A First Course in Quality Engineering – 2nd edition

K.S. Krishnamoorthi

With V. Ram Krishnamoorthi

Chapter 2: Statistics for Quality – Part 1

CHAPTER 2
STATISTICS FOR QUALITY

Why statistics in quality?

The science of statistics provides:

Means for describing populations with variability.

Methods for “estimating” population quality from samples.

And, the following methods used in quality engineering.

Major Statistical Methods Used in Quality Engineering

1. **Control charts**:

   Used for controlling processes so they produce products of uniform quality

2. **Sampling plans**:

   Used for determining acceptability of lots based on samples taken from them

3. **Designed experiments**:

   Used for determining the best combination of process parameter levels to obtain desired levels of quality characteristics

Major Statistical Methods Used in Quality Engineering (contd.)

4. **Regression Analysis**:
   Used for determining which cause variables affect, to what extent, which quality characteristics

5. **Reliability engineering**:
   Used in understanding the factors that affect the life of parts and assemblies in order to increase their life

6. **Tolerancing**:
   Used for determining allowable variability in product and process variables so the products can be produced economically while meeting customer needs.

Three Major Sections

Need to learn the fundamentals of probability and statistics

Section I:	The empirical methods of describing populations with variability - based on data gathered from populations.
Section II:	Mathematical methods of modeling such populations using probability distributions.
Section III:	Mathematical methods used for inferring population quality from sample quality

Section I: Empirical methods

These are methods based on observations obtained from the populations under study.

A few terms are first explained.

Population

Sample, Random sample (?)

Two types of data: Measurement data.

Attribute data.

Empirical methods for describing populations

Every population has variability in it.
Excessive variability in a population causes poor quality and waste.
Example 2.1:

