)$ is a cdf if and only if the following three conditions hold:

1. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \to \infty} F(x) = 1$.
2. $F(x)$ is a non-decreasing function of x.
3. $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x \to x_0^+} F(x) = F(x_0)$.

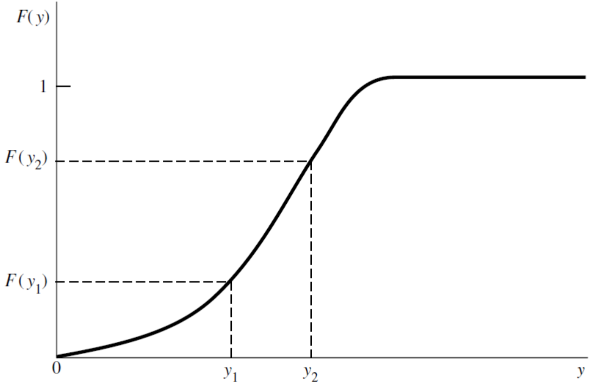
Therefore, there is a more rigorous definitions of a continuous and discrete random variable in terms of CDF.

definitions

A random variable X is **continuous** if its cdf $F(x)$ is a continuous function of x;
A random variable X is **discrete** if its cdf $F(x)$ is a step function of x.

cumulative distribution function

FIGURE **4.2**
Distribution function for a continuous random variable



probability density function

probability density function

The probability density function or pdf, $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \text{ for all } x.$$

Note.

(1) If cdf $F(x)$ is known, then pdf $f(x) = \dfrac{dF(x)}{dx} = F'(x)$.

(2) A pdf $f(x) \geq 0$.

probability density function

A function $f(x)$ is a pdf of a random variable X if and only if
(a) $f(x) \geq 0$ for all x.
(b) $\int_{-\infty}^{\infty} f(t)dt = 1$.

probability density function

Example 4.3 Let Y be a continuous random variable with pdf

$$f(y) = \begin{cases} 3y^2, & 0<y<1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the cumulative distribution function $F(y)$ of Y.

Solution.

probability density function

Example 4.4. Given $f(y) = cy^2, 0 \leq y \leq 2$, and $f(y) = 0$ elsewhere, find the value of c for which $f(y)$ is a valid density function.

Solution.

Continuous Probability Distribution

Example 4.2. Suppose

$$F(y) = \begin{cases} 0, & y \le 0; \\ y, & 0 \le y \le 1; \\ 1, & y > 1. \end{cases}$$

Find the probability density function $f(y)$.

Solution.

Continuous Probability Distribution

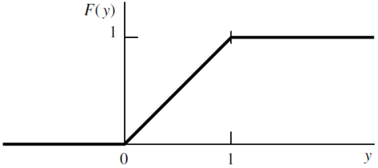


FIGURE **4.4**
Distribution function
$F(y)$ for Example 4.2

FIGURE **4.5**
Density function
$f(y)$ for Example 4.2

Remark. $F(y)$ is a continuous function of y, but $f(y)$ is discontinuous at the points $y = 0, 1$. In general, the cdf for a continuous random variable must be continuous, but the density function need not be everywhere continuous.

Continuous Probability Distribution

Theorem. If the random variable Y has density function $f(y)$ and $a < b$, then the probability that Y falls in the interval $[a, b]$ is

$$\boxed{P(a \le Y \le b) = F(b) - F(a) = \int_a^b f(t)dt.}$$

Remark. $P(Y = c) = 0$ for any constant c.

Exercise 4.14. Let Y be a continuous random variables with probability density function

$$f(y) = \begin{cases} y, & 0 \le y \le 1; \\ 2 - y & 1 < y < c; \\ 0, & \text{elsewhere.} \end{cases}$$

(1) Find c.
(2) Find $F(y)$
(3) Find $P(0.5 \le Y \le 1.5)$.

Solution.

Expected Values

Expected value

Let Y be a continuous random variable with the probability density function $f(y)$. Then the expected value of Y, denoted by $E(Y)$ or μ, is defined to be

$$\mu = E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

provided that the integral exists.

Expected Value of a Transformation

If Y is a continuous random variable and g is a function, then

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f_Y(y)dy.$$

Remark. In general, $E[g(Y)] \neq g[E(Y)]$.

Variance

$$\sigma^2 = Var(Y) = E[(Y-\mu)^2] = \int_{-\infty}^{\infty}(y-\mu)^2 f_Y(y)dy = E(Y^2) - \mu^2$$

Expected Values

Remark. The properties of expectation for continuous random variables are the same the same as those for the discrete random variables.

Properties

Let Y be a continuous random variable with probability density function $f(y)$, mean μ and variance σ^2.

1. $E(c) = c$ for any constant c.
2. $E[cg(Y)] = cE[g(Y)]$ for any function g of Y and constant c.
3. $E\left(\sum_{i=1}^{k} g_i(Y)\right) = \sum_{i=1}^{k} E[g_i(y)]$.
4. $E(c_1 Y + c_2) = c_1 E(Y) + c_2$ for any constants c_1 and c_2.
5. $Var(c) = 0$ for any constant c.
6. $Var(cY + b) = c^2 Var(Y)$ for any constant c and b.
7. If X and Y are two **independent** random variables, and g and h are two functions, then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Expected Values

Example. Refer to the previous example. Find $E(Y)$ and $E(Y^2)$.

Solution.

Other Expected Values

Moment and central moment. Let X be a continuous random variable. For each integer k,

$$\boxed{\mu_k' = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx}$$

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

Moment Generating Function

Let X be a continuous random variable. The moment generating function (mgf) of X is

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

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E\left(e^{tX}\right)$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

Other Expected Values

Remark. or **Recall.**

(1) If the mgf of a random variable X exists in a neighborhood of 0, then the distribution of X is **uniquely** determined. Also, if the MGFs of two random variables X and Y are equal for t near 0, then X and Y have the same probability distribution.

(2) Let X be a continuous random variable and g a function of X. Let $Y = g(x)$. Then

$$\boxed{M_Y(t) = E\left(e^{tg(X)}\right) = \int_{-\infty}^{\infty} e^{tg(x)} f(x)dx}$$

and thus

$$\boxed{M_{aX+b}(t) = e^{bt} M_X(at).}$$

for any constants a and b.

(3) Let X and Y be two random variables that are **independent**. Then

$$\boxed{M_{X+Y}(t) = M_X(t)M_Y(t).}$$

Other Expected Values

If X has mgf $M_X(t)$, then

$$E(X^n) = M_X^{(n)}(0),$$

where we define $M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$. That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$.

Example. Let Y be a continuous random variable with pdf

$$f(y) = \begin{cases} 2y, & 0 \le y \le 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find mgf of Y.

The Uniform Distribution

Definition. A random variable Y is said to have a uniform distribution on $[\theta_1, \theta_2], \theta_1 < \theta_2$ if its probability density distribution function is of the form

$$\boxed{f(y) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & \theta_1 \le y \le \theta_2; \\ 0, & \text{otherwise.} \end{cases}}$$

mgf

The moment generating function of the above Uniform Distribution is

$$m(t) = \frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$$

Mean and variance

$$E(Y) = \frac{\theta_1 + \theta_2}{2} \text{ and } Var(Y) = \frac{(\theta_1 - \theta_2)^2}{12}$$

The Uniform Distribution

Example. The change in depth of a river from one day to the next, measured (in feet) at a specific location, is a random variable Y with the following density function

$$\boxed{f(y) = \begin{cases} k, & -2 \leq y \leq 2; \\ 0, & \text{otherwise.} \end{cases}}$$

(a) Find k.
(b) Find the cdf of Y.
(c) Find $P(-0.2 < Y < 1.2)$.

Solution.

The Normal Distribution

Definition. A random variable Y is said to have a normal probability distribution if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the pdf of Y is

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

Theorem. The mgf of $Y \sim N(\mu, \sigma^2)$ is

$$\boxed{m(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right).}$$

Proof.

The Normal Distribution

Theorem 4.7. If Y is a normally distributed random variable with parameters μ and σ, then

$$\boxed{E(Y) = \mu \text{ and } Var(Y) = \sigma^2.}$$

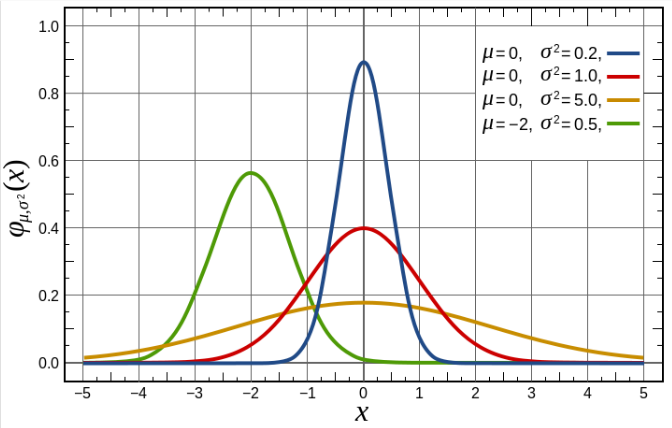
Proof.

