Probability and Statistics

Narendra Singh Yadav

Department of Mathematics & Data Science, IIIT Sri City,
Andhra Pradesh, India

February 22, 2025

Syllabus

Unit 1: Probability Theory

- Sample Space, Events, Axioms of Probability
- Other Definitions, Probability Rules
- Conditional Probabilities, Bayes' Theorem
- Independence of Events

Unit 2: Random Variable and Distribution

- Cumulative Distribution Function, Expectation, Variance, Moments
- Skewness, Kurtosis, Quantiles, Median
- Discrete and Continuous Random Variables and Their Distributions
- Properties of PDF and PMF
- Moment Generating Functions, Characteristic Function

Syllabus ...

Unit 3: Jointly Distributed Random Variable

- Conditional Distributions: Discrete and Continuous Cases
- Joint and Marginal Distribution
- Covariance and Correlation
- Bivariate and Multivariate Normal Distribution
- Functions of Random Variables and Random Vectors
- Distributions of Sums of Random Variables

Unit 4: Limit Theorems and Inequalities

- The Central Limit Theorem, Law of Large Numbers
- Boole's Inequality, Bonferroni's Inequality
- Chebyshev's and Markov's Inequality
- Cauchy-Schwartz Inequality, Jensen's Inequality

Syllabus ...

Unit 5: Statistical Inference-I

- Distributions of the Sample Mean and Variance for a Normal Population
- Chi-Square, t, and F Distributions
- Estimation, Unbiasedness, Consistency
- Method of Moments, Maximum Likelihood Estimation

Unit 6: Statistical Inference-II

- Confidence Intervals, Test of Hypothesis
- Null and Alternative Hypotheses, Critical and Acceptance Regions
- Two Types of Error, Power of the Test
- Z-test, t-test, Chi-square Test, F-test
- Computation, Simulation, and Visualization Using R or Matlab

Books

Textbooks

1. V. K. Rohatgi
   An Introduction to Probability and Statistics, 2nd Edition, Wiley, ISBN: 9788126519262
2. S. Ross
   A First Course in Probability, 8th Edition, Pearson, ISBN: 9780136033134
3. D.C. Montgomery, G.C. Runger
   Applied Statistics and Probability for Engineers, 7th Edition, Wiley, ISBN: 9781119456261
4. Michael J. Evans, Jeffrey S. Rosenthal
   Probability and Statistics, 2nd Edition, ISBN: 9781429224628

Reference Books

1. Anderson, T.W.
   Introduction to Multivariate Statistical Analysis, Wiley, ISBN: 9788126524488
2. A.M. Gun, M.K. Gupta, B. Dasgupta
   An Outline of Statistical Theory, Volume One, World Press

Evaluation Plan (Tentative)

- Mid Sem: 25
- End Sem: 40
- Schedule Quiz (2): 30
- Class Participation: 5

Class Participation

- Attendance and punctuality
- Active engagement and interaction in class
- Proper discipline (e.g., no mobile phones or laptops during class)
- Avoiding breaks or proxies

Key to Success

- Try to be a learner
- Practice regularly
- Attend all classes
- Maintain honesty and integrity

Applications of Probability and Statistics

- AI and Machine Learning
- Social Sciences
- Finance and Economics
- Government and Administration
- Biology and Physics
- Actuarial Sciences and Marketing
- Gambling

Sample space

Consider an experiment whose outcome is not predictable with certainty in advance. Although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known.

Sample space

The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S.

Examples

1. If the outcome of an experiment consists in the determination of the sex of a newborn child, then
$$S = \{g, b\}$$
where the outcome g means that the child is a girl and b that it is a boy.
2. If the experiment consists of the running of a race among the seven horses having post positions $1, 2, 3, 4, 5, 6, 7$, then
$$S = \{\text{all orderings of } (1, 2, 3, 4, 5, 6, 7)\}$$
The outcome $(2, 3, 1, 6, 5, 4, 7)$ means, for instance, that the number 2 horse is first, then the number 3 horse, then the number 1 horse, and so on.

Event

Any subset E of the sample space is known as an *event*. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred.

Example

1. In Example 1, if $E = \{g\}$, then E is the event that the child is a girl. Similarly, if $F = \{b\}$, then F is the event that the child is a boy.

For any two events E and F of a sample space S, we define the new event $E \cup F$, called the *union* of the events E and F, to consist of all outcomes that are either in E or in F or in both E and F. That is, the event $E \cup F$ will occur if either E or F occurs.

For instance, in Example 1, if $E = \{g\}$ and $F = \{b\}$, then

$$E \cup F = \{g, b\}.$$

That is, $E \cup F$ would be the whole sample space S.

In Example 2, if $E = \{\text{all outcomes starting with 6}\}$ is the event that the number 6 horse wins and $F = \{\text{all outcomes having 6 in the second position}\}$ is the event that the number 6 horse comes in second, then $E \cup F$ is the event that the number 6 horse comes in either first or second.

Event ...

- Similarly, for any two events E and F, we may also define the new event $E \cap F$, called the *intersection* of E and F, to consist of all outcomes that are in both E and F.
- In Example 2, if $E = \{\text{all outcomes ending in 5}\}$ is the event that horse number 5 comes in last and $F = \{\text{all outcomes starting with 5}\}$ is the event that horse number 5 comes in first, then the event $E \cap F$ does not contain any outcomes and hence cannot occur.
- To give such an event a name, we shall refer to it as the *null event* and denote it by $\emptyset$. Thus, $\emptyset$ refers to the event consisting of no outcomes.
- If $E \cap F = \emptyset$, implying that E and F cannot both occur, then E and F are said to be *mutually exclusive*.
- For any event E, we define the event E^c, referred to as the complement of E, to consist of all outcomes in the sample space S that are not in E.
- In Example 1, if $E = \{b\}$ is the event that the child is a boy, then $E^c = \{g\}$ is the event that it is a girl.

Axioms of probability

- The term probability refers to the study of randomness and uncertainty. In any situation in which one of a number of possible outcomes may occur, the discipline of probability provides methods for quantifying the chances associated with the various outcomes.
- The study of probability as a branch of mathematics goes back over 300 years, where it had its genesis in connection with questions involving games of chance.
- An experiment is any activity or process whose outcome is subject to uncertainty.
- We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions. For each event E of the sample space S, we define $n(E)$ to be the number of times the event E occurs. Then $P(E)$, the probability of the event E, is defined as

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

That is, $P(E)$ is defined as the (limiting) proportion of time that E occurs.

Axioms of probability ...

- For instance, if a coin is continually flipped, then the proportion of flips resulting in heads will approach some value as the number of flips increases. It is this constant limiting frequency that we often have in mind when we speak of the probability of an event.
- For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three axioms:
  Axiom 1.
  $$0 \leq P(E) \leq 1$$
  Axiom 2.
  $$P(S) = 1$$
  Axiom 3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i \cap E_j = \emptyset$ when $i \neq j$),
  $$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i), \quad n = 1, 2, \ldots, \infty.$$
  We call $P(E)$ the probability of the event E.

Axioms of probability ...

- It should be noted that if we interpret $P(E)$ as the relative frequency of the event E when a large number of repetitions of the experiment are performed, then $P(E)$ would indeed satisfy the above axioms.

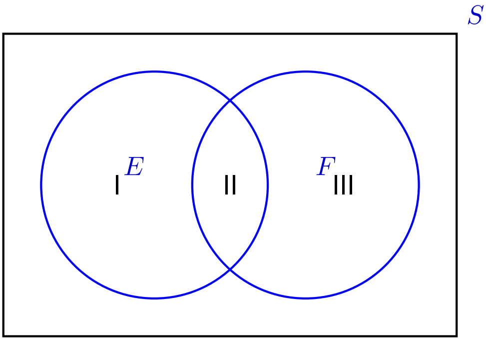
Lemma

$$P(E^c) = 1 - P(E).$$

Proof. We first note that E and E^c are always mutually exclusive, and since $E \cup E^c = S$, we have by Axioms 2 and 3 that

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c).$$

Axioms of probability ...



Lemma

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

Proof. This proposition is most easily proven by the use of a Venn diagram as shown in the above Figure As the regions I, II, and III are mutually exclusive, it follows that

$$P(E \cup F) = P(I) + P(II) + P(III)$$

$$P(E) = P(I) + P(II)$$

$$P(F) = P(II) + P(III)$$

which shows that

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

and the proof is complete since $II = E \cap F$.

Examples

Example: A total of 28 percent of males living in Nevada smoke cigarettes, 6 percent smoke cigars, and 3 percent smoke both cigars and cigarettes. What percentage of males smoke neither cigars nor cigarettes?

Example: A total of 28 percent of males living in Nevada smoke cigarettes, 6 percent smoke cigars, and 3 percent smoke both cigars and cigarettes. What percentage of males smoke neither cigars nor cigarettes?

Solution. Let E be the event that a randomly chosen male is a cigarette smoker and let F be the event that he is a cigar smoker. Then, the probability this person is either a cigarette or a cigar smoker is

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.28 + 0.06 - 0.03 = 0.31$$

Thus, the probability that the person is not a smoker is

$$1 - 0.31 = 0.69,$$

implying that 69 percent of males smoke neither cigarettes nor cigars. ■

Sample spaces having equally likely outcomes

- For a large number of experiments, it is natural to assume that each point in the sample space is equally likely to occur. That is, for many experiments whose sample space S is a finite set, say $S = \{1, 2, \ldots, N\}$, it is often natural to assume that

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\}) = p \quad \text{(say)}.$$

- Now it follows from Axioms 2 and 3 that

$$1 = P(S) = P(\{1\}) + P(\{2\}) + \cdots + P(\{N\}) = Np$$

which shows that

$$P(\{i\}) = p = \frac{1}{N}.$$

From this, it follows from Axiom 3 that for any event E,

$$P(E) = \frac{\text{Number of points in } E}{N}.$$

- In words, if we assume that each outcome of an experiment is equally likely to occur, then the probability of any event E equals the proportion of points in the sample space that are contained in E.
- Thus, to compute probabilities, it is often necessary to be able to effectively count the number of different ways that a given event can occur. To do this, we will make use of the following rule.

Basic principle of counting

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are mn possible outcomes of the two experiments.

Example: Two balls are “randomly drawn” from a bowl containing 6 white and 5 black balls. What is the probability that one of the drawn balls is white and the other is black?

Basic principle of counting

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are mn possible outcomes of the two experiments.

Example: Two balls are "randomly drawn" from a bowl containing 6 white and 5 black balls. What is the probability that one of the drawn balls is white and the other is black?