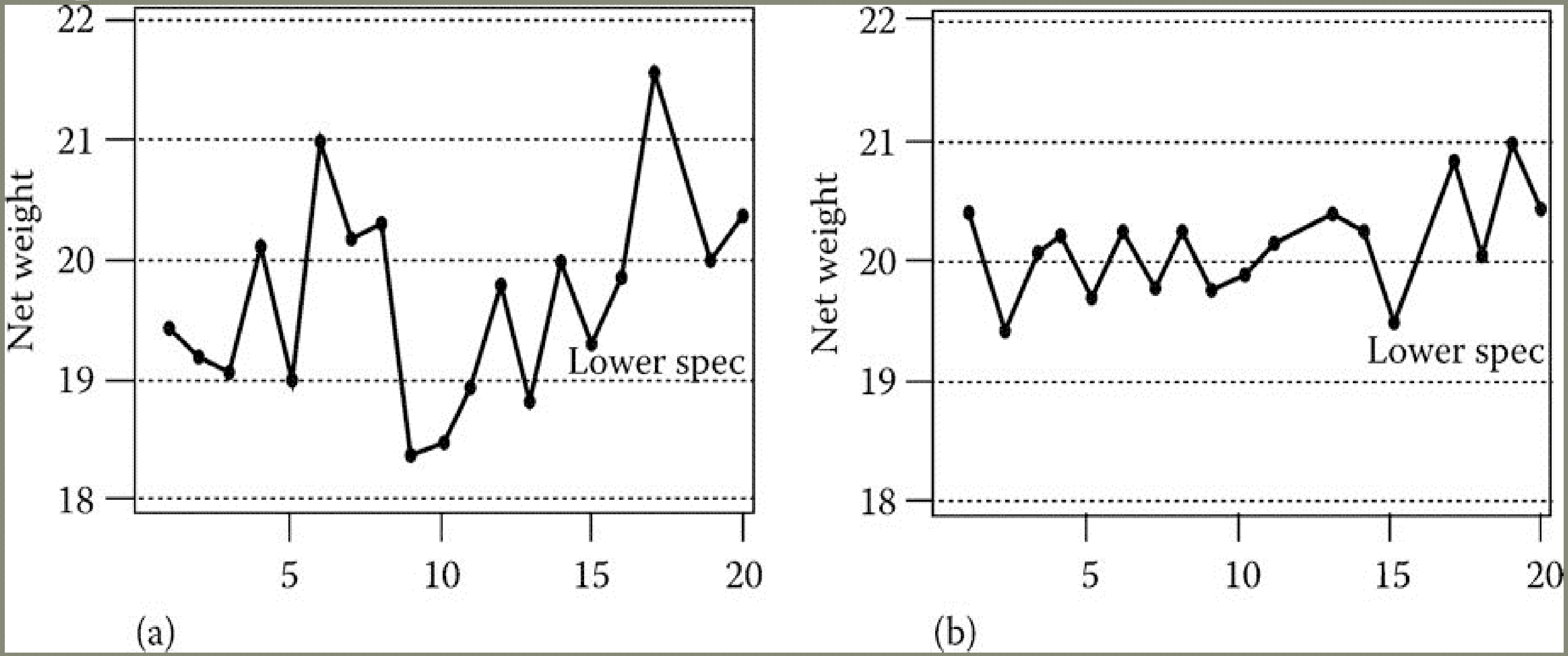


Figure 2.1 Net weight of nails in 20-lb boxes from two filling lines

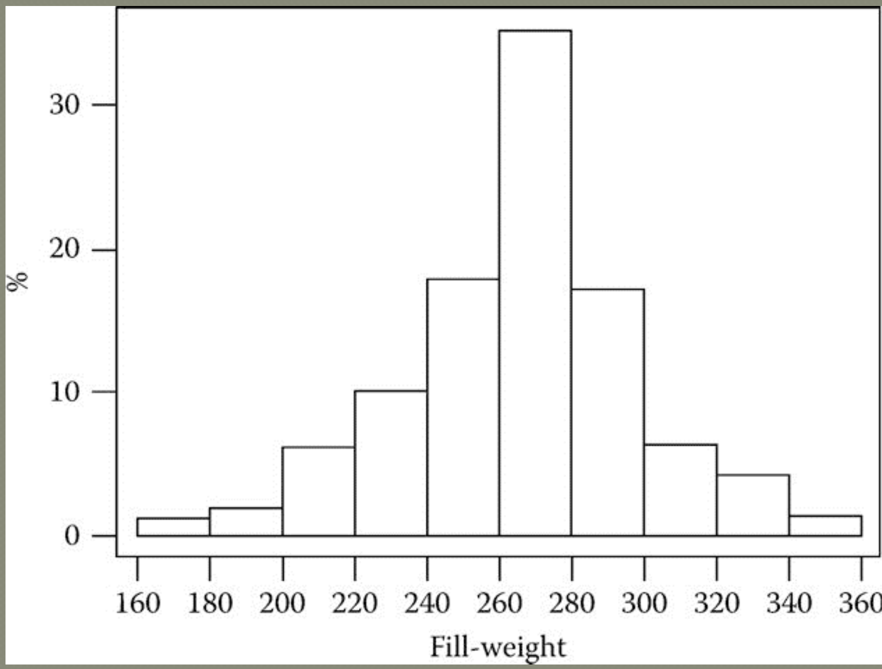
Empirical methods for describing populations (contd.)

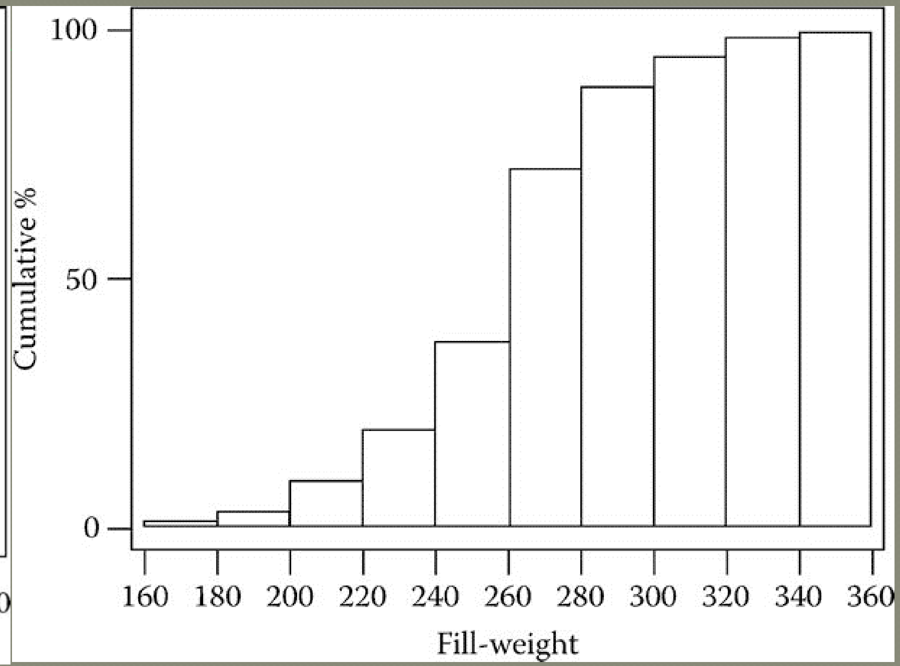
The frequency distribution is the tool used to understand and describe the variation among units in a population.

