Applied Statistics

Course Code: MAT1011

Dr. Sukanta Nayak

Department of Mathematics
School of Advanced Sciences
VIT-AP University, Amaravati
Andhra Pradesh

VIT-AP
UNIVERSITY

Outline

Module No. 1 **Data and Decisions**

Module No. 2 **Probability**

Module No. 3 **Modelling with Probability**

Module No. 4 **Exploring the Data using Correlation and Regression**

Module No. 5 **Inference for Decision Making-I**

Module No. 6 **Inference for Decision Making-II**

Module No. 1

Data and Decisions

Data,

Variable types,

Data Source: Where - How and When, Summarizing the variable.

Module No. 2

Probability

Basics of Probability,

Different types of probability,

Probability rules,

Conditional probability.

Module No. 3

Modelling with Probability

Random Variables – Expected value of Random variable, Properties of Expected Values, Bernoulli Trials, - Discrete Probability, Models- The Binomial Model, The Poisson Model - Continuous Probability, Model-Normal Model, Estimation.

Module No. 4

Exploring the Data using Correlation and Regression

Understanding Correlation,

The linear model,

Regression line.

Module No. 5

Inference for Decision Making-I

Hypothesis,
A trial as a hypothesis test,
The Reasoning of Hypothesis Testing,
p-values and decision: What to tell about a Hypothesis test (z-test),
Testing hypothesis about mean- the one sample t-test,
Comparing two means, The two sample tests, Test of hypothesis for single proportion,
Test of Hypothesis for Difference of proportion.

Module No. 6

Inference for Decision Making-II

Goodness of fit tests,
The Chi-Square Test of Homogeneity,
Comparing Two Proportions,
Chi-Square Test of Independence,
The F- Test. ANOVA: One-way analysis of variance (fixed-effect model), Two-way analysis of variance (fixed-effect model).

Module No. 1

Data and Decisions

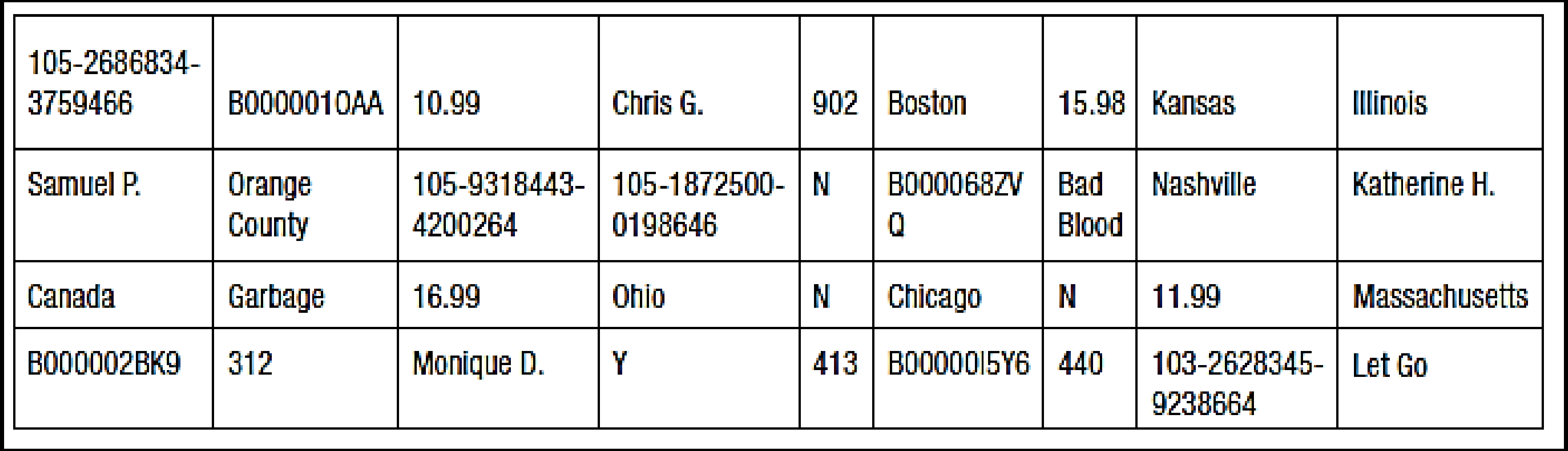
Data

- Collection of information on virtually or physically that are recorded and stored electronically, in vast digital repositories called data warehouses.
- The process of using data, especially of transactional data (data collected for recording the companies’ transactions) to make other decisions and predictions, is sometimes called data mining or predictive analytics. The more general term business analytics (or sometimes simply analytics) describes any use of statistical analysis to drive business decisions from data whether the purpose is predictive or simply descriptive.
- Credit card transactions hold the key to understanding customer behaviour.
- Netflix uses analytics on customer information both to recommend new movies and to adapt the website that customers see to individual tastes.

Data

- ➢ To understand better what data are, let’s look at some hypothetical company records that Amazon might collect:

Table 1. An example of data with no context. It’s impossible to say anything about what these values might mean without knowing their context.

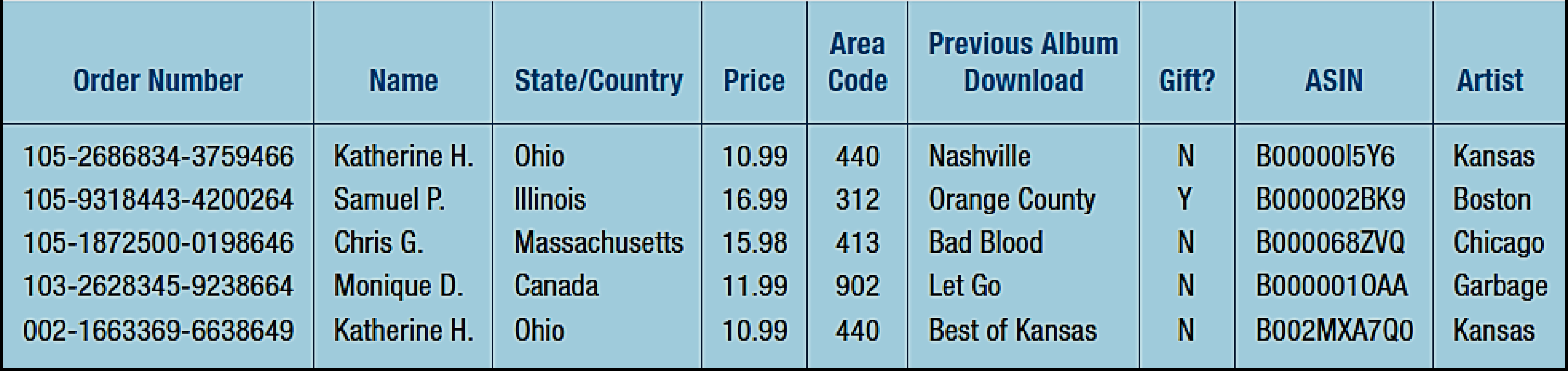


105-2686834-3759466	B0000010AA	10.99	Chris G.	902	Boston	15.98	Kansas	Illinois
Samuel P.	Orange County	105-9318443-4200264	105-1872500-0198646	N	B000068ZVQ	Bad Blood	Nashville	Katherine H.
Canada	Garbage	16.99	Ohio	N	Chicago	N	11.99	Massachusetts
B000002BK9	312	Monique D.	Y	413	B00000I5Y6	440	103-2628345-9238664	Let Go

We can make the meaning clear if we add the context of who the data are about and what was measured and organize the values into a data table such as this one.

Data

Table 2. Example of a data table. The variable names are in the top row. Typically, the Who of the table are found in the leftmost column.

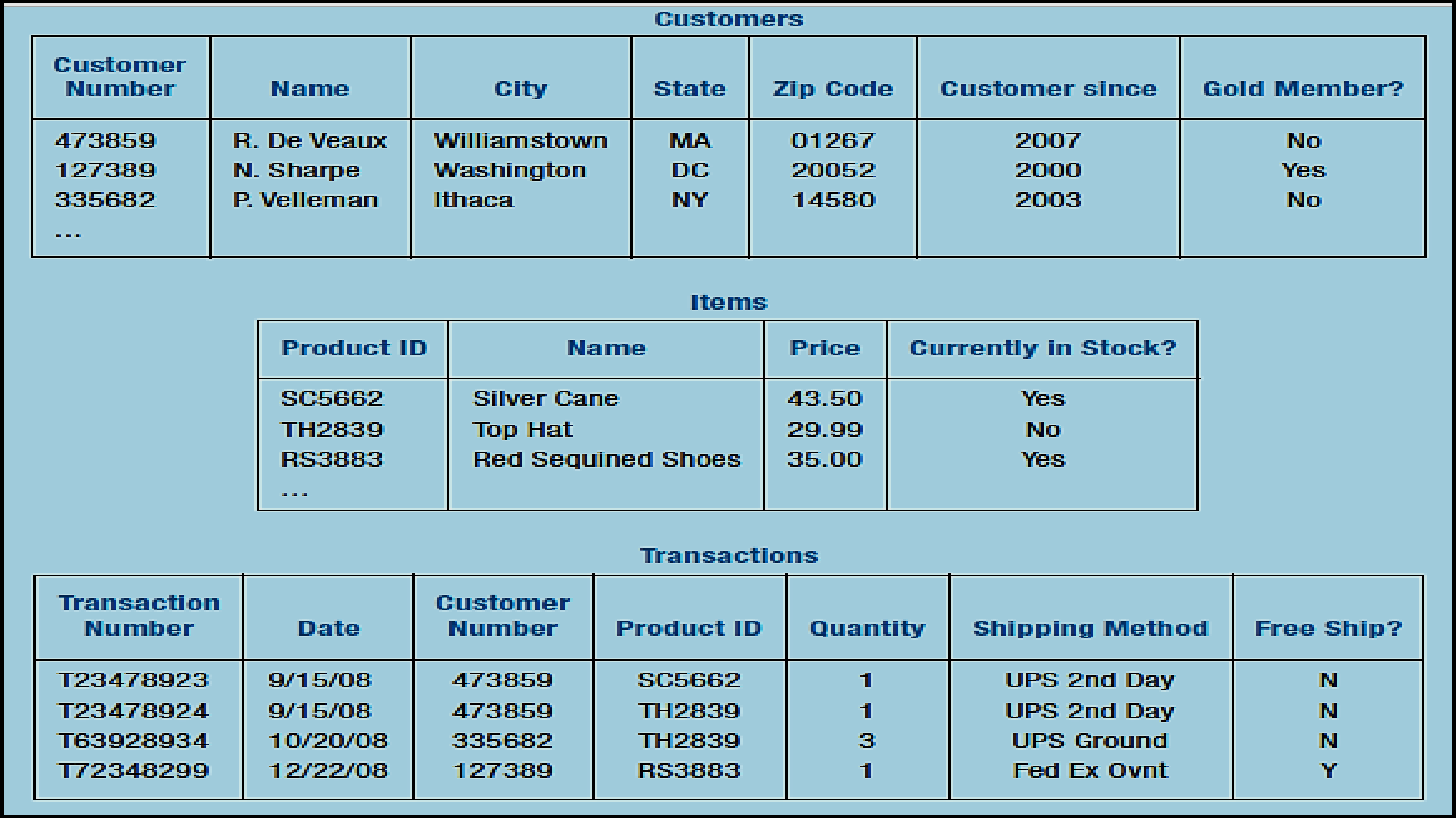


Order Number	Name	State/Country	Price	Area Code	Previous Album Download	Gift?	ASIN	Artist
105-2686834-3759466	Katherine H.	Ohio	10.99	440	Nashville	N	B00000I5Y6	Kansas
105-9318443-4200264	Samuel P.	Illinois	16.99	312	Orange County	Y	B000002BK9	Boston
105-1872500-0198646	Chris G.	Massachusetts	15.98	413	Bad Blood	N	B000068ZVQ	Chicago
103-2628345-9238664	Monique D.	Canada	11.99	902	Let Go	N	B0000010AA	Garbage
002-1663369-6638649	Katherine H.	Ohio	10.99	440	Best of Kansas	N	B002MXA7Q0	Kansas

- In general, the rows of a data table correspond to individual cases about which we’ve recorded some characteristics called variables.
- Individuals who answer a survey are referred to as respondents. People on whom we experiment are subjects or (in an attempt to acknowledge the importance of their role in the experiment) participants, but animals, plants, websites, and other inanimate subjects are often called experimental units. Often we call cases just what they are: for example, customers, economic quarters, or companies. In a database, rows are called records—in this example, purchase records. Perhaps the most generic term is cases.