Solution:
The total number of ways to choose 2 balls from the bowl is:

$$\binom{6+5}{2} = \binom{11}{2} = \frac{11 \cdot 10}{2} = 55.$$

The number of ways to choose 1 white ball and 1 black ball is:

$$\binom{6}{1} \cdot \binom{5}{1} = 6 \cdot 5 = 30.$$

Therefore, the probability that one of the drawn balls is white and the other is black is:

$$P = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}} = \frac{30}{55} = \frac{6}{11}.$$

Exercises

1. If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year ? How large need n be so that this probability is less than $1/2$?
2. A basketball team consists of 6 black and 6 white players. The players are to be paired in groups of two for the purpose of determining roommates. If the pairings are done at random, what is the probability that none of the black players will have a white roommate ?
3. From a set of n items a random sample of size k is to be selected. What is the probability a given item will be among the k selected ?

Conditional probability

Definition (Conditional probability)

If we let E and F denote, respectively, the event then the probability just obtained is called the conditional probability of E given that F has occurred, and is denoted by $P(E|F)$.

- Because we know that F has occurred, it follows that we can regard F as the new sample space and hence the probability that the event $E \cap F$ occurs will equal the probability of $E \cap F$ relative to the probability of F. That is,

$$P(E|F) = \frac{P(E \cap F)}{P(F)}. \tag{1}$$

- Note that Equation (1) is well defined only when $P(F) > 0$ and hence $P(E|F)$ is defined only when$P(F) > 0$.

Conditional probability

- $P(S|F) = \frac{P(S\cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1.$
- For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i \cap E_j = \emptyset$ when $i \neq j$),

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} E_i | F\right) &= \frac{P\left(\bigcup_{i=1}^{n} E_i \cap F\right)}{P(F)} \\
&= \frac{P\left(\bigcup_{i=1}^{n} (E_i \cap F)\right)}{P(F)} \\
&= \frac{\sum_{i=1}^{n} P\left(E_i \cap F\right)}{P(F)}, \quad n = 1, 2, \ldots, \infty.
\end{aligned}
$$

Conditional probability...

Example. The sample space S of this experiment can be taken to be the following set of 36 outcomes:

$$S = \{(i, j) \,|\, i = 1, 2, 3, 4, 5, 6,\ j = 1, 2, 3, 4, 5, 6\}.$$

Suppose now that each of the 36 possible outcomes is equally likely to occur and thus has probability $\frac{1}{36}$. Suppose further that we observe that the first die lands on side 3. Then, given this information, what is the probability that the sum of the two dice equals 8 $?$

To calculate this probability, we reason as follows: Given that the initial die is a 3, there can be at most 6 possible outcomes of our experiment, namely,

$$(3, 1),\ (3, 2),\ (3, 3),\ (3, 4),\ (3, 5),\ \text{and}\ (3, 6).$$

In addition, because each of these outcomes originally had the same probability of occurring, they should still have equal probabilities. That is, given that the first die is a 3, then the (conditional) probability of each of the outcomes $(3, 1),\ (3, 2),\ (3, 3),\ (3, 4),\ (3, 5),\ (3, 6)$ is $\frac{1}{6}$, whereas the (conditional) probability of the other 30 points in the sample space is 0. Hence, the desired probability will be $\frac{1}{6}$.

Conditional probability...

Example. A bin contains 5 defective (that immediately fail when put in use), 10 partially defective (that fail after a couple of hours of use), and 25 acceptable transistors. A transistor is chosen at random from the bin and put into use. If it does not immediately fail, what is the probability it is acceptable?

Conditional probability...

Example. A bin contains 5 defective (that immediately fail when put in use), 10 partially defective (that fail after a couple of hours of use), and 25 acceptable transistors. A transistor is chosen at random from the bin and put into use. If it does not immediately fail, what is the probability it is acceptable?
Solution. Since the transistor did not immediately fail, we know that it is not one of the 5 defectives, and so the desired probability is:

$$\begin{aligned} P\{\text{acceptable} \mid \text{not defective}\} &= \frac{P\{\text{acceptable, not defective}\}}{P\{\text{not defective}\}} \\ &= \frac{P\{\text{acceptable}\}}{P\{\text{not defective}\}}. \end{aligned}$$

Hence, assuming that each of the 40 transistors is equally likely to be chosen, we obtain that

$$P\{\text{acceptable} \mid \text{not defective}\} = \frac{25/40}{35/40} = \frac{5}{7}.$$

$\square$

Conditional probability...

Example. The organization that Jones works for is running a father-son dinner for those employees having at least one son. Each of these employees is invited to attend along with his youngest son. If Jones is known to have two children, what is the conditional probability that they are both boys given that he is invited to the dinner? Assume that the sample space S is given by

$$S = \{(b,b),(b,g),(g,b),(g,g)\}$$

and all outcomes are equally likely.

Conditional probability...

Example. The organization that Jones works for is running a father-son dinner for those employees having at least one son. Each of these employees is invited to attend along with his youngest son. If Jones is known to have two children, what is the conditional probability that they are both boys given that he is invited to the dinner? Assume that the sample space S is given by

$$S = \{(b, b), (b, g), (g, b), (g, g)\}$$

and all outcomes are equally likely.

Solution. The knowledge that Jones has been invited to the dinner is equivalent to knowing that he has at least one son. Hence, letting B denote the event that both children are boys, and A the event that at least one of them is a boy, we have that the desired probability $P(B \mid A)$ is given by

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

Since $B \cap A = \{(b, b)\}$, we have

$$P(B \mid A) = \frac{P(\{(b, b)\})}{P(\{(b, b), (b, g), (g, b)\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Exercises

1. A coin is flipped twice. Assuming that all four points in the sample space

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

   are equally likely, what is the conditional probability that both flips land on heads, given that:

   (a) the first flip lands on heads?

   (b) at least one flip lands on heads?

2. The birthday problem asks: What is the probability that, in a group of n people, at least two people share the same birthday? We assume that there are 365 days in a year, and each day is equally likely for a birthday.

Bayes formula

Let E and F be events. We may express E as

$$E = (E \cap F) \cup (F \cap F^c)$$

for, in order for a point to be in E, it must either be in both E and F or be in E but not in F. As $E \cap F$ and $E \cap F^c$ are clearly mutually exclusive, we have by Axiom 3 that

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

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

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

Problems on Bayes theorem

1. A factory produces two types of products: Product P and Product Q in a $70:30$ mix. It is known that 5% of Product P is defective, while 8% of Product Q is defective. If a randomly chosen defective product is found, what is the probability that it is Product P?
2. In a city, 60% of the vehicles are cars and 40% are motorcycles. The probability of a car being involved in an accident is 10%, while the probability of a motorcycle being involved in an accident is 5%. If an accident occurred, what is the probability that it involved a car?
3. In a population, 2% of people have a certain genetic condition. A test has been developed to detect this condition, and it correctly identifies the condition in 90% of cases. However, it also produces a false positive result in 5% of cases for people who do not have the condition. If a randomly selected person tests positive, what is the probability that they actually have the genetic condition?

Problems on Bayes theorem ...

Sr. No.	Education	Selection
1	Grad	Y
2	Grad	N
3	PG	Y
4	Grad	N
5	PG	Y
6	PG	Y
7	PG	N
8	Grad	Y
9	PG	N
10	Grad	N
11	Grad	N
12	PG	Y
13	PG	Y
14	Grad	Y
15	PG	Y

1. Probability that a randomly chosen candidate is a PG?
2. Probability that a randomly chosen candidate is Selected?
3. Probability that a randomly chosen candidate is a Grad and is selected?
4. Probability that a randomly chosen candidate is a PG and is rejected?
5. Probability that a randomly chosen candidate is a Grad when she is known to have been selected?

Application of Bayes theorem

Theorem (General Bays Theorem)

Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S.$$

Suppose now that E has occurred and we are interested in determining which one of F_j also occurred. We have that

$$P(F_j \mid E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(E \mid F_j)P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i)P(F_i)} \tag{2}$$

Equation (2) is known as extended Bayes' formula, after the English philosopher Thomas Bayes.

In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ must occur. By writing

$$E = \bigcup_{i=1}^{n} E \cap F_i$$

and using the fact that the events $E \cap F_i, i = 1, \ldots, n$ are mutually exclusive, we obtain that

$$P(E) = \sum_{i=1}^{n} P(E \cap F_i)$$

$$= \sum_{i=1}^{n} P(E|F_i)P(F_i).$$

Independent events

Definition

Two events $E, F \in S$(sample space) are said to be independent if and only if :

$$P(E \cap F) = P(E)P(F).$$

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

Example: A card is selected at random from an ordinary deck of 52 playing cards. If A is the event that the selected card is an ace and H is the event that it is a heart, then A and H are independent, since

$$P(AH) = \frac{1}{52}, \quad P(A) = \frac{4}{52}, \quad \text{and} \quad P(H) = \frac{13}{52}.$$

Lemma

If E and F are independent, then so are E and F^c.

Proof. Assume that E and F are independent. Since $E = E \cap F \cup E \cap F^c$, and $E \cap F$ and $E \cap F^c$ are obviously mutually exclusive, we have that

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

$$= P(E)P(F) + P(E \cap F^c) \quad \text{by the independence of } E \text{ and } F.$$

or equivalently,

$$P(E \cap F^c) = P(E)(1 - P(F))$$

$$= P(E)P(F^c),$$

and the result is proven.

Note: Similarly E^c and F, E^c and F^c are independent events.

Independent events ...

Definition

The three events $E, F,$ and G are said to be independent if

$$P(E \cap F \cap G) = P(E)P(F)P(G)$$

$$P(E \cap F) = P(E)P(F)$$

$$P(E \cap G) = P(E)P(G)$$

$$P(F \cap G) = P(F)P(G).$$

It should be noted that if the events E, F, G are independent, then E will be independent of any event formed from F and G.

Problem based on independent events

- Two coins are flipped, and all 4 outcomes are assumed to be equally likely. If E is the event that the first coin lands on heads and F is the event that the second lands on tails, then E and F are independent, since

$$P(E \cap F) = P(\{(H,T)\}) = \frac{1}{4},$$

whereas

$$P(E) = P(\{(H,H),(H,T)\}) = \frac{1}{2}$$

and

$$P(F) = P(\{(H,T),(T,T)\}) = \frac{1}{2}.$$

- Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Find $P(E_1 \cap F)$ and $P(E_1)$, $P(F)$.
- There is a 60 percent chance that the event A will occur. If A does not occur, there is a 10 percent chance that B will occur. What is the probability that at least one of the events A or B occur?
- Fifty-two percent of the students at a certain college are females. Five percent of the students in this college are majoring in computer science. Two percent of the students are women majoring in computer science. If a student is selected at random, find the conditional probability that
  1. this student is female, given that the student is majoring in computer science;
  2. this student is majoring in computer science, given that the student is female.