Histogram is the sample analog of the frequency distribution of a population.

How to make the histogram?
How to make the cumulative frequency distribution?
(Pl. read in the book)

Histogram and Cum Freq Dist.





Quite a bit can be learned about a population by studying the Histogram and the Cum. Freq. Distribution drawn for it.

Especially, we can see: centering and variability of the population

Examples of using a histogram

Two different populations mixed together: Example 2.3a

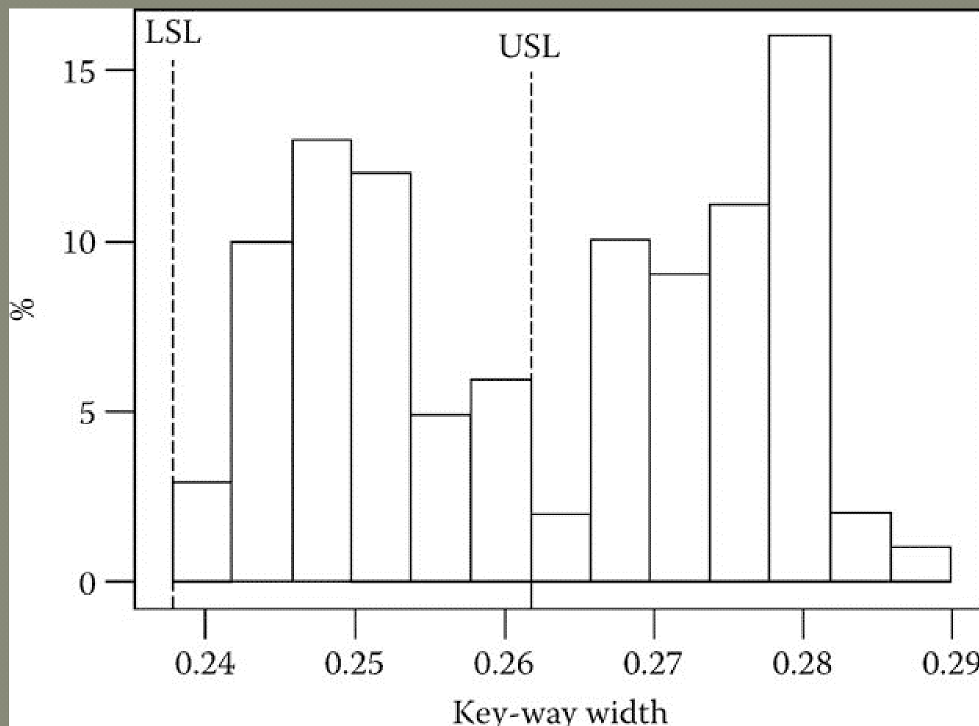
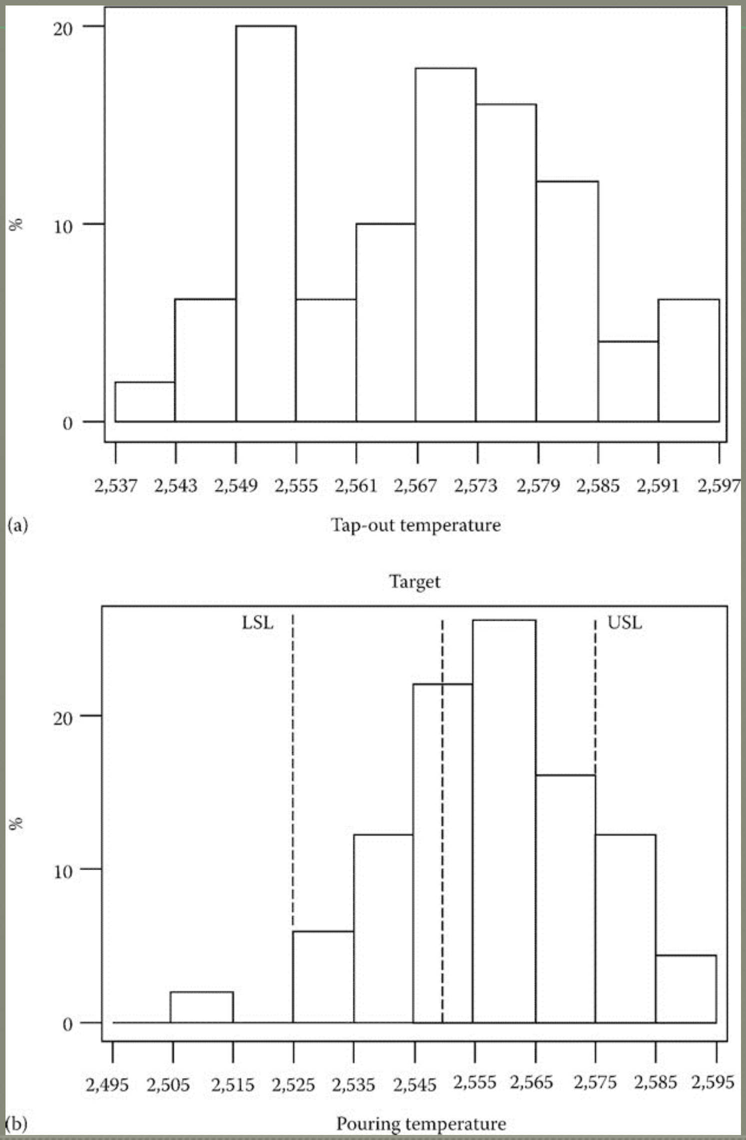