Theorem. Let $Y \sim N(\mu, \sigma^2)$. Then

$$\boxed{Z = \frac{Y - \mu}{\sigma} \sim N(0, 1).}$$

Proof.

The Normal Distribution



1. Mean = μ; Standard deviation = σ.
2. Symmetric about $x = \mu$.
3. Total area under the curve is 1.

The Normal Distribution

EXAMPLE 4.9. The achievement scores for a college entrance examination are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90?

Solution.

The Gamma Distribution

Gamma function

The Gamma function, denote by Γ is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0.$$

Properties of Gamma function.

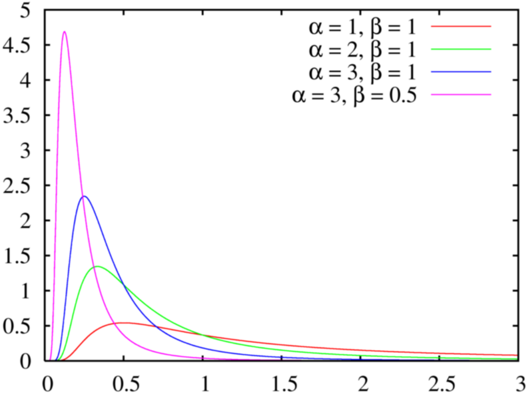
1. $0 < \Gamma(\alpha) < \infty$
2. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (Exercise 4.81)
3. $\Gamma(n) = (n - 1)!$ if n is a positive integer. (Exercise 4.82)
4. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The Gamma Distribution

Definition. A random variable Y is said to have a Gamma distribution with parameters (α, β) if the pdf of Y is

$$\boxed{f(y) = \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad 0 \leq y < \infty}.$$

α is called Shape parameter and β is called Scale parameter.



The Gamma Distribution

mgf

The moment generating function for Gamma(α, β) distributions is

$$m(t) = (1 - \beta t)^{-\alpha}.$$

Proof.

The Gamma Distribution

Theorem 4.8. If $Y \sim \text{Gamma}(\alpha, \beta)$, then

$$\boxed{\mu = E(Y) = \alpha\beta \text{ and } \sigma^2 = Var(Y) = \alpha\beta^2.}$$

Proof.

The Gamma Distribution

Exercise 4.110. If Y has a probability density function given by

$$f(y) = 4y^2 e^{-2y}, \; y > 0.$$

Find $E(Y)$ and $Var(Y)$ by inspection.

Solution.

The Gamma Distribution

Example. The weekly amount of down time Y (in hours) for an industrial machine has approximately a gamma distribution with $\alpha = 3$ and $\beta = 2$. The loss L (in dollars) to the industrial operation as a result of this down time is given by

$$L = 30Y + 2Y^2.$$

Find the expected value of L.

Solution.

Special Case of Gamma Distribution - Chi-square Distribution

Definition. Let v be a positive integer. Then the Gamma distribution with

$$\boxed{\alpha = \frac{v}{2} \qquad \beta = 2}$$

is called the Chi-square distribution with v degrees of freedom, denoted by $\chi^2(v)$.

Theorem. If $Y \sim \chi^2(v)$, then

$$\boxed{E(Y) = v \text{ and } Var(Y) = 2v.}$$

Special Case of Gamma Distribution - Exponential Distribution

Definition. The Gamma distribution with $\alpha = 1$ is called the exponential distribution with parameter β:

$$\boxed{f(y) = \frac{1}{\beta} e^{-y/\beta}, \quad 0 \le y < \infty.}$$

Theorem. If $Y \sim$exponential(β), then

$$\boxed{E(Y) = \beta \text{ and } Var(Y) = \beta^2.}$$

Remark. (1) If $Y \sim$exponential(β), then $F(y) = 1 - e^{-y/\beta}$.

(2) The exponential distribution is used to model lifetimes(analogous to the use of geometric distribution).

Special Case of Gamma Distribution - Exponential Distribution

Exercise 4.88. The magnitude of earthquakes recorded in a region of North America can be approximately modeled as having an exponential distribution with mean 4.0 on the scale.

$$f(y) = \frac{1}{\beta} e^{-y/\beta}, \;\; 0 < y < \infty$$

(1) Find the probability that an earthquake striking this region will exceed 5.0 on the scale.
(2) Of the next 10 earthquakes to strike this region, what is the probability that at least 2 will exceed 5.0 on the scale. (Hint: Let X = number of earthquakes that will exceed 5.0. Then X Binomial distribution with $n = 10$ and $p = P(Y > 5)$)

Solution.

Poisson versus Exponential Distribution

Theorem. The waiting time of a Poisson random variable has an exponential distribution. That is, if

$$X \sim \textit{Poisson}(\lambda)$$

and let T be the time to the first occurrence (waiting time), then

$$T \sim \textit{Exponential}(\frac{1}{\lambda})$$

Partial verification. $E(X) = \lambda$ and $E(T) = \frac{1}{\lambda}$.

Q. At a certain location of the expressway I-80, the number of cars exceeding the speed limit by more than ten miles per hour in half an hour is a random variable having a Poisson distribution with $\lambda = 2$. A policeman was hiding in bushes to catch speeders. What is the average time for the policeman to catch a first (or next) driver whose car exceeded the speed limit by more than ten miles per hour?

Exponential Distribution

Memoryless property. Suppose that $Y \sim$Exponential(β). If $a > 0$ and $b > 0$,

$$\boxed{P(Y > a + b | Y > a) = P(Y > b).}$$

Proof.

Beta function

Beta function

The Beta function, denote by B is defined as

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt, \quad x>0, y>0.$$

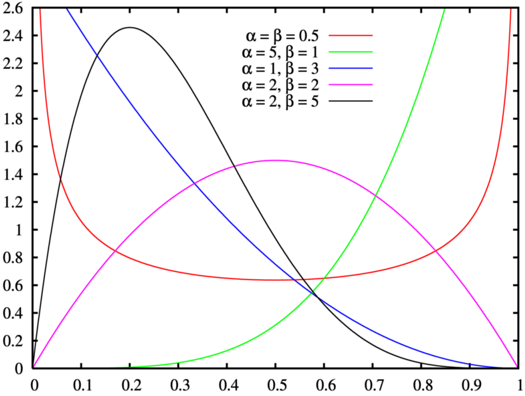
Properties of Beta function.

1. $B(x,y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$
2. $B(x,y) = B(y,x)$
3. $B(x,y) = \dfrac{(x-1)!(y-1)!}{(x+y-1)!}$ if x and y are positive integers.

The Beta Distribution

Definition. A random variable Y is said to have a Beta distribution with parameters $(\alpha, \beta), \alpha > 0, \beta > 0$ if the pdf of Y is

$$\boxed{f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, \quad 0 \le y \le 1.}$$



The Beta Distribution

Theorem 4.11. If $Y \sim \text{Beta}(\alpha, \beta)$ with $\alpha > 0, \beta > 0$, then

$$\boxed{\mu = E(Y) = \frac{\alpha}{\alpha+\beta} \text{ and } \sigma^2 = Var(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}.}$$

Remark. The mgf of a Beta distribution does not exist in closed form.

Proof.

The Beta Distribution

Exercise 4.133. The proportion of time per day that all checkout counters in a supermarket are busy is a random variable Y with a density function

$$f(y) = cy^2(1-y)^4, \quad 0 \le y \le 1.$$

Find $c, E(Y)$ and $P(Y < 0.5)$.

Solution.

The Beta Distribution

Example. During an 8-hour shift, the proportion of time Y that a sheet-metal stamping machine is down for repair has a beta distribution with $\alpha = 1$ and $\beta = 2$. The cost (in hundreds of dollars) of this downtime, due to lost of production and cost of repair, is given by

$$C = 20Y + Y^2.$$

Find the mean of C.

Solution.

Lecture 5
Multivariate Probability Distributions

MATH 311 Statistics I: Probability Theory
April, 2023

Xuemao Zhang
East Stroudsburg University

Agenda

Introduction

Probability mass function. For a discrete random variable Y, the function p defined by

$$p(y) = P(Y = y)$$

is called the *probability mass function* of Y.

CDF (cumulative distribution function) or distribution function. For any random variable X,

$$F(x) = P(X \leq x) \text{ for all } x.$$

is called the *distribution function* of X.

Probability density function. For a continuous random variable Y, the function f defined by

$$f(y) = F^{'}(y)$$

is called the *probability density function* of Y.

In this chapter, we discuss probability models that involve more than one random variable-naturally enough, called *multivariate models*.

In an experiment, it would be very unusual to observe only the values of one random variable. For example, consider an experiment designed to gain information about some health characteristics of a population of people. It would be a modest experiment indeed if the only data collected were the body weights of several people. Rather, several physical characteristics can be measured on each person. For example, temperature, height, and blood pressure, in addition to weight, might be measured.