Random variables

- When a random experiment is performed, we are often not interested in all of the details of the experimental result but only in the value of some numerical quantity determined by the result.
- For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the values of the individual dice.
- That is, we may be interested in knowing that the sum is 7 and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1).
- These quantities of interest that are determined by the result of the experiment are known as random variables.
- A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment. It is used to quantify outcomes of a random process.

Example for Random variables

Letting X denote the random variable that is defined as the sum of two fair dice, then

$$P\{X=2\}=P\{(1,1)\}=\frac{1}{36}$$

$$P\{X=3\}=P\{(1,2),(2,1)\}=\frac{2}{36}$$

$$P\{X=4\}=P\{(1,3),(2,2),(3,1)\}=\frac{3}{36}$$

$$P\{X=5\}=P\{(1,4),(2,3),(3,2),(4,1)\}=\frac{4}{36}$$

$$P\{X=6\}=P\{(1,5),(2,4),(3,3),(4,2),(5,1)\}=\frac{5}{36}$$

$$P\{X=7\}=P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}=\frac{6}{36}$$

$$P\{X=8\}=P\{(2,6),(3,5),(4,4),(5,3),(6,2)\}=\frac{5}{36}$$

$$P\{X=9\}=P\{(3,6),(4,5),(5,4),(6,3)\}=\frac{4}{36}$$

$$P\{X=10\}=P\{(4,6),(5,5),(6,4)\}=\frac{3}{36}$$

$$P\{X=11\}=P\{(5,6),(6,5)\}=\frac{2}{36}$$

$$P\{X=12\}=P\{(6,6)\}=\frac{1}{36}$$

Properties of Random variables ...

- Consider a random experiment with sample space S. A **random variable** $X(x)$ is a single-valued real function $X : S \to \mathbb{R}$ that assigns a real number called the *value of* $X(x)$ to each sample point x of S and use *r.v.* to denote the random variable.
- Clearly, a random variable is not a variable at all in the usual sense, and it is a function.
- The sample space S is termed the *domain* of the r.v. X, and the collection of all numbers [values of $X(x)$] is termed the *range* of the r.v. X. Thus the range of X is a certain subset of the set of all real numbers.

Example: In the experiment of tossing a coin once, we might define the r.v. X as

$$X(H) = 1 \quad \text{and} \quad X(T) = 0.$$

Note that we could also define another r.v., say Y or Z, with

$$Y(H) = 0, \quad Y(T) = 1 \quad \text{or} \quad Z(H) = 0, \quad Z(T) = 0.$$

Random variables ...

- A random variable can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$, and is said to be *discrete*.
- For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable.
- However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables.
- The *cumulative distribution function*, or more simply the *distribution function*, F of the random variable X is defined for any real number x by

$$F(x) = P\{X \leq x\}, \quad x \in \mathbb{R}.$$

- That is, $F(x)$ is the probability that the random variable X takes on a value that is less than or equal to x.

Random variables ...

- For example, suppose we wanted to compute $P\{a < X \leq b\}$.
- This can be accomplished by first noting that the event $\{X \leq b\}$ can be expressed as the union of the two mutually exclusive events $\{X \leq a\}$ and $\{a < X \leq b\}$.
- Therefore, applying Axiom 3, we obtain that

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

or

$$P\{a < X \leq b\} = F(b) - F(a)$$

Example: Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \leq 0, \\ 1 - \exp\{-x^2\} & x > 0. \end{cases}$$

What is the probability that X exceeds 1?

Random variables ...

Solution. The desired probability is computed as follows:

$$P\{X > 1\} = 1 - P\{X \leq 1\} = 1 - F(1) = e^{-1} = 0.368.$$

Properties of $F_X(x)$:

1. $0 \leq F(x) \leq 1$
2. $F(x_1) \leq F(x_2) \quad$ if $x_1 < x_2$
3. $\lim_{x\to\infty} F(x) = F(\infty) = 1$
4. $\lim_{x\to-\infty} F(x) = F(-\infty) = 0$
5. $\lim_{x\to a^+} F(x) = F(a^+) = F(a), \quad a^+ = \lim_{\epsilon\to 0}(a+\epsilon)$.

Random variables ...

- For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}.$$

- The probability mass function $p(a)$ is positive for at most a countable number of values of a.
- That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$p(x_i) > 0, \quad i = 1, 2, \ldots$$

$$p(x) = 0, \quad \text{all other values of } x$$

- Since X must take on one of the values x_i, we have

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

Random variables ...

- **Example:** Consider a random variable X that is equal to 1, 2, or 3. If we know that
$$p(1) = \frac{1}{2} \quad \text{and} \quad p(2) = \frac{1}{3}$$
then it follows (since $p(1) + p(2) + p(3) = 1$) that
$$p(3) = \frac{1}{6}.$$

- The cumulative distribution function F can be expressed in terms of $p(x)$ by
$$F(a) = \sum_{\text{all } x \le a} p(x).$$

Random variables ...

- If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \ldots$, then its distribution function F is a step function.
- That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i. For instance, suppose X has a probability mass function given by

$$p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{3}, \quad p(3) = \frac{1}{6}.$$

- Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1, \\ \frac{1}{2} & 1 \leq a < 2, \\ \frac{5}{6} & 2 \leq a < 3, \\ 1 & 3 \leq a. \end{cases}$$

Random variables ...

- Whereas the set of possible values of a discrete random variable is a sequence, we often must consider random variables whose set of possible values is an interval.
- Let X be such a random variable. We say that X is a *continuous random variable* if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

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

- The function $f(x)$ is called the *probability density function* of the random variable X.
-

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

Properties of Probability Density Functions

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain that

$$P\{a \le X \le b\} = \int_a^b f(x)\,dx.$$

If we let $a = b$ in the above, then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0.$$

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x)\,dx.$$

Differentiating both sides yields

$$\frac{d}{da}F(a) = f(a).$$

That is, the density is the derivative of the cumulative distribution function. A somewhat more intuitive interpretation of the density function may be obtained as follows:

$$P\left\{a - \frac{\varepsilon}{2} \le X \le a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\,dx \approx \varepsilon f(a),$$

when ε is small.

Example on PDFs

Example. Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

Solution.

(a) Since f is a probability density function, we must have that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \text{ implying that } C\int_0^2 (4x - 2x^2)\,dx = 1$$

or

$$C\left[2x^2 - \frac{2x^3}{3}\right]_{x=0}^{x=2} = 1, \text{ or } C\left[\left(2(2)^2 - \frac{2(2)^3}{3}\right) - \left(2(0)^2 - \frac{2(0)^3}{3}\right)\right] = 1$$

or

$$C\left[8 - \frac{16}{3}\right] = 1$$

or

$$C \cdot \frac{24 - 16}{3} = 1, \implies C = \frac{3}{8}.$$

Example on PDFs

Hence,

$$P\{X > 1\} = \int_1^\infty f(x)\,dx = \frac{3}{8}\int_1^2 (4x - 2x^2)\,dx.$$

Now,

$$\frac{3}{8}\int_1^2 (4x - 2x^2)\,dx = \frac{3}{8}\left[2x^2 - \frac{2x^3}{3}\right]_{x=1}^{x=2}$$

or

$$\frac{3}{8}\left[\left(2(2)^2 - \frac{2(2)^3}{3}\right) - \left(2(1)^2 - \frac{2(1)^3}{3}\right)\right]$$

or

$$\frac{3}{8}\left[8 - \frac{16}{3} - 2 + \frac{2}{3}\right]$$

or

$$\frac{3}{8}\left[6 - \frac{14}{3}\right]$$

Thus, $P\{X > 1\} = \frac{1}{2}$. $\blacksquare$

Expectation

If X is a discrete random variable taking on the possible values $x_1, x_2, \ldots$, then the *expectation* or *expected value* of X, denoted by $\mathbb{E}[X]$, is defined by:

$$\mu = \mathbb{E}[X] = \sum_i x_i P(X = x_i).$$

Example. If the probability mass function of X is given by:

$$p(0) = \frac{1}{2} = p(1), \implies \mathbb{E}[X] = 0 \cdot \left(\frac{1}{2}\right) + 1 \cdot \left(\frac{1}{2}\right) = \frac{1}{2},$$

which is just the ordinary average of the two possible values 0 and 1 that X can assume. On the other hand, if:

$$p(0) = \frac{1}{3}, \quad p(1) = \frac{2}{3}, \implies \mathbb{E}[X] = 0 \cdot \left(\frac{1}{3}\right) + 1 \cdot \left(\frac{2}{3}\right) = \frac{2}{3}.$$

This is a weighted average of the two possible values 0 and 1, where the value 1 is given twice as much weight as the value 0 since $P(1) = 2P(0)$.

Expectation of a continuous random variable

Suppose that X is a continuous random variable with probability density function f. Since, for small dx:

$$f(x)dx \approx P\{x < X < x + dx\}.$$

Hence, it is natural to define the expected value of X by:

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

Example. Suppose that you are expecting a message at some time past 5 P.M. From experience, you know that X, the number of hours after 5 P.M. until the message arrives, is a random variable with the following probability density function:

$$f(x) = \begin{cases} \frac{1}{1.5}, & \text{if } 0 < x < 1.5, \\ 0, & \text{otherwise.} \end{cases}$$

The expected amount of time past 5 P.M. until the message arrives is given by:

$$\mathbb{E}[X] = \int_{0}^{1.5} x \cdot \frac{1}{1.5} dx = 0.75.$$

Hence, on average, you would have to wait three-fourths of an hour. ■

Properties of the expected value I

- Suppose now that we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case).
- Suppose also that we are interested in calculating, not the expected value of X, but the expected value of some function of X, say $g(X)$.
- How do we go about doing this? One way is as follows.
- Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X.
- Once we have obtained the distribution of $g(X)$, we can then compute $\mathbb{E}[g(X)]$ by the definition of the expectation.

Example. Suppose X has the following probability mass function:

$$p(0) = 0.2, \quad p(1) = 0.5, \quad p(2) = 0.3.$$

Calculate $\mathbb{E}[X^2]$.
Solution. Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values $0^2, 1^2, 2^2$ with respective probabilities:

$$p_Y(0) = P\{Y = 0^2\} = 0.2,$$

Properties of the expected value II

$$p_Y(1) = P\{Y = 1^2\} = 0.5,$$

$$p_Y(4) = P\{Y = 2^2\} = 0.3.$$

Hence,

$$E[X^2] = E[Y] = 0 \cdot (0.2) + 1 \cdot (0.5) + 4 \cdot (0.3) = 1.7.$$

Example. The time, in hours, it takes to locate and repair an electrical breakdown in a certain factory is a random variable -call it X ? whose density function is given by

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

If the cost involved in a breakdown of duration x is x^3, what is the expected cost of such a breakdown?