Figure 2-4 Histogram of key-way widths on a thermostat calibration shaft

Examples of using histogram (contd.)

Example 2.3b
Figure 2.5
Observing the difference between two populations:



Numerical Methods for Describing Populations

Average: $\bar{X} = \frac{\sum_i X_i}{n}$

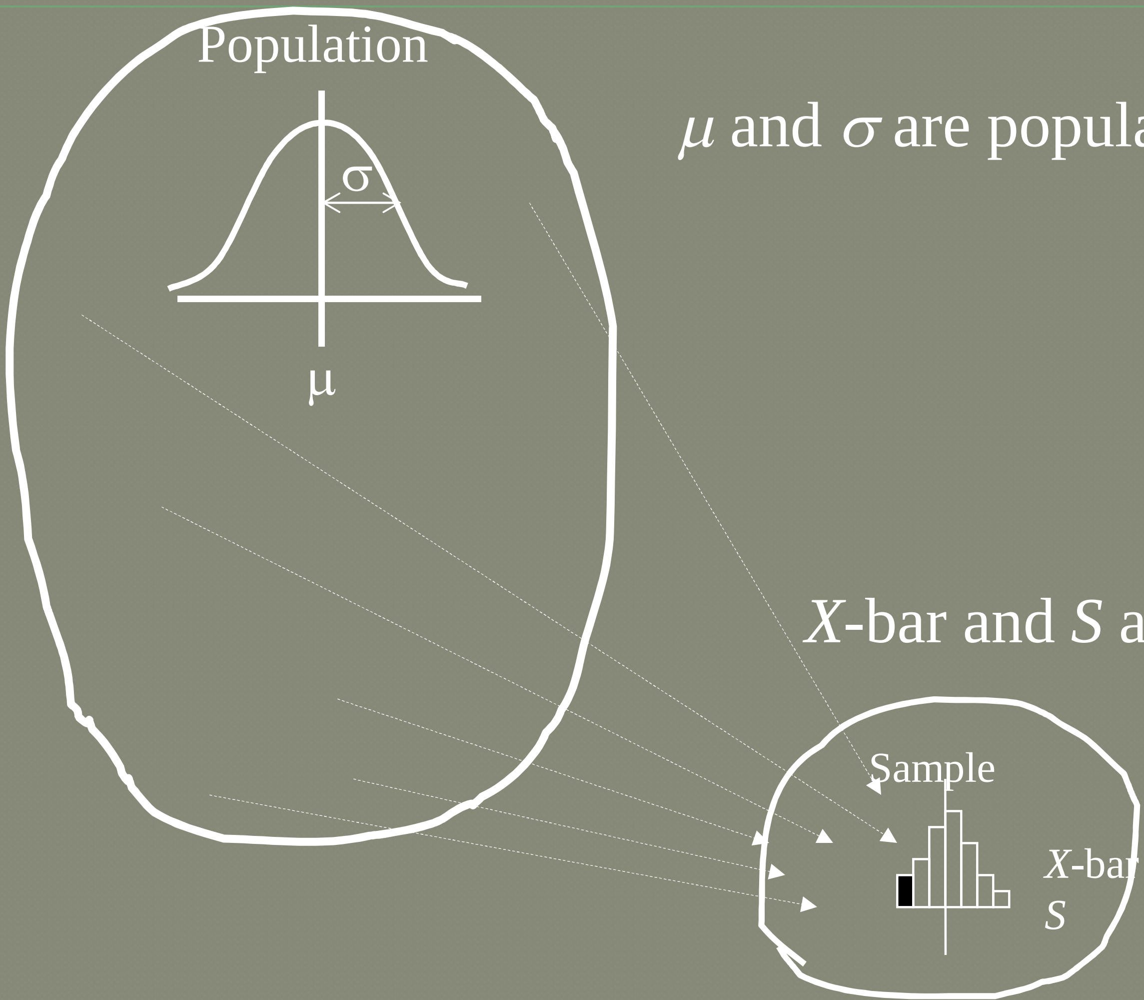
Standard deviation: $S = \sqrt{\frac{\sum \left(Xi - \bar{X}\right)^2}{n-1}} = \sqrt{\frac{n\sum Xi^2 - \left(\sum Xi\right)^2}{n(n-1)}}$