Data

- Metadata typically contains information about how, when, and where (and possibly why) the data were collected; who each case represents; and the definitions of all the variables.



Customers

Customer Number	Name	City	State	Zip Code	Customer since	Gold Member?
473859	R. De Veaux	Williamstown	MA	01267	2007	No
127389	N. Sharpe	Washington	DC	20052	2000	Yes
335682	P. Velleman	Ithaca	NY	14580	2003	No
...						

Items

Product ID	Name	Price	Currently in Stock?
SC5662	Silver Cane	43.50	Yes
TH2839	Top Hat	29.99	No
RS3883	Red Sequined Shoes	35.00	Yes
...			

Transactions

Transaction Number	Date	Customer Number	Product ID	Quantity	Shipping Method	Free Ship?
T23478923	9/15/08	473859	SC5662	1	UPS 2nd Day	N
T23478924	9/15/08	473859	TH2839	1	UPS 2nd Day	N
T63928934	10/20/08	335682	TH2839	3	UPS Ground	N
T72348299	12/22/08	127389	RS3883	1	Fed Ex Ovnt	Y

Variable Types

- When a variable names categories and answers questions about how cases fall into those categories, we call it a categorical, or qualitative, variable. When a variable has measured numerical values with units and the variable tells us about the quantity of what is measured, we call it a quantitative variable.
- There are exactly as many categories as individuals and only one individual in each category. While it’s easy to count the totals for each category, it’s not very interesting. This is an identifier variable.
- By contrast, a categorical variable that names categories that don’t have order is sometimes called nominal.
- Cross-Sectional and Time Series Data: crosssectional data, where several variables are measured at the same time point.

Data Sources: Where, How, and When

- We must know who, what, and why to analyze data. Without knowing these three, we don’t have enough to start. Of course, we’d always like to know more because the more we know, the more we’ll understand.

What can go wrong?

- Don’t label a variable as categorical or quantitative without thinking about the data and what they represent. The same variable can sometimes take on different roles.
- Don’t assume that a variable is quantitative just because its values are numbers. Categories are often given numerical labels. Don’t let that fool you into thinking they have quantitative meaning. Look at the context.
- Always be skeptical. One reason to analyze data is to discover the truth. Even when you are told a context for the data, it may turn out that the truth is a bit (or even a lot) different. The context colors our interpretation of the data, so those who want to influence what you think may slant the context. A survey that seems to be about all students may in fact report just the opinions of those who visited a fan website. The question that respondents answered may be posed in a way that influences responses.

Terminologies

- Business analytics: The process of using statistical analysis and modeling to drive business decisions.
- Case: A case is an individual about whom or which we have data.
- Cross-sectional data: Data taken from situations that vary over time but measured at a single time instant is said to be a cross-section of the time series.
- Context: The context ideally tells who was measured, what was measured, how the data were collected, where the data were collected, and when and why the study was performed.
- Categorical (or qualitative) variable: A variable that names categories (whether with words or numerals) is called categorical or qualitative.
- Data: Recorded values whether numbers or labels, together with their context.

Terminologies

- Data mining: The process of using a variety of statistical tools to analyze large data bases or data warehouses.
- Data table: An arrangement of data in which each row represents a case and each column represents a variable.
- Data warehouse: A large data base of information collected by a company or other organization usually to record transactions that the organization makes, but also used for analysis via data mining.
- Experimental unit: An individual in a study for which or for whom data values are recorded. Human experimental units are usually called subjects or participants.
- Identifier variable: A categorical variable that records a unique value for each case, used to name or identify it.

Terminologies

- Metadata: Auxiliary information about variables in a database, typically including how, when, and where (and possibly why) the data were collected; who each case represents; and the definitions of all the variables.
- Nominal variable: The term “nominal” can be applied to a variable whose values are used only to name categories.
- Ordinal variable: The term “ordinal” can be applied to a variable whose categorical values possess some kind of order.
- Participant: A human experimental unit. Also called a subject.
- Quantitative variable: A variable in which the numbers are values of measured quantities with units. Record Information about an individual in a database.
- Relational database: A relational database stores and retrieves information. Within the database, information is kept in data tables that can be “related” to each other.

Terminologies

- Respondent: Someone who answers, or responds to, a survey.
- Spreadsheet: A spreadsheet is layout designed for accounting that is often used to store and manage data tables. Excel is a common example of a spreadsheet program.
- Subject: A human experimental unit. Also called a participant.
- Time series: Data measured over time. Usually the time intervals are equally spaced or regularly spaced (e.g., every week, every quarter, or every year).
- Transactional Data: Data collected to record the individual transactions of a company or organization.
- Units: A quantity or amount adopted as a standard of measurement, such as dollars, hours, or grams.
- Variable: A variable holds information about the same characteristic for many cases.

Summary

- Identify whether a variable is being used as categorical or quantitative.
- Categorical variables identify a category for each case. Usually we think about the counts of cases that fall in each category. (An exception is an identifier variable that just names each case.)
- Quantitative variables record measurements or amounts of something; they must have units.
- Sometimes we may treat the same variable as categorical or quantitative depending on what we want to learn from it, which means some variables can’t be pigeonholed as one type or the other.
- Consider the source of your data and the reasons the data were collected. That can help you understand what you might be able to learn from the data.

Summary

- Understand that data are values, whether numerical or labels, together with their context.
- who, what, why, where, when (and how)—the W's—help nail down the context of the data.
- We must know who, what, and why to be able to say anything useful based on the data. The who are the cases. The what are the variables. A variable gives information about each of the cases. The why helps us decide which way to treat the variables.
- Stop and identify the W's whenever you have data, and be sure you can identify the cases and the variables.

Task

CATEGORY	NAME
HOME & KITCHEN	P BHAVYA
MOBILE & ACCESSORIES	ANNA LOKESH HARSHIT
PERSONAL HEALTH, GROOMING & WELLNESS	SRAVAN KUMAR BALIJEPALLI
ELECTRONICS & ACCESSORIES	BEZAWADA MURALI KRISHNA
COMPUTER & ACCESSORIES	EBBILI MANI GURU SAI
TV & APPLIANCES	VEMULA BHANU KIRAN
WOMEN’S FASHION	DIRISALA ROHITH REDDY
MEN’S FASHION	CYRIL AMBEDKAR KONDRU
KID’S FASHION	IRAGAVARAPU SRI RANGA RAMANUJA
SPORTS & FITNESS	YELLETI TEJESWARA RAO

Module No. 2

Probability

Sample Space

- Statistician is often dealing with either numerical data, representing counts or measurements, or categorical data, which can be classified according to some criterion. We shall refer to any recording of information, whether it be numerical or categorical, as an **observation**.
- Statisticians use the word experiment to describe any process that generates a set of data. A simple example of a statistical experiment is the tossing of a coin. In this experiment, there are only two possible outcomes, heads or tails.
- Another experiment might be the launching of a missile and observing of its velocity at specified times.
- The opinions of voters concerning a new sales tax can also be considered as observations of an experiment.
- The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.

Sample Space

- Each outcome in a sample space is called an element or a member of the sample space, or simply a sample point. If the sample space has a finite number of elements, we may list the members separated by commas and enclosed in braces.
- The sample space S, of possible outcomes when a coin is flipped, may be written

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

where H and T correspond to heads and tails, respectively.

- Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

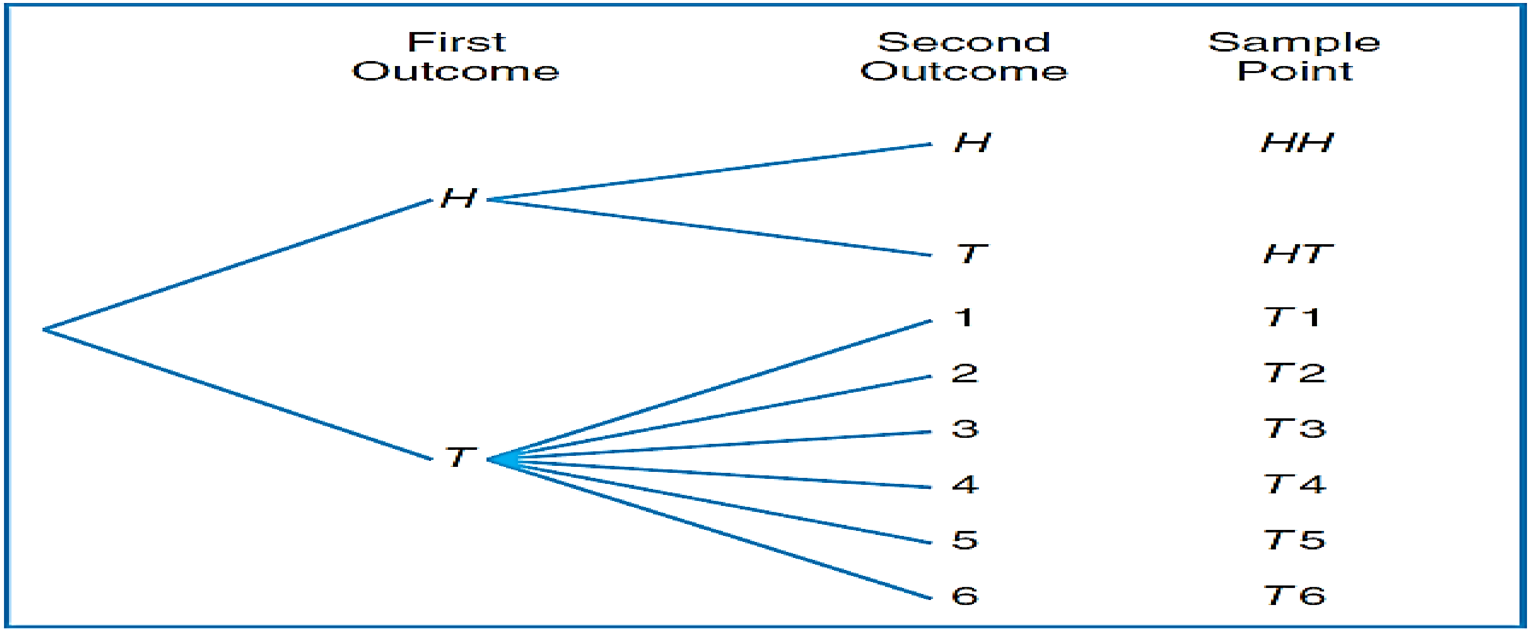
$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

- If we are interested only in whether the number is even or odd, the sample space is simply

$$S_2 = \{\text{even}, \text{odd}\}.$$

Sample Space

- In some experiments, it is helpful to list the elements of the sample space systematically by means of a tree diagram.
- An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once.



Here the sample space is $\boldsymbol{S = \{HH, HT, T1, T2, T3, T4, T5, T6\}}$.