Solution. Letting $Y = X^3$ denote the cost, we first calculate its distribution function as follows. For $0 \leq a \leq 1$,

$$F_Y(a) = P(Y \leq a)$$

Properties of the expected value III

$$= P(X^3 \le a)$$

$$= P(X \le a^{1/3})$$

$$= \int_0^{a^{1/3}} dx$$

$$= a^{1/3}.$$

By differentiating $F_Y(a)$, we obtain the density of Y,

$$f_Y(a) = \frac{1}{3}a^{-2/3}, \quad 0 \le a < 1.$$

Hence,

$$E[X^3] = E[Y] = \int_{-\infty}^{\infty} a f_Y(a)\, da$$

$$= \int_0^1 a \cdot \frac{1}{3}a^{-2/3}\, da$$

Properties of the expected value IV

$$= \frac{1}{3}\int_0^1 a^{1/3}\,da = \frac{1}{3}\cdot\frac{3}{4}a^{4/3}\Big|_0^1 = \frac{1}{4}.$$

Propositions

(a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

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

(b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

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

Mean & Moment

The expected value of a random variable X, $E[X]$, is also referred to as the *mean* or the *first moment* of X. The quantity $E[X^n], n \geq 1$, is called the *nth moment* of X. We note that

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x^n f(x)\,dx & \text{if } X \text{ is continuous.} \end{cases}$$

Examples on Mean & Moment I

Example. Find $\mathbb{E}[X]$ where X is the outcome when we roll a fair die.
Example. Find the expected value of the sum obtained when two fair dice are rolled.
Solution If X is the sum, then $\mathbb{E}[X]$ can be obtained from the formula

$$\mathbb{E}[X] = \sum_{i=2}^{12} i \cdot P(X = i).$$

However, it is simpler to name the dice, and let X_i be the value on dice i, $i = 1, 2$. As $X = X_1 + X_2$, this yields that

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

Thus, we see that $\mathbb{E}[X] = 7$.
Example. Suppose there are 20 different types of coupons and suppose that each time one obtains a coupon it is equally likely to be any one of the types. Compute the expected number of different types that are contained in a set for 10 coupons.
Solution Let X denote the number of different types in the set of 10 coupons. We compute $\mathbb{E}[X]$ by using the representation

$$X = X_1 + X_2 + \cdots + X_{20}$$

where

$$X_i = \begin{cases} 1, & \text{if at least one type } i \text{ coupon is contained in the set of 10,} \\ 0, & \text{otherwise.} \end{cases}$$

Examples on Mean & Moment II

Now,

$$\mathbb{E}[X_i] = P(X_i = 1)$$

$$= P\{\text{at least one type } i \text{ coupon is in the set of 10}\}$$

$$= 1 - P\{\text{no type } i \text{ coupons are contained in the set of 10}\}$$

$$= 1 - \left(\frac{19}{20}\right)^{10}.$$

where the last equality follows since each of the 10 coupons will (independently) not be a type i with probability $\frac{19}{20}$, Hence,

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_{20}]$$

$$= 20\left[1 - \left(\frac{19}{20}\right)^{10}\right] = 8.025.$$

Variance

Definition. If X is a random variable with mean μ, then the *variance* of X, denoted by $\text{Var}(X)$, is defined by

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2]$$

An alternative formula for $\text{Var}(X)$ can be derived as follows:

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2]$$

$$\begin{aligned} \text{Var}(X) &= E\left[X^2 - 2\mu X + \mu^2\right] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

That is,

$$\text{Var}(X) = E[X^2] - (E[X])^2,$$

or, in words, the variance of X is equal to the expected value of the square of X minus the square of the expected value of X. This is, in practice, often the easiest way to compute $\text{Var}(X)$.

Examples on Variance

Example 1. Compute $\text{Var}(X)$ when X represents the outcome when we roll a fair die.
Solution. Since $P(X=i)=\frac{1}{6}, i=1,2,3,4,5,6$, we obtain

$$E[X^2]=\sum_{i=1}^{6} i^2 P(X=i)=1^2\left(\frac{1}{6}\right)+2^2\left(\frac{1}{6}\right)+3^2\left(\frac{1}{6}\right)+4^2\left(\frac{1}{6}\right)+5^2\left(\frac{1}{6}\right)+6^2\left(\frac{1}{6}\right)=\frac{91}{6}.$$

We have $E[X]=\frac{7}{2}$, we obtain that

$$\text{Var}(X)=E[X^2]-(E[X])^2=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}.$$

Example 2. (Variance of an Indicator Random Variable). If, for some event A,

$$I=\begin{cases}1 & \text{if event } A \text{ occurs}\\ 0 & \text{if event } A \text{ does not occur}\end{cases}$$

then

$$\text{Var}(I)=E[I^2]-(E[I])^2.$$

$$=E[I]-(E[I])^2 \quad \text{since } I^2=I \text{ (as } 1^2=1 \text{ and } 0^2=0\text{)}.$$

$$=E[I](1-E[I])=P(A)(1-P(A)) \quad \text{since } E[I]=P(A).$$

Properties of Variance I

A useful identity concerning variances is that for any constants a and b,

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \tag{4.6.2}$$

To prove Equation (4.6.2), let $\mu = \mathbb{E}[X]$ and recall that $\mathbb{E}[aX + b] = a\mu + b$. Thus, by the definition of variance, we have

$$\begin{aligned}
\text{Var}(aX + b) &= \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] \\
&= \mathbb{E}[(aX + b - a\mu - b)^2] \\
&= \mathbb{E}[(aX - a\mu)^2] \\
&= \mathbb{E}[a^2(X - \mu)^2] \\
&= a^2\mathbb{E}[(X - \mu)^2] \\
&= a^2\text{Var}(X).
\end{aligned}$$

For instance, by setting $a = 0$ in Equation (4.6.2) we obtain that

$$\text{Var}(b) = 0.$$

Properties of Variance II

That is, the variance of a constant is 0. (Is this intuitive?). Similarly, by setting $a = 1$ we obtain

$$\text{Var}(X + b) = \text{Var}(X).$$

That is, the variance of a constant plus a random variable is equal to the variance of the random variable. (Is this intuitive? Think about it.) Finally, setting $b = 0$ yields

$$\text{Var}(aX) = a^2\text{Var}(X).$$

The quantity $\sigma = \sqrt{\text{Var}(X)}$ is called the *standard deviation* of X. The standard deviation has the same units as does the mean.

Remark. Analogous to the mean's being the center of gravity of a distribution of mass, the variance represents, in the terminology of mechanics, the moment of inertia.

Consider

$$\begin{aligned}\text{Var}(X + X) &= \text{Var}(2X)\\ &= 2^2\,\text{Var}(X)\\ &= 4\,\text{Var}(X)\\ &\neq \text{Var}(X) + \text{Var}(X).\end{aligned}$$

Properties of Variance III

Definition. The *covariance* of two random variables X and Y, written $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

where μ_X and μ_Y are the means of X and Y, respectively. A useful expression for $\text{Cov}(X, Y)$ can be obtained by expanding the right side of the definition. This yields

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\ &= \mathbb{E}[XY] - \mu_X \mathbb{E}[Y] - \mu_Y \mathbb{E}[X] + \mu_X \mu_Y \\ &= \mathbb{E}[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned} \tag{4.7.1}$$

From its definition we see that covariance satisfies the following properties:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X) \tag{4.7.2}$$

and

$$\text{Cov}(X, X) = \text{Var}(X). \tag{4.7.3}$$

Properties of Variance IV

Another property of covariance, which immediately follows from its definition, is that, for any constant a,

$$\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y). \tag{4.7.4}$$

Lemma

$$Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y).$$

Proof.

$$\begin{aligned}
\text{Cov}(X_1 + X_2, Y) &= \mathbb{E}[(X_1 + X_2)Y] - \mathbb{E}[X_1 + X_2]\mathbb{E}[Y] \quad \text{from Equation (4.7.1)} \\
&= \mathbb{E}[X_1 Y] + \mathbb{E}[X_2 Y] - (\mathbb{E}[X_1] + \mathbb{E}[X_2])\mathbb{E}[Y] \\
&= \mathbb{E}[X_1 Y] - \mathbb{E}[X_1]\mathbb{E}[Y] + \mathbb{E}[X_2 Y] - \mathbb{E}[X_2]\mathbb{E}[Y] \\
&= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y). \quad \blacksquare
\end{aligned}$$

Note. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{COV}(X, Y)$.

Properties of Variance V

The above lemma can be easily generalized to show that

$$\text{Cov}\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} \text{Cov}(X_i, Y). \tag{4.7.5}$$

Lemma

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j).$$

Properties of Variance VI

Proof.

$$\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \text{Cov}\left(X_i, \sum_{j=1}^{m} Y_j\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j) \quad \text{by the symmetry property Equation (4.7.2) and}$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{n} \text{Cov}(Y_j, X_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j).$$

the result now follows by again applying the symmetry property Equation (4.7.2). ■

Examples on Variance I

Example 1. Compute the variance of the sum obtained when 10 independent rolls of a fair die are made.
Solution. Letting X_i denote the outcome of the i-th roll, we have that

$$\text{Var}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \text{Var}(X_i)$$

$$= 10 \cdot \frac{35}{12}.$$

$$= \frac{175}{6} \quad \blacksquare$$

Example 2. Compute the variance of the number of heads resulting from 10 independent tosses of a fair coin.
Solution. Letting

$$I_j = \begin{cases} 1 & \text{if the } j\text{-th toss lands heads} \\ 0 & \text{if the } j\text{-th toss lands tails} \end{cases}$$

Examples on Variance II

then the total number of heads is equal to

$$\sum_{j=1}^{10} I_j$$

Hence, we have

$$\text{Var}\left(\sum_{j=1}^{10} I_j\right) = \sum_{j=1}^{10} \text{Var}(I_j)$$

Now, since I_j is an indicator random variable for an event having probability $\frac{1}{2}$, it follows that

$$\text{Var}(I_j) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

$$\text{and thus} \quad \text{Var}\left(\sum_{j=1}^{10} I_j\right) = \frac{10}{4} \blacksquare$$

Skewness I

Skewness is a statistical measure that quantifies the asymmetry of a dataset's probability distribution about its mean. It indicates whether the data distribution leans to the left, right, or is symmetric.

