Chapter 1

Overview and Descriptive Statistics

Outline

① Populations, Samples and Processes

② Pictorial and Tabular Methods in Descriptive Statistics

③ Measures of Location

④ Measures of Variability

Load the data set in the book

- Open R 3.6.1 software
- In the menu bar press File -> New script
- In the new working area, install the R package containing this data from CRAN (Comprehensive R Archive Network).

  - install.packages("Devore7")
- To load the files within an R session you may type:

  - library(Devore7)
- You have access to a vignette describing R and its use with many of the examples in the text.

  - vignette("Devore7")

Example 1.6

- The use of alcohol by college students is of great concern.
- Binge drinking : five or more drinks in a row for males and four or more for females
- x= the percentage of undergraduate students who are binge drinkers

Example 1.6

> install.packages("Devore7")

> library(Devore7)

```
> str(xmp01.05)
'data.frame':   140 obs. of  1 variable:
 $ bingePct: int  4 11 13 14 15 16 17 18 18 18 ...
> with(xmp01.05, summary(bingePct))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   4.00   32.75   43.00   42.34   53.25   68.00 
```

Stem and leaf graph

```
> with(xmp01.05, stem(bingePct))

  The decimal point is 1 digit(s) to the right of the |

  0 | 4
  0 |
  1 | 134
  1 | 5678889
  2 | 12234
  2 | 56666777889999
  3 | 0112233344
  3 | 5556666777788889999
  4 | 11122222334444
  4 | 5566666677788888999
  5 | 00111222334
  5 | 5566666777788899
  6 | 011112444
  6 | 55666778
```

Data collection

Simple random sample : A *simple random sample* consists of a sample drawn in such a way that all possible samples of the same size have equal probability of being drawn.

5 objects : a, b, c, d, e

all possible sample of size 2 : $\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{120}{2\times 6} = 10$

(a, b) (a, c) (a, d) (a, e)

(b, c) (b, d) (b, e)

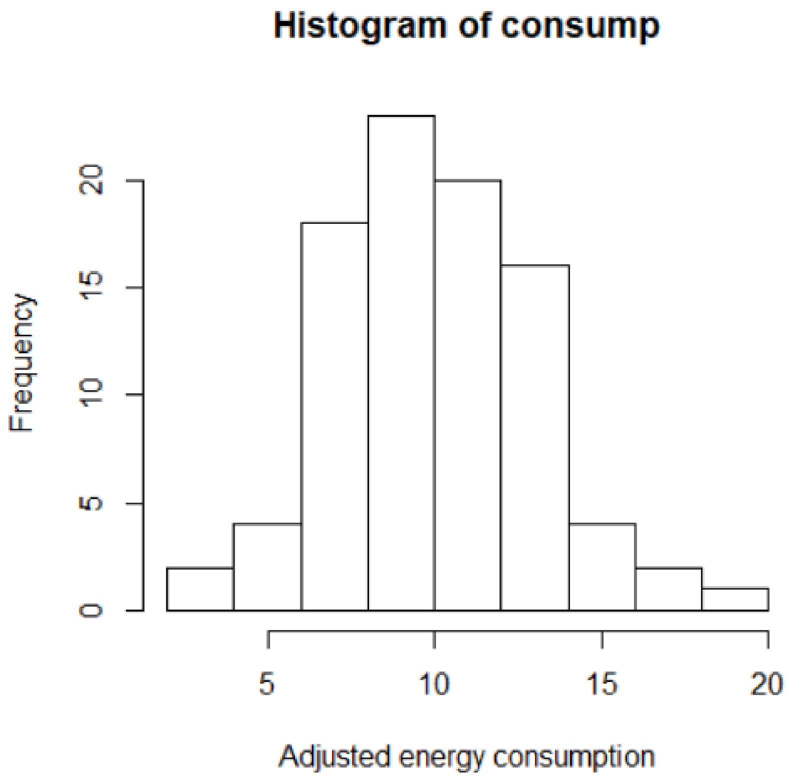
(c, d) (c, e)

(d, e)

```
> sample(1:60, 5) # select 5 random numbers from 1 to 60
[1] 15 49 58 42 12
```

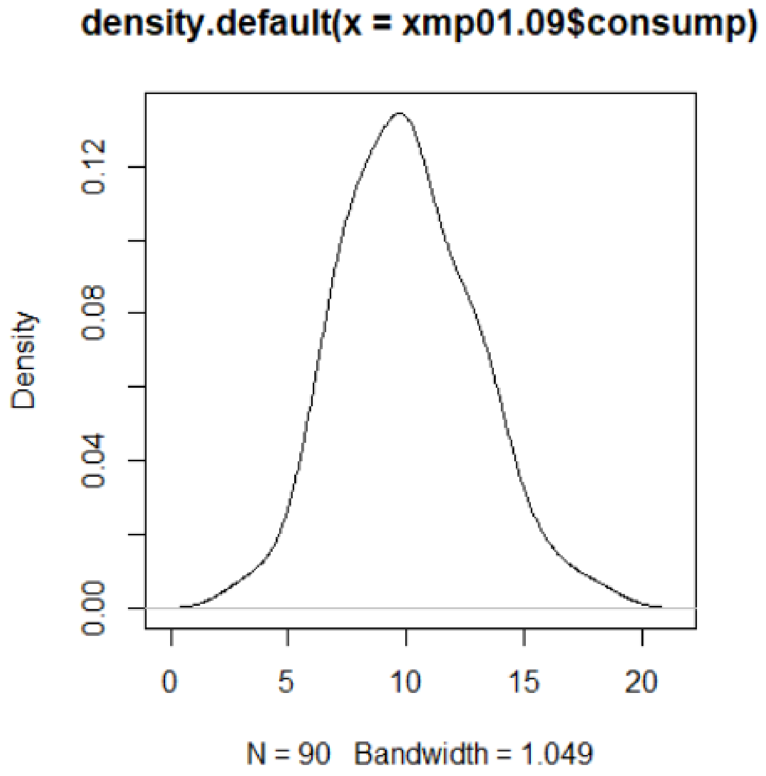
Histogram of power consumption

```
> with(xmp01.09, hist(consump, xlab="Adjusted energy consumption"))
> hist(xmp01.09$consump, xlab="Adjusted energy consumption"))
```



Density plot of power consumption

```
> d <- density(xmp01.09$consump) # returns the density data
> plot(d) # plots the results
```



Categorical data

- In a Categorical data set we only observe the category of the response.
- Example: In exercise 1.23(p.26), 60 observations of the type of health complaint are given.

  J=joint swelling, F=fatigue, B=back pain,

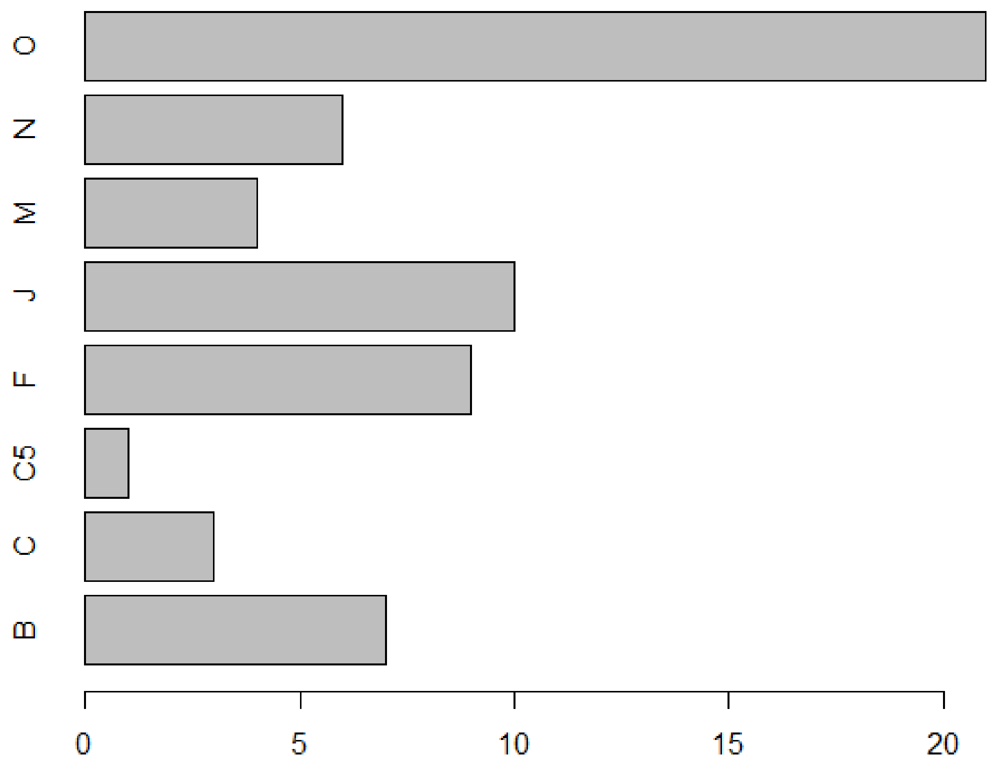
  M=muscle weakness, C=coughing,

  N=nose running/irritation, O=other
- Use of *R* : In *R* a factor is used to represent categorical data.
- Plotting : A bar chart or bar plot is a convenient way to plot such data

```
> str(ex01.29)
'data.frame':   61 obs. of  1 variable:
 $ C1: Factor w/ 8 levels "B","C","C5","F",..: 3 8 8 7 5 2
```

Bar plot for health complaints

```
> with(ex01.29, barplot(table(C1), horiz=TRUE))
```



Measures of location

Sample mean – the average data value.

$$\bar{x}=\sum_{i=1}^{n} x_i/\mathrm{n}$$

The *population mean*, μ, is the average over the whole population.

Median :

The *Sample Median* is the "middle" data value.

When the data are skewed, this is more representative than the sample mean.

order statistics : $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.

If n is odd, the median is $x_{([n+1]/2)}$, $ex)\, n = 7 :\ x_{(4)}$

If n is even, $\left[x_{(n/2)}+x_{(\frac{n}{2}+1)}\right]/2$ $ex)\, n = 10 :\ \frac{x_{(5)}+x_{(5+1)}}{2}$

Quartiles, percentiles, and quantiles

Quartiles : Just as the median divides the sorted data in half, the first and third quartiles divide the data into quarters

Percentiles and quantiles :

The p'th quantile ($0<p<1$) is any value such that a proportion p of the data is below it.

A percentile is the same but expressed as a percentage rather than a fraction.

- Common example : Scores such as the SAT(Scholastic Aptitude Test) are often converted to percentiles. e.g. 95th percentile
- Calculation : *quantile* in *R*.

Example 1.15 : A sample of n=12 recordings of Beethovens Symphony #9, yielding the following durations (min) listed in increasing order:

$$62.3, 62.8, 63.6, 65.2, 65.7, 66.4, 67.4, 68.4, 68.8, 70.8, 75.7, 79.0$$

- The mean is $\bar{x}$=(62.3+62.8+ … + 79.0)/12=68.01
- The sample median is the average of the n/2=6th and (n/2+1)=7th values from the ordered list:

  $\tilde{x}$=(66.4+67.4)/2=66.9

```
> duration <- c(62.3, 62.8, 63.6, 65.2, 65.7, 66.4, 67.4, 68.4, 68.8, 70.8, 75.7, 79.0)
> duration
 [1] 62.3 62.8 63.6 65.2 65.7 66.4 67.4 68.4 68.8 70.8 75.7 79.0
> summary(duration)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  62.30   64.80   66.90   68.01   69.30   79.00
```

● 3rd Quartile : 75th Percentiles

```
> quantile(duration, probs=0.75)
  75%
69.30
```

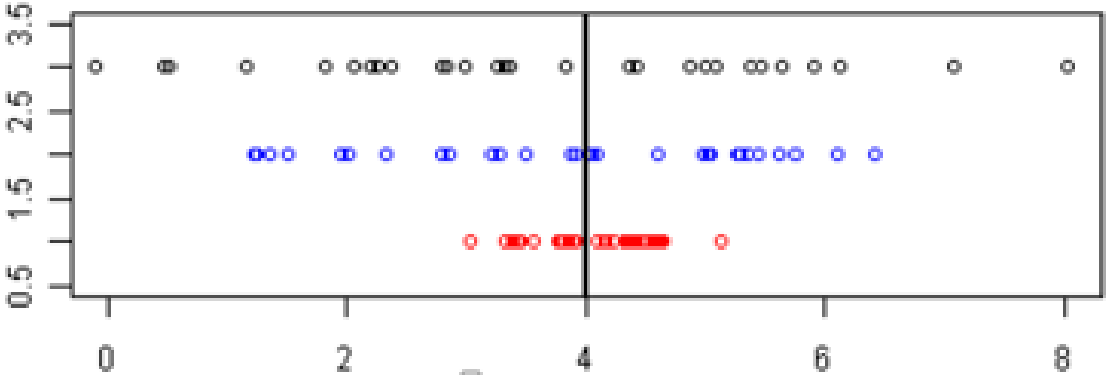
1.4 Measures of Variability

Some measures of variability

- ➢ Range (max - min)
- ➢ Interquartile Range (Q3 – Q1)
- ➢ Variance or standard deviation

Different samples or populations may have identical measures of center yet differ from one another in other important ways.

Example :



Sample Variance and standard deviation

The *Sample Variance* tells us how far away the points are from the center of the data. That is, it measures spread or dispersion.

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$$

The *Sample Standard Deviation* is the square root of sample variance.

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$$

Sample Variance

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$$

Example 1.18 : Find the variance and standard deviation of

154 142 137 133 122 126 135 135 108 120 127 134 122

- Step 1 : Form and find

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1695}{13} = 130.38$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (154 - 130.38)^2 + \cdots + (122 - 130.38)^2 = 1579.08$$

- Stem 3 : The variance is $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1579.08}{12} = 131.59$

  The standard deviation is $s = \sqrt{131.59} = 11.5$

Properties of variance

The variance and standard deviation are not affected by shifting the data by a constant but they are affected by scaling the data.

If the data are $x_1, x_2, \cdots, x_n$ and c is a constant then

① If $y_1 = x_1 + c$, $y_2 = x_2 + c, \cdots$, $y_n = x_n + c$, then $s_y^2 = s_x^2$

② If $y_1 = cx_1$, $y_2 = cx_2, \cdots$, $y_n = cx_n$, then $s_y^2 = c^2 s_x^2$ and $s_y = |c| s_x$

Notice the absolute value in the last expression. You cannot have a negative variance or a negative standard deviation.

Properties of variance

Suppose, $y_i = cx_i$,

$$\bar{x} = \;= \frac{\sum_{i=1}^{n} x_i}{n} \;,$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} cx_i}{n} = \mathrm{c}\,\frac{\sum_{i=1}^{n} x_i}{n} = c\bar{x}$$

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

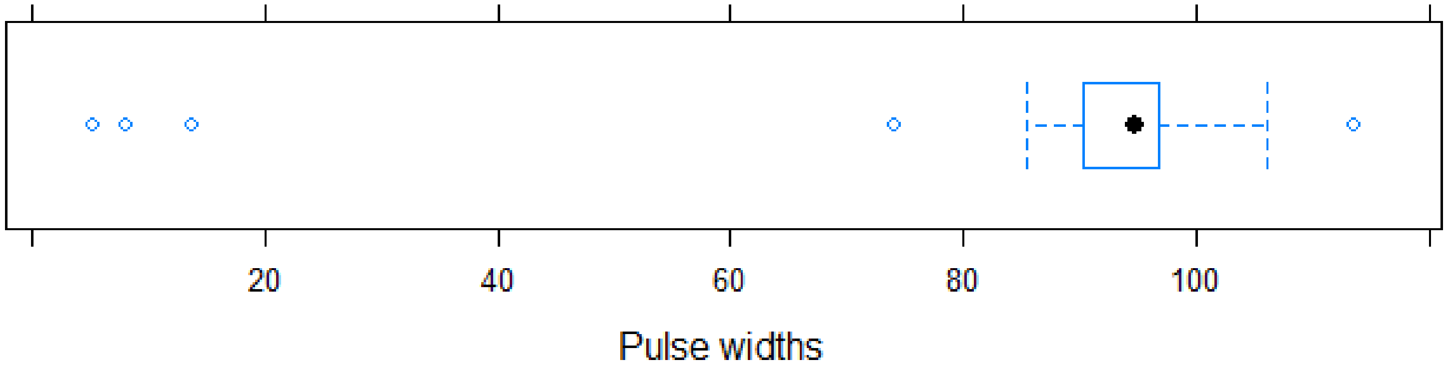
$$= \frac{\sum_{i=1}^{n}(cx_i - c\bar{x})^2}{n-1} = \frac{c^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = c^2 s_x^2$$

Boxplots

The box spans the “middle half” of the data – the region from the first quartile to the third quartile.

The whiskers extend to the minimum and maximum unless these appear to be outliers based on the *inter-quartile range.*

- boxplot(xmp01.18 $C1, xlab = "Pulse widths")



Boxplots show us location, scale, outliers and symmetry.

Creating a Boxplot

- Indicate the median and the first and third quartiles with horizontal lines.
- Find the largest sample value that is no more than 1.5 IQR above the third quartile
- Find the smallest sample value that is not more than 1.5 IQR below the first quartile.
- Extend vertical lines (whiskers) from the quartile lines to these points.
- Points more than 1.5 IQR above the third quartile, or more than 1.5 IQR below the first quartile are designated as outliers. Plot each outlier individually.