Event

- An **event** is a subset of a sample space.
- Given the sample space $S = \{t|t > 0\}$, where t is the life in years of a certain electronic component, then the event A that the component fails before the end of the fifth year is the subset

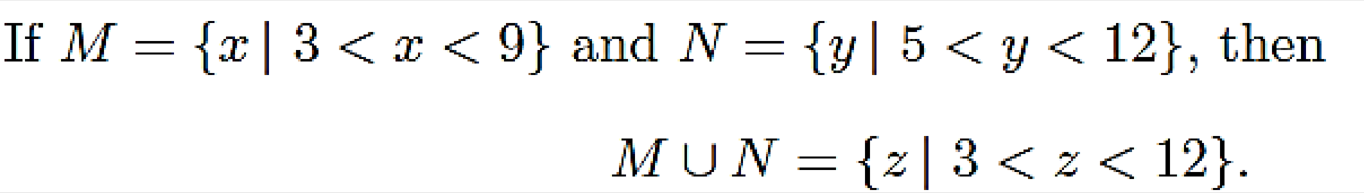
$$\boldsymbol{A = \{t|0 \leq t \leq 5\}}.$$

- It is conceivable that an event may be a subset that includes the entire sample space S or a subset of S called the null set and denoted by the symbol ϕ, which contains no elements at all.
- The **complement of an event *A*** with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.
- Consider the sample space $S = \{\text{book, cell phone, mp3, paper, stationery, laptop}\}$.

Let $A = \{\text{book, stationery, laptop, paper}\}$. Then the complement of A is $A' = \{\text{cell phone, mp3}\}$.

Event

- The intersection of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.
- Two events A and B are mutually exclusive, or disjoint, if $A \cap B = \phi$, that is, if A and B have no elements in common.
- The union of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.



If $M = \{x \mid 3 < x < 9\}$ and $N = \{y \mid 5 < y < 12\}$, then

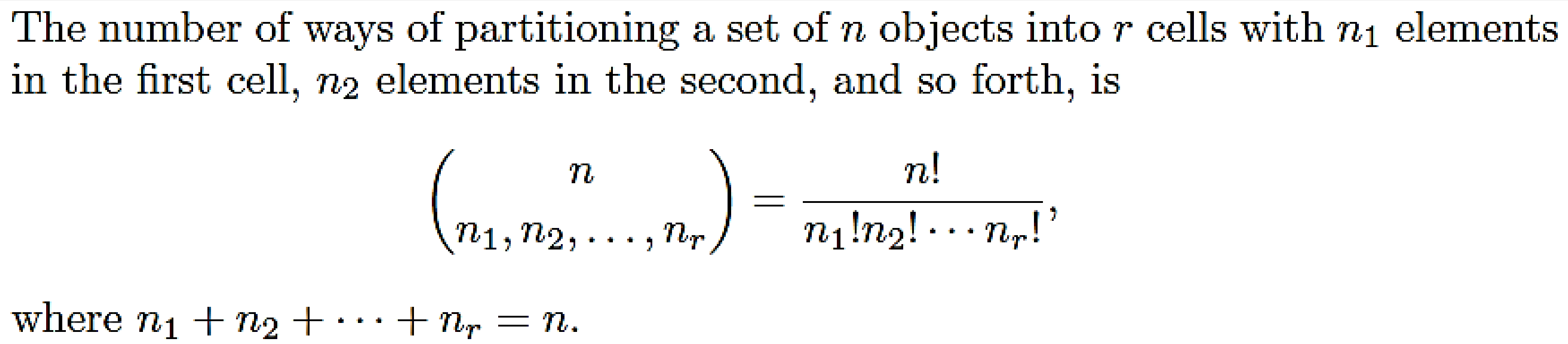
$$M \cup N = \{z \mid 3 < z < 12\}.$$

Counting Sample Points

- If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.
- If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.
- A permutation is an arrangement of all or part of a set of objects.
- The number of permutations of n objects is $n!$.
- The number of permutations of n distinct objects taken r at a time is $n_{P_r} = \frac{n!}{(n-r)!}$.
- The number of permutations of n objects arranged in a circle is $(n - 1)!$.

Counting Sample Points

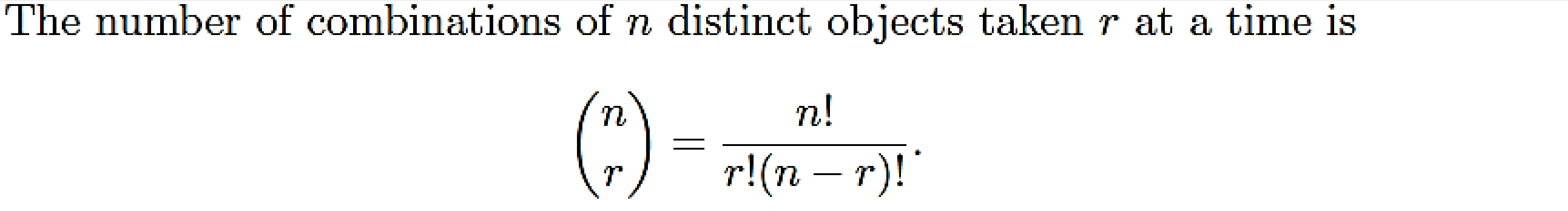
- The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, ..., n_k of a kth kind is $\dfrac{n!}{n_1!n_2!\cdots n_k!}$.



The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!},$$

where $n_1 + n_2 + \cdots + n_r = n$.



The number of combinations of n distinct objects taken r at a time is

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

Probability of an Event

The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

Furthermore, if A_1, A_2, A_3, ... is a sequence of mutually exclusive events, then

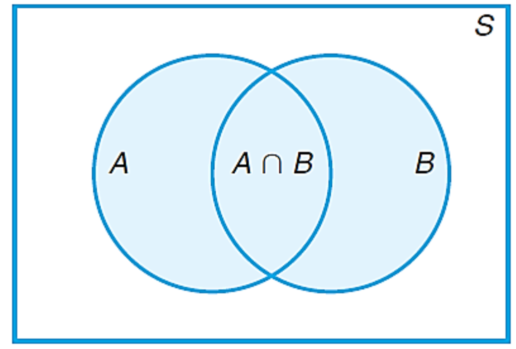
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots .$$

If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}.$$

If A and B are two events, then

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



Probability of an Event

If A and B are mutually exclusive, then

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

If $A_1,\ A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

If $A_1,\ A_2, \ldots, A_n$ is a partition of sample space S, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(S) = 1.$$

For three events A, B, and C,

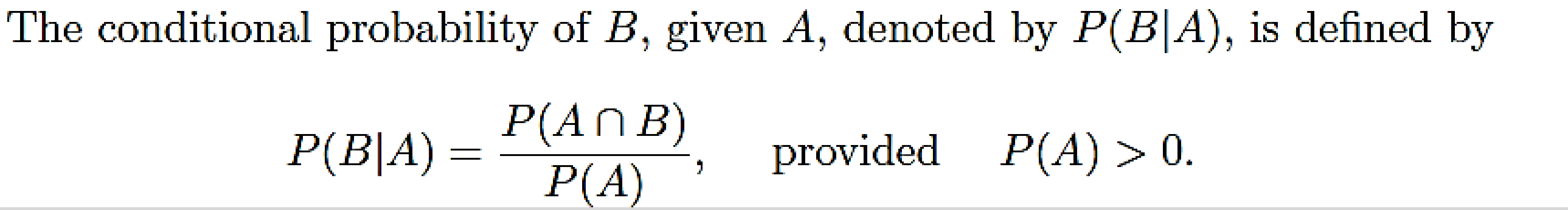
$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\qquad - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

If A and A' are complementary events, then

$$P(A) + P(A') = 1.$$

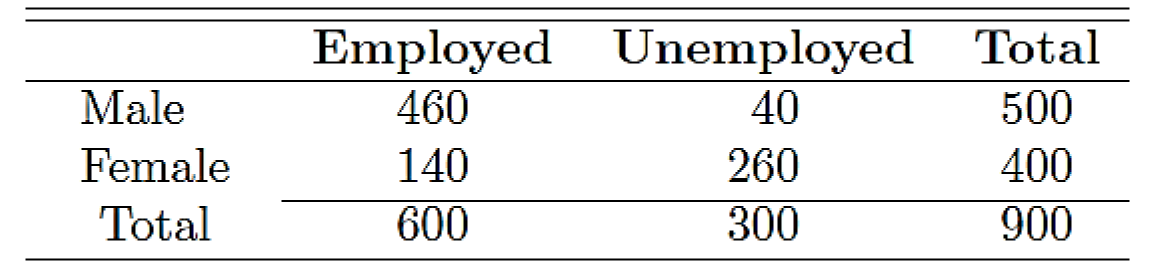
Conditional Probability, Independence, and the Product Rule

The probability of an event B occurring when it is known that some event A has occurred is called a conditional probability and is denoted by $P(B|A)$. The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

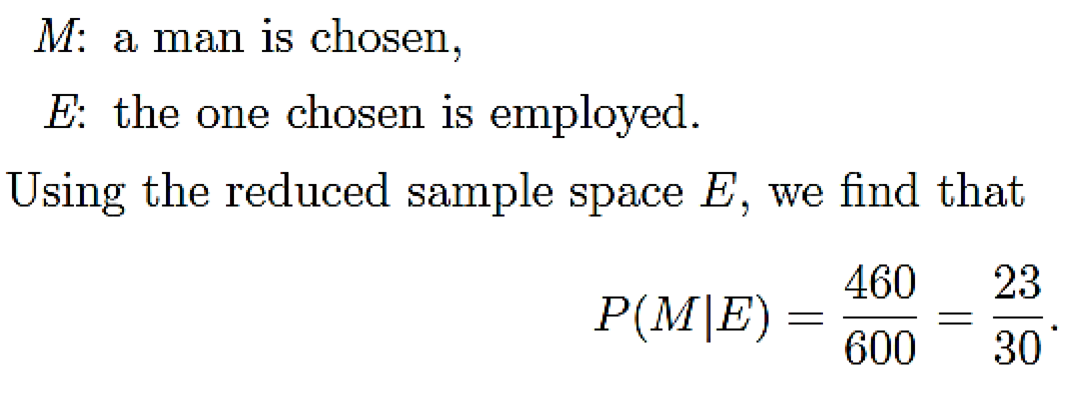


The conditional probability of B, given A, denoted by $P(B|A)$, is defined by

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



	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

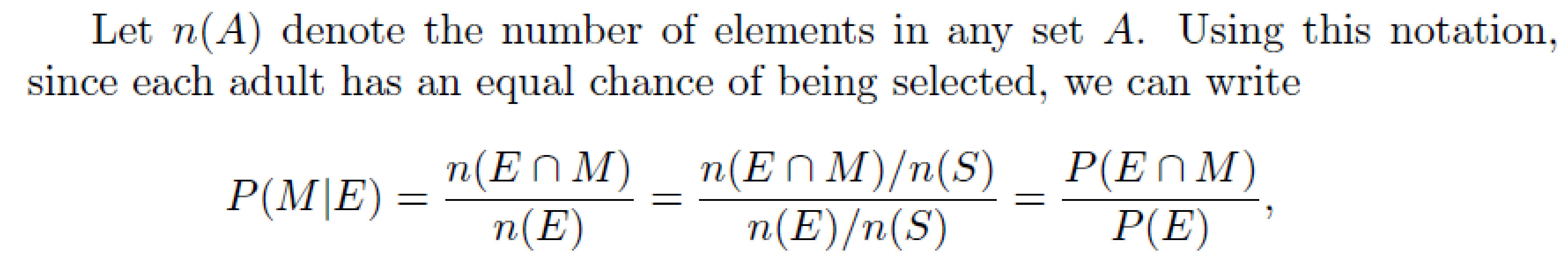


M: a man is chosen,

E: the one chosen is employed.