- **Positive Skewness (Right-Skewed):** The distribution has a longer or fatter tail on the right side. Most data points are concentrated on the left, and the mean is greater than the median.
  - *Example:* Income distribution in many countries, where a small proportion of individuals earn very high incomes.
- **Negative Skewness (Left-Skewed):** The distribution has a longer or fatter tail on the left side. Most data points are concentrated on the right, and the mean is less than the median.
  - *Example:* Age of retirement, where most people retire around a certain age, but a few retire much earlier.
- **Zero Skewness (Symmetrical):** The distribution is perfectly symmetrical, where the mean, median, and mode are equal.
  - *Example:* Heights of adult men in an ideal scenario.

Skewness II

Definition. Let X be a random variable with mean $\mu = \mathbb{E}[X]$ and standard deviation $\sigma = \sqrt{\mathbb{E}[(X-\mu)^2]}$. The skewness of X is defined as:

$$\text{Skewness}(X) = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3}.$$

Alternatively, for large datasets:

$$Sk = \frac{\mu_3}{\sigma^3}$$

Where:

- $\mu_3 = \frac{1}{n}\sum(x_i - \bar{x})^3$: The third central moment.
- σ: Standard deviation.

Interpretation of Skewness Values

- $Sk > 0$: Positive skewness (right-skewed).
- $Sk < 0$: Negative skewness (left-skewed).
- $Sk = 0$: Symmetrical distribution.

Positive Skewness (Right-Skewed) *Dataset:* $\{2, 3, 5, 8, 50\}$
Here, the mean is significantly greater than the median.

$$xf(x)$$

Skewness III

Negative Skewness (Left-Skewed) *Dataset:* {1, 1, 2, 3, 4, 10}
Here, the mean is less than the median.

$$xf(x)$$

Zero Skewness (Symmetrical) *Dataset:* {1, 2, 3, 4, 5}
Here, the mean equals the median, indicating symmetry.

$$xf(x)^2); Symmetrical$$

Kurtosis I

Definition. Let X be a random variable with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathbb{E}[(X-\mu)^2]$. The kurtosis of X is defined as:

$$\text{Kurtosis}(X) = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4}.$$

Types of Kurtosis Kurtosis is often compared to the kurtosis of a normal distribution, which is 3. Based on this comparison:

- **Mesokurtic:** A distribution with kurtosis equal to 3 (e.g., normal distribution).
- **Leptokurtic:** A distribution with kurtosis greater than 3, indicating heavy tails and a sharp peak.
- **Platykurtic:** A distribution with kurtosis less than 3, indicating light tails and a flatter peak.

Quantiles I

Definition. Let X be a random variable with cumulative distribution function (CDF) $F(x)$. The p-quantile ($0 < p < 1$) of X is defined as the value q_p such that:

$$F(q_p) = P(X \leq q_p) = p,$$

where $F(x)$ is the CDF of X.
Quantiles divide the distribution into parts based on proportions:

- **Median:** The 0.5-quantile ($p = 0.5$) divides the distribution into two equal halves.
- **Quartiles:** Divide the distribution into four equal parts. The first quartile (Q_1) corresponds to $p = 0.25$, and the third quartile (Q_3) corresponds to $p = 0.75$.
- **Percentiles:** Divide the distribution into 100 equal parts.
- **Deciles:** Divide the distribution into 10 equal parts.

Median

Definition. Let X be a random variable with cumulative distribution function (CDF) $F(x)$. The median of X is defined as the value m such that:

$$F(m) = P(X \leq m) = 0.5.$$

For a continuous random variable, this can be expressed as:

$$\int_{-\infty}^{m} f(x)\, dx = 0.5,$$

where $f(x)$ is the probability density function (PDF) of X.

Properties of the Median

- The median minimizes the sum of absolute deviations: $\min_m \mathbb{E}[|X - m|]$.
- The median is robust to outliers, unlike the mean.
- In symmetric distributions (e.g., normal), the median equals the mean.

Mode

Definition. Let X be a random variable. The mode of X is defined as the value x_{mode} that maximizes the probability density function (PDF) $f(x)$ for continuous random variables or the probability mass function (PMF) $P(X = x)$ for discrete random variables.

$$x_{\text{mode}} = \arg\max_{x} f(x) \quad \text{(for continuous variables)},$$

$$x_{\text{mode}} = \arg\max_{x} P(X = x) \quad \text{(for discrete variables)}.$$

Properties of the Mode

- A distribution can have one mode (**unimodal**), two modes (**bimodal**), or more (**multimodal**).
- The mode may not always exist or be unique, especially in distributions with a flat or uniform density.
- The mode is not sensitive to extreme values or outliers.

Moment generating functions I

The moment generating function $\phi(t)$ of the random variable X is defined for all values t by

$$\phi(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{if } X \text{ is continuous} \end{cases}$$

We call $\phi(t)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(t)$. For example,

$$\phi'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt}\left(e^{tX}\right)\right] = E[Xe^{tX}]$$

Hence,

$$\phi'(0) = E[X]$$

Similarly,

$$\phi''(t) = \frac{d}{dt}\phi'(t)$$

Moment generating functions II

$$= \frac{d}{dt} E[Xe^{tX}]$$

$$= E\left[\frac{d}{dt}(Xe^{tX})\right]$$

$$= E[X^2 e^{tX}]$$

and so

$$\phi''(0) = E[X^2]$$

In general, the nth derivative of $\phi(t)$ evaluated at $t = 0$ equals $E[X^n]$; that is,

$$\phi^n(0) = E[X^n], \quad n \geq 1$$

An important property of moment generating functions is that the *moment generating function of the sum of independent random variables is just the product of the individual moment generating functions*. To see this, suppose that X and Y are independent and have moment generating

Moment generating functions III

functions $\phi_X(t)$ and $\phi_Y(t)$, respectively. Then $\phi_{X+Y}(t)$, the moment generating function of $X+Y$, is given by

$$\phi_{X+Y}(t) = E[e^{t(X+Y)}]$$

$$= E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = \phi_X(t)\phi_Y(t)$$

where the next to the last equality follows since X and Y, and thus e^{tX} and e^{tY}, are independent. Another important result is that the *moment generating function uniquely determines the distribution*. That is, there exists a one-to-one correspondence between the moment generating function and the distribution function of a random variable.

Characteristic Function

Definition . Let X be a random variable (RV). The complex-valued function ϕ defined on $\mathbb{R}$ by

$$\phi(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)], \quad t \in \mathbb{R} \tag{3}$$

where $i = \sqrt{-1}$ is the imaginary unit, is called the *characteristic function* (CF) of the RV X. Clearly,

$$\phi(t) = \sum_k (\cos tk + i \sin tk) P(X = k) \tag{4}$$

in the discrete case, and

$$\phi(t) = \int_{-\infty}^{\infty} \cos(tx) f(x)\, dx + i \int_{-\infty}^{\infty} \sin(tx) f(x)\, dx \tag{5}$$

in the continuous case.

Characteristic Function ...

Example 10. Let X be a normal RV with PDF

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}. \tag{6}$$

Then,

$$\phi(t) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx + i \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \sin(tx) e^{-x^2/2} dx. \tag{7}$$

Note that $\sin(tx)$ is an odd function, and so is $\sin(tx)e^{-x^2/2}$. Thus, the second integral on the right-side vanishes, and we have

$$\phi(t) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx. \tag{8}$$

Using known results,

$$\phi(t) = \left(\frac{2}{\sqrt{2\pi}}\right) \int_{0}^{\infty} \cos(tx) e^{-x^2/2} dx = e^{-t^2/2}, \quad t \in \mathbb{R}. \tag{9}$$

Remark 4. Unlike a moment generating function (MGF), which may not exist for some distributions, a characteristic function (CF) always exists, making it a much more convenient tool. In fact, it is easy to see that ϕ is continuous on $\mathbb{R}$, satisfies $|\phi(t)| \leq 1$ for all t, and $\phi(-t) = \overline{\phi(t)}$, where $\overline{\phi}$ is the complex conjugate of ϕ. Thus, ϕ is the CF of $-X$. Moreover, ϕ uniquely determines the distribution function (DF) of the RV X.

Markov's inequality

Proposition. If X is a random variable that takes only nonnegative values, then for any value $a > 0$

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

Proof. We give a proof for the case where X is continuous with density f.

$$E[X] = \int_0^\infty x f(x)\, dx = \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx$$

$$\geq \int_a^\infty x f(x)\, dx \geq \int_a^\infty a f(x)\, dx$$

$$= a \int_a^\infty f(x)\, dx = a P\{X \geq a\}$$

and the result is proved.

Chebyshev's inequality I

Theorem

If X is a random variable with mean μ and variance σ^2, then for any value $k > 0$,

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

Proof.

Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}. \quad (4.9.1)$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, Equation (4.9.1) is equivalent to

$$P(|X - \mu| \geq k) \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2},$$

and the proof is complete. □

Chebyshev's inequality II

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.

Examples based on Chebyshev's inequality

Example

Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution. Let X be the number of items that will be produced in a week:

(a) By Markov's inequality

$$P(X > 75) \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) By Chebyshev's inequality

$$P(|X - 50| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{25}{100} = \frac{1}{4}.$$

Hence,

$$P(|X - 50| < 10) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

and so the probability that this week's production will be between 40 and 60 is at least 0.75.

Remarks on Chebyshev's inequality

By replacing k by $k\sigma$ in Equation (4.9.1), we can write Chebyshev's inequality as

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Thus, it states that the probability a random variable differs from its mean by more than k standard deviations is bounded by $1/k^2$. We will end this section by using Chebyshev's inequality to prove the weak law of large numbers, which states that the probability that the average of the first n terms in a sequence of independent and identically distributed random variables differs from its mean by more than ε goes to 0 as n goes to infinity.

The weak law of large numbers

Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$p\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty$$

Proof. We shall prove the result only under the additional assumption that the random variables have a finite variance σ^2. Now, as

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \text{and} \quad \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

it follows from Chebyshev's inequality that

$$p\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

and the result is proved.

Examples I

Example 1.
The median, like the mean, is important in predicting the value of a random variable. Whereas it was shown in the text that the mean of a random variable is the best predictor from the point of view of minimizing the expected value of the square of the error, the median is the best predictor if one wants to minimize the expected value of the absolute error. That is, $E[|X-c|]$ is minimized when c is the median of the distribution function of X. Prove this result when X is continuous with distribution function F and density function f.
Hint: Write

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

$$= \int_{-\infty}^{c} |x-c|\, f(x)\, dx + \int_{c}^{\infty} |x-c|\, f(x)\, dx$$

$$= \int_{-\infty}^{c} (c-x) f(x)\, dx + \int_{c}^{\infty} (x-c) f(x)\, dx$$

$$= c\,F(c) - \int_{-\infty}^{c} x\, f(x)\, dx + \int_{c}^{\infty} x\, f(x)\, dx - c[1-F(c)]$$

Now, use calculus to find the minimizing value of c.