Bivariate and Multivariate Probability Distributions

Definition. An n-dimensional **random vector** is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean space.

Definition. The **joint probability mass function** for discrete random variables X and Y is the function p defined by

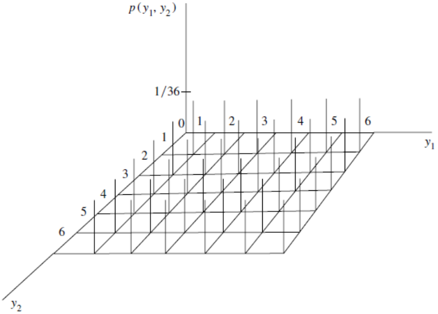
$$\boxed{p(x, y) = P(X = x, Y = y)}$$

where $(X = x, Y = y)$ denotes the intersection of the events $(X = x)$ and $(Y = y)$.

Example. Consider the experiment of tossing a pair of dice and define

$$\begin{aligned} Y_1 &: \text{ The number of dots appearing on die 1.} \\ Y_2 &: \text{ The number of dots appearing on die 2.} \end{aligned}$$

FIGURE 5.1 Bivariate probability function; y_1 = number of dots on die 1, y_2 = number of dots on die 2



Bivariate and Multivariate Probability Distributions

Theorem 5.1. If Y_1 and Y_2 are discrete random variables with joint pmf $p(y_1, y_2)$, then

1. $p(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\sum\limits_{y_1, y_2} p(y_1, y_2) = 1$, where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities.

EXAMPLE 5.1. A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let Y_1 denote the number of customers who choose counter 1 and Y_2, the number who select counter 2. Find the joint probability function of Y_1 and Y_2.

Bivariate and Multivariate Probability Distributions

Definition. The **joint distribution function** or **joint CDF** for any random variables X and Y is the function F defined by

$$\boxed{F(x, y) = P(X \leq x, Y \leq y)}$$

EXAMPLE 5.2. Consider the random variables Y_1 and Y_2 of Example 5.1. Find $F(-1, 2)$ and $F(1.5, 2)$.

Solution.

Bivariate and Multivariate Probability Distributions

Bivariate cumulative distribution functions satisfy a set of properties similar to those specified for univariate cumulative distribution functions.

Theorem. If Y_1 and Y_2 are random variables with joint distribution function $F(y_1, y_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0.$
2. $F(\infty, \infty) = 1.$
3. If $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0$$

Joint probability density function. Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a nonnegative function $f(y_1, y_2)$, such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2 dt_1,$$

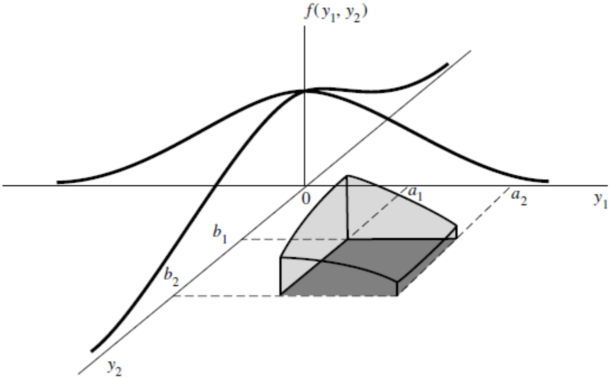
for all $-\infty < y_1 < \infty, -\infty < y_2 < -\infty$, then Y_1 and Y_2 are said to be **jointly continuous** random variables. The function $f(y_1, y_2)$ is called the **joint probability density function**.

Bivariate and Multivariate Probability Distributions

Theorem. If Y_1 and Y_2 are jointly continuous random variables with a joint density function given by $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 dy_1 = 1$.

FIGURE 5.2
A bivariate density function $f(y_1, y_2)$



Bivariate and Multivariate Probability Distributions

If Y_1 and Y_2 are jointly continuous and $F(y_1, y_2)$ is the joint cdf, then

$$f(y_1, y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} F(y_1, y_2)$$

if $F(y_1, y_2)$ is differentiable.

A pair (X, Y) of continuous random variables is described by its joint density function f: For any region R,

$$\boxed{P((X, Y) \in R) = \int \int_R f(x, y)\, dxdy}.$$

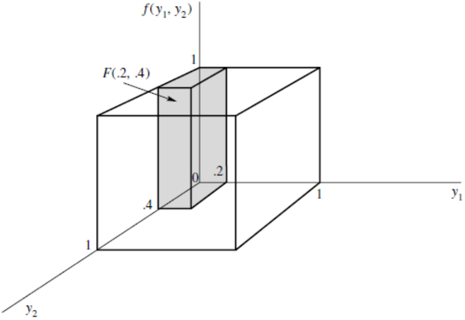
Bivariate and Multivariate Probability Distributions

Example 5.3 (bivariate uniform). Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions within the unit square and of equal area are considered, the particle is equally likely to be in either region. Let Y_1 and Y_2 denote the coordinates of the particle's location. A reasonable model for the relative frequency histogram for Y_1 and Y_2 is the bivariate analogue of the univariate uniform density function:

$$f(y_1, y_2) = \begin{cases} 1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a. Sketch the probability density surface.
b. Find $F(0.2, 0.4)$.
c. Find $P(0.1 \leq Y_1 \leq 0.3, 0 \leq Y_2 \leq 0.5)$.

FIGURE 5.3 Geometric representation of $f(y_1, y_2)$, Example 5.3



Bivariate and Multivariate Probability Distributions

Solution.

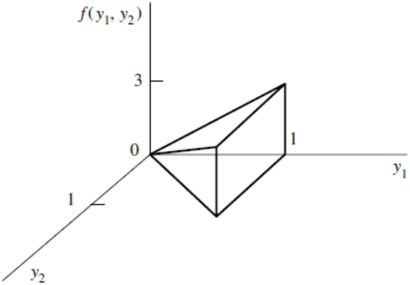
Bivariate and Multivariate Probability Distributions

Example 5.4 (triangle). Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables take on values between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose that the joint density function for Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

A sketch of this function is given in Figure 5.4. Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

FIGURE 5.4 The joint density function for Example 5.4



Multivariate Probability Distributions

Introduction

Bivariate and Multivariate Probability Distributions

Marginal and Conditional Probability Distributions

Independent Random Variables

Expected Values

Covariance

Expected Value and Variance of Linear Functions of Random Variables

Multinomial Probability Distribution

Bivariate and Multivariate Probability Distributions

Solution.

Multivariate Probability Distributions

we can define a probability function (or probability density function) for the intersection of n events $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$.

The joint pmf corresponding to the discrete case is given by

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n).$$

The joint density function of $Y_1, Y_2, \ldots, Y_n$ is given by $f(y_1, y_2, \ldots, y_n)$ such that

$$P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) = F(y_1, y_2, \ldots, y_n)$$
$$= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) dt_n \cdots dt_1.$$

Exercise 5.8 (A non-uniform distribution on the square). Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 4xy, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability $P(X \leq \frac{3}{4}, Y \leq \frac{3}{4})$.

Solution.

Exercise 5.11 (The triangle). Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 1, & 0 \leq y \leq 1, y-1 \leq x \leq 1-y \\ 0, & \text{elsewhere.} \end{cases}$$

(1) Calculate $P(X \leq \frac{3}{4}, Y < \frac{3}{4})$.
(2) Calculate $P(X \geq Y)$.

Solution.

Exercise 5.7 (Two exponential-type random variables). Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{8}xe^{-(x+y)/2}, & x \geq 0, y \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

(1) Calculate $P(X \leq Y)$.
(2) Calculate $P(X + Y \leq 1)$.

Solution.

Marginal Probability Distributions

Definition.

a. Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. Then the marginal probability functions of Y_1 and Y_2, respectively, are given by

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2) \quad \text{and} \quad p_2(y_2) = \sum_{y_1} p(y_1, y_2).$$

b. Let Y_1 and Y_2 be jointly continuous random variables with joint density function $f(y_1, y_2)$. Then the marginal density functions of Y_1 and Y_2, respectively, are given by

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_2 \quad \text{and} \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_1.$$

Example 5.5. From a group of three Republicans, two Democrats, and one independent, a committee of two people is to be randomly selected. Let Y_1 denote the number of Republicans and Y_2 denote the number of Democrats on the committee. Find the joint probability function of Y_1 and Y_2 and then find the marginal probability function of Y_1.

Marginal Probability Distributions

Solution.