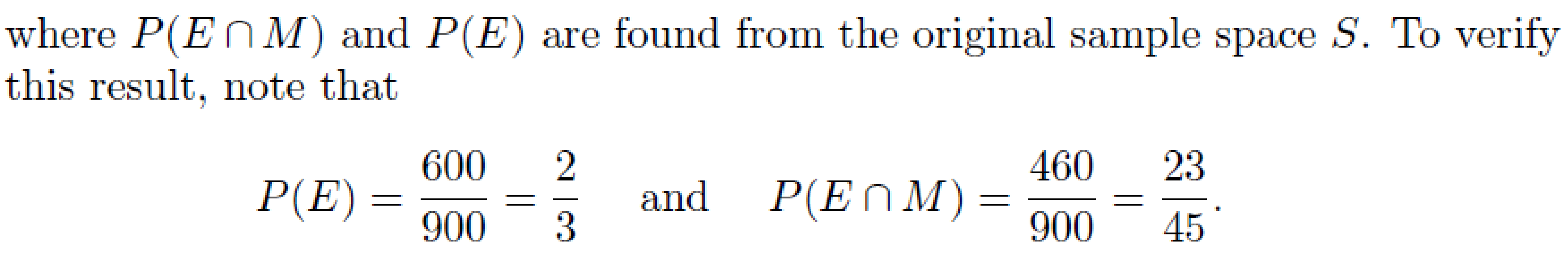
Using the reduced sample space E, we find that

$$P(M|E) = \frac{460}{600} = \frac{23}{30}.$$



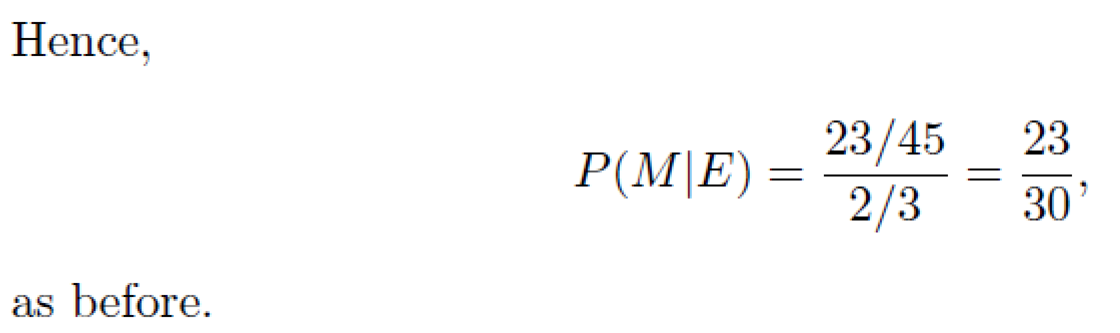
Let $n(A)$ denote the number of elements in any set A. Using this notation, since each adult has an equal chance of being selected, we can write

$$P(M|E) = \frac{n(E \cap M)}{n(E)} = \frac{n(E \cap M)/n(S)}{n(E)/n(S)} = \frac{P(E \cap M)}{P(E)},$$



where $P(E \cap M)$ and $P(E)$ are found from the original sample space S. To verify this result, note that

$$P(E) = \frac{600}{900} = \frac{2}{3} \quad \text{and} \quad P(E \cap M) = \frac{460}{900} = \frac{23}{45}.$$



Hence,

$$P(M|E) = \frac{23/45}{2/3} = \frac{23}{30},$$

as before.

Question

The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane (a) arrives on time, given that it departed on time, and (b) departed on time, given that it has arrived on time.

Answer

(a) 0.94

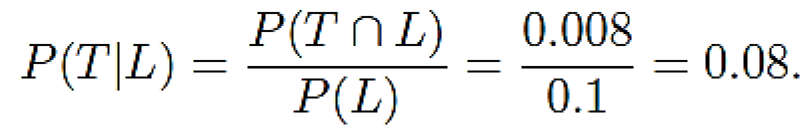
(b) 0.95

Question

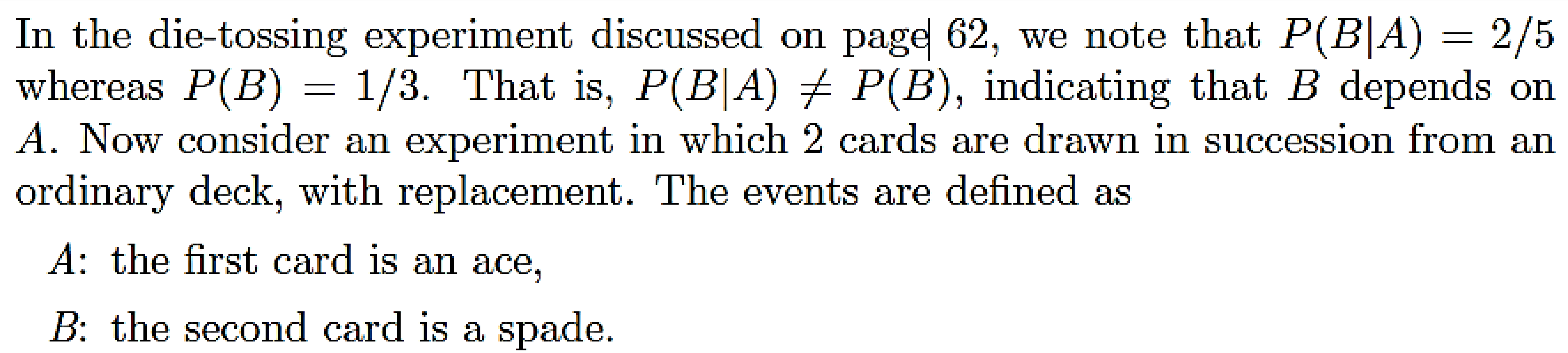
The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Answer

Consider the events L: length defective, T: texture defective. Given that the strip is length defective, the probability that this strip is texture defective is given by



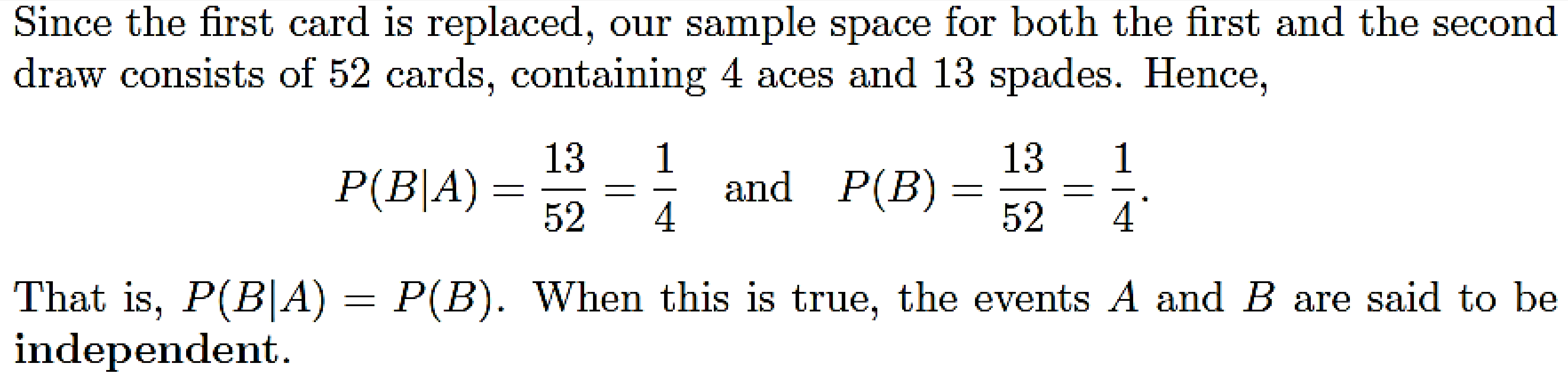
$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08.$$



In the die-tossing experiment discussed on page 62, we note that $P(B|A) = 2/5$ whereas $P(B) = 1/3$. That is, $P(B|A) \neq P(B)$, indicating that B depends on A. Now consider an experiment in which 2 cards are drawn in succession from an ordinary deck, with replacement. The events are defined as

A: the first card is an ace,

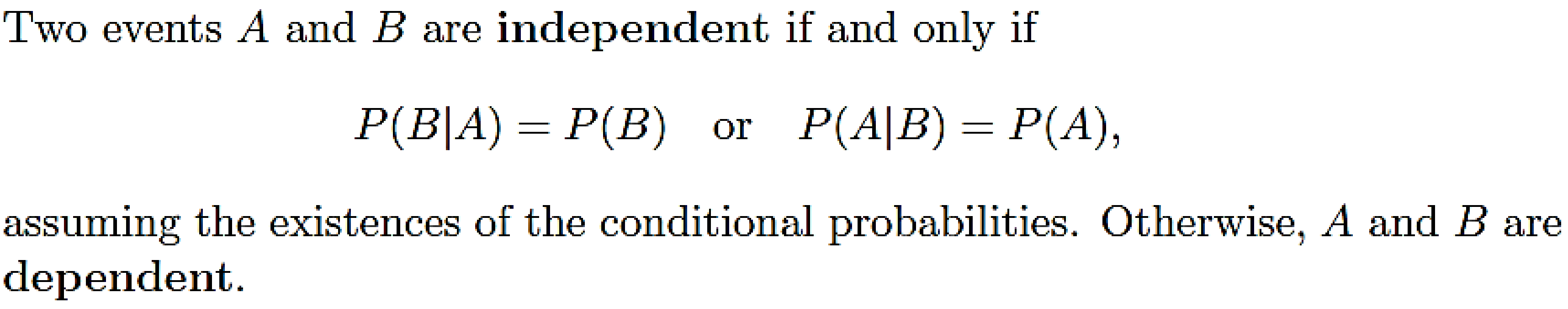
B: the second card is a spade.



Since the first card is replaced, our sample space for both the first and the second draw consists of 52 cards, containing 4 aces and 13 spades. Hence,

$$P(B|A) = \frac{13}{52} = \frac{1}{4} \quad \text{and} \quad P(B) = \frac{13}{52} = \frac{1}{4}.$$

That is, $P(B|A) = P(B)$. When this is true, the events A and B are said to be **independent**.