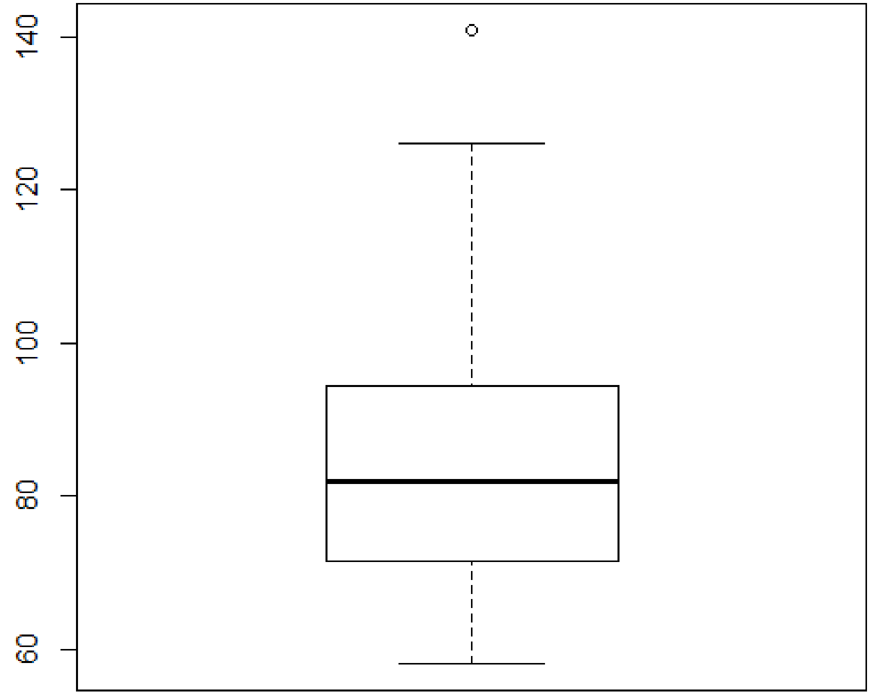
Example

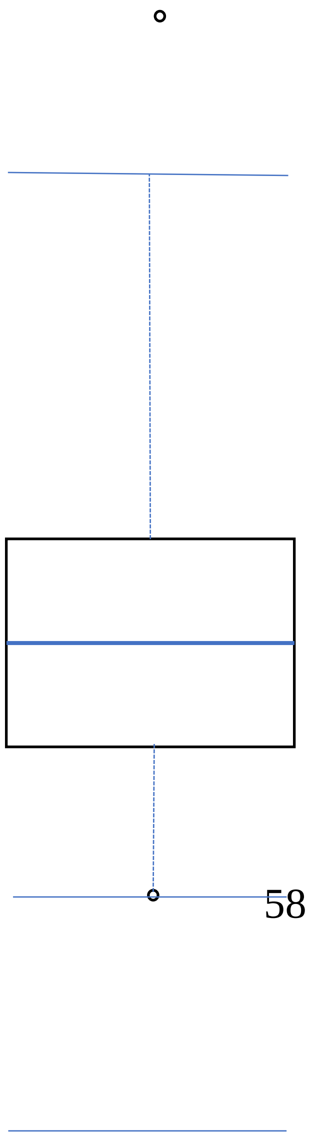
- Consider the following sample data

61, 58, 78, 71, 72, 92, 66, 83, 70, 83, 78, 81, 74, 85, 89, 98, 97, 141, 126, 112

- The following R commands for the above data yielded the following result.

```
> x <-  c(61,58,78,71,72,92,66,83,70,83,78,81,74,85,89,98, 97,141,126,112)
> summary(x)
   Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
  58.00   71.75       82.00        85.75     93.25     141.00
> sort(x)
 [1]  58  61  66  70  71  72  74  78  78  81  83  83  85  89  92  97  98  112
126 141
> boxplot(x)
```





141

$93.25 + 1.5 \times (93.25 - 71.75) = 125.5$

3rd quartile : 93.25

Median : 82.00

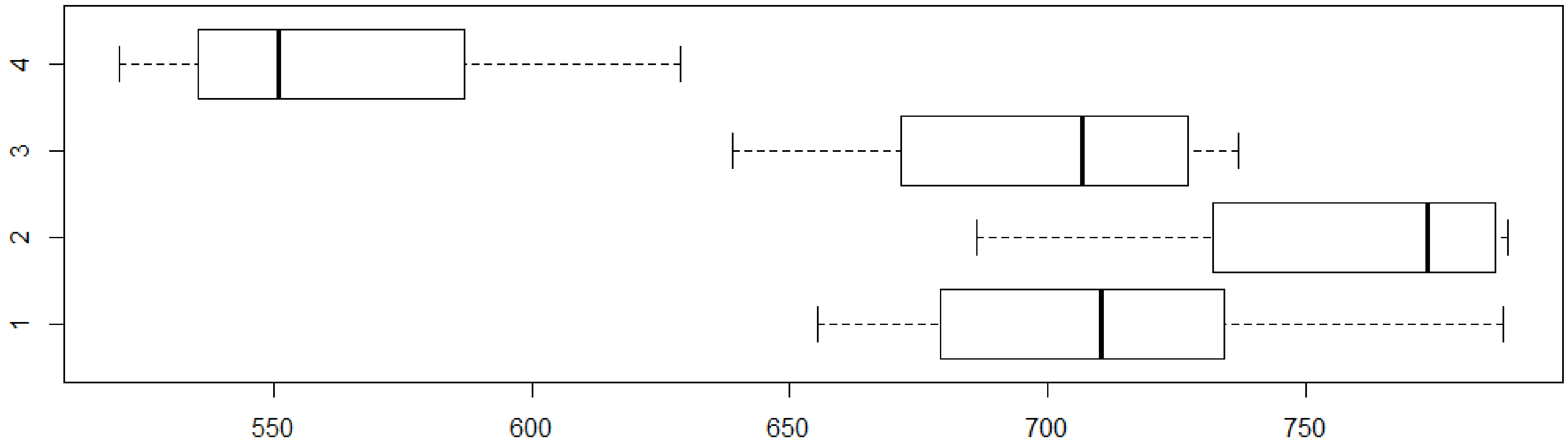
1st quartile : 71.75

58

$71.75 - 1.5 \times (93.25 - 71.75) = 39.5$

Comparative Boxplots

- When we have two or more groups of observations of a numeric variable, *comparative boxplots* allow us to examine differences between the groups in location, scale, or symmetry.
- with(xmp10.01, boxplot(C1 ~ as.factor(C2), horizontal=T))



Chapter 2

Probability

Outline

① Sample Spaces and Events

② Axioms, Interpretations, and Properties of Probability

③ Counting Techniques

④ Conditional Probability

⑤ Independence

Basic concepts in probability

Experiment : situations for which the outcome occurs randomly

Sample space Ω : the set of all possible outcomes

Event : any subset of the sample space.

We say an event occurs if the outcome of the experiment falls in the subset.

Examples

Toss a coin : sample space : {H, T}. event of “head” : {H}.

Toss a coin twice : sample space : {HH, HT, TH, TT}. event of getting one head : {HT, TH}.

Venn diagram : A pictorial representation of events.

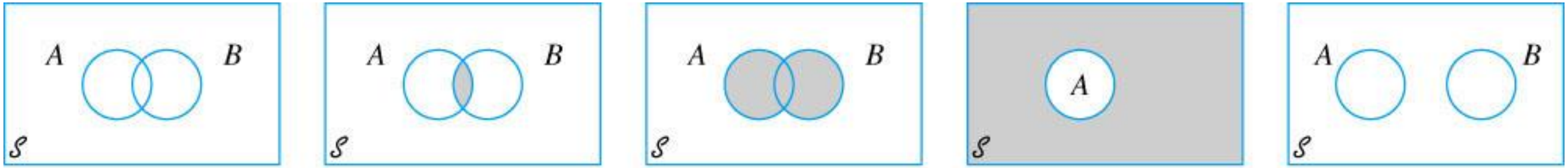
Empty event : The event that never occurs, ϕ.

Union : $A \cup B$ occurs if either A occurs or B occurs or both occur.

Intersection : $A \cap B$ occurs if both A and B occur.

Disjoint event : A and B are disjoint if $A \cap B = \phi$.

Complement : A′ occurs if A does not occur.



(a) Venn diagram of events A and B

(b) Shaded region is $A \cap B$

(c) Shaded region is $A \cup B$

(d) Shaded region is A'

(e) Mutually exclusive events

Some Relations from Set Theory

Example : Let S={0, 1, 2, 3, 4, 5, 6}, A={0, 1, 2, 3, 4} and B={1, 3, 5}

$A' = \{5, 6\}$

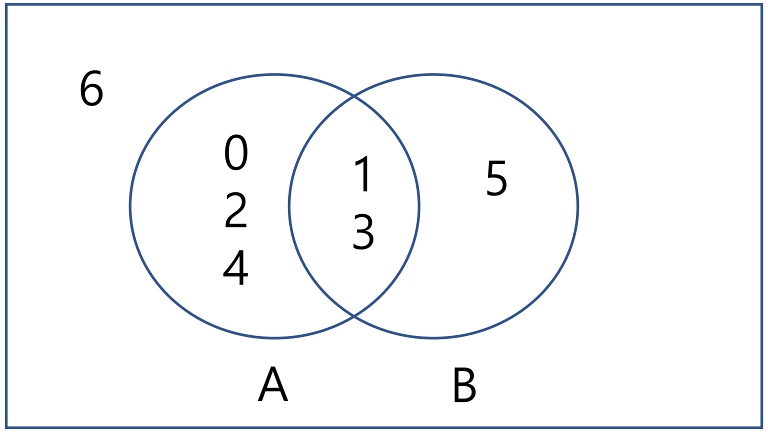
$B' = \{0, 2, 4, 6\}$

$A \cup A' = \{0, 1, 2, 3, 4, 5, 6\}$

$B \cup B' = \{0, 1, 2, 3, 4, 5, 6\}$

$A \cap B = \{1, 3\}$

$A \cup B = \{0, 1, 2, 3, 4, 5\}$



Axioms …

Given an experiment and a sample space S, a number P(A), called the probability of the event A, is assigned to each event A.

P(A) measures the chance that A will occur.

P(A) should satisfy the following axioms (basic properties) of probability.

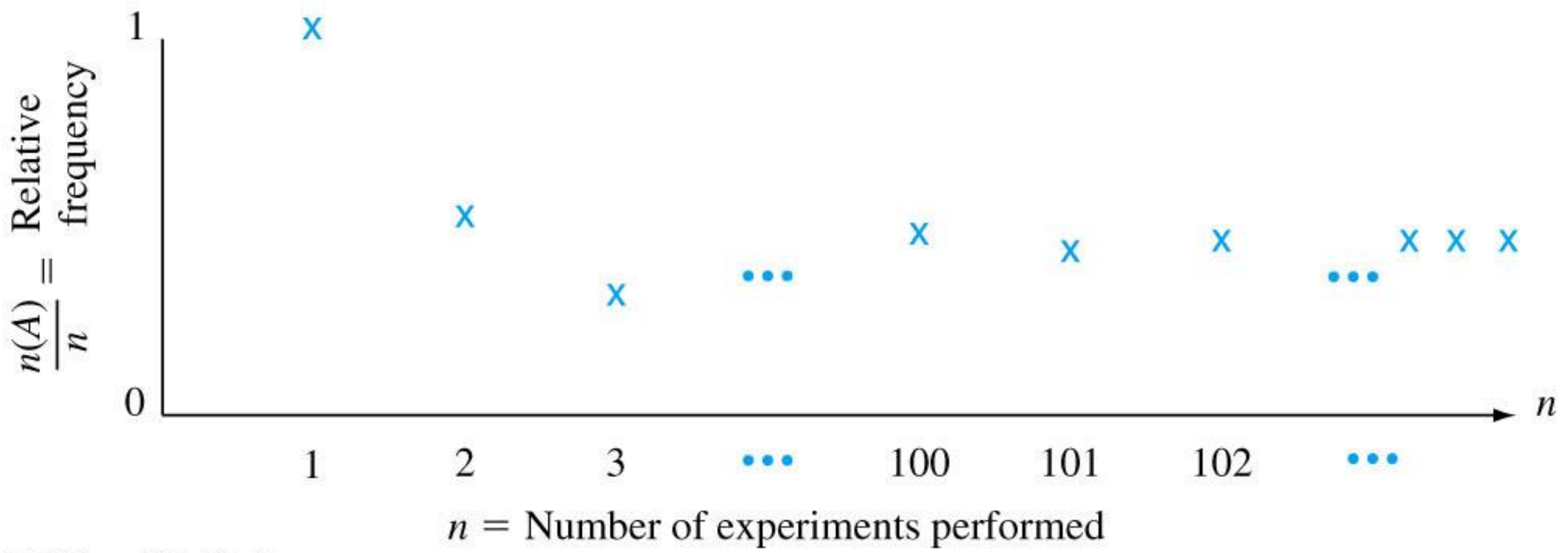
Interpretation : long term relative frequency

Axiom 1 : For any event A, $\mathrm{P}(A) \geq 0$.

Axiom 2 : $\mathrm{P}(\Omega) = 1$.

Axiom 3 : IF $A_1, A_2, \ldots,$ is a collection of mutually exclusive events,

Then, $\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \sum_i P(A_i)$



- The objective interpretation of probability :

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

Simple Properties of Probability

Property 1 : $P(A) = 1 - P(A')$ $\quad$ $P(\Omega) = P(A \cup A') = P(A) + P(A') = 1.$

Property 2 : For any two events A and B

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

Proof) $P(A \cup B) = P(A) + P(B \cap A') = P(A) + [P(B) - P(A \cap B)]$

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

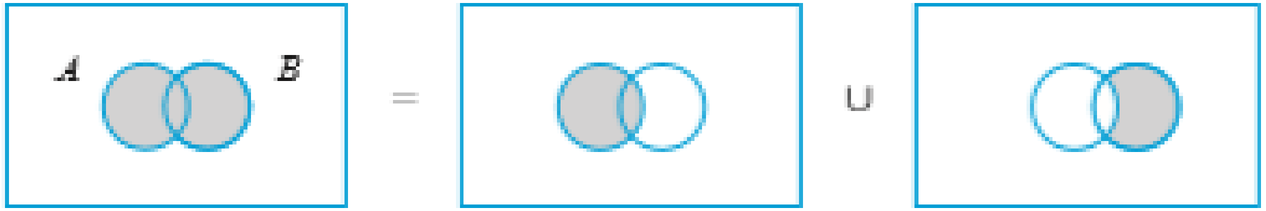


Figure 2.4 Representing A ∪ B as a union of disjoint events

Simple Properties of Probability

(Cf Example 2.14) : In a certain residential suburb,

60% of all households subscribe to the metropolitan newspaper published in a nearby city.

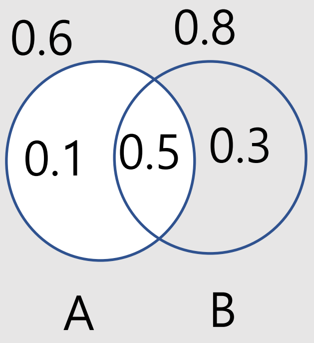
80% subscribe to the local afternoon paper.

50% of all households subscribe to both papers.

If a household is selected at random, what is the probability that it subscribes to

(1) at least one of the two newspapers and (2) exactly one of the two newspapers?

Answer :



A = {subscribe to the metropolitan newspaper } $P(A) = 0.6$

B = {subscribe to the local afternoon paper} $P(B) = 0.8$, $P(A \cap B) = 0.5$

(1) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.8 + 0.5 = 0.9$

(2) $P(A \cup B) - P(A \cap B) = 0.9 - 0.5 = 0.4$

Combinations

Given n distinct objects, any unordered subset of size k of the objects is called a combination.

The number of combinations of size k from the n objects is denoted by $\binom{n}{k}$.

Example : A bridge hand consists of any 13 cards selected from a 52-card

There are $\binom{52}{13}$ different bridge hands, which works out to be approximately 635 billion.

```
> choose(52, 13)
[1] 635013559600
```

Conditional Probability

Example : In a small company there are 24 female employees and 26 male employees. Twelve of the female employees and eighteen of the male employees are in favor of a proposal. Randomly ask one person in the company.

① What is the probability that this person is in favor of the proposal?

② If we have the information that the person asked is a female employee, what is the probability that this person is in favor of the proposal?

Answer : 1. (12+18)/(24+26)=0.6. 2. 12/24=0.5

	Favor	Not favor	Total
Female	12	12	24
Male	18	8	26
Total	30	20	50

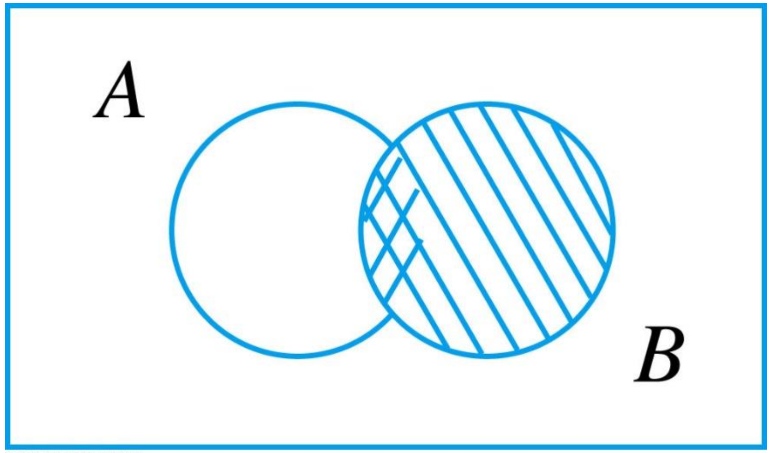
The probability calculated given some information, or under certain condition, is called conditional probability, and is denoted by P(A|B).