Examples II

Example 2.
We say that m_p is the *100p percentile* of the distribution function F if

$$F(m_p) = p$$

Find m_p for the distribution having density function

$$f(x) = 2e^{-2x}, \quad x \geq 0$$

Example 3. A community consists of 100 married couples. If 50 members of the community die, what is the expected number of marriages that remain intact? Assume that the set of people who die is equally likely to be any of the $\binom{200}{50}$ groups of size 50.
Hint: For $i = 1, \ldots, 100$ let

$$X_i = \begin{cases} 1 & \text{if neither member of couple } i \text{ dies} \\ 0 & \text{otherwise} \end{cases}$$

Example 4.
Compute the expectation and variance of the number of successes in n independent trials, each of which results in a success with probability p. Is independence necessary?

Example 4. Suppose that X is equally likely to take on any of the values 1, 2, 3, 4. Compute (a) $E[X]$ and (b) $\mathrm{Var}(X)$.

Bernoulli and binomial random variables

Suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or a "failure" is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the probability mass function of X is given by

$$\begin{cases} P\{X = 0\} = 1 - p \\ P\{X = 1\} = p \end{cases} \tag{10}$$

where $p, 0 \leq p \leq 1$, is the probability that the trial is a "success."
Bernoulli random variables. A random variable X is said to be a Bernoulli random variable if its probability mass function is given by Equations (10) for some $p \in (0, 1)$.
Its expected value is

$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = p.$$

Bernoulli and binomial random variables ...

That is, the expectation of a Bernoulli random variable is the probability that the random variable equals 1. Suppose now that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1-p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a **binomial** random variable with parameters (n, p). The probability mass function of a binomial random variable with parameters n and p is given by

$$P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n. \tag{11}$$

It can be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n-i$ failures is, by the assumed independence of trials, $p^i(1-p)^{n-i}$.

Bernoulli and binomial random variables ...

Equation (11) then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n-i$ failures ? which can perhaps most easily be seen by noting that there are $\binom{n}{i}$ different selections of the i trials that result in successes. For instance, if $n = 5$, $i = 2$, then there are $\binom{5}{2}$ choices of the two trials that are to result in successes ? namely, any of the outcomes

$$
\begin{array}{lll}
(s,s,f,f,f) & (f,s,s,f,f) & (f,f,s,s,f) \\
(s,f,s,f,f) & (f,s,f,s,f) & \\
(s,f,f,s,f) & (f,s,f,f,s) & (f,f,s,f,s) \\
(s,f,f,f,s) & (f,f,f,s,s) &
\end{array}
$$

where the outcome (f,s,f,s,f) means, for instance, that the two successes appeared on trials 2 and 4. Since each of the $\binom{5}{2}$ outcomes has probability $p^2(1-p)^3$, we see that the probability of a total of 2 successes in 5 independent trials is $\binom{5}{2}p^2(1-p)^3$.

Bernoulli and binomial random variables ...

As a check, note that, by the binomial theorem, the probabilities sum to 1; that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1.$$

The probability mass function of three binomial random variables with respective parameters $(10, .5)$, $(10, .3)$, and $(10, .6)$ are presented in Figure 5.1. The first of these is symmetric about the value .5, whereas the second is somewhat weighted, or *skewed*, to lower values and the third to higher values.

Example 1. It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned?

The Poisson random variable

A random variable X, taking on one of the values $0, 1, 2, \ldots$, is said to be a Poisson random variable with parameter $\lambda, \lambda > 0$, if its probability mass function is given by

$$P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, \ldots.$$

The symbol e stands for a constant approximately equal to 2.7183. The above equation defines a probability mass function, since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1.$$

Mean and Variance of Poisson random variable

The mean and variance of a Poisson random variable, let us first determine its moment generating function.

$$\phi(t) = E[e^{tX}] = \sum_{i=0}^{\infty} e^{ti} e^{-\lambda} \frac{\lambda^i}{i!}$$

$$= e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} = \exp(\lambda(e^t - 1)) e^{\lambda e^t}.$$

Differentiation yields

$$\phi'(t) = \lambda e^t \exp(\lambda(e^t - 1)).$$

$$\phi''(t) = (\lambda e^t)^2 \exp(\lambda(e^t - 1)) + \lambda e^t \exp(\lambda(e^t - 1))$$

Evaluating at $t = 0$ gives that

$$E[X] = \phi'(0) = \lambda$$

$$\text{Var}(X) = \phi''(0) - (E[X])^2$$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda$$

Thus both the mean and the variance of a Poisson random variable are equal to the parameter λ.

Observation on Poisson random variable

The Poisson random variable has a wide range of applications in a variety of areas because it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small. To see this, suppose that X is a binomial random variable with parameters (n, p) and let $\lambda = np$. Then

$$P\{X = i\} = \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i}.$$

$$= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}.$$

$$= \frac{n(n-1)\ldots(n-i+1)\lambda^i}{n^i i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}.$$

Observation on Poisson random variable...

Now, for n large and p small,

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1)\dots(n-i+1)}{n^i} \approx 1, \quad \left(1-\frac{\lambda}{n}\right)^i \approx 1.$$

Hence, for n large and p small,

$$P\{X=i\} \approx e^{-\lambda}\frac{\lambda^i}{i!}.$$

In other words, if n independent trials, each of which results in a *success* with probability p, are performed, then when n is large and p small, the number of successes occurring is approximately a Poisson random variable with mean $\lambda = np$.

Examples on Poisson random variable

- The number of misprints on a page (or a group of pages) of a book.
- The number of people in a community living to 100 years of age.
- The number of wrong telephone numbers that are dialed in a day.
- The number of transistors that fail on their first day of use.
- Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

Examples on Poisson random variable

- The number of misprints on a page (or a group of pages) of a book.
- The number of people in a community living to 100 years of age.
- The number of wrong telephone numbers that are dialed in a day.
- The number of transistors that fail on their first day of use.
- Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.
  Solution. Let X denote the number of accidents occurring on the stretch of highway in question during this week. Because it is reasonable to suppose that there are a large number of cars passing along that stretch, each having a small probability of being involved in an accident, the number of such accidents should be approximately Poisson distributed. Hence,

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-3}\frac{3^0}{0!} = 1 - e^{-3} \approx 0.9502.$$

Degenerate Distribution

A degenerate random variable X at point k is defined as:

$$P(X = k) = 1, \quad \text{and} \quad P(X \neq k) = 0$$

Equivalently, its cumulative distribution function (CDF) is given by:

$$F(x) = \begin{cases} 0, & x < k \\ 1, & x \geq k. \end{cases}$$

The mean and variance of a degenerate random variable X at point k are:

$$E[X] = k.$$

$$\text{Var}(X) = 0.$$

Moment generating function $\phi(t) = e^{tk}$.

Uniform Distribution on n Points

Definition. A discrete uniform random variable X takes values $x_1, x_2, \ldots, x_n$ with equal probability:

$$P(X = x_i) = \frac{1}{n}, \quad \text{for } i = 1, 2, \ldots, n.$$

Cumulative Distribution Function (CDF). The cumulative distribution function (CDF) of X is given by:

$$F(x) = \begin{cases} 0, & x < x_1, \\ \frac{k}{n}, & x_k \le x < x_{k+1}, \quad k = 1, 2, \ldots, n-1, \\ 1, & x \ge x_n. \end{cases}$$

Mean (Expected Value). The expected value of X is:

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i \frac{1}{n}.$$

Thus,

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

If X takes values $1, 2, \ldots, n$, then:

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Uniform Distribution on n Points...

Variance. The variance of X is given by:

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

First, we compute $E[X^2]$:

$$E[X^2] = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

Now, using $E[X] = \frac{n+1}{2}$, we get:

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2.$$

Expanding:

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}.$$

Simplifying:

$$\text{Var}(X) = \frac{(n+1)(n-1)}{12}.$$

Uniform Distribution on n Points

Moment Generating Function (MGF). The moment generating function (MGF) is defined as:

$$\phi(t) = E[e^{tX}].$$

For a discrete uniform distribution:

$$\phi(t) = \frac{1}{n} \sum_{i=1}^{n} e^{ti}.$$

This sum is a geometric series:

$$\phi(t) = \frac{1}{n} \cdot \frac{e^t(1 - e^{tn})}{1 - e^t}, \quad \text{for } t \neq 0.$$

For $t = 0$, we get $\phi(0) = 1$, which is expected.

The hypergeometric random variable

A bin contains $N + M$ batteries, of which N are of acceptable quality and the other M are defective. A sample of size n is to be randomly chosen (without replacements) in the sense that the set of sampled batteries is equally likely to be any of the $\binom{N+M}{n}$ subsets of size n. If we let X denote the number of acceptable batteries in the sample, then

$$P\{X = i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}, \quad i = 0, 1, \ldots, \min(N, n)^*$$

Any random variable X whose probability mass function is given by the above equation is said to be a *hypergeometric* random variable with parameters N, M, n.

Example based on hypergeometric random variable

Example. The components of a 6-component system are to be randomly chosen from a bin of 20 used components. The resulting system will be functional if at least 4 of its 6 components are in working condition. If 15 of the 20 components in the bin are in working condition, what is the probability that the resulting system will be functional?

Solution. If X is the number of working components chosen, then X is hypergeometric with parameters $15, 5, 6$. The probability that the system will be functional is

$$P\{X \geq 4\} = \sum_{i=4}^{6} P\{X = i\}$$

$$= \frac{\binom{15}{4}\binom{5}{2} + \binom{15}{5}\binom{5}{1} + \binom{15}{6}\binom{5}{0}}{\binom{20}{6}}$$

$$\approx 0.8687.$$

The uniform random variable I

A random variable X is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its probability density function is given by

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

We know that

$$\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} dx = 1.$$

The uniform distribution arises in practice when we suppose a certain random variable is equally likely to be near any value in the interval $[\alpha, \beta]$. The probability that X lies in any subinterval of $[\alpha, \beta]$ is equal to the length of that subinterval divided by the length of the interval $[\alpha, \beta]$. This follows since when $[a, b]$ is a subinterval of $[\alpha, \beta]$,

$$P\{a < X < b\} = \frac{1}{\beta - \alpha} \int_{a}^{b} dx = \frac{b-a}{\beta - \alpha}.$$

The uniform random variable II

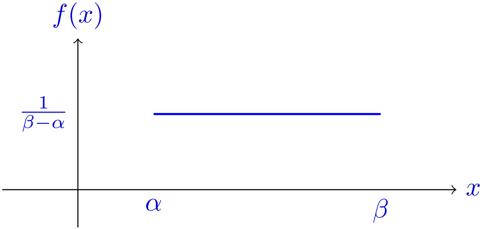


Figure 1: Graph of $f(x)$ for a uniform $[\alpha, \beta]$.