Two events A and B are **independent** if and only if

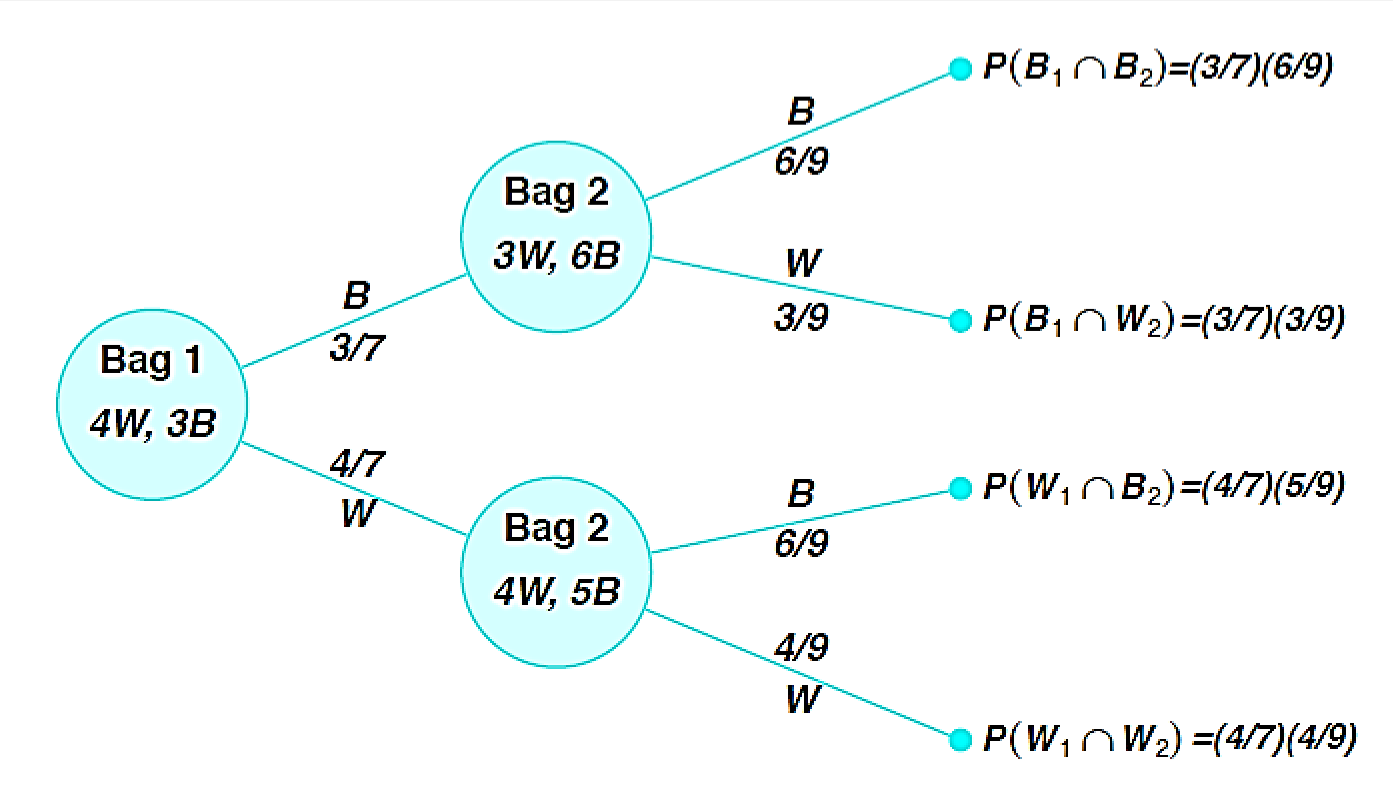
$$P(B|A) = P(B) \quad \text{or} \quad P(A|B) = P(A),$$

assuming the existences of the conditional probabilities. Otherwise, A and B are **dependent**.

If in an experiment the events A and B can both occur, then

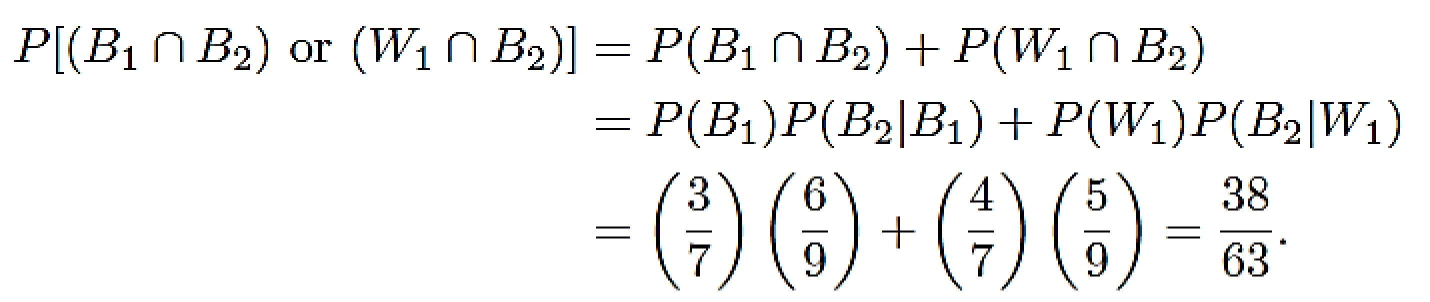
$$P(A \cap B) = P(A)P(B|A), \text{ provided } P(A) > 0.$$

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

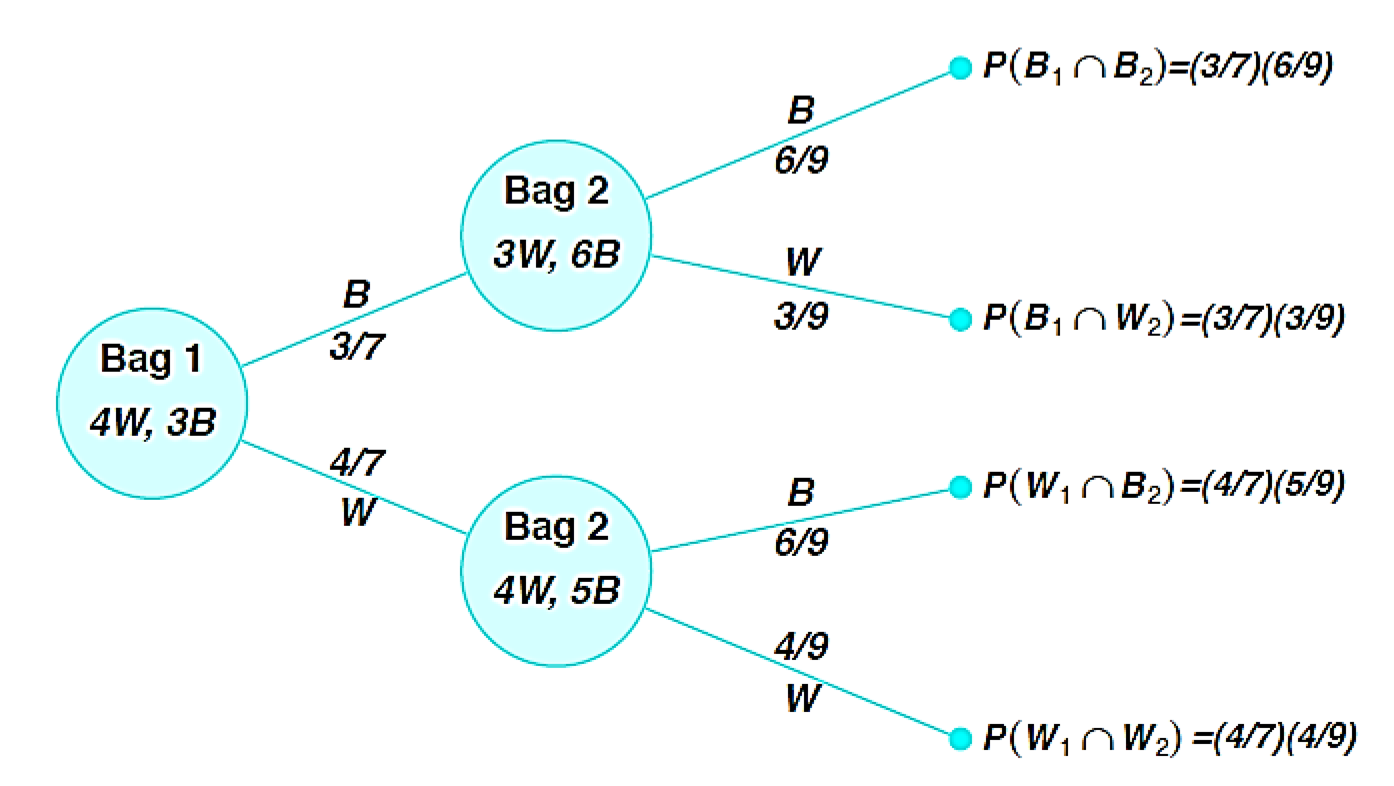


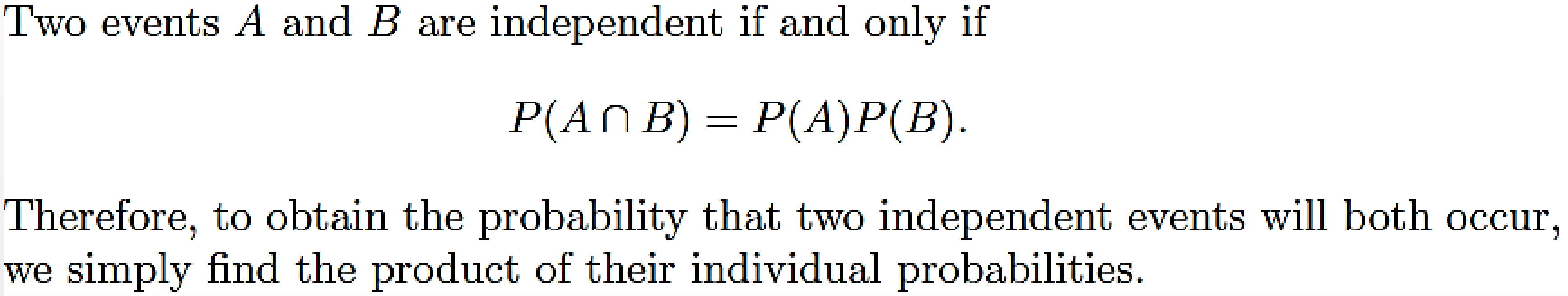
Bag 1
4W, 3B
B
3/7
4/7
W
Bag 2
3W, 6B
Bag 2
4W, 5B
B
6/9
W
3/9
B
6/9
4/9
W
$P(B_1 \cap B_2)$=(3/7)(6/9)
$P(B_1 \cap W_2)$=(3/7)(3/9)
$P(W_1 \cap B_2)$=(4/7)(5/9)
$P(W_1 \cap W_2)$ =(4/7)(4/9)

Let B_1, B_2, and W_1 represent, respectively, the drawing of a black ball from bag 1, a black ball from bag 2, and a white ball from bag 1.



$$\begin{aligned} P[(B_1 \cap B_2) \text{ or } (W_1 \cap B_2)] &= P(B_1 \cap B_2) + P(W_1 \cap B_2) \\ &= P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\ &= \left(\frac{3}{7}\right)\left(\frac{6}{9}\right) + \left(\frac{4}{7}\right)\left(\frac{5}{9}\right) = \frac{38}{63}. \end{aligned}$$