Calculating Conditional Probability

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

Conditional probability can be seen to be the probability with respect to a reduced sample space.

We illustrate the conditional probability with the Venn diagram.



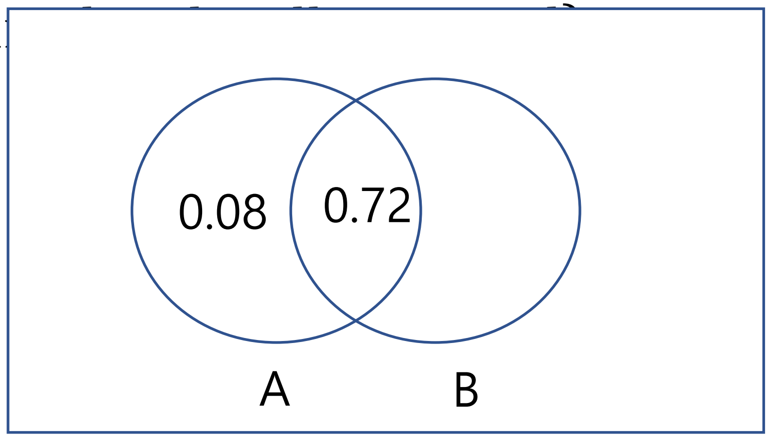
Examples

Example 1 : If the probability that a research project will be well planned is 0.8 and the probability that it will be well planned and well executed is 0.72, what is the probability that a well planned research project will be well executed?

$$A = \{\text{a research project is well planned}\}: \mathrm{P(A)} = 0.8$$

$$B = \{\text{a research project is well executed}\}$$

$A \cap B = \{$a research project is well plan



$$P(A \cap B) = 0.72$$

Answer : $P(B|A) = \frac{P(A \cap B)}{P(A)}$ =0.72/0.8=0.9

More on Conditional Probability

Sometimes the conditional probability can be determined easily, so we can use the conditional probability to calculate probability.

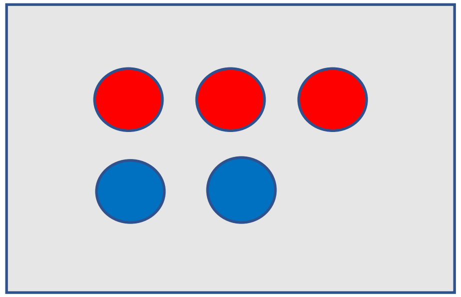
The multiplication rule : $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

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

Example : There are 3 red balls and 2 blue balls in a box.

Randomly take 2 balls from the box.

What is the probability that both are red?



Answer : $R_1 = \{\text{firtst ball is red}\}$, $R_2 = \{\text{second ball is red}\}$,

$$P(R_1 \cap R_2) = P(R_1)\ P(R_2|R_1) = (3/5)(2/4) = 0.3 \quad \text{or} \quad \frac{\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10} = 0.3$$

Another Example

Example : Three students have only one ticket to the super bowl. They will decide who is going in the following way: they will ask someone else to put the ticket in one of three boxes. One of them will pick and open one of the boxes. If the ticket is in the box, he gets it: otherwise, he is out, and the second person picks, and so on. Does it matter who picks first?

Answer : $P(1^{st}$ person get the ticket$)=1/3$.

The second person will get the ticket if 1^{st} person does not get the ticket, and the 2^{nd} person pick the right one when facing the two remaining boxes. The probability is $(2/3)(1/2)=1/3$.

Example 2.29

A chain of video stores sells three different brands of DVD(Digital Versatile Disc) players.

Of its DVD player sales, 50% are brand 1, 30% are brand 2, and 20% are brand 3.

Each manufacturer offers a 1-year warranty.

It is known that 25%, 20%, 10% of brand 1 , brand 2, brand 3's DVD players require warranty repair work.

1. What is the probability that a randomly selected purchaser has bought a brand 1 DVD player that will need repair while under warranty?
2. What is the probability that a randomly selected purchaser has a DVD player that will need repair while under warranty?
3. If a customer returns to the store with a DVD player that needs warranty repair work, what is the probability that it is a brand 1 DVD player? A brand 2 DVD player? A brand 3 DVD player?

Sol)

Let A_i={brand i is purchased}.   B={needs repair}

1. $P(A_1 \cap B)$=$P(B|A_1) \cdot P(A_1) = 0.25 \times 0.5$=0.125

2. P(B)=P[(brand 1 and repair) or (brand 2 and repair) or (brand 3 and repair)]

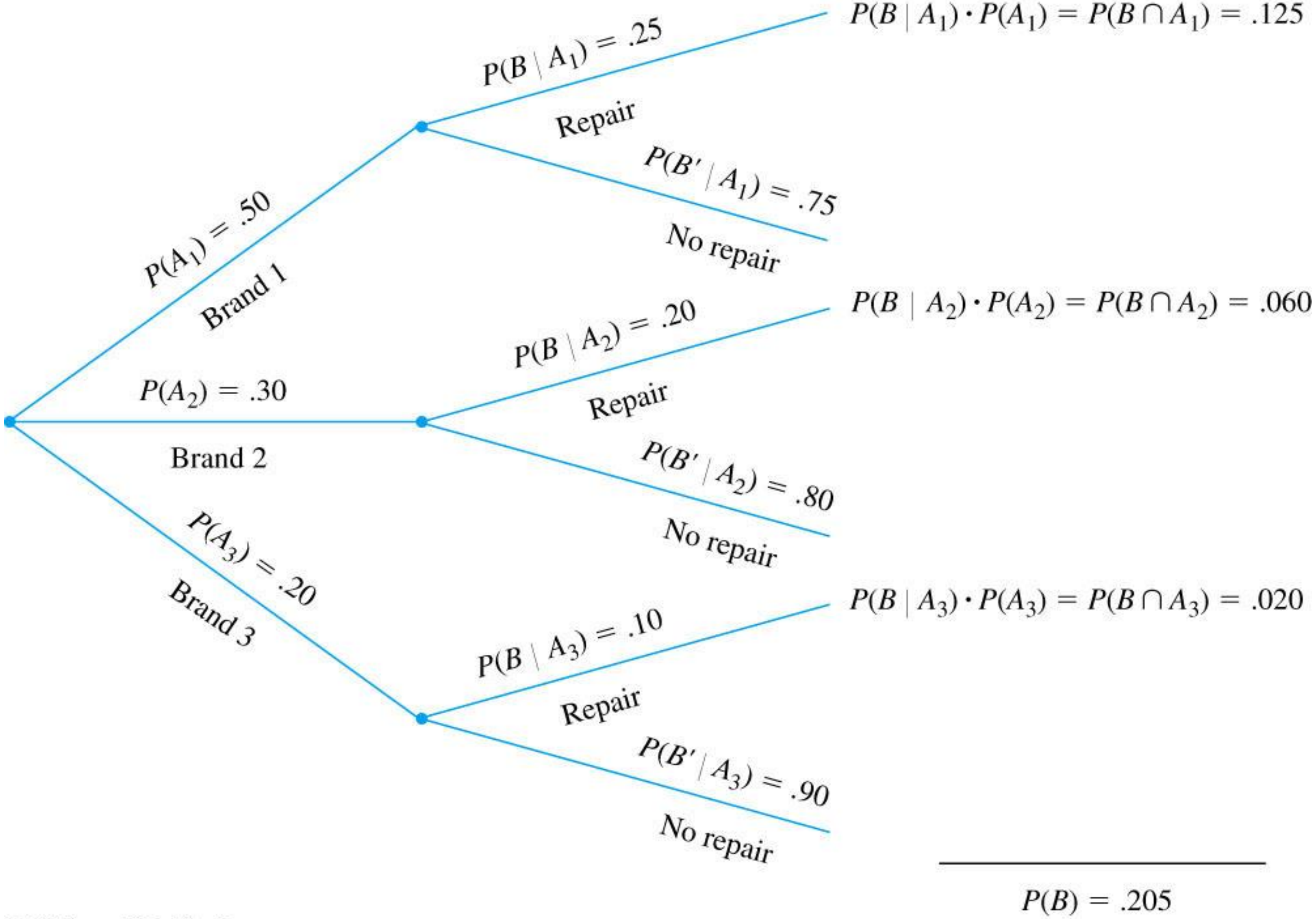
$$=P(A_1 \cap B)+P(A_2 \cap B)+P(A_3 \cap B) = 0.25 \times 0.5 + 0.2 \times 0.3 + 0.1 \times 0.2$$

= 0.125+0.06+0.02=0.205

3. $P(A_1|B)$=$\frac{P(A_1 \cap B)}{P(B)}$=$\frac{0.125}{0.205}$ =0.61

$P(A_2|B)$=$\frac{P(A_2 \cap B)}{P(B)}$=$\frac{0.060}{0.205}$ = 0.29

$P(A_3|B)$=$\frac{P(A_3 \cap B)}{P(B)}$=$\frac{0.020}{0.205}$ = 0.10



$P(A_1) = .50$
Brand 1
$P(A_2) = .30$
Brand 2
$P(A_3) = .20$
Brand 3
$P(B \mid A_1) = .25$
Repair
$P(B' \mid A_1) = .75$
No repair
$P(B \mid A_2) = .20$
Repair
$P(B' \mid A_2) = .80$
No repair
$P(B \mid A_3) = .10$
Repair
$P(B' \mid A_3) = .90$
No repair
$P(B \mid A_1) \cdot P(A_1) = P(B \cap A_1) = .125$
$P(B \mid A_2) \cdot P(A_2) = P(B \cap A_2) = .060$
$P(B \mid A_3) \cdot P(A_3) = P(B \cap A_3) = .020$
$P(B) = .205$

The Law of Total Probability

Let $A_1, \cdots, A_k$ be mutually exclusive and exhaustive events.

Then for any other event B,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_k \cap B)$$

$$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

$$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

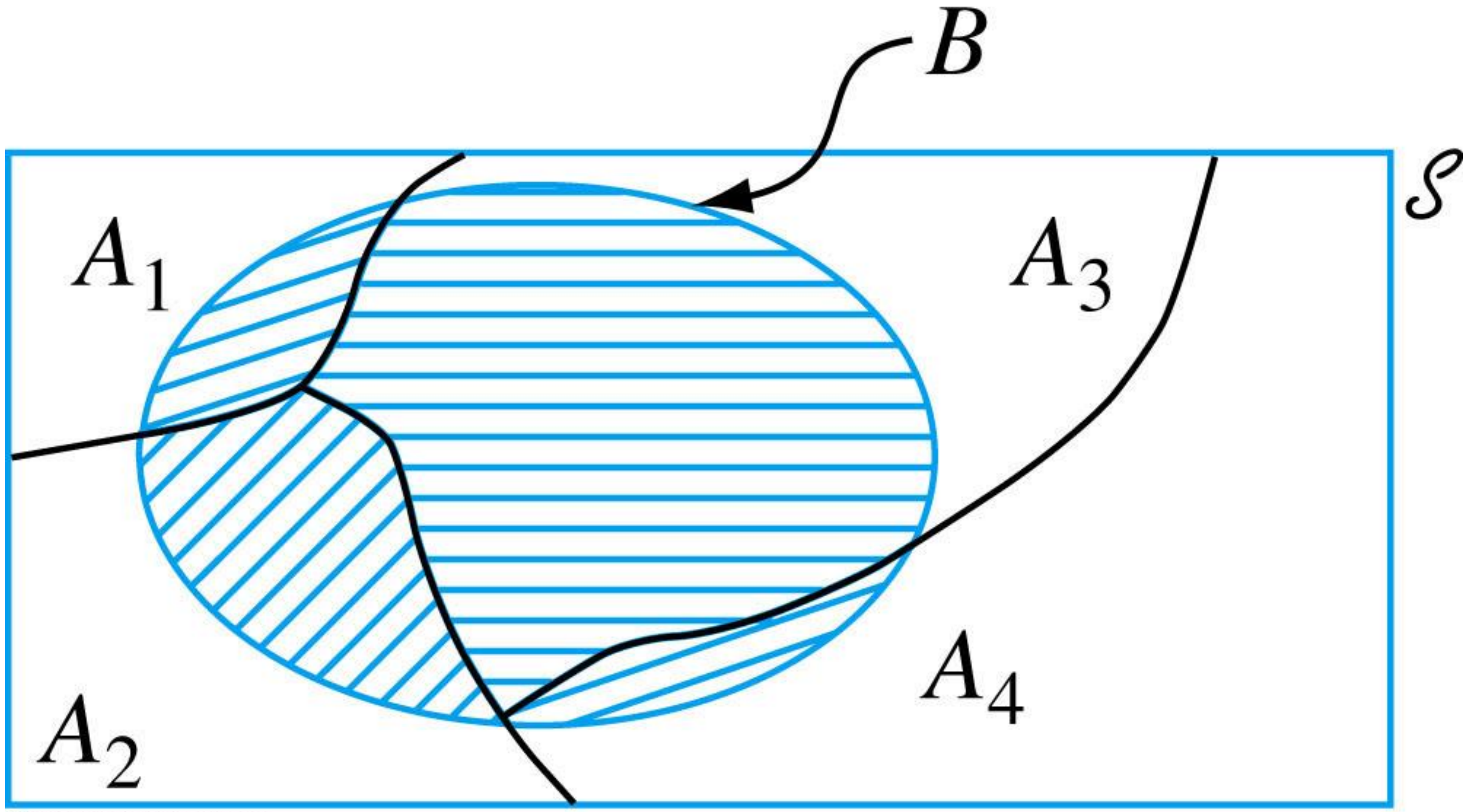


Fig. 2-11, p. 72

Bayes' Theorem

Let A_1, A_2,$\cdots$,A_k be a collection of k mutually exclusive and exhaustive events with *prior probabilities* P(A_i) $(\mathrm{i} = 1, \cdots, \ k)$.

Then for any other event B for which P(B)>0, the *posterior probability* of A_j given that B has occurred is

$$P\left(A_j \middle| B\right) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}\ , \ \mathrm{j=1,\ \cdots,\ k}$$

Example 2.31 Incidence of a rare disease

- Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed.
- When an individual has the disease, the test yields a positive result 99% of the time.
- When an individual does not have the disease, the test will show a positive test result 2% of the time.
- If a randomly selected individual has a positive test result, what is the probability that the individual has the disease?

Sol) Let A_1={individual has the disease}, A_2={individual does not have the disease}, and B={positive test result}.

Then, $P(A_1) = 0.001, P(A_2) = 0.999$, $P(B|A_1) = 0.99$, $P(B|A_2) = 0.02$,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$

$$= 0.99 \times 0.001 + 0.02 \times 0.999 = 0.00099 + 0.01998 = 0.02097$$

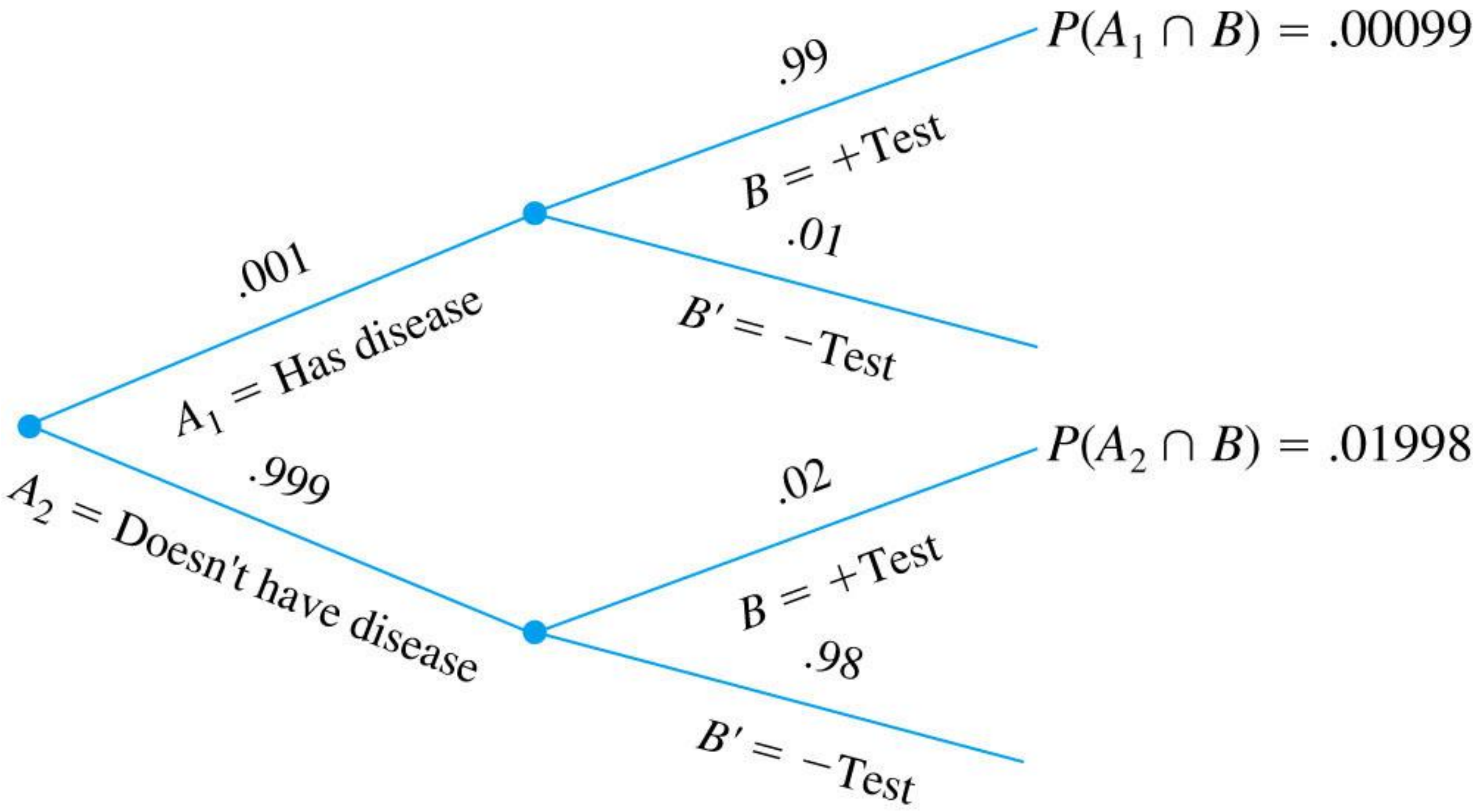
From this we have,

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{0.00099}{0.02097} = 0.047$$

This result seems counterintuitive: the diagnostic test appears very accurate, and we expect that someone with a positive test result is highly likely to have the disease.