The uniform random variable III

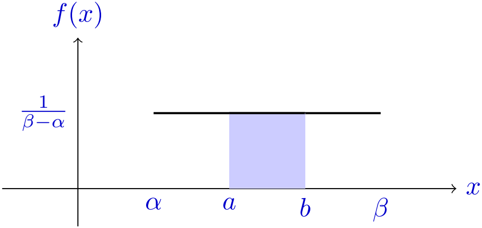


Figure 2: Probabilities of a uniform random variable.

Compute the CDF of Uniform distribution: The cumulative distribution function is given by:

$$F(x) = \begin{cases} 0, & x < \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \alpha \leq x \leq \beta \\ 1, & x > \beta \end{cases}$$

Example based on uniform random variable

Example If X is uniformly distributed over the interval $[0, 10]$, compute the probability that

- (a) $2 < X < 9$,
- (b) $1 < X < 4$,
- (c) $X < 5$,
- (d) $X > 6$.

Example. Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

- (a) less than 5 minutes for a bus;
- (b) at least 12 minutes for a bus.

Mean and Variance UDF

The mean of a uniform $[\alpha, \beta]$ random variable is

$$E[X] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)} = \frac{\alpha + \beta}{2}.$$

The variance is computed as follows:

$$E[X^2] = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x^2 dx = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}.$$

Hence,

$$\text{Var}(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 = \frac{4(\beta^2 + \alpha\beta + \alpha^2) - 3(\alpha^2 + 2\alpha\beta + \beta^2)}{12}$$

$$= \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{12}$$

$$= \frac{(\beta - \alpha)^2}{12}$$

Normal random variables

A random variable is said to be normally distributed with parameters μ and σ^2, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density is

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

The normal density $f(x)$ is a bell-shaped curve that is symmetric about μ and that attains its maximum value of $\frac{1}{\sqrt{2\pi}\sigma} \approx 0.399/\sigma$ at $x = \mu$.

Compute CDF: The cumulative distribution function (CDF) of a normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ is given by:

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, dt.$$

Alternatively, it is often expressed in terms of the standard normal CDF $\Phi(z)$:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} \, dt.$$

Normal random variables ...

To compute $E[X]$ note that

$$E[X-\mu] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Letting $y = (x-\mu)/\sigma$ gives that

$$E[X-\mu] = \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ye^{-y^2/2}\,dy$$

But

$$\int_{-\infty}^{\infty} ye^{-y^2/2}\,dy = -e^{-y^2/2}\Big|_{-\infty}^{\infty} = 0$$

showing that $E[X-\mu] = 0$, or equivalently that

$$E[X] = \mu.$$

Normal random variables ...

We now compute $\text{Var}(X)$ as follows:

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

$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma^2 y^2 e^{-y^2/2}\,dy$$

With $u = y$ and $dv = ye^{-y^2/2}\,dy$, the integration by parts formula

$$\int u\,dv = uv - \int v\,du$$

yields that

$$\int_{-\infty}^{\infty} y^2 e^{-y^2/2}\,dy = -ye^{-y^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}\,dy$$

$$= \int_{-\infty}^{\infty} e^{-y^2/2}\,dy.$$

Normal random variables ...

Hence, We get

$$\text{Var}(X) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy = \sigma^2,$$

where the preceding used that $\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$ is the density function of a normal random variable with parameters $\mu = 0$ and $\sigma = 1$, so its integral must equal 1.

Thus μ and σ^2 represent, respectively, the mean and variance of the normal distribution.

A very important property of normal random variables is that if X is normal with mean μ and variance σ^2, then for any constants a and $b, b \neq 0$, the random variable $Y = a + bX$ is also a normal random variable with parameters

$$E[Y] = E[a + bX] = a + bE[X] = a + b\mu$$

and variance

$$\text{Var}(Y) = \text{Var}(a + bX) = b^2 \text{Var}(X) = b^2\sigma^2.$$

Normal random variables ...

It follows from the foregoing that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

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

is a normal random variable with mean 0 and variance 1. Such a random variable Z is said to have a *standard*, or *unit*, normal distribution. Let $\Phi(\cdot)$ denote its distribution function. That is,

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

This result that $Z = (X - \mu)/\sigma$ has a standard normal distribution when X is normal with parameters μ and σ^2 is quite important, for it enables us to write all probability statements about X in terms of probabilities for Z. For instance, to obtain $P(X < b)$, we note that X will be less than b if and only if $(X - \mu)/\sigma$ is less than $(b - \mu)/\sigma$, and so

$$P\{X < b\} = P\left\{\frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right\} = \Phi\left(\frac{b - \mu}{\sigma}\right).$$

Normal random variables ...

Similarly, for any $a < b$,

$$P\{a < X < b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right\}$$

$$= P\left\{\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right\}$$

$$= P\left\{Z < \frac{b-\mu}{\sigma}\right\} - P\left\{Z < \frac{a-\mu}{\sigma}\right\}$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Normal random variables ...

$$0 f_Z(z) - x x^2/2)/sqrt(2*pi);^2/2)/sqrt(2*pi);^2/2)/sqrt(2*pi); \quad P(Z < -x) \quad P(Z$$

Normal random variables ...

This has been accomplished by an approximation and the results are presented in Table A of the Appendix, which tabulates $\Phi(x)$ (to a 4-digit level of accuracy) for a wide range of nonnegative values of x.
While Table A tabulates $\Phi(x)$ only for nonnegative values of x, we can also obtain $\Phi(-x)$ from the table by making use of the symmetry (about 0) of the standard normal probability density function. That is, for $x > 0$, if Z represents a standard normal random variable, then

$$\Phi(-x) = P(Z < -x)$$

$$= P(Z > x) \quad \text{by symmetry}$$

$$= 1 - \Phi(x)$$

Thus, for instance,

$$P(Z < -1) = \Phi(-1) = 1 - \Phi(1) = 1 - .8413 = .1587$$

Normal random variables ...

Example If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find
(a) $P(X < 11)$;
(b) $P(X > -1)$;
(c) $P(2 < X < 7)$.

Solution (a).

$$P(X < 11) = P\left(\frac{X-3}{4} < \frac{11-3}{4}\right) = \Phi(2) = .9772$$

Solution (b).

$$P(X > -1) = P\left(\frac{X-3}{4} > \frac{-1-3}{4}\right) = P(Z > -1)$$

Solution (c).

$$P(2 < X < 7) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right) = \Phi(1) - \Phi(-1/4)$$

$$= \Phi(1) - (1 - \Phi(1/4))$$

MGF of a normal random variable

Let us now compute the moment generating function of a normal random variable. To start, we compute the moment generating function of a standard normal random variable Z.

$$E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2tx)/2}\, dx$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$$= e^{t^2/2}$$

MGF of Normal random variables ...

Now, if Z is a standard normal, then $X = \mu + \sigma Z$ is normal with mean μ and variance σ^2. Using the preceding, its moment generating function is

$$E[e^{tX}] = E[e^{t(\mu+\sigma Z)}] = E[e^{t\mu} e^{t\sigma Z}]$$

$$= e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} e^{(\sigma t)^2/2} = e^{\mu t + \sigma^2 t^2/2}.$$

Exponential random variables

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

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

is said to be an *exponential* random variable (or, more simply, is said to be exponentially distributed) with parameter λ. The cumulative distribution function $F(x)$ of an exponential random variable is given by

$$F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda y}\, dy = 1 - e^{-\lambda x}, \quad x \geq 0.$$

The exponential distribution often arises, in practice, as being the distribution of the amount of time until some specific event occurs. For instance, the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number are all random variables that tend in practice to have exponential distributions.

Exponential random variables

The moment generating function of the exponential is given by

$$\phi(t) = E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

Differentiation yields

$$\phi'(t) = \frac{\lambda}{(\lambda - t)^2}$$

$$\phi''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

and so

$$E[X] = \phi'(0) = \frac{1}{\lambda}.$$

Exponential random variables

$$\text{Var}(X) = \phi''(0) - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Thus, λ is the reciprocal of the mean, and the variance is equal to the square of the mean. The key property of an exponential random variable is that it is memoryless, where we say that a nonnegative random variable X is *memoryless* if

$$P(X > s + t \mid X > t) = P(X > s) \quad \text{for all } s, t \geq 0.$$

The gamma distribution

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, $\alpha > 0$, if its density function is given by

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

where

$$\Gamma(\alpha) = \int_0^\infty \lambda e^{-\lambda x} (\lambda x)^{\alpha - 1} dx$$

$$= \int_0^\infty e^{-y} y^{\alpha - 1} dy \quad \text{(by letting } y = \lambda x\text{).}$$

The gamma distribution ...

The integration by parts formula

$$\int u dv = uv - \int v du$$

yields, with $u = y^{\alpha-1}$, $dv = e^{-y}dy$, $v = -e^{-y}$, that for $\alpha > 1$,

$$\int_0^\infty e^{-y}y^{\alpha-1}dy = -e^{-y}y^{\alpha-1}\Big|_{y=0}^{y=\infty} + \int_0^\infty e^{-y}(\alpha-1)y^{\alpha-2}dy$$

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

or

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1). \tag{5.7.1}$$

The gamma distribution ...

When α is an integer- say, $\alpha = n$-we can iterate the foregoing to obtain that

$$\Gamma(n) = (n-1)\Gamma(n-1)$$

$$= (n-1)(n-2)\Gamma(n-2) \quad \text{by letting } \alpha = n-1 \text{ in Equation (5.7.1)}$$

$$= (n-1)(n-2)(n-3)\Gamma(n-3) \quad \text{by letting } \alpha = n-2 \text{ in Equation (5.7.1)}$$

$$\vdots$$

$$= (n-1)!\Gamma(1).$$

Because

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

The gamma distribution ...

we see that

$$\Gamma(n) = (n-1)!$$

The function $\Gamma(\alpha)$ is called the *gamma function*. It should be noted that when $\alpha = 1$, the gamma distribution reduces to the exponential with mean $1/\lambda$. The moment generating function of a gamma random variable X with parameters (α, λ) is obtained as follows:

$$\phi(t) = E[e^{tX}]$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{tx} e^{-\lambda x} x^{\alpha-1} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(\lambda-t)x} x^{\alpha-1} dx$$

$$= \left(\frac{\lambda}{\lambda-t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-y} y^{\alpha-1} dy \quad \text{[by } y = (\lambda-t)x\text{]}$$

$$= \left(\frac{\lambda}{\lambda-t}\right)^\alpha, \tag{5.7.2}$$

The gamma distribution ...

where the final equality used that $e^{-y}y^{\alpha-1}/\Gamma(\alpha)$ is a density function, and thus integrates to 1. Differentiation of Equation (5.7.2) yields

$$\phi'(t) = \frac{\alpha\lambda^{\alpha}}{(\lambda - t)^{\alpha+1}}$$

$$\phi''(t) = \frac{\alpha(\alpha+1)\lambda^{\alpha}}{(\lambda - t)^{\alpha+2}}$$

Hence,

$$E[X] = \phi'(0) = \frac{\alpha}{\lambda} \tag{5.7.3}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \phi''(0) - \left(\frac{\alpha}{\lambda}\right)^2$$

$$= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}. \tag{5.7.4}$$

The gamma distribution ...