The average represents the center point (location on the *x*-axis) around which the data (population) is distributed.

The standard deviation represents the amount of variability (dispersion) in the data about the center point.

These two measures adequately describe a population that has the symmetric, bell shaped (normal) distribution.

Population parameter vs. sample statistics



μ and σ are population ***parameters***

X-bar and S are sample ***statistics***

Stem & Leaf Diagram

Cum. Counts	Stems	Leaves
1	17	5
2	18	7
3	19	7
6	20	058
9	21	045
13	22	0138
19	23	114555
26	24	2235688
37	25	00013447888
(21)	26	000012334455555577899
42	27	01124444566788
28	28	0000113367
18	29	0346899
11	30	0178
7	31	78
5	32	18
3	33	47
1	34	6

Figure 2.7 Stem and Leaf Diagram of the amount of deodorant in cans

Box and Whisker plots

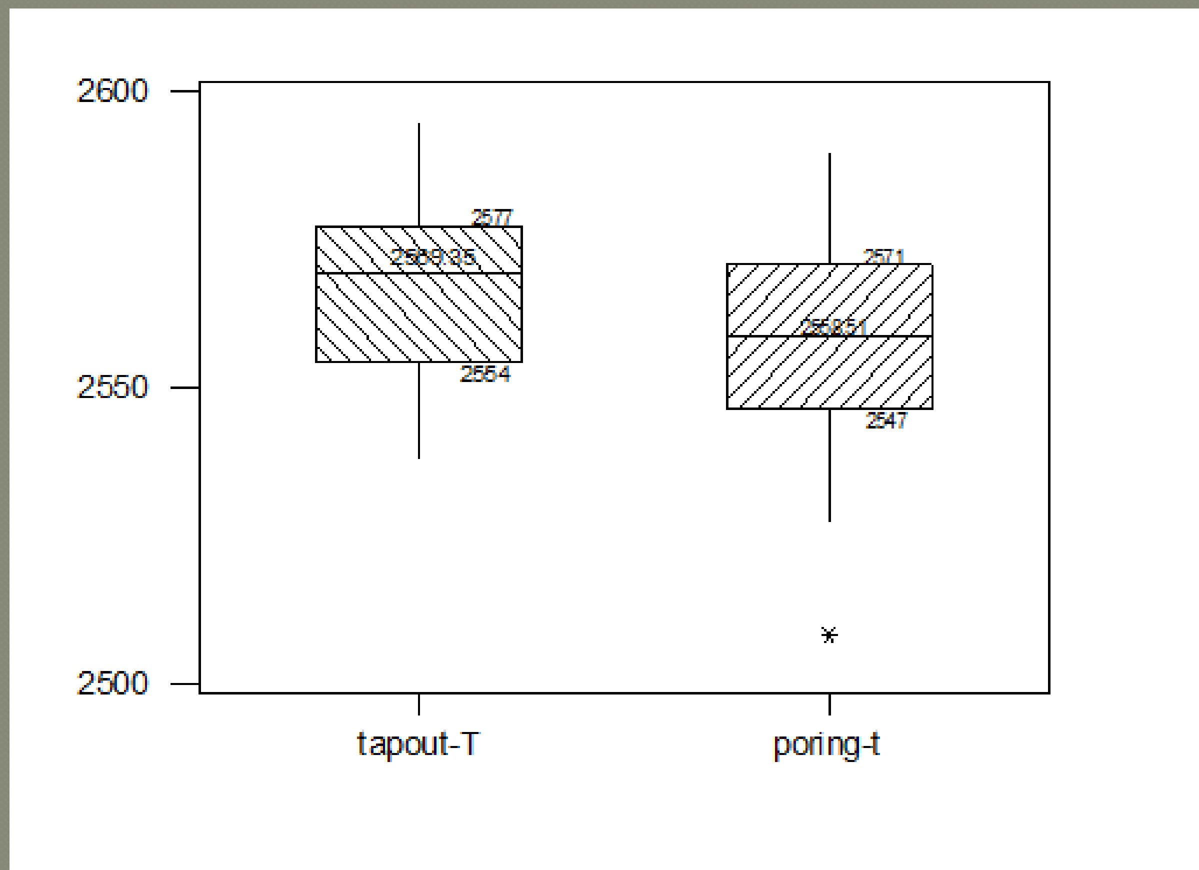


Figure 2-8b Box and Whisker plots of pouring and tap-out temperatures

Other numerical measures to describe populations

Measures of location:

The Median (X) is the middle value in the population.

The Mode (M) is the value that occurs most frequently in the data.

Mid-range = $(X_{max}+X_{min})/2$

Measures of dispersion:

The Range: $R = X_{max} - X_{min}$.

Inter Quartile Range: $IQR = X_{75} - X_{25}$.

(X_{75} and X_{25} are respectively the 75th and 25th percentiles in the data.)

Mathematical Models for describing populations

Mathematical models are idealized mathematical functions chosen to represent the graph of the frequency distribution.

These are called the probability distribution functions.

In order to discuss these models of distributions we need to define several terms: probability, random variable, etc.

We begin with probability.

Probability

An experiment is a clearly defined procedure that results in observations.

A single performance of an experiment is called a trial and each trial results in an outcome or observation.

The experiments we are dealing with here are called the random experiments in that, the outcome in any one trial of the experiment cannot be predicted with certainly.

However, all possible outcomes of the experiment are known.

Probability (contd.)

The set of all possible outcomes of a random experiment is called the sample space and is denoted by S.