But the computed conditional probability is only 0.047.

The probability of having the disease has increased by a multiplicative factor of 47 (from prior 0.001 to posterior 0.047)


.001
A_1 = Has disease
.999
A_2 = Doesn't have disease
.99
B = +Test
.01
B′ = −Test
.02
B = +Test
.98
B′ = −Test
$P(A_1 \cap B) = .00099$
$P(A_2 \cap B) = .01998$

Bayesian Theorem

Lie-detector tests are administered to employees in sensitive positions.

Let + (−) denote the event that the lie-detector reading is positive (negative).

Let T (L) denote the event that the subject is telling the truth (lie).

According to studies of lie-detector reliability, $P(+|L) = 0.88, P(-|T) = 0.86$.

Now suppose that lie-detector tests are administered to screen employees for security reasons and that the vast majority have no reason to lie, so that P(T)=0.99.

A subject produces a positive response on the lie detector.

What is the probability that he is in fact telling the truth?

Sol) P(T)=0.99, P(L)=0.01, $P(+|L) = 0.88, P(-|T) = 0.86$, $P(+|T) = 0.14$.

$P(+) = P(+\cap T) + P(+\cap L) = P(+|T)P(T) + P(+|L)P(L) =$ 0.14x0.99+0.88x0.01

=0.1386+0.0088=0.1474

$P(T|+) = \frac{P(+\cap T)}{P(+)} = \frac{0.1386}{0.1474} = 0.9402$

Independent Events

Two events A and B are independent if

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

or, equivalently,

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

Events $A_1,\ A_2, \cdots, A_n$ are mutually independent if for every $k \geq 2$

and every subset of indices $i_1,\ i_2, \cdots, i_k$,

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_k}\right)$$

Example

Example : Electrical engineers are considering two alternative systems for sending messages.

A message consists of a word that is either a 0 or a 1.

However, because of random noise in the channel, a 1 that could be received as a 0 and vice versa. That is, there is a small probability, p, that

$$P(1 \rightarrow 0) = p$$

$$P(0 \rightarrow 1) = p$$

One scheme is to send a single digit.

A second scheme is to repeat the selected digit three times in succession.

At the receiving end, the majority rule will be used to decode. Compare the probabilities that a transmitted 1 will be received as a 1 under the two schemes.

Answer : for the second scheme, the probability is

$$(1-p)^3 + 3p(1-p)^2$$

When $p = 0.01, (1-p)^3 + 3p(1-p)^2 = 0.999702$

Chapter 3

Chapter 3 – Discrete Random Variables and Probability Distributions

Outline

① Random variables

② Discrete random variables and distributions

③ Expected values of discrete random variables

④ Binomial probability distribution

⑤ Hypergeometric and negative binomial distributions

⑥ Poisson probability distribution

Random Variables

A random variable is any rule that associates a number with each outcome in the sample space S.

Bernoulli experiment - in an experiment with a binary outcome, such as success/failure.

We will define the random variable :

success → 1, failure → 0.

Counts – the total number of times that a particular event occurs.

ex) the number of heads in 20 coin tosses

Measurements –

Binge percentage, duration time of Beethovens Symphony #9

are numeric outcomes on a continuous scale.

Discrete random variables

Definition : A random variable that can only assume distinct values is said to be discrete. Usually these represent a count.

Bernoulli : A Bernoulli experiment provides a 0/1 response

Binomial : A binomial random variable gives the number of successes in n independent, identical trials. Possible values are $0, 1, \cdots, n$.

ex) the number of heads in 20 coin tosses

Geometric : Number of objects tested until a success.

Possible values are $1, 2, 3, \cdots$.

ex) the number of newborn child until a boy (B) is born.

B, GB, GGB, GGGB,

Note that discrete random variables can have a finite range or an infinite range.

Continuous random variables

Definition : A random variable that can assume any value in a finite or infinite interval is said to be continuous.

Measurements : The height above sea level in the United States

The result could be any value in the range [-290, 14500].

Time to failure : The result is any positive number.

the lifetime of an light bulb (0, 10000)

Round-off error : Round-off error is generally modeled as a uniform continuous distribution.

When rounding off to nearest integer, the error is uniform in [-0.5, 0.5].

Probability mass function

A discrete distribution is described by giving its probability mass function(pmf) either as a table or as a function.

Properties :

- For any x, $p(x) = P(X = x)$
- In R, the pmf's for various distributions are functions whose names start with d, such as dbinom, dgeom, dpois.

The probability that the number of heads is 3 when we toss coin 10 times.

> dbinom(3, 10, 0.5) #0.1171875 : $P(X = 3) = \binom{10}{3} 0.5^3 (1 - 0.5)^{10-3} = 0.1172$

> factorial(10)/(factorial(3)*factorial(7))*0.5^3*0.5^7 #0.1171875

- $p(x) \geq 0, -\infty < x < \infty.$
- $\sum_x p(x) = 1$

Probability Distributions for Discrete Random Variables

Example 3.7:

A Department of Statistics has a lab with six computers reserved for statistics majors.

Let X denote the number of these computers that are in use at a particular time of day.

Suppose that the probability distributions of X is as given in the following table.

x	0	1	2	3	4	5	6
p(x)	0.05	0.10	0.15	0.25	0.20	0.15	0.10

ex) the number of newborn child until a boy (B) is born. P(B)=p

Distribution	Functions			
Beta	pbeta	qbeta	dbeta	rbeta
Binomial	pbinom	qbinom	dbinom	rbinom
Cauchy	pcauchy	qcauchy	dcauchy	rcauchy
Chi-Square	pchisq	qchisq	dchisq	rchisq
Exponential	pexp	qexp	dexp	rexp
F	pf	qf	df	rf
Gamma	pgamma	qgamma	dgamma	rgamma
Geometric	pgeom	qgeom	dgeom	rgeom
Hypergeometric	phyper	qhyper	dhyper	rhyper
Logistic	plogis	qlogis	dlogis	rlogis
Log Normal	plnorm	qlnorm	dlnorm	rlnorm
Negative Binomial	pnbinom	qnbinom	dnbinom	rnbinom
Normal	pnorm	qnorm	dnorm	rnorm
Poisson	ppois	qpois	dpois	rpois
Student t	pt	qt	dt	rt
Studentized Range	ptukey	qtukey	dtukey	rtukey
Uniform	punif	qunif	dunif	runif
Weibull	pweibull	qweibull	dweibull	rweibull

Probability computation by R

Consider a standard normal distribution :

$$P(X < 1) = 0.8413$$

```
> pnorm(1)
[1] 0.8413447
```

$$P(-1 < X < 1) = P(X < 1) - P(X < -1) = 0.8413 - (1 - 0.8413) = 0.6826$$

```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
```

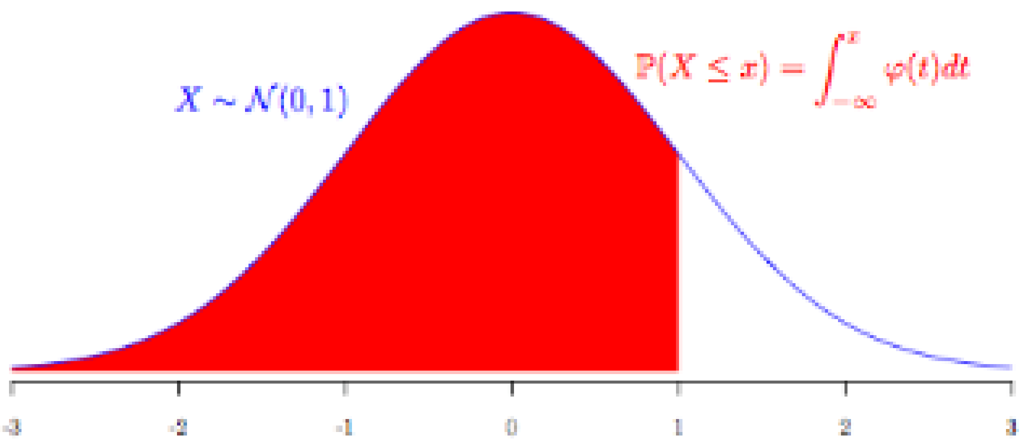
$$P(X < 1) = 0.8413$$

```
> qnorm(0.8413)
[1] 0.9998151
```

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \rightarrow f(1) = \frac{1}{\sqrt{2\pi}} e^{-1^2/2} = 0.2419707$$

```
> 1/sqrt(2*pi)*exp(-1/2)
[1] 0.2419707
> dnorm(1)
[1] 0.2419707
```

Normal Distribution Table



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

Probability Distributions for Discrete Random Variables

Example 3.7:

A Department of Statistics has a lab with six computers reserved for statistics majors.

Let X denote the number of these computers that are in use at a particular time of day.

Suppose that the probability distributions of X is as given in the following table.

x	0	1	2	3	4	5	6
p(x)	0.05	0.10	0.15	0.25	0.20	0.15	0.10

- $P(X \leq 2) = P(X = 0 \; or \; 1 \; or \; 2) = p(0) + p(1) + p(2) = 0.05 + 0.10 + 0.15 = 0.30$

- $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.30 = 0.70$

- $P(2 \leq X \leq 5) = P(X = 2, 3, 4, or \; 5) = 0.15 + 0.25 + 0.20 + 0.15 = 0.75$

- $P(2 < X < 5) = P(X = 3 \; or \; 4) = 0.25 + 0.20 = 0.45$

Parameters of a distribution

Frequently we describe an entire family of distributions that depend upon one or more parameters.

Bernoulli : A Bernoulli probability distribution depends upon the success probability of the trial. If we call this α then the probability function is

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Geometric : The total number of trials until the first "success" occurs.

The parameter is p, the probability of success on a given trial.

$$p(x;p) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

A Parameter of a Probability Distribution – Example 3.12

Starting at a fixed time, observe the gender of each newborn child at a certain hospital until a boy (B) is born.

Let $p = P(B)$ and assume that successive births are independent.

Define the random variable X = number of births observed. Then

$$p(1) = P(X = 1) = P(B) = p$$

$$p(2) = P(X = 2) = P(GB) = P(G)P(B) = (1 - p)p$$

$$p(3) = P(X = 3) = P(GGB) = P(G)P(G)P(B) = (1 - p)^2 p$$

Continuing this way, a general formula emerges :

$$p(x) = \begin{cases} (1 - p)^{x-1} p & \text{if } x = 1, 2, 3, \cdots \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

The parameter p can assume any value between 0 and 1. Expression (2) describes the family of geometric distributions.

Cumulative Distribution Function

Definition : The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with pmf $p(x)$ is

$$F(x) = P(X \leq x)$$

Interpretation : The cdf accumulates the probability to the left of x. For a discrete random variable its graph is a step function.

R functions : In R, the cdf's for various distributions are functions whose names start with p , such as pbinom, pgeom, ppois.

- The probability that the number of heads is less than or equal to 2 when we toss coin 10 times :

$$P(X \leq 2) = \sum_{x=0}^{2} \binom{10}{x} 0.5^x (1-0.5)^{10-x}$$

$$= \binom{10}{0} 0.5^0 (1-0.5)^{10} + \binom{10}{1} 0.5^1 (1-0.5)^9 + \binom{10}{2} 0.5^2 (1-0.5)^8$$

$$= 0.000977 + 0.00977 + 0.0439 = 0.054647$$

```
> pbinom(2, 10, 0.5)

[1] 0.0546875
```

The Cumulative Distribution Function – Example 3.13

A store carries flash drives with either 1GB, 2GB, 4GB, 8GB, or 16GB of memory.

The accompanying table gives the distribution of Y= the amount of memory in a purchased drive:

y	1	2	4	8	16
p(y)	0.05	0.10	0.35	0.40	0.10

Let's first determine $F(y)$ for each of the five possible values of Y:

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = 0.05$$

$$F(2) = P(Y \leq 2) = P(Y = 1\ or\ 2) = p(1) + p(2) = 0.15$$

$$F(4) = P(Y \leq 4) = p(1) + p(2) + p(4) = 0.50$$

$$F(8) = P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = 0.90$$

$$F(16) = P(Y \leq 16) = 1$$

The Cumulative Distribution Function – Example 3.13

y	1	2	4	8	16
p(y)	0.05	0.10	0.35	0.40	0.10

For any other number y, $F(y)$ will equal the value of F at the closest possible value of Y to the left of y . Example

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2) = F(2) = 0.15$$

$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4) = F(4) = 0.50$$

- If y is less than 1, $F(y)$=0; if y is at least 16, $F(y)$=1

The cdf is given by

$$F(y) = \begin{cases} 0 & \text{if } y < 1 \\ 0.05 & \text{if } 1 \leq y < 2 \\ 0.15 & \text{if } 2 \leq y < 4 \\ 0.50 & \text{if } 4 \leq y < 8 \\ 0.90 & \text{if } 8 \leq y < 16 \\ 1 & \text{if } y \geq 16 \end{cases}$$

Cumulative Distribution Function

Definition : The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with pmf $p(x)$ is

$$F(x) = P(X \leq x)$$

Interpretation : The cdf accumulates the probability to the left of x. For a discrete random variable its graph is a step function.

R functions : In R, the cdf's for various distributions are functions whose names start with p , such as pbinom, pgeom, ppois.

- The probability that the number of heads is less than or equal to 2 when we toss coin 10 times :

$$P(X \le 2) = \sum_{x=0}^{2} \binom{10}{x} 0.5^x (1-0.5)^{10-x}$$

$$= \binom{10}{0} 0.5^0 (1-0.5)^{10} + \binom{10}{1} 0.5^1 (1-0.5)^9 + \binom{10}{2} 0.5^2 (1-0.5)^8$$

$$= 0.000977 + 0.00977 + 0.0439 = 0.054647$$

```
> pbinom(2, 10, 0.5)
[1] 0.0546875
```

The Cumulative Distribution Function – Example 3.13

A store carries flash drives with either 1GB, 2GB, 4GB, 8GB, or 16GB of memory.

The accompanying table gives the distribution of Y= the amount of memory in a purchased drive:

y	1	2	4	8	16
p(y)	0.05	0.10	0.35	0.40	0.10

Let's first determine $F(y)$ for each of the five possible values of Y:

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = 0.05$$

$$F(2) = P(Y \leq 2) = P(Y = 1 \ or\ 2) = p(1) + p(2) = 0.15$$

$$F(4) = P(Y \leq 4) = p(1) + p(2) + p(4) = 0.50$$

$$F(8) = P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = 0.90$$

$$F(16) = P(Y \leq 16) = 1$$

The Cumulative Distribution Function – Example 3.13

y	1	2	4	8	16
p(y)	0.05	0.10	0.35	0.40	0.10

For any other number y, $F(y)$ will equal the value of F at the closest possible value of Y to the left of y. Example

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2) = F(2) = 0.15$$

$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4) = F(4) = 0.50$$

- If y is less than 1, $F(y)$=0; if y is at least 16, $F(y)$=1

The cdf is given by

$$F(y) = \begin{cases} 0 & \text{if } y < 1 \\ 0.05 & \text{if } 1 \leq y < 2 \\ 0.15 & \text{if } 2 \leq y < 4 \\ 0.50 & \text{if } 4 \leq y < 8 \\ 0.90 & \text{if } 8 \leq y < 16 \\ 1 & \text{if } y \geq 16 \end{cases}$$

Expected value

Definition : The expected value of a discrete random variable is the weighted sum of the outcomes using the probabilities as weights.