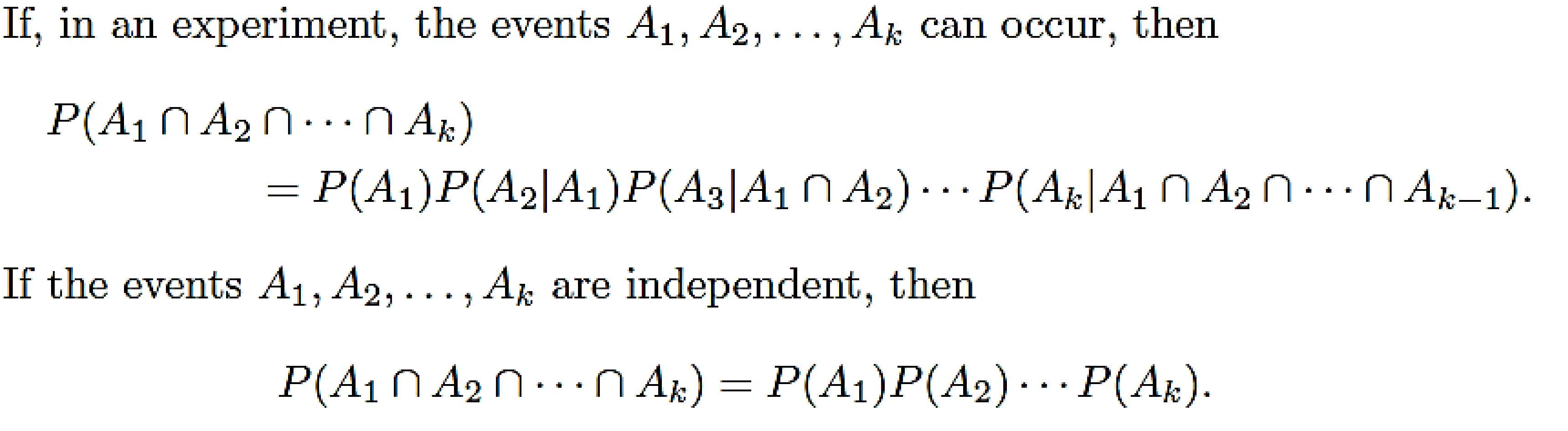

Bag 1
4W, 3B
B
3/7
4/7
W
Bag 2
3W, 6B
Bag 2
4W, 5B
B
6/9
W
3/9
B
6/9
4/9
W
$P(B_1 \cap B_2)$=(3/7)(6/9)
$P(B_1 \cap W_2)$=(3/7)(3/9)
$P(W_1 \cap B_2)$=(4/7)(5/9)
$P(W_1 \cap W_2)$ =(4/7)(4/9)



Two events A and B are independent if and only if

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

Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

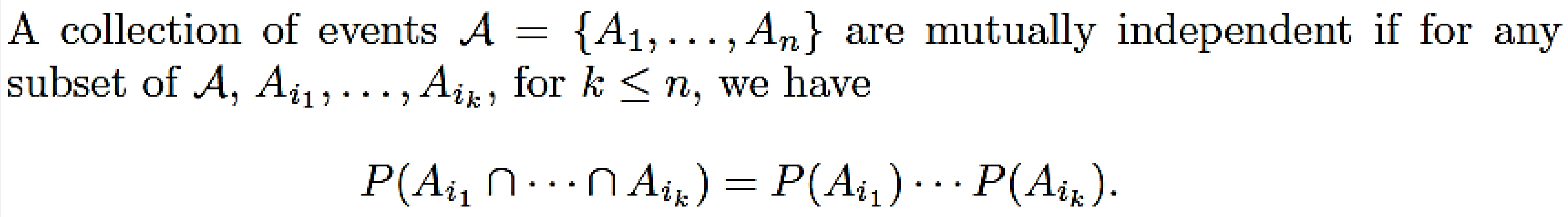


If, in an experiment, the events $A_1, A_2, \ldots, A_k$ can occur, then

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_k) \\ &\qquad = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}). \end{aligned}$$

If the events $A_1, A_2, \ldots, A_k$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k).$$



A collection of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ are mutually independent if for any subset of $\mathcal{A}$, $A_{i_1}, \ldots, A_{i_k}$, for $k \leq n$, we have

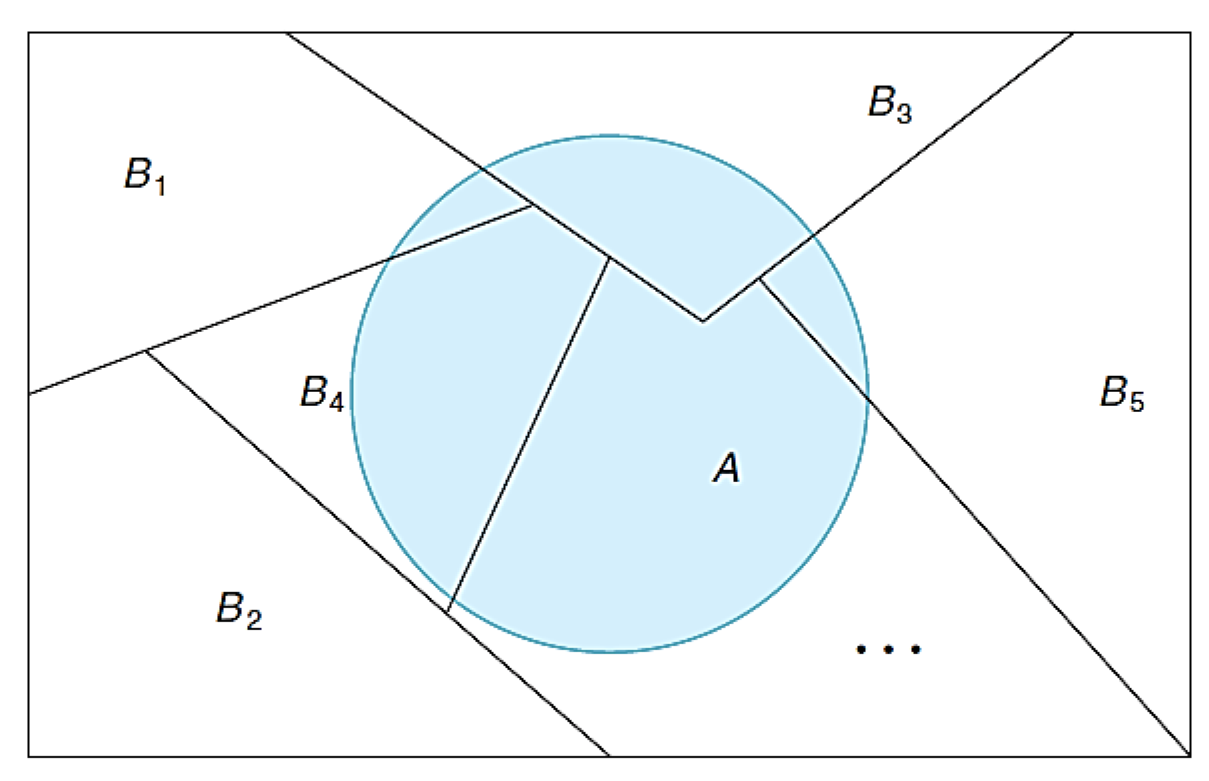
$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

Bayes' Rule

Theorem of total probability

If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A of S,

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



Bayes' Rule

Question

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Answer

Consider the following events:

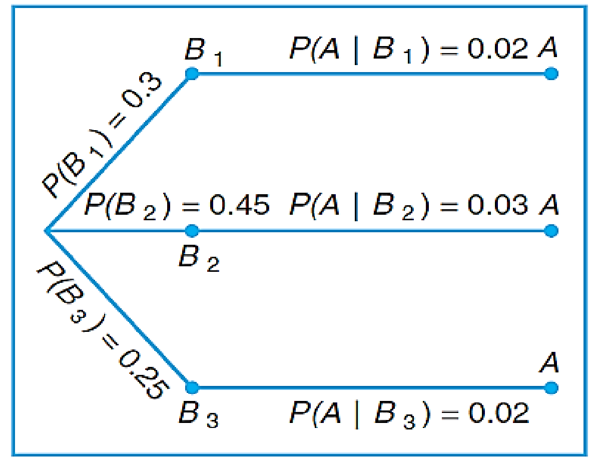
A: the product is defective,

B_1: the product is made by machine B1,

B_2: the product is made by machine B2,

B_3: the product is made by machine B3.

Applying the rule of elimination, we can write

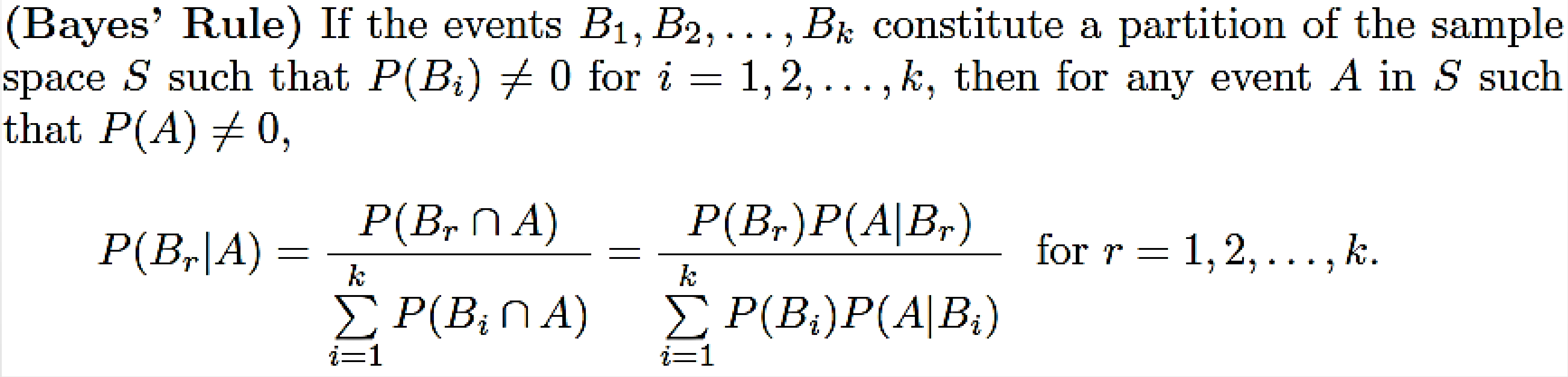


$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

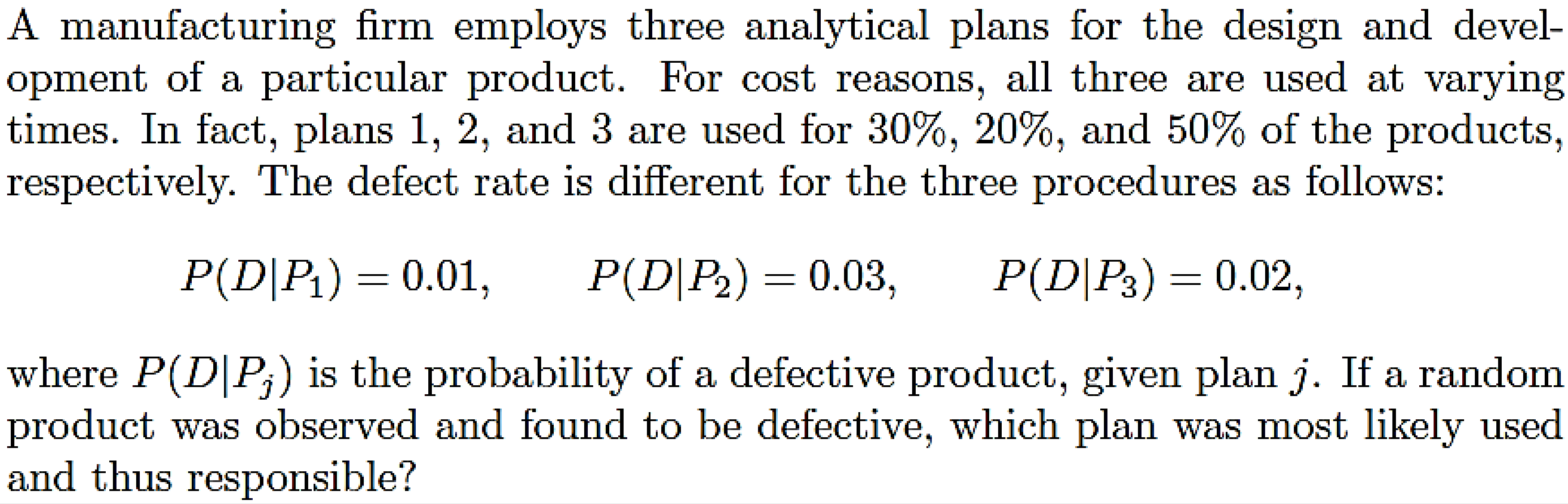
Bayes' Rule



(Bayes' Rule) If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum\limits_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum\limits_{i=1}^{k} P(B_i)P(A|B_i)} \quad \text{for } r = 1, 2, \ldots, k.$$