Each element of a sample space is called a sample point.

An event is a subset of the sample space such that all the elements in it follow a common rule.

Probability (contd.)

Examples of Experiments, Sample space and Events

Experiment	Sample space	Example events
a. Toss a die and observe the number	{1, 2, 3, 4, 5, 6}	A:{# less than 4} = {1,2,3}
b. Toss two coins and observe the faces	{HH, HT, TH, TT}	B: {At least one head} = {HH, HT, TH)
c. Toss two coins and count the # of heads	{0, 1, 2}	C: {No head} = {0}
d. Pick a sample of 10 bulbs and count the # of defectives: D	{0, 1, 2, …, 10}	E: {D<3} = {0,1,2}

Definition of probability

Probability of an event A is a number between 0 and 1. It represents the chance the event A occurs when the related experiment is performed.

$$0 \leq P(A) \leq 1$$

$P(\Phi) = 0,$ where Φ denotes the null event, the event that cannot occur.

$P(S) = 1,$ because when an experiment is performed any one outcome should occur.

Computing Probability of an event

Two methods: 1. Method of analysis of experiment

2. Method of relative frequency

Method of Analysis is explained in steps in the following example.

Example:

A card is drawn from a deck. What is the probability the card has a number (not a picture)?

Computing probability- Method of analysis

Step 1: Formulate S

S = {Hearts Number, Hearts Picture, Clubs Number, Clubs Picture, Diamond Number, Diamond Picture, Spade Number, Spade Picture}

Step 2: Assign weights to the elements to reflect their chance of occurrence.

We could assign the following weights to the outcomes from our knowledge of the experiment.

	{9/52,	4/52,	9/52,	4/52,	9/52,	4/52,	9/52,	4/52}
S =	{HN,	HP,	CN,	CP,	DN,	DP,	SN,	SP}

Computing probability-Method of analysis (contd.)

Step 3: Calculate $P(A)$ = Sum of weights of elements in A.

The event A: {the card is a number} = {HN, CN, DN, SN}

Therefore $P(A) = 9/52 + 9/52 + 9/52 + 9/52 = 36/52$

Computing probability- a special case

If the sample space of an experiment consists of elements that are all equally likely, then:

$$P(A) = \frac{\#\text{ofelements} \subset \text{A}}{\#\text{ofelements} \subset \text{S}}$$

Example:

When a coin is tossed three times, what is the probability that all tosses will show the same face?