If X is a discrete r.v. with possible values in D then

$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)$ (cf) Discrete r.v. $E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)$

The expected values for the distributions in examples 3.16

x	**1**	**2**	**3**	**4**	**5**	**6**	**7**
p(x)	0.01	0.03	0.13	0.25	0.39	0.17	0.02

```
> prob <- c(0.01, 0.03, 0.13, 0.25, 0.39, 0.17, 0.02)
> sum(prob)
[1] 1
```

> sum(1:7 * prob) # $\sum_{x=1}^{7} x \cdot p(x) = 1 \times 0.01 + 2 \times 0.03 + \cdots + 7 \times 0.02 = 4.57$

[1] 4.57

Expected value

The expected values for the distributions in examples 3.19 :

X= number of children born up to and including the first boy

$$p(x) = \begin{cases} (1-p)^{x-1}p & \text{if } x = 1, 2, 3, \cdots \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

> sum(1:40 * dgeom(0:39, prob=0.5)) # $\sum_{x=1}^{40} x \cdot p(x)$

[1] 2

In R, dgeom(x, prob=0.5) = $(1-p)^x p \qquad x = 0, 1, 2, 3, \cdots$

> dgeom(3, 0.5) # 0.0625

> p(x=4)= (1-0.5)^3*0.5= 0.0625

Expected Values - Examples

A Department of Statistics has a lab with six computers reserved for statistics majors. Let X denote the number of these computers that are in use at a particular time of day.

Suppose that the probability distributions of X is as given in the following table.

x	0	1	2	3	4	5	6
p(x)	0.05	0.10	0.15	0.25	0.20	0.15	0.10

From the pmf of X,

$$\mu = 0 \times p(0) + 1 \times p(1) + 2 \times p(2) + \cdots + 6 \times p(6)$$

$$= 0(0.05) + 1(0.10) + 2(0.15) + 3(0.25) + 4(0.20) + 5(0.15) + 6(0.10)$$

$$= 0 + 0.1 + 0.3 + 0.75 + 0.8 + 0.75 + 0.6 = 3.3$$

Expected Values - Examples

Find the expected value of the following

- $X \sim$ Bernoulli(p) with pmf

$$p(x) = P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & \mathrm{x} = 0 \end{cases}$$

$$E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1 - p) + 1(p) = p$$

Expected Values - Examples

- X= number of children born up to and including the first boy

$$p(x) = P(X = x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, 3, \cdots \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \sum_D x \cdot p(x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1}$$

$$= p \sum_{x=1}^{\infty} x(1-p)^{x-1}$$

$$= p \sum_{x=1}^{\infty} \left[-\frac{d}{dp}(1-p)^x \right]$$ (If $\mathrm{f(x) = h(g(x))}$, then $\mathrm{f'(x) = h'(g(x)) \times g'(x)}$))

$$= -p \left[\frac{d}{dp} \sum_{x=1}^{\infty} (1-p)^x \right]$$ (c.f. $1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}$)

$$= -p \left[\frac{d}{dp} \frac{1-p}{p} \right]$$ (c.f. $(1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{1-p}{p}$)

Expected value of a function of $\boldsymbol{X}$

Definition : If the r.v. X has set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $\mathrm{E}[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Special case $E[aX + b] = a \cdot E(X) + b$ or $\mu_{aX+b} = a \cdot \mu_X$+b

Expected value of a function of $\boldsymbol{X}$

Definition : If the r.v. X has set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $\mathrm{E}[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Special case $E[aX + b] = a \cdot E(X) + b$ or $\mu_{aX+b} = a \cdot \mu_X + \mathrm{b}$

Proof) $E[aX + b] = \sum_{x \in D} (aX + b) \cdot p(x)$

$$= \sum_{x \in D} aX \cdot p(x) + \sum_{x \in D} b \cdot p(x)$$

$$= a \sum_{x \in D} X \cdot p(x) + b \sum_{x \in D} p(x)$$

$$= a \cdot E(X) + b$$

Variance of a distribution

Definition : Let X have pmf $p(x)$ and expected value μ.

Then the variance of X, denoted by $V(X)$ or σ_X^2 or just σ^2, is

$$V(X) = \sum_{x \in D}(x-\mu)^2 \cdot p(x) = \mathrm{E}[(X-\mu)^2]$$

The standard deviation of X is

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

Shortcut : $V(X) = \sigma^2 = \sum_{x \in D} x^2 \cdot p(x) - \mu^2 = \mathrm{E}[X^2] - [E(X)]^2$

$$\begin{aligned} V(X) = \mathrm{E}[(X-\mu)^2] &= \sum_{x \in D}(x-\mu)^2 \cdot p(x) \\ &= \sum_{x \in D}(x^2 - 2\mu x + \mu^2) \cdot p(x) \\ &= \sum_{x \in D} x^2 \cdot p(x) - 2\mu \sum_{x \in D} x \cdot p(x) + \mu^2 \sum_{x \in D} p(x) \\ &= \sum_{x \in D} x^2 \cdot p(x) - 2\mu \cdot \mu + \mu^2 \\ &= \sum_{x \in D} x^2 \cdot p(x) - \mu^2 \end{aligned}$$

Variance of a distribution

Linear combinations :

$$V[aX+b] = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2$$

$$\sigma_{aX+b} = |a| \cdot \sigma_X.$$

These imply that $\sigma_{aX}^2 = a^2 \cdot \sigma_X^2$ and $\sigma_{X+b}^2 = \sigma_X^2$

$$\begin{aligned} V[aX+b] &= \mathrm{E}[(aX+b-(a\mu+b))^2] = \mathrm{E}[a^2(X-\mu)^2] \\ &= \sum_{x\in D} a^2(x-\mu)^2 \cdot p(x) = a^2 \sum_{x\in D} (x-\mu)^2 \cdot p(x) \\ &= a^2 \mathrm{E}[(X-\mu)^2] \\ &= a^2 \cdot \sigma_X^2 \end{aligned}$$

Binomial Probability Distribution

Definition : A binomial experiment satisfies the following conditions

① The experiment consists of n trials

② The trials are identical, and each trial can result in one of two possible outcomes called success (S) or failure (F).

③ The trials are independent, so the outcome on any particular trial does not affect the other outcomes.

④ The probability of success, called p, does not vary from trial to trial

A count of the total number of successes in n trials of a binomial experiment is a binomial random variable with pmf

$$b(x : n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

$\binom{n}{x}$: number of possible occurances xoxxoxoxxo,

oxxxoxoxxo

Binomial distribution in R

Description : density, distribution function, quantile function and random number generation for the binomial distribution with parameters 'size' and 'prob'

dbinom(3, 10, 0.5) , pbinom(3, 10, 0.5) ,qbinom(0.3, 10, 0.5) , rbinom(100, 10, 0.5)

Density function : $\binom{n}{x} p^x (1-p)^{n-x}$

Distribution function : $F(x) = p(X \le x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}$

Quantile function : the value such that the probability of the random variable being less than or equal to that value equals the given probability

$$Q(p) = \inf\{x \in R : p \le F(x)\}$$

Ex) $b(x : 10, 0.5)$: $p(x) = \binom{10}{x} 0.5^x (1-0.5)^{10-x}$

Q(0.3) : 4

$F(3) = \sum_{y=0}^{3} \binom{10}{x} 0.5^x (1-0.5)^{10-x} = 0.172, F(4) = \sum_{y=0}^{4} \binom{10}{x} 0.5^x (1-0.5)^{10-x} = 0.377$

```
> qbinom(0.3, 10, 0.5) # 4
```

Binomial distribution in R

Usage

dbinom(x, size, prob, log=FALSE)

pbinom(x, size, prob, lower.tail=TRUE, log.p=FALSE)

qbinom(p, size, prob, lower.tail=TRUE, log.p=FALSE)

rbinom(n, size, prob)

Arguments

x : vector of possible values

p : vector of probabilities

n : number of random values

size : number of trials

prob : probability of success on each trial

Calculations with binomial distributions

Example 3.32 :

Suppose that 20% of all copies of a textbook fail a certain binding strength test.

Let X, denote the number of books failed the test among 15 randomly selected copies.

Then X has a binomial distribution with n=15 and p=0.2

The probability that at most 8 books fail the test is :

$$F(8) = p(X \leq 8) = \sum_{y=0}^{8} \binom{15}{y} 0.2^{y} (1 - 0.2)^{15-y}$$

> sum(dbinom(0:8, size=15, prob=0.2)) # 0.999215

> pbinom(8, 15, 0.2) # 0.999215

- The probability that exactly 8 books fail is : $p(X = 8)$

> dbinom(8, 15, 0.2) # 0.003454764

> pbinom(8, 15, 0.2)-pbinom(7, 15, 0.2) # 0.003454764

Calculations with binomial distributions

- The probability that at least 8 fail is : $p(8 \le X) = 1 - p(X \le 7)$

```
> sum(dbinom(8:15, 15, 0.2)) # 0.00423975
> pbinom(7.5, 15, 0.2, lower=FALSE) # 0.00423975
> 1-pbinom(7, 15, 0.2) # 0.00423975
```

- The probability that between 4 and 7 books, inclusive, fail is : $p(4 \le X \le 7)$

```
> sum(dbinom(4:7, 15, 0.2)) # 0.3475981
```

- The mean : $E(X) = \sum_{x \in D} x \cdot p(x)$

```
> sum(0:15 * dbinom(0:15, 15, 0.2)) # 3
> mean = 0
> for(i in 0:15){
+   mean = mean + i * dbinom(i, 15, 0.2)
+ }
> mean #3
```

Calculations with binomial distributions

- The variance $:V(X) = \sum_{x \in D}(x-\mu)^2 \cdot p(x)$
- sum((0:15 - 15*0.2)^2 * dbinom(0:15, 15, 0.2)) # 2.4

```
> var = 0
> for(i in 0:15){
+   var = var + (i-15*0.2)^2 * dbinom(i, 15, 0.2)
+ }
> var # 2.4
```

- Mean = $n \cdot p$, variance = $n \cdot p \cdot (1-p)$

```
> c(expected = 15*0.2, variance=15*0.2*0.8)
expected variance
     3.0      2.4
```

Mean and variance of binomial distributions

```
m_binom <- function(n, p){
    mean = 0
    for(i in 0:n){
        mean = mean + i * dbinom(i, n, p)
    }
    return(mean )
}
m_binom(15, 0.2)  # 3

v_binom <- function(n, p){
    var = 0
    for(i in 0:n){
        var = var + (i-n*p)^2 * dbinom(i, n, p)
    }
    return(var)
}
v_binom(15, 0.2) # 2.4
```

Binomial R.V. Mean and Variance

- Mean : $\mu_X = np$
- Variance : $\sigma_X^2 = np(1-p)$

When $X_1, X_2, \cdots, X_n$ are independent and follow Bernoulli distribution,

$$X = X_1 + X_2 + \cdots + X_n$$

follows $\mathrm{Bin}(n, p)$. Since,

$$p(x) = P(X_i = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

$$E(X_i) = 1 \times p + 0(1-p) = p$$

$$Var(X_i) = \sum_{x \in D} (x - \mu)^2 \cdot p(x)$$

$$= (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

Binomial R.V. Mean and Variance

Therefore,

$$
\begin{aligned}
E(X) &= E(X_1 + X_2 + \cdots + X_n) \\
&= E(X_1) + E(X_2) + \cdots + E(X_n) = np \\
Var(X) &= Var(X_1 + X_2 + \cdots + X_n) \\
&= Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p)
\end{aligned}
$$

Poisson distribution

pmf :

$$P(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad x = 0, 1, 2, \cdots \ \lambda > 0$$

R functions :

dpois(x, lambda, log=FALSE)

ppois(x, lambda, lower.tail=TRUE, log.p=FALSE)

qpois(p, lambda, lower.tail=TRUE, log.p=FALSE)

rpois(n, lambda)

Properties : $E(X) = V(X) = \lambda$

E(X) of Poisson distribution

$$E(X) = \sum_{x\geq 0} x \cdot p(x)$$

$$= \sum_{x\geq 0} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x\geq 1} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x\geq 1} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x\geq 0} \frac{\lambda^{x}}{x!}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

Taylor Series Expansions of Exponential Function

$$= \lambda$$

$$(e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots)$$

V(X) of Poisson distribution

$$V(X) = \mathrm{E}[X^2] - [E(X)]^2$$

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

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

$$E[X(X-1)] = \sum_{x\geq 0} x(x-1)\cdot p(x)$$

$$= \sum_{x\geq 0} x(x-1)\cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x\geq 2} x(x-1)\cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x\geq 2} \frac{\lambda^{x-2}}{(x-2)!}$$

$$= \lambda^2 e^{-\lambda} \sum_{x\geq 0} \frac{\lambda^x}{x!}$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$$

Example 3.39

Let X denote the number of creatures captured in a trap. Suppose that X has a Poisson distribution with λ=4.5, so on average traps will contain 4.5 creatures.

The probability that a trap contains exactly five creature is

$$P(X = 5) = \frac{e^{-4.5}4.5^5}{5!}=0.1708$$

```
> dpois(5, 4.5)
[1] 0.1708269
```

The probability that a trap has at most five creatures is

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4.5}4.5^x}{x!}=0.7029$$

```
> sum(dpois(0:5, 4.5))
[1] 0.7029304
> ppois(5, 4.5)
[1] 0.7029304
```

Approximating binomials

Formal statement : In the binomial probability mass function if $n \to \infty$ and $p \downarrow 0$ in such a way that np approaches a finite value λ then

$$b(x; n, p) \to p(x;\ \lambda)$$

Practical application : For n large and p small you can approximate the binomial with the Poisson by setting $\lambda = np$. (This is not necessary in R)

Poisson limit theorem

$$\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}=\frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{when } n\to\infty \text{ and np}=\lambda$$

$$\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}\cong\lim_{n\to\infty}\frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$=\lim_{n\to\infty}\frac{n^x+O(n^{x-1})}{x!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$=\lim_{n\to\infty}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Since

$$\lim_{\mathrm{n}\to\infty}\left(1-\frac{\lambda}{\mathrm{n}}\right)^{\mathrm{n}}=\mathrm{e}^{-\lambda} \text{ and } \lim_{\mathrm{n}\to\infty}\left(1-\frac{\lambda}{\mathrm{n}}\right)^{-\mathrm{x}}=1$$

$$\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}\cong\frac{e^{-\lambda}\lambda^x}{x!}$$

Example 3.40

The probability of any given page containing at least one typographical error is 0.005 and errors are independent from page to page.

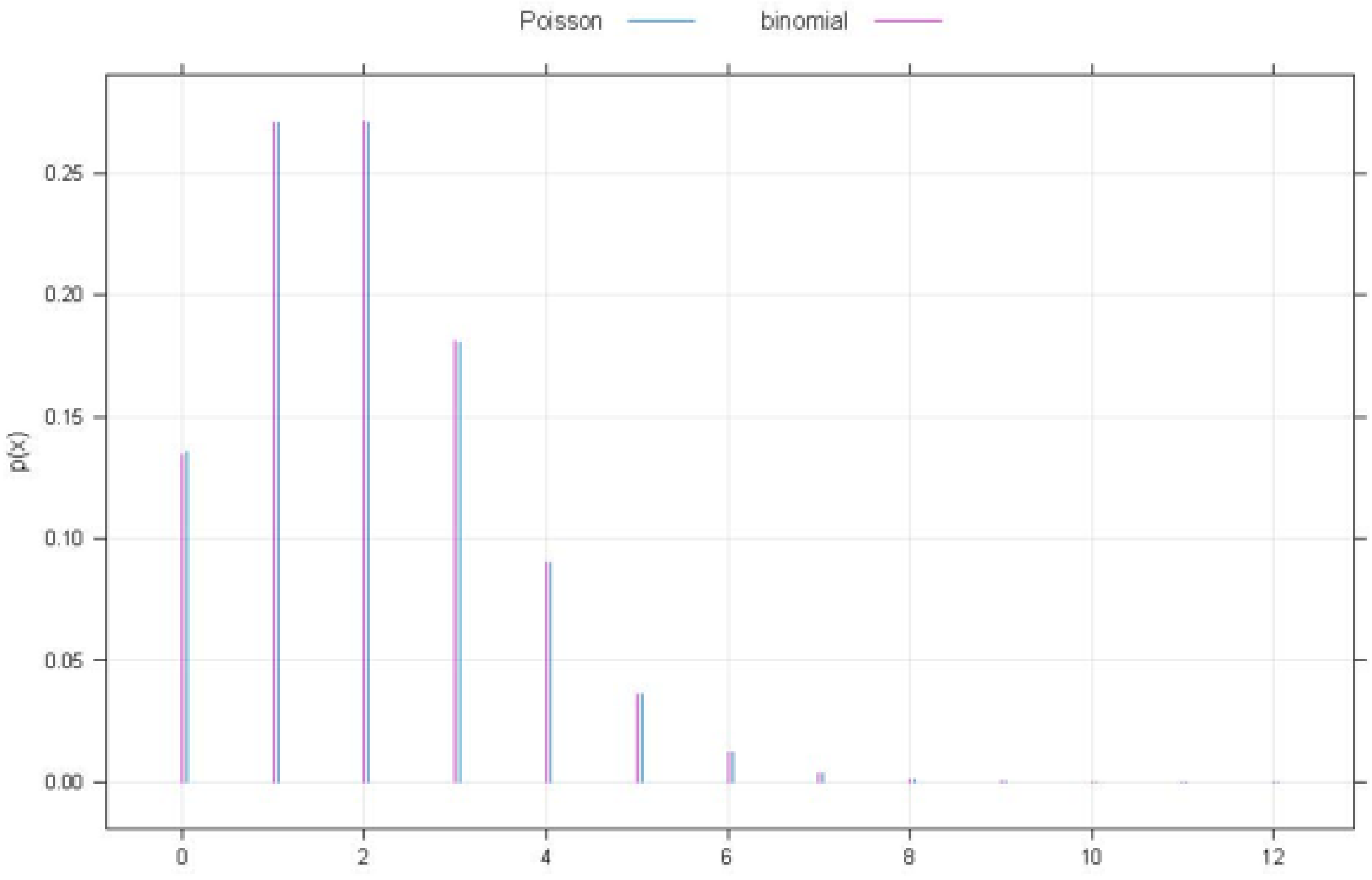
What is the probability that one of its 400-page novels will contain exactly one page with errors?