An important property of the gamma is that if X_1 and X_2 are independent gamma random variables having respective parameters (α_1, λ) and (α_2, λ), then $X_1 + X_2$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$. This result easily follows since

$$\phi_{X_1+X_2}(t) = E[e^{t(X_1+X_2)}]$$

$$= \phi_{X_1}(t)\phi_{X_2}(t)$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} \quad \text{from Equation (5.7.2)}$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2}$$

which is seen to be the moment generating function of a gamma $(\alpha_1 + \alpha_2, \lambda)$ random variable. Since a moment generating function uniquely characterizes a distribution, the result entails. The foregoing result easily generalizes to yield the following proposition.

The gamma distribution ...

Proposition If $X_i, i = 1, \ldots, n$ are independent gamma random variables with respective parameters (α_i, λ), then $\sum_{i=1}^{n} X_i$ is gamma with parameters $\sum_{i=1}^{n} \alpha_i, \lambda$.

Example

The lifetime of a battery is exponentially distributed with rate λ. If a stereo cassette requires one battery to operate, then the total playing time one can obtain from a total of n batteries is a gamma random variable with parameters (n, λ).

Jointly distributed random variables

For a given experiment, we are often interested not only in probability distribution functions of individual random variables but also in the relationships between two or more random variables. For instance, in an experiment into the possible causes of cancer, we might be interested in the relationship between the average number of cigarettes smoked daily and the age at which an individual contracts cancer. To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

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

Jointly distributed random variables ...

The distribution function of X -call it F_X- can be obtained from the joint distribution function F of X and Y as follows:

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F(x, \infty)$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(y) = F(\infty, y)$$

In the case where X and Y are both discrete random variables whose possible values are, respectively, $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, we define the *joint probability mass function* of X and Y, $p(x_i, y_j)$, by

$$p(x_i, y_j) = P(X = x_i, Y = y_j).$$

Jointly distributed random variables ...

The individual probability mass functions of X and Y are easily obtained from the joint probability mass function by the following reasoning. Since Y must take on some value y_j, it follows that the event $\{X = x_i\}$ can be written as the union, over all j, of the mutually exclusive events $\{X = x_i, Y = y_j\}$. That is,

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

and so, using Axiom 3 of the probability function, we see that

$$P(X = x_i) = P\left(\bigcup_j \{X = x_i, Y = y_j\}\right) = \sum_j P(X = x_i, Y = y_j)$$

$$= \sum_j p(x_i, y_j).$$

Jointly distributed random variables ...

Similarly, we can obtain $P(Y = y_j)$ by summing $p(x_i, y_j)$ over all possible values of x_i, that is,

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p(x_i, y_j).$$

Hence, specifying the joint probability mass function always determines the individual mass functions. However, it should be noted that the reverse is not true. Namely, knowledge of $P(X = x_i)$ and $P(Y = y_j)$ does not determine the value of $P(X = x_i, Y = y_j)$.

Example. Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability mass function of X and Y, $p(i, j) = P(X = i, Y = j)$, is given by

$$p(i, j) = \frac{\binom{3}{i}\binom{4}{j}\binom{5}{3-i-j}}{\binom{12}{3}}.$$

Jointly distributed random variables ...

We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane)

$$P\{(X, Y) \in C\} = \iint\limits_{(x,y)\in C} f(x, y)\, dx\, dy$$

The function $f(x, y)$ is called the *joint probability density function* of X and Y. If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from Equation (4.3.3) that

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y)\, dx\, dy.$$

Jointly distributed random variables ...

Because

$$F(a,b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\}$$

$$= \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\,dx\,dy$$

it follows, upon differentiation, that

$$f(a,b) = \frac{\partial^2}{\partial a\,\partial b} F(a,b)$$

wherever the partial derivatives are defined. Another interpretation of the joint density function is obtained as follows:

$$P\{a < X < a+da, b < Y < b+db\} = \int_{b}^{b+db} \int_{a}^{a+da} f(x,y)\,dx\,dy$$

$$\approx f(a,b)da\,db$$

Jointly distributed random variables ...

when da and db are small and $f(x, y)$ is continuous at a, b. Hence $f(a, b)$ is a measure of how likely it is that the random vector (X, Y) will be near (a, b). If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$\begin{aligned} P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\} &= \int_A \int_{-\infty}^{\infty} f(x, y)\, dy\, dx \\ &= \int_A f_X(x)\, dx \end{aligned}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

is thus the probability density function of X. Similarly, the probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Jointly distributed random variables ...

Example. The joint density function of X and Y is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Compute (a) $P\{X > 1, Y < 1\}$; (b) $P\{X < Y\}$; and (c) $P\{X < a\}$.
Solution. (a)

$$P\{X > 1, Y < 1\} = \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\, dx\, dy = \int_0^1 2e^{-2y} \left(-e^{-x} \Big|_1^\infty \right) dy$$

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

Jointly distributed random variables ...

(b)

$$P\{X < Y\} = \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy$$

$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy = \int_0^\infty 2e^{-2y}(1-e^{-y})\,dy$$

$$= \int_0^\infty 2e^{-2y}\,dy - \int_0^\infty 2e^{-3y}\,dy = 1 - \frac{2}{3} = \frac{1}{3}.$$

(c)

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-2y}e^{-x}\,dy\,dx = \int_0^a e^{-x}\,dx = 1 - e^{-a}.$$

Jointly distributed random variables ...

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \tag{4.3.7}$$

In other words, X and Y are independent if, for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent. It can be shown by using the three axioms of probability that Equation (4.3.7) will follow if and only if for all a, b

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

Hence, in terms of the joint distribution function F of X and Y, we have that X and Y are independent if

$$F(a, b) = F_X(a)F_Y(b) \quad \text{for all } a, b.$$

Jointly distributed random variables ...

When X and Y are discrete random variables, the condition of independence Equation (4.3.7) is equivalent to

$$p(x, y) = p_X(x)p_Y(y) \quad \text{for all } x, y \tag{4.3.8}$$

where p_X and p_Y are the probability mass functions of X and Y. The equivalence follows because, if Equation (4.3.7) is satisfied, then we obtain Equation (4.3.8) by letting A and B be, respectively, the one-point sets $A = \{x\}$, $B = \{y\}$. Furthermore, if Equation (4.3.8) is valid, then for any sets A, B,

$$P\{X \in A, Y \in B\} = \sum_{y \in B} \sum_{x \in A} p(x, y)$$

$$= \sum_{y \in B} \sum_{x \in A} p_X(x)p_Y(y) = \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) = P(Y \in B)P(X \in A)$$

and thus Equation (4.3.7) is established.

Jointly distributed random variables ...

In the jointly continuous case, the condition of independence is equivalent to

$$f(x,y) = f_X(x) f_Y(y) \quad \text{for all } x, y$$

Loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other. Random variables that are not independent are said to be dependent.

Example. Suppose that X and Y are independent random variables having the common density function

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

Solution. We start by determining the distribution function of X/Y. For $a > 0$,

$$F_{X/Y}(a) = P\{X/Y \leq a\} = \iint\limits_{x/y \leq a} f(x,y)\, dx\, dy$$

$$= \iint e^{-x} e^{-y}\, dx\, dy.$$

Jointly distributed random variables ...

$$= \int_0^\infty \int_0^{ay} e^{-x} e^{-y}\, dx\, dy = \int_0^\infty (1 - e^{-ay}) e^{-y}\, dy$$

$$= \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right]_0^\infty = 1 - \frac{1}{a+1}$$

Differentiation yields that the density function of X/Y is given by

$$f_{X/Y}(a) = \frac{1}{(a+1)^2}, \quad 0 < a < \infty.$$

We can also define joint probability distributions for n random variables in exactly the same manner as we did for $n = 2$. For instance, the joint cumulative probability distribution function $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}.$$

Jointly distributed random variables ...

If these random variables are discrete, we define their joint probability mass function $p(x_1, x_2, \ldots, x_n)$ by

$$p(x_1, x_2, \ldots, x_n) = P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$$

Further, the n random variables are said to be jointly continuous if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the joint probability density function, such that for any set C in n-space

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \int \cdots \int_{(x_1, \ldots, x_n) \in C} f(x_1, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n$$

In particular, for any n sets of real numbers $A_1, A_2, \ldots, A_n$

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \int_{A_1} \cdots \int_{A_n} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

Jointly distributed random variables ...

The concept of independence may, of course, also be defined for more than two random variables. In general, the n random variables $X_1, X_2, \ldots, X_n$ are said to be independent if, for all sets of real numbers $A_1, A_2, \ldots, A_n$,

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

Jointly distributed random variables ...

As before, it can be shown that this condition is equivalent to

$$P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$

$$= \prod_{i=1}^{n} P\{X_i \leq a_i\} \quad \text{for all } a_1, a_2, \ldots, a_n$$

Finally, we say that an infinite collection of random variables is independent if every finite subcollection of them is independent.

Jointly distributed random variables ...

Example. Suppose that the successive daily changes of the price of a given stock are assumed to be independent and identically distributed random variables with probability mass function given by

$$P(\text{daily change is } i) = \begin{cases} -3 & \text{with probability } 0.05 \\ -2 & \text{with probability } 0.10 \\ -1 & \text{with probability } 0.20 \\ 0 & \text{with probability } 0.30 \\ 1 & \text{with probability } 0.20 \\ 2 & \text{with probability } 0.10 \\ 3 & \text{with probability } 0.05 \end{cases}$$

Then the probability that the stockâs price will increase successively by 1, 2, and 0 points in the next three days is

$$P\{X_1 = 1, X_2 = 2, X_3 = 0\} = (.20)(.10)(.30) = .006$$

where we have let X_i denote the change on the ith day.