Table 5.2 Joint probability function for Y_1 and Y_2, Example 5.5

	y_1			
y_2	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

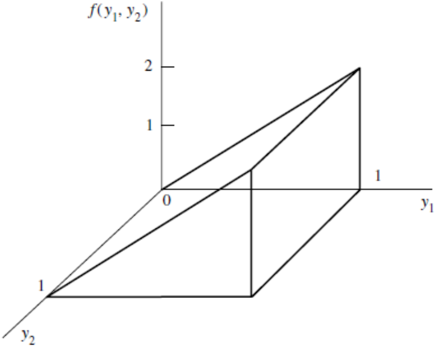
Marginal Probability Distributions

Example 5.6. Suppose that the joint density function for Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Sketch $f(y_1, y_2)$ and find the marginal density functions for Y_1 and Y_2.

FIGURE 5.6 Geometric representation of $f(y_1, y_2)$, Example 5.6



Marginal Probability Distributions

Solution.

Conditional Probability Distributions

Often times when two random variables, (X, Y), are observed, the values of the two variables are related. Knowledge about the value of X gives us some information about the value of Y even if it does not tell us the value of Y exactly. Conditional probabilities regarding Y given knowledge of the X value can be computed using the joint distribution of (X, Y).

Conditional discrete probability function. If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then the conditional discrete probability function of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{p(y_1, y_2)}{p_2(y_2)},$$

provided that $p_2(y_2) > 0$.

Conditional Probability Distributions

Example 5.7. Refer to Example 5.5 and find the conditional distribution of Y_1 given that $Y_2 = 1$. That is, given that one of the two people on the committee is a Democrat, find the conditional distribution for the number of Republicans selected for the committee

Solution.

Conditional Probability Distributions

Example. The joint probability distribution of two discrete random variables X and Y is given by

		X		
		1	2	3
Y	1	1/10	1/10	2/10
	2	1/10	2/10	3/10

Find the conditional pmf $p(y|2)$, which is $P(Y = y|X = 2), y = 1, 2$.

Solution.

Conditional Probability Distributions

Conditional distribution function. If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$, then the conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2).$$

Conditional pdf. Let Y_1 and Y_2 be jointly continuous random variables with joint density $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density of Y_1 given $Y_2 = y_2$ is given by

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

and, for any y_1 such that $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_2 = y_2$ is given by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

Conditional Probability Distributions

Example 5.8. A soft-drink machine has a random amount Y_2 in supply at the beginning of a given day and dispenses a random amount Y_1 during the day (with measurements in gallons). It is not resupplied during the day, and hence $Y_1 \leq Y_2$. It has been observed that Y_1 and Y_2 have a joint density given by

$$f(y_1, y_2) = \begin{cases} 1/2, & 0 \leq y_1 \leq y_2 \leq 2, \\ 0, & \text{elsewhere.} \end{cases}$$

That is, the points (y_1, y_2) are uniformly distributed over the triangle with the given boundaries. Find the conditional density of Y_1 given $Y_2 = y_2$. Evaluate the probability that less than 1/2 gallon will be sold, given that the machine contains 1.5 gallons at the start of the day.

Solution.

Conditional Probability Distributions

Exercise 5.24. Uniform distribution on the square.
Let Y_1 and Y_2 be the bivariate analogue of the univariate uniform density function:

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $f(y_1)$, $f(y_2)$ and $f(y_1 | Y_2 = y_2)$.

Solution.

Conditional Probability Distributions

Exercise 5.29 (The triangle).
Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 1, & 0 \le y \le 1, y-1 \le x \le 1-y \\ 0, & \text{elsewhere.} \end{cases}$$

Find $f(x)$, $f(y)$ and $f(x|Y=y)$.

Solution.

Independent Random Variables

Sometimes, knowledge about X in the conditional probability of y gives us no further information about Y.

Recall Two events A and B are independent if $P(A \bigcap B) = P(A) \times P(B)$. If two variables Y_1 and Y_2 are independent, for any two sets A and B, we would like to have

$$P(Y_1 \in A, Y_2 \in B) = P(Y_1 \in A)P(Y_2 \in B).$$

Definition.
Let Y_1 have distribution function $F_1(y_1)$, Y_2 have distribution function $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be **independent** if and only if

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

for every pair of real numbers (y_1, y_2). If Y_1 and Y_2 are not independent, they are said to be **dependent**.

Theorem. If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

for all pairs of real numbers (y_1, y_2).
Remark.

- Two discrete random variables Y_1 and Y_2 are independent if and only if $p(y_1|y_2) = p(y_1)$ for all values of y_2.

Independent Random Variables

Example 5.9. For the die-tossing problem of Section 5.2, show that Y_1 and Y_2 are independent.

Solution.

Example 5.10. Refer to Example 5.5. Is the number of Republicans in the sample independent of the number of Democrats? (Is Y_1 independent of Y_2?)

Table 5.2 Joint probability function for Y_1 and Y_2, Example 5.5

	y_1			
y_2	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

Independent Random Variables

Revisit **Example.** The joint probability distribution of two discrete random variables X and Y is given by

		X		
		1	2	3
Y	1	1/10	1/10	2/10
	2	1/10	2/10	3/10

Are X and Y independent? Justify your answer.

Solution.

Independent Random Variables

Theorem. If Y_1 and Y_2 are continuous random variables with joint pdf $f(y_1, y_2)$ and marginal probability density functions $f_1(y_1)$ and $f_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

for all pairs of real numbers (y_1, y_2).

Remarks.

- Y_1 and Y_2 are independent if and only if $f(y_1|y_2)$ is not related to y_2 (not a function of y_2.)
- If the ranges of Y_1 and Y_2 are depending on each other, Y_1 and Y_2 are NOT independent.

Example 5.11. Let

$$f(y_1, y_2) = \begin{cases} 6y_1y_2^2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that Y_1 and Y_2 are independent.

Solution.

Independent Random Variables

Example 5.12. Let

$$f(y_1, y_2) = \begin{cases} 2, & 0 \le y_2 \le y_1 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that Y_1 and Y_2 are dependent.

Solution.

Independent Random Variables

Theorem 5.5. Let Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every x and y,

$$f(x, y) = g(x)h(y)$$

where $g(x)$ is a nonnegative function of x alone and $h(y)$ is a nonnegative function of y alone.

Remark: The key benefit of the result given in Theorem 5.5 is that we **do not actually need to derive the marginal densities**.

Indicator Function. Let y be a variable and A be a set. Then the indicator function $1_A(y)$ is defined as

$$1_A(y) = \begin{cases} 1, & y \in A \\ 0, & y \notin A \end{cases}$$

Independent Random Variables

EXAMPLE 5.13 Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 2x, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Are X and Y independent?

Solution.

Exercise 5.55 (The triangle).
Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 1, & 0 \leq y \leq 1, y-1 \leq x \leq 1-y \\ 0, & \text{elsewhere.} \end{cases}$$

Are X and Y independent?

Solution.

Expected Values

Definition. Let $g(Y_1, Y_2, \ldots, Y_k)$ be a function of the discrete random variables, $Y_1, Y_2, \ldots, Y_k$, which have joint pmf $p(y_1, y_2, \ldots, y_k)$. Then the expected value of $g(Y_1, Y_2, \ldots, Y_k)$ is

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \sum_{y_k} \cdots \sum_{y_1} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k).$$

If $Y_1, Y_2, \ldots, Y_k$ are continuous random variables with joint density function $f(y_1, y_2, \ldots, y_k)$, then

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) f(y_1, y_2, \ldots, y_k) dy_1 \cdots dy_k.$$

Remark. The definition can be used to **calculate the expected value of a marginal distribution without deriving the marginal distribution**.

Expected Values

Example 5.15.
Let Y_1 and Y_2 have the joint density function:

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_1 Y_2)$.

Solution.

Expected Values

Example 5.17.
Let Y_1 and Y_2 have the joint density function:

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_1)$, $Var(Y_1)$ and $E(Y_2)$.

Solution.

Expected Values

Example 5.19.
Let Y_1 and Y_2 have the joint density function:

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_1 Y_2)$.

Solution.

Expected Values

Theorem 5.7. and 5.8 Let $g(Y_1, Y_2)$ and $h(Y_1, Y_2)$ be two functions of the random variables Y_1 and Y_2 and let c be a constant. Then

$$E[cg(Y_1, Y_2) + h(Y_1, Y_2)] = cE[g(Y_1, Y_2)] + E[h(Y_1, Y_2)].$$

Example 5.20. Refer to Example 5.4. The random variable $Y_1 - Y_2$ denotes the proportional amount of gasoline remaining at the end of the week. Find $E(Y_1 - Y_2)$.

Solution.

Expected Values

Theorem 5.9. Let X and Y be two **independent** random variables. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then

$$E[g(X)h(Y)] = [Eg(x)]E[h(y)],$$

provided that the expectations exist.

Proof.

Expected Values

Example 5.21. Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 2(1-x), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

By investigating the form of the joint density function given there, we can see that X and Y are independent. Find $E(XY)$ by using the result that $E(XY) = E(X)E(Y)$ if X and Y are independent.