the number X of pages containing at least one error is a binomial rv with $n = 400$ and $p = 0.005$

$$P(X = 1) = b(1; 400, 0.005) \approx p(1; 2) = \frac{e^{-2}2^1}{1!}=0.270671$$

```
> dbinom(1, 400, 0.005) # 0.2706694
> dpois(1, lambda=2) # 0.2706706
> sum(dbinom(0:3, 400, 0.005)) # 0.8575767
> sum(dpois(0:3, lambda=2)) # 0.8571235
```

Comparison plot of binomial and Poisson approximation



Sums of independent Poisson random variables

- https://llc.stat.purdue.edu/2014/41600/notes/prob1805.pdf
- Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2

Define $\lambda = \lambda_1 + \lambda_2$ and $\mathrm{Z} = X + Y$

$p_z(z) = P(Z = z)$

$= \sum_{j=0}^{z} P(X = j \ \ and\ Y = z - j)$ so $X + Y = z$

$= \sum_{j=0}^{z} P(X = j\ )P(Y = z - j)$ since $X\ and\ Y$ are independent

$= \sum_{j=0}^{z} \frac{e^{-\lambda_1}{\lambda_1}^{j}}{j!} \frac{e^{-\lambda_2}{\lambda_2}^{z-j}}{(z-j)!}$

$= \sum_{j=0}^{z} \frac{z!}{j!(z-j)!} \frac{e^{-\lambda_1}{\lambda_1}^{j} e^{-\lambda_2}{\lambda_2}^{z-j}}{z!}$

$= \frac{e^{-\lambda}}{z!} \sum_{j=0}^{z} \binom{Z}{j} {\lambda_1}^{j} {\lambda_2}^{z-j}$

$= \frac{e^{-\lambda}}{z!} (\lambda_1 + \ \lambda_2)^{z} = \frac{e^{-\lambda}\lambda^{z}}{z!}$ using binomial expansion

So $\mathrm{Z} = X + Y$ follows Poisson distribution.

- When X and Y follows a uniform distribution,

$$p(x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6$$

$Z = X + Y$ has following distribution, which is not uniform

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$\begin{aligned}\mathrm{P}(Z = 4) &= P(X = 1\ and\ Y = 3) + P(X = 2\ and\ Y = 2) + P(X = 3\ and\ Y = 1)\\ &= P(X = 1)P(Y = 3) + P(X = 2)P(Y = 2) + P(X = 3)P(Y = 1)\\ &= \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} + \frac{1}{6} \times \frac{1}{6} = \frac{3}{36}\end{aligned}$$

Binomial R.V. Mean and Variance

- Mean : $\mu_X = np$
- Variance : $\sigma_X^2 = np(1-p)$

When $X_1, X_2, \cdots, X_n$ are independent and follow Bernoulli distribution,

$$X = X_1 + X_2 + \cdots + X_n$$

1 0 1 0 0 0 1 0 0 0

follows $\text{Bin}(n, p)$. Since,

$$p(x) = P(X_i = x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

$$E(X_i) = 1 \times p + 0(1-p) = p$$

$$Var(X_i) = \sum_{x \in D} (x-\mu)^2 \cdot p(x)$$

$$= (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

Binomial Probability Distribution

Definition : A binomial experiment satisfies the following conditions

① The experiment consists of n trials

② The trials are identical, and each trial can result in one of two possible outcomes called success (S) or failure (F).

③ The trials are independent, so the outcome on any particular trial does not affect the other outcomes.

④ The probability of success, called p, does not vary from trial to trial

A count of the total number of successes in n trials of a binomial experiment is a binomial random variable with pmf

$$b(x : n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

$\binom{n}{x}$: number of possible occurances XOXXOXOXXO,

OXXXOXOXXO

Sums of independent Poisson random variables

- https://llc.stat.purdue.edu/2014/41600/notes/prob1805.pdf
- Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2

Define $\lambda = \lambda_1 + \lambda_2$ and $\mathrm{Z} = X + Y$

$$p_z(z) = P(Z = z)$$

$$= \sum_{j=0}^{z} P(X = j \ \ and\ Y = z - j) \quad \text{so } X + Y = z$$

$$= \sum_{j=0}^{z} P(X = j\ )P(Y = z - j) \quad \text{since } X\ and\ Y \text{ are independent}$$

$$= \sum_{j=0}^{z} \frac{e^{-\lambda_1} {\lambda_1}^j}{j!} \frac{e^{-\lambda_2} {\lambda_2}^{z-j}}{(z-j)!}$$

$$= \sum_{j=0}^{z} \frac{z!}{j!(z-j)!} \frac{e^{-\lambda_1} {\lambda_1}^j e^{-\lambda_2} {\lambda_2}^{z-j}}{z!}$$

$$= \frac{e^{-\lambda}}{z!} \sum_{j=0}^{z} \binom{Z}{j} {\lambda_1}^j {\lambda_2}^{z-j}$$

$$= \frac{e^{-\lambda}}{z!} (\lambda_1 + \ \lambda_2)^z = \frac{e^{-\lambda} \lambda^z}{z!} \quad \text{using binomial expansion}$$

So $\mathrm{Z} = X + Y$ follows Poisson distribution.

Moment generating function of independent Poisson random variables**

- https://www.probabilitycourse.com/chapter6/6_1_3_moment_functions.php (example 6.6)

Theorem. Moment generating function of a Poisson random variable X is

$$M_X(t) = E(e^{tX}) = e^{\lambda(e^t - 1)} \quad for\ -\infty < t < \infty$$

Proof)

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

https://en.wikipedia.org/wiki/Moment-generating_function**

Distribution	Moment-generating function $M_X(t)$	Characteristic function $\varphi(t)$
Degenerate δ_a	e^{ta}	e^{ita}
Bernoulli $P(X=1)=p$	$1-p+pe^t$	$1-p+pe^{it}$
Geometric $(1-p)^{k-1}\,p$	$\frac{pe^t}{1-(1-p)e^t}$ $\forall t<-\ln(1-p)$	$\frac{pe^{it}}{1-(1-p)\,e^{it}}$
Binomial $B(n,p)$	$\left(1-p+pe^t\right)^n$	$\left(1-p+pe^{it}\right)^n$
Negative Binomial $NB(r,p)$	$\frac{(1-p)^r}{(1-pe^t)^r}$	$\frac{(1-p)^r}{(1-pe^{it})^r}$
Poisson $Pois(\lambda)$	$e^{\lambda(e^t-1)}$	$e^{\lambda(e^{it}-1)}$
Uniform (continuous) $U(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$

Important Properties of moment generating function**

- https://stats.stackexchange.com/questions/34956/proof-that-moment-generating-functions-uniquely-determine-probability-distributi
- If two distributions have the same moment generating function, then they are identical at almost all points. That is, if for all values of t

$$M_X(t) = M_Y(t),$$

Then

$$F_X(x) = F_Y(y)$$

for all values of x (or equivalently X and Y have the same distribution)

Important Properties of moment generating function**

- http://www.milefoot.com/math/stat/rv-moments.htm
- Calculations of moments

$$E(X^n) = M_X^{(n)}(0) = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}$$

$$M_X(t) = E(e^{tX}) = \sum_{x \in D} e^{tx} \cdot p(x)$$

$$= \sum_{x \in D} \left(1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^nx^n}{n!} + \cdots\right) \cdot p(x)$$

$$= 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \frac{t^3E(X^3)}{3!} + \cdots + \frac{t^nE(X^n)}{n!} + \cdots$$

$$M_X{}'(t) = \mathrm{E(X)} + \frac{2tE(X^2)}{2!} + \frac{3t^2E(X^3)}{3!} + \cdots + \frac{nt^{n-1}E(X^n)}{n!} + \cdots \qquad M_X{}'(t)|_{t=0} = E(X)$$

$$M_X{}''(t) = E(X^2) + \frac{3\times 2tE(X^3)}{3!} + \cdots + \frac{n(n-1)t^{n-2}E(X^n)}{n!} + \cdots \qquad M_X{}''(t)|_{t=0} = E(X^2)$$

Important Properties of moment generating function**

- Calculations of moments

$$E(X^n) = M_X^{(n)}(0) = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}$$

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

$$= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^nx^n}{n!} + \cdots\right) f(x)dx$$

$$= 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \frac{t^3E(X^3)}{3!} + \cdots + \frac{t^nE(X^n)}{n!} + \cdots$$

$$M_X{}'(t) = \mathrm{E(X)} + \frac{2tE(X^2)}{2!} + \frac{3t^2E(X^3)}{3!} + \cdots + \frac{nt^{n-1}E(X^n)}{n!} + \cdots \qquad M_X{}'(t)|_{t=0} = E(X)$$

$$M_X{}''(t) = E(X^2) + \frac{3\times 2tE(X^3)}{3!} + \cdots + \frac{n(n-1)t^{n-2}E(X^n)}{n!} + \cdots \qquad M_X{}''(t)|_{t=0} = E(X^2)$$

Mean and variance of Poisson distribution**

- https://online.stat.psu.edu/stat414/node/83/
- The mean of a Poisson random variable X is λ

$$M_X(t) = e^{\lambda(e^t-1)}$$

$$E(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.[e^{\lambda(e^t-1)}\lambda e^t]\right|_{t=0}$$

$$= e^{\lambda(e^0-1)}\lambda e^0 = \lambda$$

- The variance of a Poisson random variable X is λ

$$E(X^2) = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \left.\frac{d}{dt}[e^{\lambda(e^t-1)}\lambda e^t]\right|_{t=0}$$

$$= \left.[e^{\lambda(e^t-1)}\lambda e^t\lambda e^t + e^{\lambda(e^t-1)}\lambda e^t]\right|_{t=0}$$

$$= \left.[e^{\lambda(e^0-1)}\lambda e^0\lambda e^0 + e^{\lambda(e^0-1)}\lambda e^0]\right|_{t=0} = \lambda^2 + \lambda$$

$$V(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Sums of independent random variables**

- Let $X_1, X_2, \ldots, X_n$ are n independent random variables, and the random variable Y is defined as

$$Y = X_1 + X_2 + \cdots + X_n$$

Then

$$\begin{aligned} M_Y(t) &= E\left[e^{t(X_1 + X_2 + \cdots + X_n)}\right] \\ &= E[e^{tX_1} e^{tX_2} \cdots e^{tX_n}] \\ &= E[e^{tX_1}] E[e^{tX_2}] \cdots E[e^{tX_n}] \\ &= M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) \end{aligned}$$

Sums of independent Poisson random variables**

- https://proofwiki.org/wiki/Sum_of_Independent_Poisson_Random_Variables_is_Poisson
- Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2

$$M_X(t) = e^{\lambda_1(e^t - 1)}$$

$$M_Y(t) = e^{\lambda_2(e^t - 1)}$$

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= e^{\lambda_1(e^t-1)}\, e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

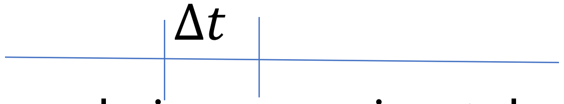
- This is the moment generating function of a random variable with distribution Poisson($\lambda_1 + \lambda_2$)

So, $M_{X+Y} \sim$ Poisson($\lambda_1 + \lambda_2$)

Poisson process
(https://www.pp.rhul.ac.uk/~cowan/stat/notes/PoissonNote.pdf)

A Poisson process is generated by discrete events that occur over time (or some other continuum such as distance) subject to the conditions

① There exists a parameter $\lambda > 0$ such that, for short time intervals of length Δt, the probability of exactly one event is approximately $\lambda \cdot \Delta t$



② The probability of more than one event in very short intervals is approximately zero.

③ The number of event in an interval is independent of the number of events prior to this interval. That is, the process has “no memory”. We also say that the process is restartable.

The parameter λ is called the rate of the process. The number of events in an interval of length t has a Poisson distribution with parameter λt.

$$P_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Poisson Process

- https://www.probabilitycourse.com/chapter11/11_1_2_basic_concepts_of_the_poisson_process.php
- When certain events happen at a certain rate, but completely at random

Example

- The number of car accidents at a site or in an area
- The number of earthquakes in a certain area
- The requests for individual documents on a web server

ex) Your web server gets an average of 2 requests each second

What's the probability of exactly 5 requests in a second?

$$P(X = 5) = \frac{e^{-2}2^5}{5!} \approx 0.0361$$

- https://web.stanford.edu/class/archive/cs/cs109/cs109.1178/lectures/08-poisson.pdf

p.31, p.32

Poisson Process

- ex) The earthquake

  There are an average of 2.8 major earthquakes in the world each year.

  What's the probability of >1 major earthquakes next year?

$$P(X > 1) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \frac{e^{-2.8}2.8^0}{0!} - \frac{e^{-2.8}2.8^1}{1!} = 1 - 0.06 - 0.17 = 0.77$$

- ex) The number of customers arriving at a grocery store can be modeled by a Poisson process with intensity $\lambda = 10$ customers per hour.

  Find the probability that there are 2 customers between 10:0 and 10:20.

Sol) 10 cutomers/hour → 10/3 customers/(1/3 hour) $X \sim Poisson(\frac{10}{3})$

$$P(X = 2) = \frac{e^{-10/3}(\frac{10}{3})^2}{2!} \approx 0.2$$

Example 3.42

Suppose pulses arrive at a counter at an average rate of six per minute, so that λ=6.

To find the probability that in a 0.5-min interval at least one pulse is received, note that the number of pulses in such an interval has a Poisson distribution with parameter $\lambda t = 6(0.5) = 3$.

Then with $X =$ the number of pulses received in the 30-sec interval

Sol) $P(1 \leq X) = 1 - P(X = 0) = 1 - \frac{e^{-3}3^0}{0!}$=0.950

Poisson Process**

- The counting process $N(t)$: the number of arrivals from time 0 to time t
- Let λ be fixed. $N(t)$ is called a Poisson process with rates λ if all the following conditions hold.

1. $N(0) = 0$:

2. $N(t)$ has independent increments:

3. The number of arrivals in any interval of length $t > 0$ has Poisson(λt) distribution

$$P_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Second Definition of the Poisson Process**

- The counting process $N(t)$: the number of arrivals from time 0 to time t
- Let λ be fixed. $N(t)$ is called a Poisson process with rates λ if all the following conditions hold.

1. $N(0) = 0$:
2. $N(t)$ has independent and stationary increments
3. We have

$$P(N(\Delta) = 0) = 1 - \lambda\,\Delta + \mathrm{o}(\Delta)$$

$$P(N(\Delta) = 1) = \lambda\,\Delta + \mathrm{o}(\Delta)$$

$$P(N(\Delta) \geq 2) = \mathrm{o}(\Delta)$$

Here $\mathrm{o}(\Delta)$ shows a function that is negligible compared to Δ, as $\Delta \rightarrow 0$. More precisely,

$\mathrm{g}(\Delta) = \mathrm{o}(\Delta)$ means that

$$\lim_{\Delta \to 0} \frac{\mathrm{g}(\Delta)}{\Delta} = 0 \qquad \text{(ex: } \mathrm{g}(\Delta) = \Delta^2\text{)}$$

Second Definition of the Poisson Process**

$$N(t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

$$P(N(\Delta) = 0) = e^{-\lambda\Delta} = 1 - \lambda\Delta + \frac{\lambda^2}{2}\Delta^2 - \cdots \text{ (Taylor series)}$$

If Δ is small,

$$P(N(\Delta) = 0) = 1 - \lambda\Delta + \mathrm{o}(\Delta)$$

$$P(N(\Delta) = 1) = e^{-\lambda\Delta}\lambda\Delta = \lambda\Delta\left(1 - \lambda\Delta + \frac{\lambda^2}{2}\Delta^2 - \cdots\right)$$

$$= \lambda\Delta + \left(-\lambda^2\Delta^2 + \frac{\lambda^3}{2}\Delta^3 - \cdots\right) = \lambda\Delta + \mathrm{o}(\Delta)$$

$$P(N(\Delta) \geq 2) = 1 - P(N(\Delta) = 0) - P(N(\Delta) = 1)$$

$$= \mathrm{o}(\Delta)$$

Chapter 4

Chapter 4 – Continuous Random Variables and Probability Distributions

Outline

① Continuous random variables and probability density functions

② The Cumulative Distribution Function and Expected Values

③ The Normal Distribution

④ The Gamma Distributions and its relatives

⑤ Probability plots

Continuous Random Variables

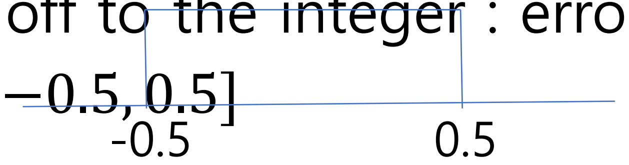
Definition A random variable that can (theoretically) assume any value in a finite or infinite interval is said to be continuous.