Question



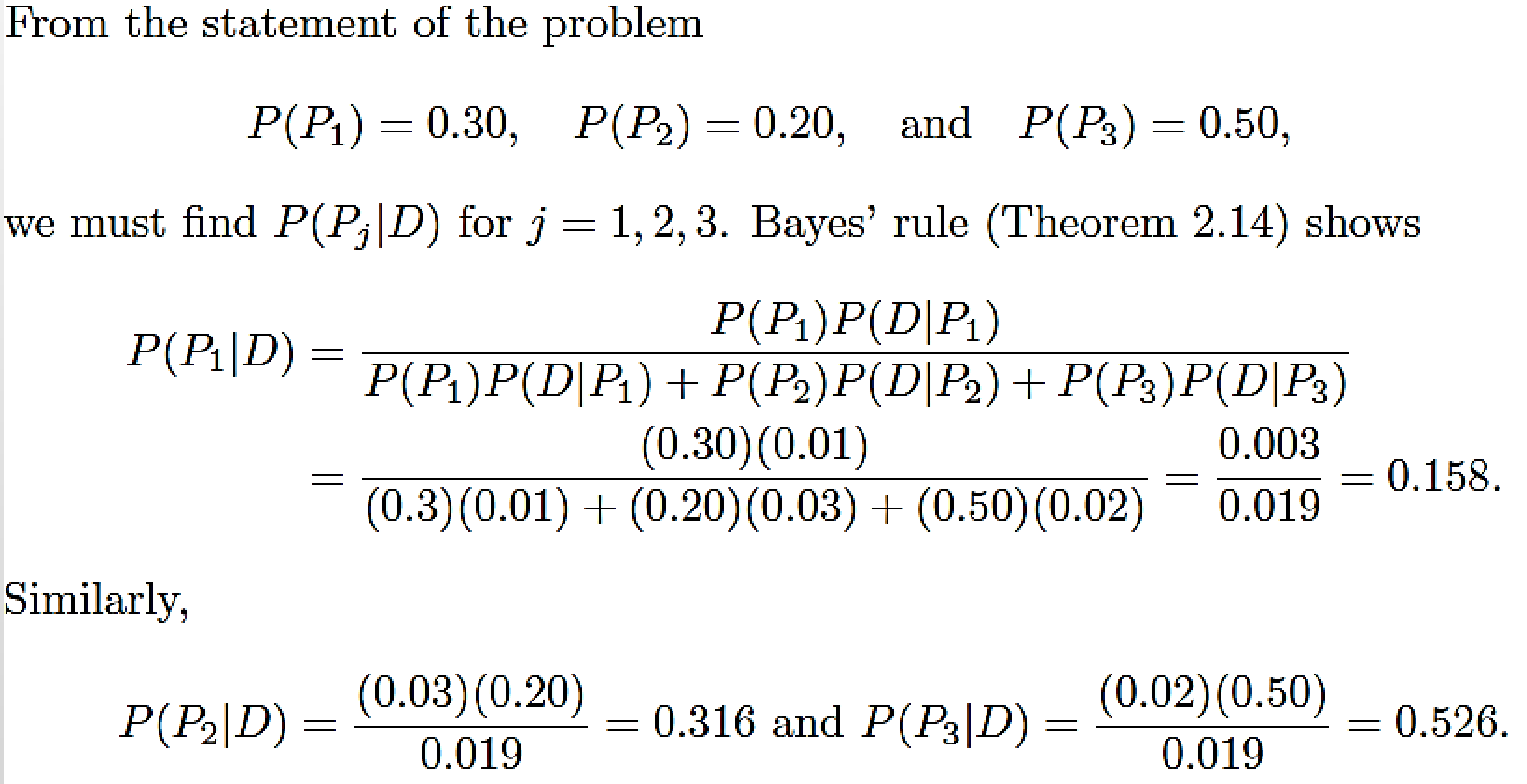
A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$P(D|P_1) = 0.01, \qquad P(D|P_2) = 0.03, \qquad P(D|P_3) = 0.02,$$

where $P(D|P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

Bayes' Rule

Answer



From the statement of the problem

$$P(P_1) = 0.30, \quad P(P_2) = 0.20, \quad \text{and} \quad P(P_3) = 0.50,$$

we must find $P(P_j|D)$ for $j = 1, 2, 3$. Bayes' rule (Theorem 2.14) shows

$$\begin{aligned} P(P_1|D) &= \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)} \\ &= \frac{(0.30)(0.01)}{(0.3)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \frac{0.003}{0.019} = 0.158. \end{aligned}$$

Similarly,

$$P(P_2|D) = \frac{(0.03)(0.20)}{0.019} = 0.316 \text{ and } P(P_3|D) = \frac{(0.02)(0.50)}{0.019} = 0.526.$$

The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3.

Module No. 3

Modelling with Probability

Random Variable

A random variable is a function that associates a real number with each element in the sample space.

Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Random Variable

Question

A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

Answer

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

Random Variable

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a discrete sample space.

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space.

The set of ordered pairs $(x, f(x))$ is a probability function, probability mass function, or probability distribution of the discrete random variable X, if, for each possible outcome x,

(i) $f(x) \geq 0$,

(ii) $\sum_x f(x) = 1$,

(iii)$P(X = x) = f(x)$.

Random Variable

The cumulative distribution function $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \qquad \text{for } -\infty < x < \infty$$

The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

(i) $f(x) \geq 0$, for all $x \in R$,

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

(iii)$P(a < X < b) = \int_{a}^{b} f(x) dx.$

Random Variable

The cumulative distribution function $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \qquad \text{for } -\infty < x < \infty$$

Expected value of Random variable

Let X be a random variable with probability distribution $f(x)$. The mean, or expected value of X is

$$\mu = E(X) = \sum_{x} xf(x)$$

if X is discrete, and

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

if X is continuous.

Question

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Expected value of Random variable

Answer

Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 1, 2, 3.$$

$f(0) = \frac{1}{35}$, $f(1) = \frac{12}{35}$, $f(2) = \frac{18}{35}$, and $f(3) = \frac{4}{35}$.

$$\mu = E(X) = 0 \times \left(\frac{1}{35}\right) + 1 \times \left(\frac{12}{35}\right) + 2 \times \left(\frac{18}{35}\right) + 3 \times \left(\frac{4}{35}\right)$$

$$= \frac{12}{7} = 1.7.$$

Expected value of Random variable

Question

Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \dfrac{20000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

Answer

$$\boldsymbol{\mu = E(X) = \int_{100}^{\infty} x \frac{20000}{x^3} dx = \int_{100}^{\infty} \frac{20000}{x^2} dx = 200}$$

Variance and Covariance of Random Variables

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

Means and Variances of Linear Combinations of Random Variables

If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

Setting $a = 0$, we see that $E(b) = b$.

Setting $b = 0$, we see that $E(aX) = aE(X)$.

Binomial and Multinomial Distributions

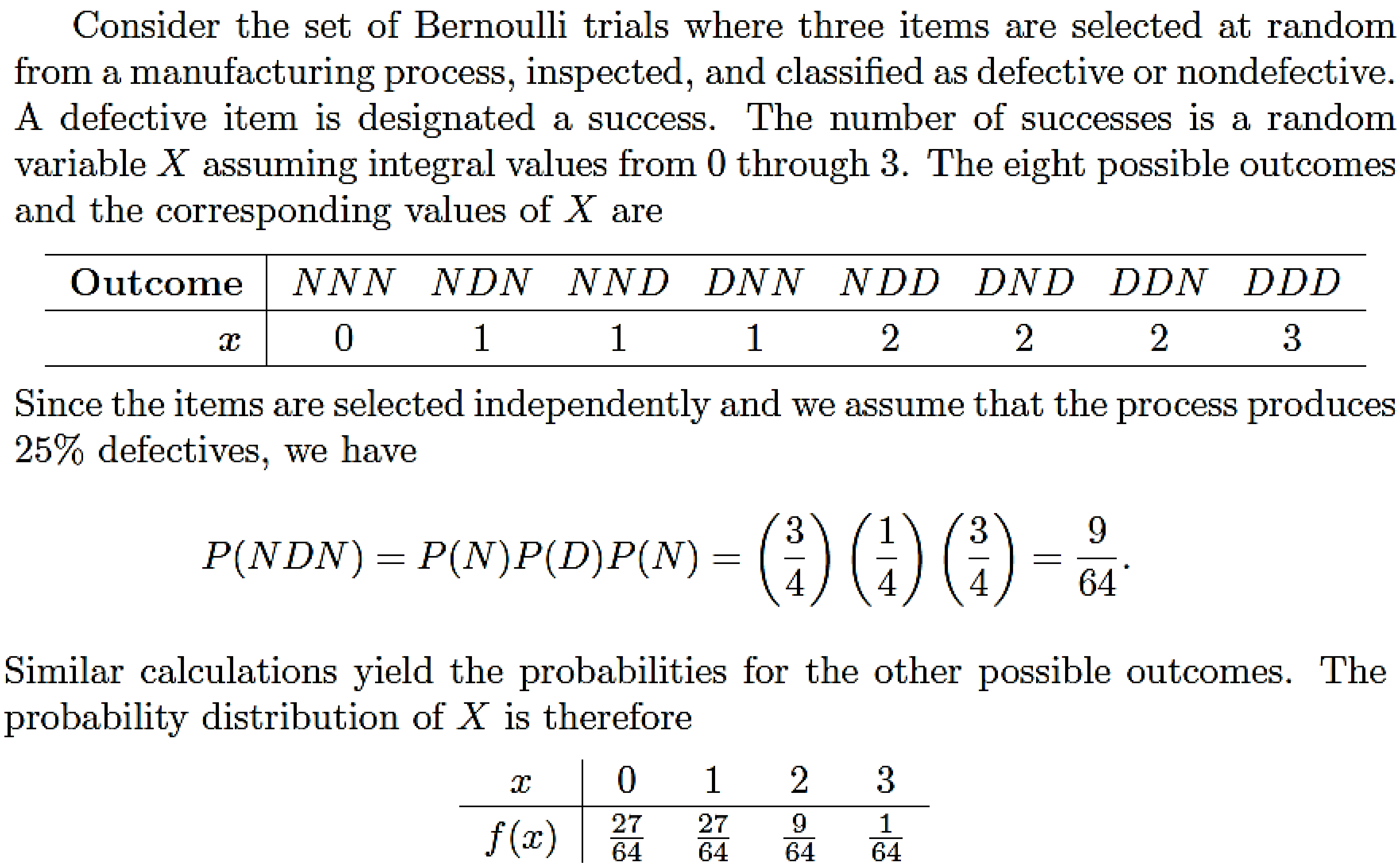
An experiment often consists of repeated trials, each with two possible outcomes that may be labelled success or failure. The most obvious application deals with the testing of items as they come off an assembly line, where each trial may indicate a defective or a non-defective item. We may choose to define either outcome as a success. The process is referred to as a **Bernoulli process**. Each trial is called a **Bernoulli trial**.

The Bernoulli Process

The Bernoulli process must possess the following properties:

1. The experiment consists of repeated trials.
2. Each trial results in an outcome that may be classified as a success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials are independent.