Expected Values

Conditional Expectations. Given the value y of a random variable Y,

$$\boxed{E(X|Y=y) = \begin{cases} \sum_x x p_{X|y}(x|y), & \text{discrete;} \\ \int_x x f_{X|y}(x|y)dx, & \text{continuous.} \end{cases}}$$

Note. $E(X|Y)$ is a r.v. since it is a function of Y and it is easy to define $E(g(X)|Y)$ for a function g.

Expected Values

Revisit **Example.** The joint probability distribution of two discrete random variables X and Y is given by

		X		
		1	2	3
Y	1	1/10	1/10	2/10
	2	1/10	2/10	3/10

Find $E(Y|X=2)$.

Solution.

Expected Values

Total Expectation Theorem (Law of Iterated Expectations):

$$\boxed{E[E(X|Y)] = E(X).}$$

Proof.

Expected Values

Conditional Variances. Conditional variance of X given the value y of Y,

$$\boxed{Var(X|Y=y) = E\left\{[X - E(X|Y=y)]^2\right\}}$$

Note. $Var(X|Y)$ is a r.v. since it is a function of Y
Law of Total Variance:

$$\boxed{Var(X) = E\left[Var(X|Y)\right] + Var\left[E(X|Y)\right].}$$

Proof.

Expected Values

Example. An insect lays a large number of eggs, each surviving with probability p. On average, how many eggs will survive? Assume that the number of eggs has a Poisson distribution with parameter λ.

Solution.

Expected Values

Example. Suppose X has a binomial distribution with parameters n and p, where p itself is random and has a Beta distribution with parameters α and β. Find $E(X)$ and $Var(X)$.

Solution.

Expected Values

Example 5.33. A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting Y, the number of defectives. If p denotes the probability of observing a defective, then Y has a binomial distribution, assuming that a large number of items are produced by the line. But p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $1/4$. Find the expected value and variance of Y.

Solution.

Covariance

Covariance. The covariance of X and Y is the expected value of $(X - \mu_X)(Y - \mu_Y)$. That is,

$$\boxed{Cov(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]}$$

Theorem. For any random variables (X, Y) we have

$$\boxed{Cov(X, Y) = E(XY) - \mu_X\mu_Y.}$$

Proof.

Covariance

Linear correlation coefficient. The linear correlation coefficient of X and Y is defined to be,

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

where σ_X and σ_Y are the standard deviations of X and Y, respectively.

Theorem. The linear correlation coefficient satisfies

$$\boxed{-1 \leq \rho \leq 1}$$

The linear correlation ρ measures how closely X and Y are related linearly. For a strong correlation $|\rho| \approx 1$.

Covariance

Example 5.22. Suppose that the joint density function for X and Y is given by

$$f(x,y) = \begin{cases} 3x, & 0 \le y \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $Cov(X, Y)$.

Solution.

Covariance

Example 5.23. Suppose that the joint density function for X and Y is given by

$$f(x, y) = \begin{cases} 2x, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $Cov(X, Y)$.

Solution.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Covariance

Theorem. If X and Y are independent random variables, then $Cov(X, Y) = 0$. Thus, independent random variables must be uncorrelated. However, **the reverse is not true**.

Theorem. If X and Y are both normal random variables, and $Cov(X, Y) = 0$. Then X and Y are independent.

Remark. If X and Y are both normal random variables, X and Y are independent $\Leftrightarrow$ $Cov(X, Y) = 0$.

Example. Let Y_1 and Y_2 be discrete random variables with joint probability distribution as shown in Table 5.3. Show that Y_1 and Y_2 are dependent but have zero covariance.

Table 5.3 Joint probability distribution, Example 5.24

	y_1		
y_2	-1	0	$+1$
-1	$1/16$	$3/16$	$1/16$
0	$3/16$	0	$3/16$
$+1$	$1/16$	$3/16$	$1/16$

Covariance

Solution.

Expected Value and Variance of Linear Functions of Random Variables

Theorem 5.12. Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_j) = \xi_j$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then

a. $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$

b. $Var(U_1) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + 2 \sum \sum_{1 \le i < j \le n} a_i a_j Cov(Y_i, Y_j)$

c. $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j)$

Remark. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$. Then

$$E(\overline{Y}) = \mu, \quad Var(\overline{Y}) = \frac{\sigma^2}{n}.$$

Expected Value and Variance of Linear Functions of Random Variables

Example 5.25. Let Y_1, Y_2, and Y_3 be random variables, where $E(Y_1) = 1, E(Y_2) = 2, E(Y_3) = -1, V(Y_1) = 1, V(Y_2) = 3, V(Y_3) = 5, Cov(Y_1, Y_2) = -0.4, Cov(Y_1, Y_3) = 1/2$, and $Cov(Y_2, Y_3) = 2$. Find the expected value and variance of $U = Y_1 - 2Y_2 + Y_3$. If $W = 3Y_1 + Y_2$, find $Cov(U, W)$.

Solution.

Expected Value and Variance of Linear Functions of Random Variables

Example 5.26. Refer to Examples 5.4 and 5.20. In Example 5.20, we were interested in $Y_1 - Y_2$, the proportional amount of gasoline remaining at the end of a week. Find the variance of $Y_1 - Y_2$. The joint density function for Y_1 and Y_2 is

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \leq y_2 \leq y_1 \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Solution.

Expected Value and Variance of Linear Functions of Random Variables

Example (Binomial). Count the total number of successes X from n independent repetitions of a success/fail experiment with probability of success p. Find $E(X)$ and $Var(X)$.

Solution.

Expected Value and Variance of Linear Functions of Random Variables

Example 5.28. The number of defectives Y in a sample of n = 10 items selected from a manufacturing process follows a binomial probability distribution. An estimator of the fraction defective in the lot is the random variable $\widehat{p} = Y/n$. Find the expected value and variance of $\widehat{p}$.

Solution.

Multinomial Probability Distribution

Multinomial experiment. A multinomial experiment possesses the following properties:

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcome of a single trial falls into cell i, is p_i, $i = 1, 2, \dots, k$ and remains the same from trial to trial. Notice that $\sum_{i=1}^{k} p_i = 1$.
4. The trials are independent.
5. The random variables of interest are $Y_1, Y_2, \dots, Y_k$, where Y_i equals the number of trials for which the outcome falls into cell i . Notice that $\sum_{i=1}^{k} Y_i = n$.

Multinomial distribution. Let $Y_1, Y_2, \dots, Y_k$ denote the number of times that the mutually exclusive and exhaustive outcomes $C_1, C_2, \dots, C_k$ occur in n independent and identical trials. Let $p_i = P(C_i)$ such that $\sum_{i=1}^{k} p_i = 1$, then

$$P(Y_1 = y_1, \dots, Y_k = y_k) = \frac{n!}{y_1! \cdots y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}$$

where $\sum_{i=1}^{k} y_i = n$. Then the random vector $(Y_1, Y_2, \dots, Y_k)$ with the above joind pmf is said to have a **multinomial distribution** with n trials and cell probabilities $p_1, \dots, p_k$.

Multinomial Probability Distribution

Theorem. If $Y_1, Y_2, \ldots, Y_k$ have a multinomial distribution with parameters n and $p_1, \ldots, p_k$, then

$$\boxed{\begin{array}{l} \textbf{1. } E(Y_i) = np_i,\ V(Y_i) = np_i(1 - p_i). \\ \textbf{2. } Cov(Y_s, Y_t) = -np_s p_t, \text{ if } s \neq t. \end{array}}$$

Proof.

Matrix Approach

Expectation of a Random Vector: Suppose we have a n-dimensional vector, $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$. Then the expected value of $\mathbf{Y}$, denoted by $E(\mathbf{Y})$, is defined by

$$E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}.$$

That is, the expected value of a random vector is a vector whose elements are the expected values of the random variables that are the elements of the random vector.

Expectation of a Random Matrix: Similarly, the expected value of a random matrix is defined to be a matrix whose elements are the expected values of the corresponding random variables in the original matrix.

Matrix Approach

Variance-Covariance Matrix of a Random Vector Suppose we have a n-dimensional vector, $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$. Then the Variance-Covariance Matrix of $\mathbf{Y}$, denoted by $Var(\mathbf{Y})$, is defined by

$$Var(\mathbf{Y}) = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{bmatrix}.$$

Note.

- The Variance-Covariance Matrix $Var(\mathbf{Y})$ is symmetric since $Cov(Y_i, Y_j) = Cov(Y_j, Y_i)$.
- $Var(\mathbf{Y}) = E\{[\mathbf{Y} - E(\mathbf{Y})][\mathbf{Y} - E(\mathbf{Y})]'\}$ = $$E\left\{\begin{bmatrix} Y_1 - E(Y_1) \\ Y_2 - E(Y_2) \\ \vdots \\ Y_n - E(Y_n) \end{bmatrix} [Y_1 - E(Y_1), Y_2 - E(Y_2), \cdots, Y_n - E(Y_n)]\right\}$$

Multivariate Normal Distribution

To use matrix notation, we define the following matrices:

$$\boxed{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}}$$

Definition. A random vector

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

is said to have a $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution if its pdf is given by

$$f(\mathbf{y}) = f(y_1, \dots, y_n) = \left(\frac{1}{2\pi}\right)^{n/2} \left[\frac{1}{\det \boldsymbol{\Sigma}}\right]^{1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right].$$

If $n = 2$, the distribution is called Bivariate Normal Distribution. Let Y_1 and Y_2 have a bivariate normal distribution, then

$$\boxed{\boldsymbol{\mu} = (\mu_1, \mu_2)', \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}}$$

Multivariate Normal Distribution

Theorem (Linear combination). Let $X \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$Y = CX \sim MVN(C\boldsymbol{\mu}, C\boldsymbol{\Sigma}C^T),$$

where C is a non-singular matrix.