Measurements : Let X be the depth measurement at a randomly chosen locations of a lake. Then X is a continuous random variable

Time to failure : The result is potentially any positive number.

Round-off error : Round-off error in calculations is generally modeled as a uniform continuous distribution

When rounded off to the integer : error is uniformly distributed in $[-0.5, 0.5]$



Probability density function

Definition : A probability density function (pdf) of a continuous random variable X is a function $f(x)$ such that for any two numbers $a \leq x \leq b$,

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

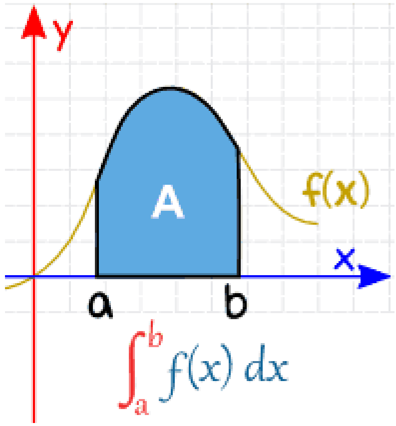
That is, the probability that X takes on a value in the interval $[a, b]$ is the area under the graph of the density function. (ex : P(height ∈ $(170, 175)$))

For any number $c, P(X = c) = 0$.

Conditions for pdf :

$f(x) \geq 0$ for all x

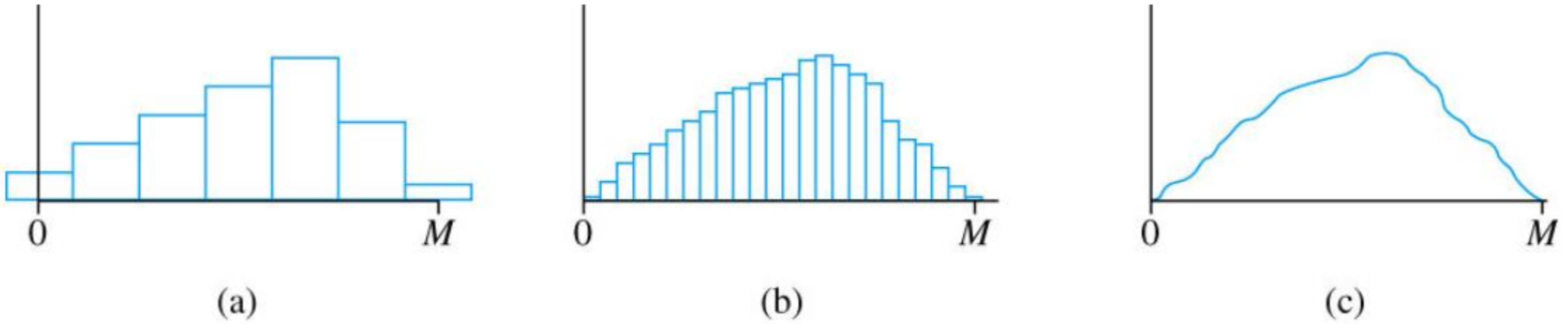
$\int_{-\infty}^{\infty} f(x)dx = 1$



Definite Integral

Motivating probability density function

- The probability density function can be viewed as a limit of discrete histograms.
- Consider the lake depth measurement example.
- We “discretize” X by measuring the depth to the nearest meter, nearest centimeter, and so on.
- As we measure depth more and more finely, the resulting sequence of histograr



Uniform distribution

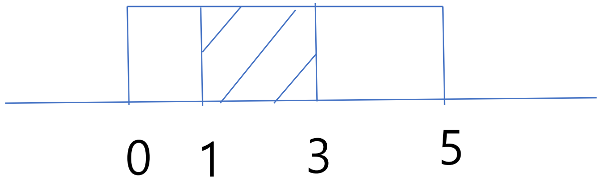
A continuous random variable X is said to have a uniform distribution on the interval $[A, B]$ if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & if\ A \le x \le B \\ 0 & \text{otherwise} \end{cases}$$

Example : Assume the waiting time at a bus stop is uniformly distributed on the interval [0, 5]. The probability that it is between 1 and 3 minutes is

$$P(1 \le X \le 3) = \int_1^3 f(x)dx = \int_1^3 \frac{1}{5} dx = \left[\frac{x}{5}\right]_1^3 = \frac{2}{5} \qquad \left(\int x^a dx = \frac{x^{a+1}}{a+1} + c\right)$$

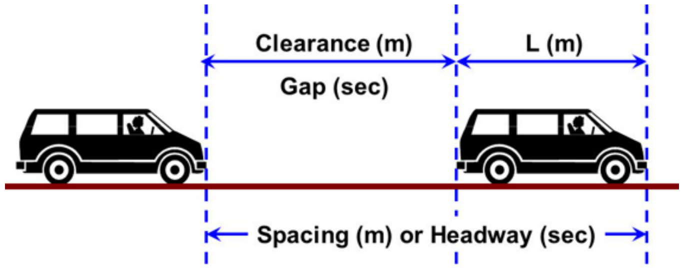
```
> punif(3, min=0, max=5)-punif(1, min=0, max=5)
[1] 0.4
> diff(punif(c(1,3), min=0, max=5))
[1] 0.4
```



Example 4.5

Let X be the time headway for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The pdf of X can be approximated by

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$



This is a density function since it is non-negative, and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0.5}^{\infty} 0.15e^{-0.15(x-0.5)}dx$$

$$= \left[0.15\frac{e^{-0.15(x-0.5)}}{-0.15}\right]_{0.5}^{\infty} = \left[-e^{-0.15(x-0.5)}\right]_{0.5}^{\infty} \qquad \left(\int e^{ax}dx = \frac{e^{ax}}{a} + c\right)$$

$$= 0 - (-e^{0}) = 1$$

The probability that headway time is at most 5 sec is

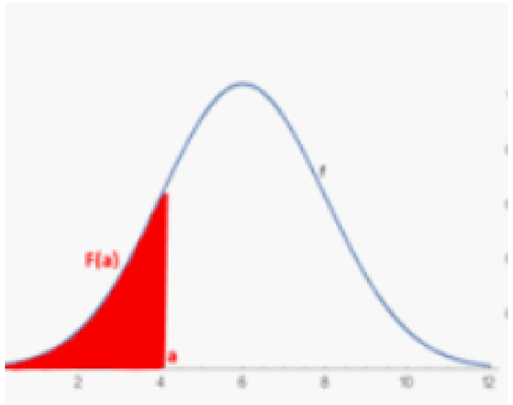
$$P(X \leq 5) = \int_{0.5}^{5} 0.15e^{-0.15(x-0.5)}dx = \left[-e^{-0.15(x-0.5)}\right]_{0.5}^{5}$$

$$= -e^{-0.15(5-0.5)} + 1 = 1 - e^{-0.675} = 0.491$$

The cumulative distribution function

The *cumulative distribution function* $F(X)$ for a continuous random variable X is defined for every number x by

$$F(X) = \mathrm{P}(\mathrm{X} \leq x) = \int_{-\infty}^{x} f(y)dy$$



For each x, $F(X)$ is the area under the density curve to the left of x.

From this we see that $f(x) = F'(x)$ at every x at which $F'(x)$ exists.

Leibniz integration rule :

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t,x)dt = f(b(x),x)\cdot\frac{d}{dx}\,b(x) - f(a(x),x)\cdot\frac{d}{dx}\,a(x) + \int_{a(x)}^{b(x)}\frac{\partial}{\partial x}f(t,x)dt$$

or $$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t)dt = f\big(b(x)\big)\cdot\frac{d}{dx}\,b(x) - f\big(a(x)\big)\cdot\frac{d}{dx}\,a(x) + \int_{a(x)}^{b(x)}\frac{d}{dx}f(t)dt$$

$$F'(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt = f(x)\cdot\frac{d}{dx}\,x - f(-\infty)\cdot\frac{d}{dx}\,(-\infty) + \int_{-\infty}^{x}\frac{d}{dx}f(t)dt = f(x)$$

Computing probabilities using the cdf

Example : Let X have a uniform distribution on $[A, B]$. For $A \leq x \leq B$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A} dy = \left[\frac{y}{B-A}\right]_{A}^{x} = \frac{x-A}{B-A} \qquad f(x; A, B) = \begin{cases} \frac{1}{B-A} & if\ A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

For $x \leq A, F(x) = 0$. For $x \geq B, F(x) = 1$.

If we are given this $F(x)$ to begin with, we can get $f(x)$ by taking the derivative.

Let X be a continuous r.v. with pdf $f(x)$ and cdf $F(x)$. Then for any number a,

$$\mathrm{P}(\mathrm{X} \geq a) = 1 - \mathrm{P}(\mathrm{X} \leq a) = 1 - F(a)$$

And for any two number a and b with $a < b$

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

Example 4.7

Suppose the pdf of the magnitude X of a dynamic load on a bridge is given by

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$$

For any number between 0 and 2,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y\right) dy = \left[\frac{y}{8} + \frac{3}{16}y^2\right]_0^x = \frac{x}{8} + \frac{3}{16}x^2 \qquad \left(\int x^a dx = \frac{x^{a+1}}{a+1} + c\right)$$

Thus

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$

Example 4.7

The probability that the load is between 1 and 1.5 is $\left(F(x) = \frac{x}{8} + \frac{3}{16}x^2\right)$

$$P(1 \le X \le 1.5) = F(1.5) - F(1)$$

$$= \left(\frac{1.5}{8} + \frac{3}{16}1.5^2\right) - \left(\frac{1}{8} + \frac{3}{16}1^2\right) = 0.297$$

The probability that the load exceeds 1 is

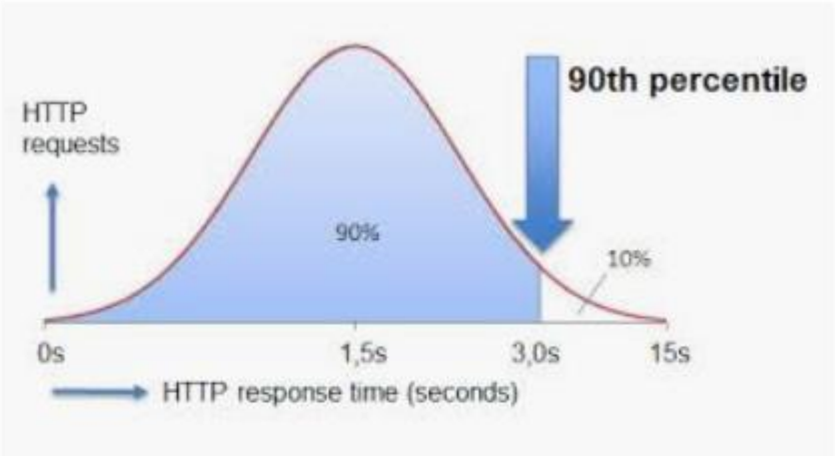
$$P(X > 1) = 1 - F(1) = 1 - \left(\frac{1}{8} + \frac{3}{16}1^2\right) = 0.688$$

Percentiles of a continuous distribution

- Let p be a number between 0 and 1.
- The (100p)th percentile of the distribution of a continuous random variable X, denoted by $\eta(p)$, is defined by

$$p = \mathrm{F}(\eta(p)) = \mathrm{P}(\mathrm{X} \leq \eta(p))$$

- That is, $\eta(p)$ is that value such that $100p\%$ of the area under the graph of $f(x)$ lies to the left of $\eta(p)$.
- The median $\tilde{\mu}$ is the 50th percentile.



Example 4.9

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous r.v. X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & if\ 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- The cdf of sales for any x between 0 and 1 is

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} \frac{3}{2}(1 - y^2)dy = \frac{3}{2}\left[y - \frac{y^3}{3}\right]_0^x = \frac{3}{2}(x - \frac{x^3}{3})$$

- The (100p)th percentile of this distribution satisfies the equation

$\left(\int x^a dx = \frac{x^{a+1}}{a+1} + c\right)$

$$p = \mathrm{F}\big(\eta(p)\big) = \frac{3}{2}(\eta(p) - \frac{\eta(p)^3}{3})$$

- To obtain the median (50th percentile, p=0.50), solve the equation

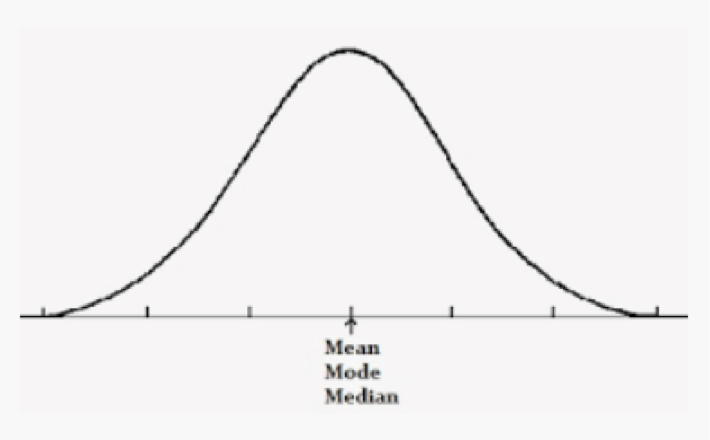
$$0.5 = \frac{3}{2}\left(\eta - \frac{\eta^3}{3}\right) \rightarrow 3\eta - \eta^3 = 1 \rightarrow \eta^3 - 3\eta + 1 = 0$$

Mean of a continuous random variable

- A different characterization of the center of the distribution is the expected value or mean of X

$$\mu_X = \mathrm{E(X)} = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

- A symmetric continuous distribution – which means that the density curve to the left of some point is a mirror image of the density curve to the right of that point – has both median $\tilde{\mu}$ and mean μ_X equal to the point of symmetry.



Example 4.9

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & if\ 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The cdf is then, for $0 \le x \le 1$

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)dy = \frac{3}{2}\left[y - \frac{y^3}{3}\right]_0^x = \frac{3}{2}(x - \frac{x^3}{3})$$

The (100p)th percentile satisfies

$$p = \mathrm{F}\big(\eta(p)\big) = \frac{3}{2}(\eta(p) - \frac{\eta(p)^3}{3})$$

Example 4.9 (cont'd)

For the median $\tilde{\mu}$ $(p = 0.5)$, the equation is $\tilde{\mu}^3 - 3\tilde{\mu} + 1 = 0$. The solution is $\tilde{\mu} = 0.347$.

$$0.5 = \frac{3}{2}\left(\tilde{\mu} - \frac{\tilde{\mu}^3}{3}\right) \rightarrow 1 = 3\tilde{\mu} - \tilde{\mu}^3$$

The mean is

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)dx \ = \frac{3}{2}\int_0^1 (x - x^3)dx$$

$$= \frac{3}{2}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1 = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{4}\right) = \frac{3}{8}$$

Polynomial Equation Calculator :

https://www.symbolab.com/solver/polynomial-equation-calculator/

Solving systems of nonlinear equations with Broyden or Newton

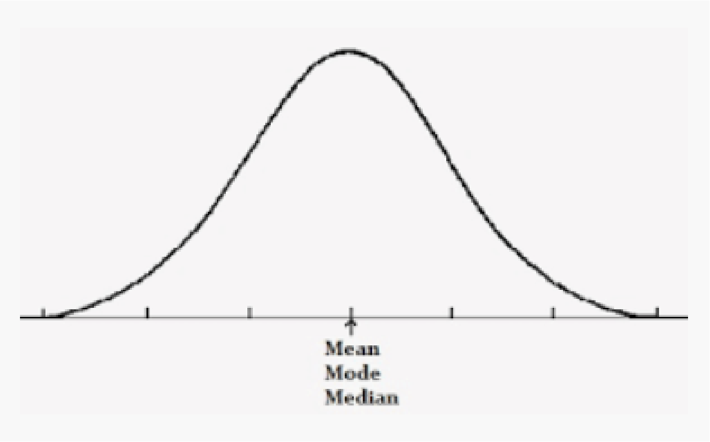
```
> install.packages("nleqslv")
> library(nleqslv)
> target <- function(x) {
+     z = x[1]^3-3*x[1]+1
+}

# Usage
> xstart <- c(0.1)

> nleqslv(xstart, target, method="Newton")
$x
[1]  0.3472964
```

Mean of a continuous random variable

- A different characterization of the center of the distribution is the expected value or mean of X

$$\mu_X = \mathrm{E(X)} = \int_{-\infty}^{\infty} x \cdot f(x)dx \qquad \text{cf) Discrete r.v. : } E[X] = \sum_{x \in D} x \cdot p(x)$$

- A symmetric continuous distribution – which means that the density curve to the left of some point is a mirror image of the density curve to the right of that point – has both median $\tilde{\mu}$ and mean μ_X equal to the point of symmetry.