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

The elements are equally likely.

If A is the event that all tosses have the same face,

then A = {HHH, TTT}

P(A) = 2/8 = 1/4

Computing probability - relative frequency method

Perform the experiment N number of times. If event A occurs n times out of N, then relative frequency of A:

$$f_A = n/N$$

Then,

$$P(A) = \lim \xrightarrow[N\to\infty]{} f_A$$

That is, *P(A)* is given by f_A provided *N* is large.

Computing probability-relative frequency method (contd.)

Example 2.7: Data on number of students from different regions in a college of engineering:

	From Illinois Outside Chicago	From Chicago	From Other Areas	Total
Boys	13	55	10	78
Girls	10	10	2	22
Total	23	65	12	100

P(student comes from Chicago) = 65/100 = 0.65.
P(student is girl and is from Ill. outside Chicago) = 10/100 = 0.1

Is 100 trials large enough? How large is large?

Theorems on Probability

Addition Theorem of Probability

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

If A and B are mutually exclusive, i.e., there are no common elements between them, $P(A \cap B) = 0$, and

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

Theorems on Probability (contd.)

If $A_1, A_2, \ldots, A_k$ are all mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \ldots\ldots\ldots \cup A_k) = P(A_1) + P(A_2) + \ldots\ldots\ldots + P(A_k)$$

Complement Theorem of Probability

$$P(A^c) = 1 - P(A)$$

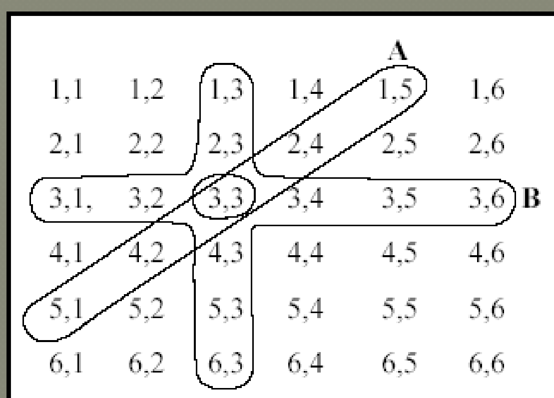
In many situations we may be interested in the Event A, but it will be easier to compute the probability of the Event A^c. Then this theorem is useful.

Theorems on Probability (contd.)

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

Example 2.11: Find the probability that, when two dice are thrown, the total equals 6, given one of the numbers is 3.

Define *A*: {total equals 6} *B*: {one of the numbers is 3}



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{11/36} = 1/11$$

Theorems of probability

Independent Events:
If A and B are independent,

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

The Multiplication Theorems of Probability:
If A and B are any two events in a sample space,

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

If A and B are independent

$$\boxed{P(A \cap B) = P(A) \times P(B)}$$

Independent events

Example 2.13

A box contains 7 black balls and 5 white balls. Two balls are drawn with replacement. What is the probability both are black?

Let B_1 be the event 1st fall is black. Then $P(B_1) = 7/12$

Let B_2 be the event 2nd ball is black. Then $P(B_2) = 7/12$

$$P(B_1 \cap B_2) = P(B_1) \times P(B_2) = \frac{7}{12} \times \frac{7}{12} = \frac{49}{144}$$

(because of independence)

Independent events (contd.)

Example 2.14

A box contains 7 black balls and 5 white balls. If two balls are drawn without replacement what is the probability both are black?

Let B_1 be the event 1st ball is black: $P(B_1) = 7/12$

Let B_2 be the event 2nd ball is black.

B_1 and B_2 are not independent

We have to use conditional probability: $P(B_2 | B_1) = 6/11$

$$P(B_1 \cap B_2) = P(B_1) \times P(B_2 | B_1) = \frac{7}{12} \times \frac{6}{11} = \frac{42}{132}$$