Theorem (Marginal distributions). Let $X \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The marginal distribution of any set of component X is multivariate normal with means, variance and covariance obtained by taking the corresponding components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively.

Theorem (Conditional distributions). Let X be a n-dimensional random vector and Y be an m-dimensional random vector. Suppose

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim MVN_{n+m}\left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \quad \text{with} \quad \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T,$$

then

$$X|Y = y \sim MVN_n(\boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(y - \boldsymbol{\mu}_Y), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Multivariate Normal Distribution

Theorem (Marginal distributions). Let Y_1 and Y_2 have a bivariate normal distribution. Then

(a). The marginal distribution of Y_1 is normal with mean μ_1 and variance σ_1^2.

(b). The marginal distribution of Y_2 is normal with mean μ_2 and variance σ_2^2.

Theorem (Conditional distributions). Let Y_1 and Y_2 have a bivariate normal distribution. Then the conditional distribution of Y_1 given that $Y_2 = y_2$ is a normal distribution with mean

$$\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)$$

and variance

$$\sigma_1^2(1 - \rho^2).$$

Lecture 6

Functions of Random Variables

MATH 311 Statistics I: Probability Theory
December, 2015

Xuemao Zhang
East Stroudsburg University

Agenda

Introduction

Often, if we are able to model a phenomenon in terms of a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ with cdf $F_{\mathbf{X}}(\mathbf{x})$, we will also be concerned with the behavior of a function of $\mathbf{X}$, say function $g(\cdot)$. $g(\mathbf{X})$, a transformation of $\mathbf{X}$, is also a random vector/variable. There are 3 different methods to get the probability distribution of the new random vector $g(\mathbf{X})$:

1. The Method of Distribution Functions
2. The Method of Transformations
3. The Method of Moment Generating Functions

- As we progress from method 1 to 3, these methods become more mathematically sophisticated and easy to apply, but can handle fewer cases.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Introduction

Method of Distribution Functions:

- This method is typically used when the $\mathbf{X}$'s have continuous distributions.

Method of Transformations:

- Need the density function of the random vector $\mathbf{Y}$.
- $\mathbf{Y} = g(\mathbf{X})$ needs to be monotonic (increasing or decreasing) in nature.

Method of Moment Generating Functions:

- Easiest transformation to use.
- $\mathbf{Y} = g(\mathbf{X})$ can only be in the form $Y = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$.
- Need to find an general form of the MGF to convert $\mathbf{Y}$ back into a distribution.
  - $-$Addition of i.i.d. normal variables gives a normal distribution
  - $-$Addition of expoential distributions with same mean, gives gamma distribution.

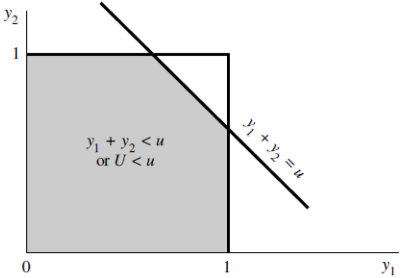
The Method of Distribution Functions

Let U be a function of the random variables $Y_1, Y_2, \ldots, Y_n$.

Step 1. Identify the variable U in terms of the variables $(y_1, y_2, \ldots, y_n)$.

$$U = h(y_1, y_2, \ldots, y_n).$$

Step 2. Find the region (R) in relation to the variables that define $U \leq u$.



Step 3. Find $F_U(u) = P(U \leq u)$ by integrating $f(y_1, y_2, \ldots, y_n)$ over the region $U \leq u$.

$$F_U(u) = \int \cdots \int_{U \leq u} f(y_1, y_2, \ldots, y_n) dy_1 \cdots dy_n.$$

Step 4. Find the density function $f_U(u)$ by differentiating $F_U(u)$. Thus, $f_U(u) = dF_U(u)/du$.

The Method of Distribution Functions

Example. 6.1 Suppose that Y has density function given by

$$f(y) = \begin{cases} 2y, & 0 \le y \le 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = 3Y - 1$.

Solution.

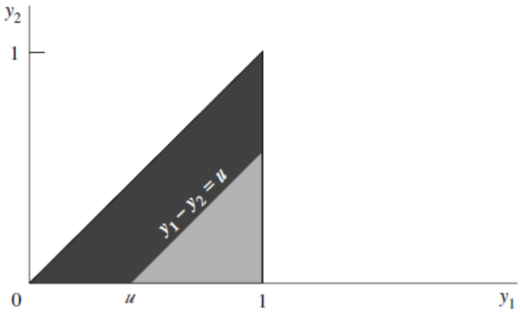
The Method of Distribution Functions

Example. 6.2 Suppose that the joint density function of Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \leq y_2 \leq y_1 \leq 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = Y_1 - Y_2$,

FIGURE 6.1 Region over which $f(y_1, y_2)$ is positive, Example 6.2

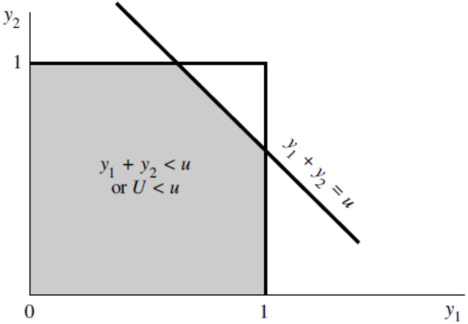


The Method of Distribution Functions

Solution.

The Method of Distribution Functions

Example. 6.3 Let (Y_1, Y_2) denote a random sample of size n = 2 from the uniform distribution on the interval (0, 1). Find the probability density



function for $U = Y_1 + Y_2$.

The Method of Distribution Functions

FIGURE 6.4
The region
$y_1 + y_2 \leq u$ for
$0 \leq u \leq 1$

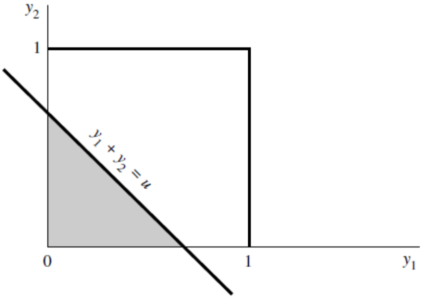
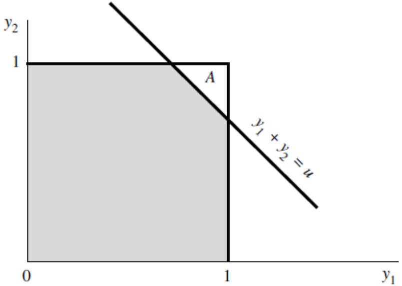


FIGURE 6.5
The region
$y_1 + y_2 \leq u$,
$1 < u \leq 2$



The Method of Distribution Functions

Solution.

The Method of Distribution Functions

Example. Square Transformation Let Y be a continuous random variable with distribution function $F_Y(y)$ and density function $f_Y(y)$. Find the probability density function for $U = Y^2$.

Solution.

The Method of Distribution Functions

Example 6.4. Let Y have probability density function given by

$$f(y) = \begin{cases} \frac{y+1}{2}, & -1 \leq y \leq 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the density function for $U = Y^2$.

Solution.

The Method of Distribution Functions

Example 6.5. Let U be a uniform random variable on the interval (0, 1). Find a transformation $G(U)$ such that $G(U)$ possesses an exponential distribution with mean β.

Solution.

The Method of Transformations

Univariate Transformation. If X is a random variable with cdf $FX(x)$, then any function of X, say $g(X)$, is also a random variable. Often $g(X)$ is of interest itself and we write $Y = g(X)$ to denote the new random variable $g(X)$. Since Y is a function of X, we can describe the probabilistic behavior of Y in terms of that of X. That is, for any set A,

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A)).$$

Univariate Transformation - Discrete. If X is a discrete random variable, $Y = g(X)$ is also a discrete random variable. The pmf for Y is

$$p_Y(y) = P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} P_X(x).$$

Theorem. Let X have cdf $F_X(x)$ and $Y = g(X)$.

a. If g is an increasing function, $F_Y(y) = F_X(g^{-1}(y))$ for all y.

b. If g is a decreasing function and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for all y.