Binomial and Multinomial Distributions



Consider the set of Bernoulli trials where three items are selected at random from a manufacturing process, inspected, and classified as defective or nondefective. A defective item is designated a success. The number of successes is a random variable X assuming integral values from 0 through 3. The eight possible outcomes and the corresponding values of X are

Outcome	NNN	NDN	NND	DNN	NDD	DND	DDN	DDD
x	0	1	1	1	2	2	2	3

Since the items are selected independently and we assume that the process produces 25% defectives, we have

$$P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}.$$

Similar calculations yield the probabilities for the other possible outcomes. The probability distribution of X is therefore

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

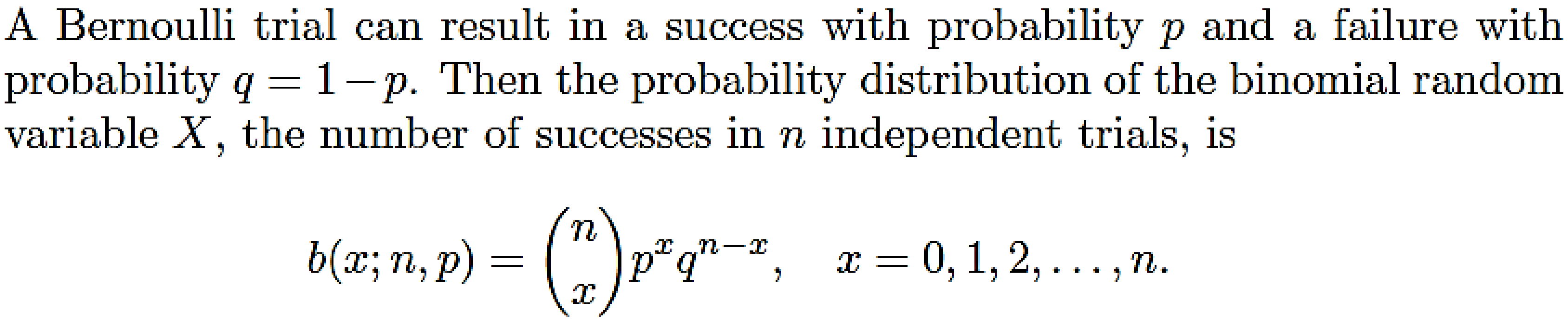
Binomial and Multinomial Distributions

Binomial Distribution

The number X of successes in n Bernoulli trials is called a **binomial random variable.** The probability distribution of this discrete random variable is called the **binomial distribution,** and its values will be denoted by $b(x; n, p)$ since they depend on the number of trials and the probability of a success on a given trial.

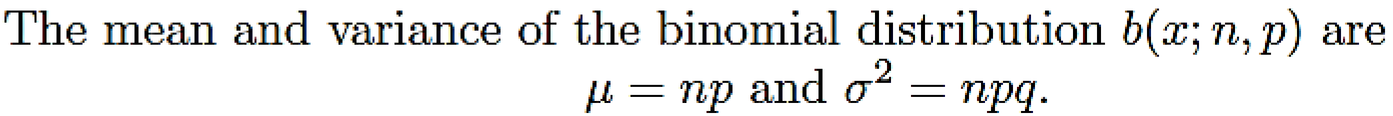
Let us now generalize the above illustration to yield a formula for $b(x; n, p)$. That is, we wish to find a formula that gives the probability of x successes in n trials for a binomial experiment. First, consider the probability of x successes and $n - x$ failures in a specified order. Since the trials are independent, we can multiply all the probabilities corresponding to the different outcomes. Each success occurs with probability p and each failure with probability $q = 1 - p$. Therefore, the probability for the specified order is $p^x q^{n-x}$. We must now determine the total number of sample points in the experiment that have x successes and $n - x$ failures. This number is equal to the number of partitions of n outcomes into two groups with x in one group and $n - x$ in the other and is written $\binom{n}{x}$ as introduced in Section 2.3. Because these partitions are mutually exclusive, we add the probabilities of all the different partitions to obtain the general formula, or simply multiply $p^x q^{n-x}$ by $\binom{n}{x}$.

Binomial and Multinomial Distributions



A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1-p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$



The mean and variance of the binomial distribution $b(x; n, p)$ are

$$\mu = np \text{ and } \sigma^2 = npq.$$

Binomial and Multinomial Distributions

If the chance that any one of 5 telephone lines is busy at any instant is 0.01, what is the probability that all the lines are busy? What is the probability that more than 3 lines are busy?

Binomial and Multinomial Distributions

The average percentage of failures in a certain examination is 40. What is the probability that out of a group of 6 candidates, at least 4 pass in the examination?

Poisson Distribution

- Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments**.
- The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. For example, a Poisson experiment can generate observations for the random variable X representing the number of cars passing through the Mandadam cross per hour, the number of days college is closed due to heavy rain in the rainy season.
- The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances, X might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page.
- A Poisson experiment is derived from the Poisson process and possesses the following properties.

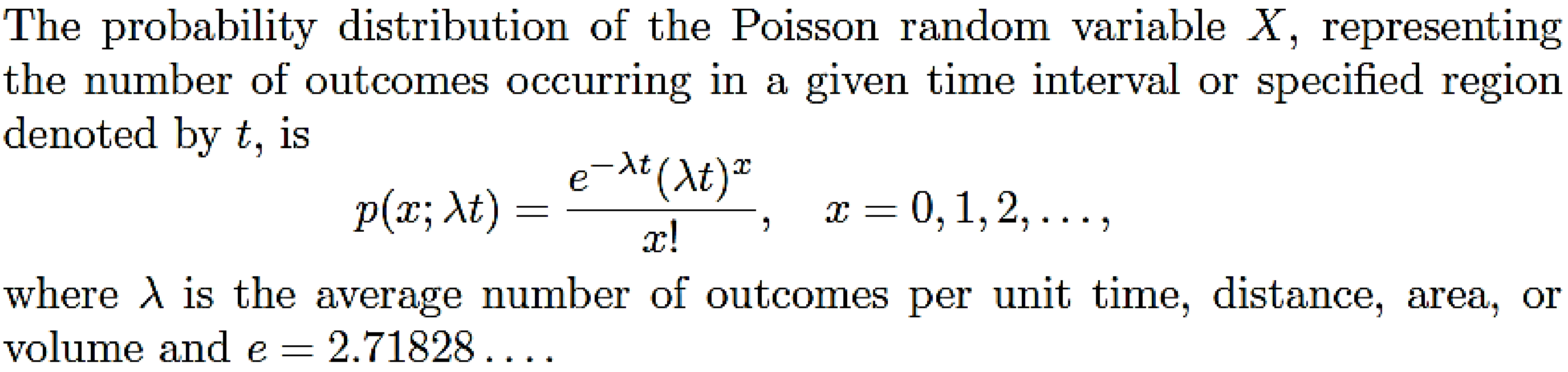
Properties of the Poisson Process

- The number of outcomes occurring in one time interval or specified region of space is independent of the number that occur in any other disjoint time interval or region. **In this sense we say that the Poisson process has no memory**.
- The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
- The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

Poisson Process

- The number X of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson distribution**.
- The mean number of outcomes is computed from $\mu = \lambda t$, where t is the specific “time,” “distance,” “area,” or “volume” of interest.
- Since the probabilities depend on λ, the rate of occurrence of outcomes, we shall denote them by $p(x; \lambda t)$.

Poisson Process



The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where λ is the average number of outcomes per unit time, distance, area, or volume and $e = 2.71828\ldots$.

Table A.2 contains Poisson probability sums,

$$P(r;\lambda t) = \sum_{x=0}^{r} p(x;\lambda t)$$

for selected values of λt ranging from 0.1 to 18.0.

Next. Table A.2 is presented for illustration.

Table A.2 Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ								
r	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865
4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977
5				1.0000	1.0000	1.0000	0.9999	0.9998	0.9997
6							1.0000	1.0000	1.0000

	μ								
r	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8		1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9			1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10				0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11				1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12					1.0000	0.9999	0.9997	0.9992	0.9980
13						1.0000	0.9999	0.9997	0.9993
14							1.0000	0.9999	0.9998
15								1.0000	0.9999
16									1.0000

Table A.2 (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ								
r	**5.5**	**6.0**	**6.5**	**7.0**	**7.5**	**8.0**	**8.5**	**9.0**	**9.5**
0	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885
6	0.6860	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649
7	0.8095	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687
8	0.8944	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918
9	0.9462	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218
10	0.9747	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453
11	0.9890	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520
12	0.9955	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364
13	0.9983	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981
14	0.9994	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400
15	0.9998	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665
16	0.9999	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823
17	1.0000	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911
18		1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957
19			1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980
20						0.9999	0.9998	0.9996	0.9991
21						1.0000	0.9999	0.9998	0.9996
22							1.0000	0.9999	0.9999
23								1.0000	0.9999
24									1.0000

Table A.2 (continued) Poisson Probability Sums $\sum_{x=0}^{r} p(x;\mu)$

	μ								
r	**10.0**	**11.0**	**12.0**	**13.0**	**14.0**	**15.0**	**16.0**	**17.0**	**18.0**
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
3	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317
25		0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554
26		1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718
27			0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827
28			1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897
29				1.0000	0.9999	0.9996	0.9989	0.9973	0.9941
30					0.9999	0.9998	0.9994	0.9986	0.9967
31					1.0000	0.9999	0.9997	0.9993	0.9982
32						1.0000	0.9999	0.9996	0.9990
33							0.9999	0.9998	0.9995
34							1.0000	0.9999	0.9998
35								1.0000	0.9999
36									0.9999
37									1.0000

Example Problem on Poisson Process

Q. During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Ans. Using the Poisson distribution with $x = 6$ and $\lambda t = 4$ and referring to Table A.2, we have

$$p(6;4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042.$$

Example Problem of Poisson Process

Q. Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

Ans. Let X be the number of tankers arriving each day. Then, referring to Table A.2, we have

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513 = 0.0487.$$

Note:

Both the *mean* and the *variance* of the Poisson distribution $p(x;\ \lambda t)$ are $\boldsymbol{\lambda t}$.

Relationship between Binomial and Poisson Distributions

Let X be a binomial random variable with probability distribution $b(x;n,p)$. When $n \to \infty$, $p \to 0$, and $np \overset{n\to\infty}{\longrightarrow} \mu$ remains constant,

$$b(x;n,p) \overset{n\to\infty}{\longrightarrow} p(x;\mu).$$

Q. In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on one day?

(b) What is the probability that there are at most three days with an accident?

Ans. Let X be a binomial random variable with $n = 400$and $p = 0.005$. Thus, $np = 2$. Using the Poisson approximation,

(a) $P(X = 1) = e^{-2}2^1 = 0.271$ and

(b) $P(X \le 3) = \sum_{x=0}^{3} e^{-2}2^x/x! = 0.857.$

Example problem of Poisson Distribution

Q. In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

Ans. Here the value of $n = 8000$ and $p = 0.001$. Since, p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using $\mu = 8000 \times 0.001 = 8$.

Hence, if X be the number of bubbles, then

$$P(X < 7) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx p(x; 8) = 0.3134.$$

References

1. Norean R. Sharpe, Richard D. De Veaux, Paul F. Velleman, Business Statistics, Fourth Edition, Pearson Education. 2019.
2. R.E. Walpole, R.H. Mayers, S.L. Mayers and K.Ye, Probability and Statistics for engineers and scientists, 9th Edition, Pearson Education, 2018.
3. Douglas A. Wolfe, Grant Schneider, Intuitive Introductory Statistics, Springer, 2017.
4. Miller & Freund's, Probability and statistics for engineers, 8th edition, Pearson publication, 2018.
5. Richard I. Levin, David S. Rubin, Masood H. Siddiqui, Sanjay Rastogi, Statistics for Management, 8th Edition, Pearson Publications, 2018.
6. S. C. Gupta and V. K. Kapoor, Fundamentals of Applied Statistics, S. Chand, 2006.

THE END