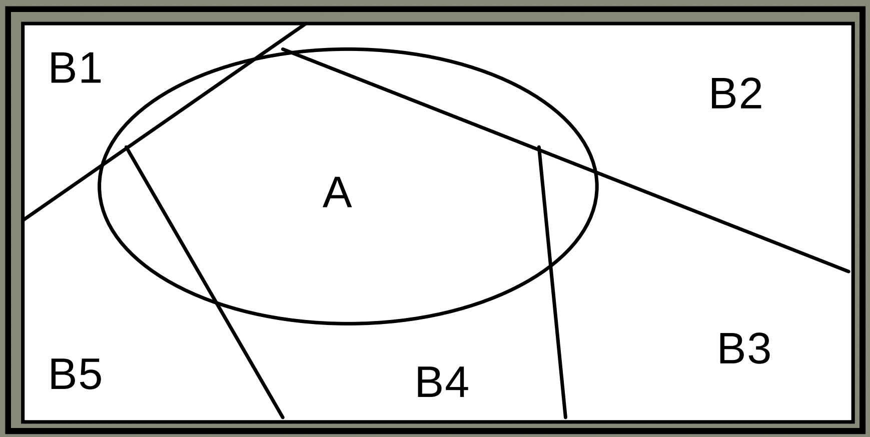
Theorem on Total Probability

Theorem on Total Probability

Let B_1, B_2, ... B_k be partitions of a sample space S such that $(B_1 \cup B_2 \cup .... \cup B_k) = S$ and $(B_i \cap B_j)$ = ø (null set) for any pair i and j (see diagram below).

If A is an event of interest, then:

$P(A) = P(A \mid B_1)\,P(B_1) + P(A \mid B_2)\,P(B_2) + ... + P(A \mid B_k)P(B_k)$



Theorem on Total Probability (contd.)

Example 2.15

$P(EE) = 0.26 \quad P(ME) = 0.25 \quad P(CE) = 0.18 \quad P(IE) = 0.12$

$P(MfG) = 0.19$

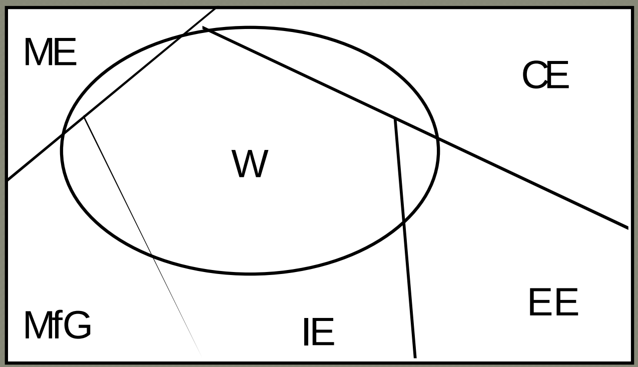
$P(W|EE) = 0.05 \quad P(W|ME) = 0.10 \quad P(W|CE) = 0.08 \quad P(W|IE) = 0.45$

$P(W|MfG) = 0.04$

$P(W)=P(W|EE)P(EE)+P(W|ME)P(ME)+P(W|CE)P(CE)+P(W|IE)P(IE)$

$+P(W|MfG)P(MfG)$

$= (0.05)(0.26) + (0.10)(.25) + (0.08)(0.18) + (0.45)(0.12) + (0.04)(0.19)=0.114$



Counting sample points

Counting sample points in the sample space:

Multiplication theorem:

If operation 1 can be performed in n_1 ways, operation 2 in n_2 ways, …, operation k in n_k ways, then the k operations can be performed together in $n_1 \times n_2 \times \ldots \times n_k$ number of ways.

Example:
When three dies are thrown and their up-faces observed, how many sample points are there in S?

$$6 \times 6 \times 6 = 216$$

Counting sample points (contd.)

Permutations are arrangements.

Number permutations of n objects taken r at a time:

$$n\,\mathrm{Pr} = \frac{n!}{(n-r)!}$$

Combinations are just groupings.

Number of combinations of n objects taken r at a time

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \qquad \text{for } r \le n$$

Final example

Example 2.21

A box contains 20 pencils out of which 16 are good and 4 are bad.

a) How many samples of 4 can be drawn from the box?

$$\binom{20}{4} = \frac{20!}{4!16!} = 4{,}845$$

b) How many samples of 4 are possible with exactly one defective?

$$\binom{16}{3} \text{x} \binom{4}{1} = \frac{16!}{3!3!} \text{x} 4 = 2{,}240$$

Final example (contd.)

c) If a sample of 4 is drawn from the box, what is the probability that exactly one of them will be defective?

P(1 defective) = $$\frac{\binom{16}{3}\binom{4}{1}}{\binom{20}{4}} = \frac{560 \times 4}{4{,}845} = 0.462$$

d) What is the probability that a sample of 4 will have no more than 1 defective?

P (no more than 1 defective) = P (0 def) + P (1 def) =

$$\frac{\binom{16}{4}\binom{4}{0} + \binom{16}{3}\binom{4}{1}}{\binom{20}{4}} = \frac{1820 + 2240}{4845} = \frac{4060}{4865} = 0.838$$