The Method of Transformations

Theorem. Let Y have probability density function $f_Y(y)$. If $h(y)$ is either increasing or decreasing for all y such that $f_Y(y) > 0$, then $U = h(Y)$ has density function

$$f_U(u) = f_Y(h^{-1}(u))\left|\frac{dh^{-1}(u)}{du}\right|$$

FIGURE **6.8**
An increasing function

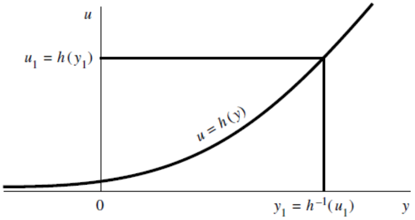
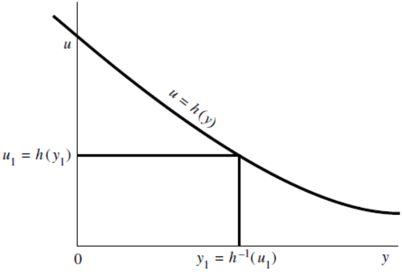


FIGURE **6.9**
A decreasing function



The Method of Transformations

Theorem. Let Y have pdf $f_Y(y)$, let $U = h(Y)$ be defined on the sample space A of Y. Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of A such that $P(Y \in A_0) = 0$ and $f_Y(y)$ is continuous on each A_i. Further, suppose there exist functions $h_1(y), \ldots, h_k(y)$, defined on $A_1, \ldots, A_k$, respectively, satisfying

i. $h(y) = h_i(y)$, for $y \in A_i$,

ii. $h_i(y)$ is monotone on A_i,

iii. the set $\{u : u = h_i(y), y \in A_i\}$ is the same for each $i = 1, \ldots, k$,

iv. $h_i^{-1}(u)$ has a continuous derivative on the set in (iii), for each $i = 1, \ldots, k$.

Then

$$f_U(u) = \sum_{i=1}^{k} f_Y(h^{-1}(u)) \left| \frac{dh_i^{-1}(u)}{du} \right|.$$

The Method of Transformations

Basic Steps for Method of Transformations:

Let U be a function of the random variables $Y_1, Y_2, \ldots, Y_n$.

Step 1. Identify the variable $U = h(y)$ in terms of the variable y.

Step 2. If $U = h(y)$ is not monotonic, find the partition $A_0, A_1, \ldots, A_k$ such that $U = h(y)$ is monotonic on each A_i, $i = 1, \ldots, k$ and satisfying the conditions (i) - (iii) in the Theorem.

Step 3. Rearrange $U = h(y)$ to find $y = h_i^{-1}(u)$, $i = 1, \ldots, k$.

Step 4. Find the derivative of $h_i^{-1}(u)$ and take their absolute values

$$\left|\frac{dh_i^{-1}(u)}{du}\right|.$$

Step 5. Plug $y = h_i^{-1}(u)$ into the density function of Y, and simplify the expression:

$$f_U(u) = \sum_{i=1}^{k} f_Y(h^{-1}(u))\left|\frac{dh_i^{-1}(u)}{du}\right|.$$

The Method of Transformations

Example. 6.6 Suppose that Y has density function given by

$$f(y) = \begin{cases} 2y, & 0 \le y \le 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = 3Y - 1$ by the transformation method.

Solution.

The Method of Transformations

Example. 6.7 Suppose that Y has density function given by

$$f(y) = \begin{cases} 2y, & 0 \leq y \leq 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = -4Y + 3$ by the transformation method.

Solution.

The Method of Transformations

Example. 6.8 Let Y_1 and Y_2 have a joint density function given by

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & 0 \le y_1, 0 \le y_2; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = Y_1 + Y_2$.
Remark. Another method to solve the problem is to use the method of multivariate transformations.

Solution.

The Method of Transformations

Example. 6.9 Let Y_1 and Y_2 have a joint density function given by

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \le y_1 \le 1, 0 \le y_2 \le 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = Y_1 Y_2$ and find $E(U)$.

Solution.

The Method of Transformations

Example. Normal Chi-square relationship Let X have the standard normal distribution. Find the probability density function for $Y = X^2$.

Solution.

Bivariate Transformations

How do we find the density of a combination of two random variables?

This type of problem is solved in two steps. First we transform the pair (X, Y) into another pair (U, V), where U is the combination that we are after. Then we use the *joint density* of (U, V) to get the *marginal density* of U. The rule for transforming between pairs of variables is similar to the rule for single variables.

Theorem. Bivariate Transformations Let (X, Y) have the joint density $f(x, y)$. Let

$$U = h_1(X, Y) \quad \text{and} \quad V = h_2(X, Y),$$

where we assume that this transformation is **one-to-one** with inverse (with an abuse of notation, denote the inverse of of (h_1, h_2) by (h_1^{-1}, h_2^{-1}))

$$X = h_1^{-1}(U, V) \quad \text{and} \quad Y = h_2^{-1}(U, V).$$

If $h_1^{-1}(u, v)$ and $h_2^{-1}(u, v)$ have continuous partial derivatives with respect to u and v and Jacobian

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Then the joint density of (U, V) is given by

$$g(u, v) = f(h_1^{-1}(u, v), h_2^{-1}(u, v)) \cdot |J|,$$

where $|J|$ is the absolute value of the Jacobian.

Multivariate Transformations

Theorem (Multivariate Transformations). Let $(Y_1, Y_2, \ldots, Y_k)$ be jointly continuous random variables with joint density $f(y_1, y_2, \ldots, y_k)$. Let

$$U_1 = h_1(Y_1, Y_2, \ldots, Y_k), \ldots U_k = h_k(Y_1, Y_2, \ldots, Y_k),$$

where we assume that the multivariate transformation is **one-to-one** with inverse (with an abuse of notation, denote the inverse of of $(h_1, \ldots, h_k)$ by $(h_1^{-1}, \ldots, h_k^{-1})$)

$$Y_1 = h_1^{-1}(U_1, U_2, \ldots, U_k), \ldots Y_k = h_k^{-1}(U_1, U_2, \ldots, U_k).$$

If $h_1^{-1}(u_1, u_2, \ldots, u_k), \ldots h_k^{-1}(u_1, u_2, \ldots, u_k)$ have continuous partial derivatives with respect to $u_1, u_2, \ldots, u_k$ and Jacobian

$$J = \det \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & \frac{\partial y_1}{\partial u_2} & \cdots & \frac{\partial y_1}{\partial u_k} \\ \frac{\partial y_2}{\partial u_1} & \frac{\partial y_1}{\partial u_2} & \cdots & \frac{\partial y_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_k}{\partial u_1} & \frac{\partial y_k}{\partial u_2} & \cdots & \frac{\partial y_k}{\partial u_k} \end{pmatrix}$$

Then the joint density of $U_1, U_2, \ldots, U_k$ is given by

$$g(u_1, u_2, \ldots, u_k) = f\left(h_1^{-1}(u_1, u_2, \ldots, u_k), \cdots, h_k^{-1}(u_1, u_2, \ldots, u_k)\right) \cdot |J|,$$

where $|J|$ is the absolute value of the Jacobian.

Multivariable Transformations

Basic Procedure for Finding Multivariable Transformations Using Jacobians: Consider the problem of finding the disribution of a transformation

$$\boxed{U_1 = h_1(Y_1, Y_2, \ldots, Y_k).}$$

Step 1. Identify the variable $U_1 = h(y)$ in terms of the variable $y_1, y_2, \ldots, y_k$.

$$\boxed{u_1 = h_1(y_1, y_2, \ldots, y_k).}$$

Step 2. Create other variables $U_2, \ldots, U_k$ pertaining to the variables $Y_1, Y_2, \ldots, Y_k$ so the number of new variables is the same as the number of old variables.

$$\begin{cases} u_1 = h_1(y_1, y_2, \ldots, y_k) \\ u_2 = h_2(y_1, y_2, \ldots, y_k) \\ \vdots \\ u_k = h_k(y_1, y_2, \ldots, y_k) \end{cases}$$

Multivariable Transformations

Remark. The new variables have to have a one-to-one correspondence to the old variables, so one point in the old coordinate system is mapped to only one point in the new coordinate system. Since we are only concerned with the marginal density of U_1, the other variables can be anything we want. For most problems, we can equate the other new variables to one of the old variables. That is,

$$\begin{cases} u_1 = h_1(y_1, y_2, \ldots, y_k) \\ u_2 = y_2 \\ \vdots \\ u_k = y_k \end{cases}$$

Some other examples:

$$\boxed{\begin{array}{cc} \begin{cases} U = X + Y \\ V = X - Y \end{cases}, & \begin{cases} U = XY \\ V = Y \end{cases} \\ \begin{cases} U = X/Y \\ V = Y \end{cases}, & \begin{cases} U = X/(X+Y) \\ V = X + Y \end{cases} \end{array}}$$

Multivariable Transformations

Step 3. Find the inverse of $u_1, u_2, \dots, u_k$, so that $y_1, y_2, \dots, y_k$ can be found in terms of $u_1, u_2, \dots, u_k$ (with an abuse of notation, denote the inverse of of $(h_1, \dots, h_k)$ by $(h_1^{-1}, \dots, h_k^{-1})$):

$$\begin{aligned} y_1 &= h_1^{-1}(u_1, u_2, \dots, u_k) \\ y_2 &= h_2^{-1}(u_1, u_2, \dots, u_k) \\ &\vdots \\ y_k &= h_k^{-1}(u_1, u_2, \dots, u_k) \end{aligned}$$

Step 4. Find the Jacobian matrix from the variable functions found in step 3.

$$J = \det \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & \frac{\partial y_1}{\partial u_2} & \cdots & \frac{\partial y_1}{\partial u_k} \\ \frac{\partial y_2}{\partial u_1} & \frac{\partial y_1}{\partial u_2} & \cdots & \frac{\partial y_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_k}{\partial u_1} & \frac{\partial y_k}{\partial u_2} & \cdots & \frac{\partial y_k}{\partial u_k} \end{pmatrix}$$

Multivariable Transformations

Step 5. Find the absolute value of the Jacobian matrix, $|J|$.

Step 6. Next we will find the density function of $u_1, u_2, \ldots, u_k$ by substituting our inverse functions of $y_1, y_2, \ldots, y_k$ in our $f(y_1, y_2, \ldots, y_k)$ and multiplying by $|J|$.

$$g(u_1, u_2, \ldots, u_k) = f\left(h_1^{-1}(u_1, u_2, \ldots, u_k), \cdots, h_k^{-1}(u_1, u_2, \ldots, u_k)\right) \cdot |J|.$$

Step 7. Integrate out the extraneous variables to find the marginal density function of U_1.

$$g_{U_1}(u_1) = \int \cdots \int g(u_1, u_2, \ldots, u_k) du_2 \cdots du_k.$$

Multivariable Transformations

Example. 6.13. Let Y_1 and Y_2 be independent standard normal random variables. If $U_1 = Y_1 + Y_2$ and $U_2 = Y_1 1 Y_2$, both U_1 and U_2 are linear combinations of independent normally distributed random variables, and Theorem 6.3 implies that U_1 is normally distributed with mean 0+0 = 0 and variance 1+1 = 2. Similarly, U_2 has a normal distribution with mean 0 and variance 2. What is the joint density of U_1 and U_2?

Solution.

Multivariable Transformations

Example. 6.14. Let Y_1 and Y_2 be independent exponential random variables, both with mean $\beta > 0$. Find the density function of

$$U = \frac{Y_1}{Y_1 + Y_2}.$$

Solution.

Multivariable Transformations

Example. 6.8 Let Y_1 and Y_2 have a joint density function given by

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & 0 \le y_1, 0 \le y_2; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = Y_1 + Y_2$.

Solution.

Multivariable Transformations

Example. 6.9 Let Y_1 and Y_2 have a joint density function given by

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for $U = Y_1 Y_2$.

Solution.

The Method of Moment-Generating Functions

If the transformation is a linear combination of n independent random variables $Y_1, \ldots, Y_n$ with easily obtained mgfs, it is easier to use the Method of Moment-Generating Functions to obtain the distribution of the transformed variable due to the **uniqueness property** of a mgf for a random variable.

Theorem 6.2. Let $Y_1, Y_2, \ldots, Y_n$ be mutually independent random variables with moment generating functions $m_{Y_1}(t), m_{Y_2}(t), \ldots, m_{Y_n}(t)$, respectively. If $U = c_1Y_1 + c_2Y_2 + \cdots + c_nY_n$, where $c_1, c_2, \ldots, c_n$ are constants, then the mgf of U is

$$M_U(t) = m_{Y_1}(c_1t) \cdot m_{Y_2}(c_2t) \cdots m_{Y_n}(c_nt).$$

Basic Procedure for Finding Multivariable Transformations Using the MGF Method: Consider the problem of finding the disribution of a transformation

$$\boxed{U = c_1Y_1 + c_2Y_2 + \cdots + c_nY_n,}$$

where $c_1, c_2, \ldots, c_n$ are real numbers.

Step 1. Find the moment generating function of U:

$$M_U(t) = m_{Y_1}(c_1t) \cdot m_{Y_2}(c_2t) \cdots m_{Y_n}(c_nt).$$

Step 2. Simplify the expression and arrange it in terms of one or more of the distribution functions in the back of the book. That is the distribution of U.

The Method of Moment-Generating Functions

Example. 6.10. Suppose that Y is a normally distributed random variable with mean μ and variance σ^2. Find the distribution of

$$Z = \frac{Y - \mu}{\sigma}.$$

Solution.

The Method of Moment-Generating Functions

Example. 6.11. Let Z be a normally distributed random variable with mean 0 and variance 1. Use the method of moment-generating functions to find the probability distribution of Z^2.

Solution.

The Method of Moment-Generating Functions

Example. 6.12 Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random variables, with the density function for Y_i given by

$$f(y_i) = \frac{1}{\theta} e^{-y_i/\theta} 1_{(0,\infty)}(y_i), i = 1, 2, \ldots, n.$$

Find the probability density function of $U = Y_1 + Y_2 + \ldots + Y_n$.

Solution.

The Method of Moment-Generating Functions

Theorem 6.3. Let $Y_1, Y_2, \ldots, Y_n$ be independent normally distributed random variables with $E(Y_i) = \mu_i$ and $V(Y_i) = \sigma_i^2$, for $i = 1, 2, \ldots, n$, and let $c_1, c_2, \ldots, c_n$ be constants. If

$$U = \sum_{i=1}^{n} c_i Y_i = c_1 Y_1 + c_2 Y_2 + \cdots + c_n Y_n.$$

Then U is a normally distributed random variable with

$$E(U) = \sum_{i=1}^{n} c_i \mu_i \text{ and } V(U) = \sum_{i=1}^{n} c_i^2 \sigma_i^2 .$$

Proof.

The Method of Moment-Generating Functions

Theorem 6.4. Let $Y_1, Y_2, \ldots, Y_n$ be independent normally distributed random variables with $E(Y_i) = \mu_i$ and $V(Y_i) = \sigma_i^2$, for $i = 1, 2, \ldots, n$. Define

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}.$$

Then $\sum_{i=1}^{n} Z_i^2$ has χ^2 distribution with n degrees of freedom.

Proof.

Order Statistics

Many functions of random variables of interest in practice depend on the relative magnitudes of the observed variables. For instance, we may be interested in the fastest time in an automobile race or the heaviest mouse among those fed on a certain diet. Thus, we often order observed random variables according to their magnitudes. The resulting ordered variables are called order statistics.

Definition. The order statistics of a random sample $Y_1, Y_2, \ldots, Y_n$ are the sample values placed in ascending order. They are denoted by $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$. The order statistics are random variables that satisfy

$$Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}.$$

In particular,

$$\boxed{\begin{aligned} Y_{(1)} &= \min_{1 \leq i \leq n} Y_i \\ Y_{(2)} &= \text{second smallest} X_i \\ &\vdots \\ Y_{(n)} &= \max_{1 \leq i \leq n} Y_i \end{aligned}}$$

Order Statistics

Theorem. If the variables $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with common cdf $F(y)$ and common pdf $f(y)$, respectively. Then the densities of $Y_{(1)}$, $Y_{(n)}$, respectively, are:

$$\boxed{g_{(1)}(y) = n[1 - F(y)]^{n-1} f(y)}$$

$$\boxed{g_{(n)}(y) = n[F(y)]^{n-1} f(y)}$$

Proof.

Order Statistics

Theorem 6.5. Let $Y_1, Y_2, \ldots, Y_n$ be independent identically distributed continuous random variables with common cdf $F(y)$ and common pdf $f(y)$. If $Y_{(k)}$ denotes the kth-order statistic, then the density function of $Y_{(k)}$ is given by

$$\boxed{g_{(k)}(y) = \frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}[1-F(y)]^{n-k}f(y)}$$

Order Statistics

Example 6.16. Electronic components of a certain type have a length of life Y, with probability density given by

$$f(y) = (1/100)e^{-y/100}1_{(0,\infty)}(y).$$

(Length of life is measured in hours.) Suppose that two such components operate independently and in series in a certain system (hence, the system fails when either component fails). Find the density function for X, the length of life of the system.

Solution.

Order Statistics

Example 6.17. Suppose that the components in Example 6.16 operate in parallel (hence, the system does not fail until both components fail). Find the density function for X, the length of life of the system.

Solution.

Order Statistics

Example 6.18. Suppose that $Y_1, Y_2, \ldots, Y_5$ denotes a random sample from a uniform distribution defined on the interval (0, 1). That is,

$$f(y) = 1 \cdot 1_{[0,1]}(y).$$

Find the density function for the second-order statistic. Also, give the joint density function for the second- and fourth-order statistics.

Solution.