Example 4.9

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & if\ 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The cdf is then, for $0 \le x \le 1$

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)dy = \frac{3}{2}\left[y - \frac{y^3}{3}\right]_0^x = \frac{3}{2}(x - \frac{x^3}{3})$$

The (100p)th percentile satisfies

$$p = \mathrm{F}\big(\eta(p)\big) = \frac{3}{2}(\eta(p) - \frac{\eta(p)^3}{3})$$

Example 4.9 (cont'd)

For the median $\tilde{\mu}$ $(p = 0.5)$, the equation is $\tilde{\mu}^3 - 3\tilde{\mu} + 1 = 0$. The solution is $\tilde{\mu} = 0.347$.

$$0.5 = \frac{3}{2}\left(\tilde{\mu} - \frac{\tilde{\mu}^3}{3}\right) \rightarrow 1 = 3\tilde{\mu} - \tilde{\mu}^3$$

The mean is

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)dx = \frac{3}{2}\int_0^1 (x - x^3)dx$$

$$= \frac{3}{2}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_0^1 = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{4}\right) = \frac{3}{8}$$

Polynomial Equation Calculator :

https://www.symbolab.com/solver/polynomial-equation-calculator/

Expected value of a function of a rv

If X is a continuous rv with pdf $f(x)$ and $h(X)$ is any function of X, then

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

For $h(X)$ a linear function, $E(aX + b) = aE(X) + b.$

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) \cdot f(x)dx$$

$$= \int_{-\infty}^{\infty} ax \cdot f(x)dx + \int_{-\infty}^{\infty} b \cdot f(x)dx$$

$$= a\int_{-\infty}^{\infty} x \cdot f(x)dx + b\int_{-\infty}^{\infty} f(x)dx = aE(X) + b$$

cf) Discrete r.v. : $E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$

Example 4.11

Two sepcies are competing for control of a certain resource.

Let X be the proportion controlled by species 1 and suppose X has a uniform distribution $[0,1]$.

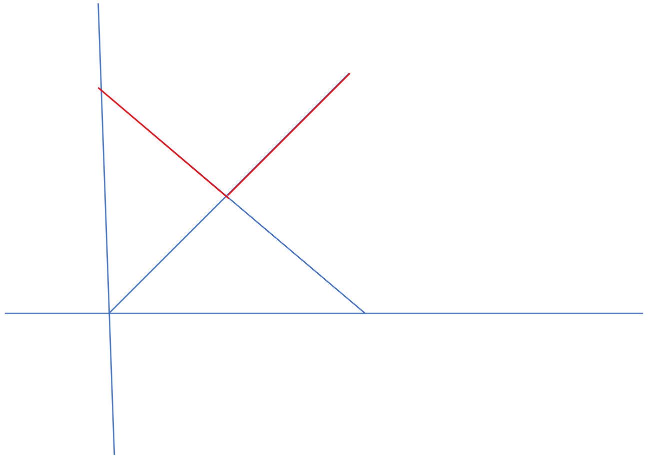
Then the species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1-X)$$

$$E[h(X)] = \int_{-\infty}^{\infty} \max(X, 1-X) \cdot f(x)dx$$

$$= \int_0^{1/2} (1-x)dx + \int_{1/2}^{1} x dx$$

$$= \left[x - \frac{x^2}{2}\right]_0^{1/2} + \left[\frac{x^2}{2}\right]_{1/2}^{1} = 3/4$$



Variance of a continuous random variable

- The variance of a continuous random variable X with pdf $f(x)$ and mean μ is

$$\sigma_X^2 = \mathrm{V(X)} = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty}(x-\mu)^2 \cdot f(x)dx$$

- The standard deviation (SD) of X is $\sigma_X = \sqrt{V(X)}$.
- The variance or standard deviation tell us how “spread out” the distribution is.

cf) Discrete r.v. : $V(X) = \mathrm{E}[(X-\mu)^2] = \sum_{x\in D}(x-\mu)^2 \cdot p(x)$

A shortcut formula for variance

$$\mathrm{V}(\mathrm{X}) = \mathrm{E}(X^2) - [E(X)]^2$$

$$V(X) = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) \cdot f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - 2\mu \int_{-\infty}^{\infty} x \cdot f(x)dx + \mu^2 \int_{-\infty}^{\infty} f(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - 2\mu \cdot \mu + \mu^2$$

$$= \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \mu^2$$

$$= \mathrm{E}(X^2) - [E(X)]^2$$

A shortcut formula for variance

$$\mathrm{V(X)} = \mathrm{E}(X^2) - [E(X)]^2$$

For the X in Example 4.9, $f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & if\ 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

$$= \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)dx$$

$$= \frac{3}{2}\int_0^1 (x^2 - x^4)dx = \frac{3}{2}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_0^1 = \frac{3}{2}\left(\frac{1}{3} - \frac{1}{5}\right) = \frac{1}{5}$$

$$\mathrm{V(X)} = \mathrm{E}(X^2) - [E(X)]^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = 0.59$$

Exercise (4.2) 11

Let X denote the amount of time a book on two-hour reserve is actually checked out, and suppose the cdf is

$$F(x) = \begin{cases} 0 & if\ x < 0 \\ \frac{x^2}{4} & 0 \le x \le 2 \\ 1 & if\ 2 < x \end{cases}$$

Use the cdf to obtain the following:

① $P(X \le 1)$

② $P(0.5 \le X \le 1)$

③ $P(X > 1.5)$

④ The median checkout duration $\tilde{\mu}$ [solve 0.5=$\mathrm{F}(\tilde{\mu})$]

⑤ $\mathrm{F}'(\mathrm{x})$ to obtain the density function $f(x)$

⑥ $E(X)$

⑦ $V(X)$ and σ_X

⑧ If the borrower is charged an amount $h(X) = X^2$ when checkout duration is X, compute the expected charge $E[h(X)]$

Exercise (4.2) 11

Solution)

① $P(X \leq 1) = F(1) = \frac{1^2}{4} = 0.25$

② $P(0.5 \leq X \leq 1) = F(1) - F(0.5) = 0.25 - 0.0625 = 0.1875$

③ $P(X > 1.5) = 1 - F(1.5) = 1 - \frac{1.5^2}{4} = 0.4375$

④ $0.5 = F(\tilde{\mu}) = \frac{\tilde{\mu}^2}{4} \Rightarrow \tilde{\mu}^2 = 2 \Rightarrow \tilde{\mu} = \sqrt{2} = 1.4142$

⑤ $F'(x)$ to obtain the density function $f(x)$: $F'(x) = \frac{d}{dx}\frac{x^2}{4} = \frac{x}{2}$

$$f(x) = \begin{cases} \frac{x}{2} & 0 \leq x \leq 2 \\ 0 & otherwise \end{cases}$$

Exercise (4.2) 11

6. $\mathrm{E(X)} = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{0}^{2} x \cdot \frac{x}{2} dx \; = \int_{0}^{2} \frac{x^2}{2} dx \; = \left[\frac{x^3}{6}\right]_0^2 = \frac{8}{6} = 1.333$

7. $\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_{0}^{2} x^2 \cdot \frac{x}{2} dx \; = \int_{0}^{2} \frac{x^3}{2} dx \; = \left[\frac{x^4}{8}\right]_0^2 = \frac{16}{8} = 2$

$\mathrm{V(X)}{=}\mathrm{E}(X^2) - [\mathrm{E(X)}]^2 = 2 - 1.33^2 = 0.2222$

$\sigma_{\mathrm{X}} = \sqrt{0.2222} = 0.4714$

8. $E[h(X)] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_{0}^{2} x^2 \cdot \frac{x}{2} dx \; = \left[\frac{x^4}{8}\right]_0^2 = \frac{16}{8} = 2$

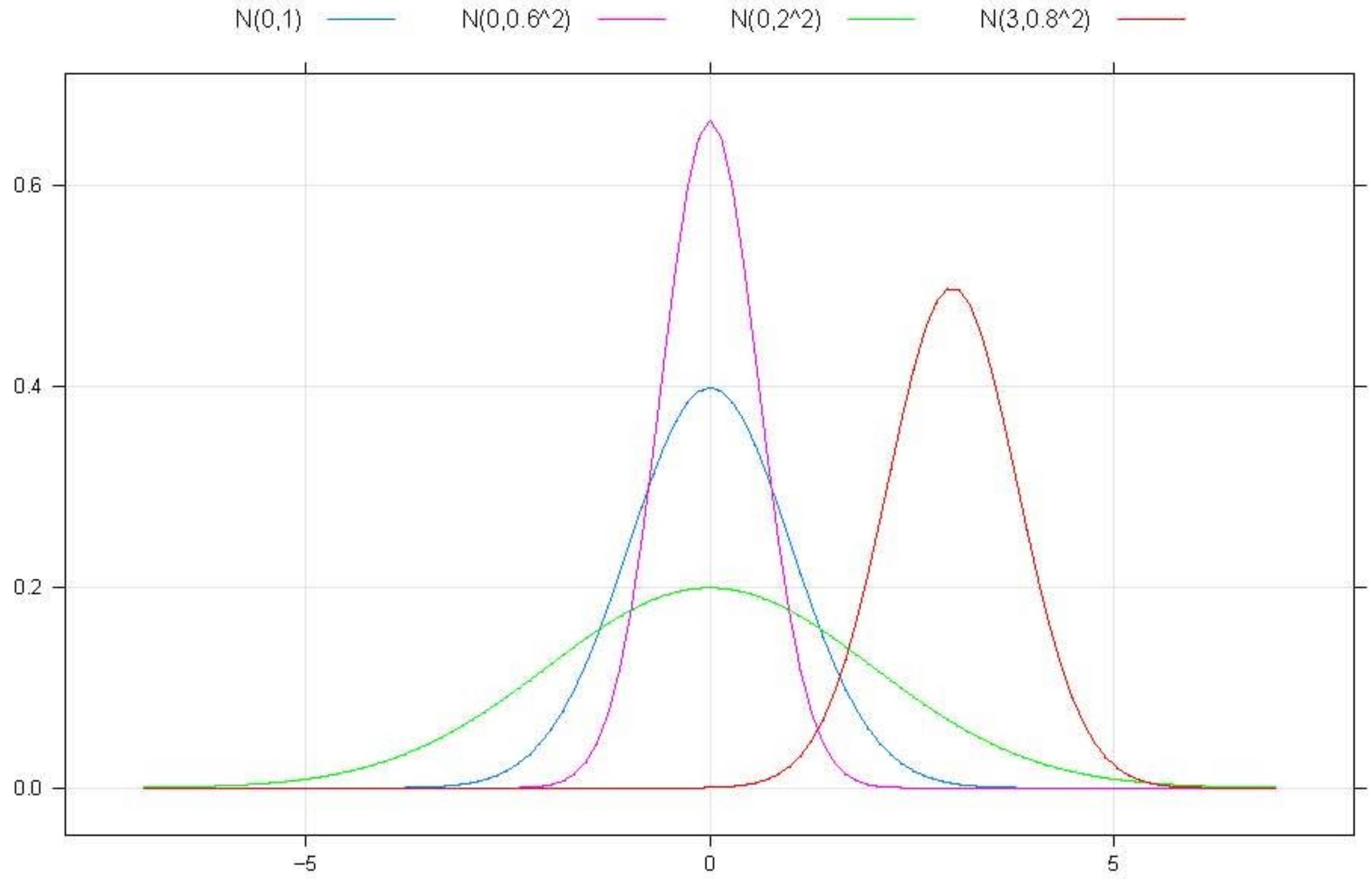
The normal distribution

- A continuous rv X is said to have a normal distribution with parameters μ and σ, where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is

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

- A shorthand notation is $X \sim N(\mu, \sigma^2)$
- It can be shown that $\mathrm{E}(X) = \mu$, and $\mathrm{V}(X) = \sigma^2$

Normal density curves



Mean of Normal Distribution**

$$\mathrm{E(X)} = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xexp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) exp(-t^2) dt$$

substituting $t = \frac{x-\mu}{\sqrt{2}\sigma} \quad dt = \frac{dx}{\sqrt{2}\sigma}$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{-\infty}^{\infty} texp(-t^2) dt + \ \mu \int_{-\infty}^{\infty} exp(-t^2) dt \right)$$

Here $\int_{-\infty}^{\infty} te^{-t^2} dt = \frac{1}{2} \int_{-\infty}^{\infty} e^{-u} du = \left[-\frac{1}{2} e^{-u}\right]_{-\infty}^{\infty} = \left[-\frac{1}{2} e^{-t^2}\right]_{-\infty}^{\infty}$

$$u = t^2 \rightarrow du = 2tdt \rightarrow tdt = \frac{1}{2} du$$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \left[-\frac{1}{2} exp(-t^2)\right]_{-\infty}^{\infty} + \mu\sqrt{\pi}\right)$$

Gaussian integral

$$= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} = \mu$$

Polynomial Equation Calculator :

https://www.symbolab.com/solver/polynomial-equation-calculator/

Variance of Normal Distribution**

$$\mathrm{V(X)} = \mathrm{E}[(X-\mu)^2] = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^2 \cdot exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

$$= \frac{\sigma}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2\sigma^2 exp(-z^2/2)dz$$ (substituting $z = \frac{x-\mu}{\sigma}$, $dz = \frac{dx}{\sigma}$, $dx = \sigma\, dz$)

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 exp(-z^2/2)dz$$

Integration by parts for definite integrals : $\int_a^b udv = uv\Big|_a^b - \int_a^b vdu$

Using integration by parts with $u = z, dv = zexp(-z^2/2)dz \rightarrow du = dz, v = -exp(-z^2/2)$

$$\mathrm{V(X)} = \frac{\sigma^2}{\sqrt{2\pi}}\left(-zexp(-z^2/2)\Big|_{-\infty}^{\infty}\right) - \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -exp(-z^2/2)dz$$

$= \sigma^2$ (since the total integral of N(0,1), $\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz = 1$)

https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading6a.pdf (Example 9 and 10)

The Moment Generating Function of the Normal Distribution**

- https://www.le.ac.uk/users/dsgp1/COURSES/MATHSTAT/6normgf.pdf

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

$$f(z;0{,}1) = \frac{1}{\sqrt{2\pi}} exp(-z^2/2)$$

$$M_z(t) = E[e^{zt}] = \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} exp(-z^2/2)dz$$

$$exp(zt)exp(-z^2/2) = exp\left(-\frac{z^2}{2} + zt\right)$$

$$= exp\left(-\frac{1}{2}(z^2 - 2zt + t^2 - t^2)\right) =$$

$$exp\left(-\frac{1}{2}(z-t)^2\right) exp\left(\frac{t^2}{2}\right)$$

$$M_z(t) = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}(z-t)^2\right) dz = e^{\frac{t^2}{2}} \qquad (1)$$

- the expression under the integral is the $N(z, \mu = t, \sigma^2 = 1)$ probability density function which integrates to unity.

The Moment Generating Function of the Normal Distribution**

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

` $= \int_{-\infty}^{\infty} e^{(z\sigma+\mu)t} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$ (substituting $z = \frac{x-\mu}{\sigma}$, $x = z\sigma + \mu, \mathrm{dx} = \sigma dz$)

$= e^{\mu t} \int_{-\infty}^{\infty} e^{z\sigma t} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$ ($(1) : M_z(t) = \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right) dz = e^{\frac{t^2}{2}}$)

$$= e^{\mu t} e^{\frac{\sigma^2 t^2}{2}} = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)}$$

- the final equality follows in view of the final equation of (1)

Mean and variance of the normal distribution**

- https://online.stat.psu.edu/stat414/lesson/16/16.4

$$M_X(t) = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)}$$

$$M_X'(0) = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)} \cdot \left(\mu + \frac{\sigma^2 2t}{2}\right)\Bigg|_{t=0}$$

$$\frac{d}{dx}\left(e^{f(x)}\right) = e^{f(x)} f'(x)$$

$$= 1(\mu + 0) = \mu$$

$$M_X''(0) = e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)} \cdot \left(\mu + \frac{\sigma^2 2t}{2}\right) \cdot \left(\mu + \frac{\sigma^2 2t}{2}\right) + e^{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)} \cdot \sigma^2\Bigg|_{t=0}$$

$$= 1(\mu \cdot \mu + \sigma^2) = \mu^2 + \sigma^2$$

$$\frac{d}{dx}(f \times g) = f'g + fg'$$

$$\mathrm{V(X)} = \mathrm{E}(X^2) - [E(X)]^2$$

$$= M_X''(0) - [M_X'(0)]^2$$

$$= \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

The Moment Generating Function of the Sum of Normal Distributions**

Let $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N\left(\mu_y, \sigma_y^2\right)$ be two independently distributed normal variables. Then their sum is also a normally distributed random variable:

$$Z = X + Y \sim N\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$$

Proof)

$$M_z(t) = M_X(t)M_Y(t) = \exp(\mu_x t + \frac{1}{2}\sigma_x^2 t^2)\ \exp(\mu_y t + \frac{1}{2}\sigma_y^2 t^2)$$

$$= \exp\left\{(\mu_x + \mu_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2\right\}$$

Which is the moment generating function of a normal distribution.

cf) The moment generating function of the sum of independent random variables X and Y.

$$M_{X+Y}(t) = E\left[e^{(X+Y)t}\right] = E[e^{Xt}e^{Yt}] = E[e^{Xt}]E[e^{Yt}] = M_X(t)M_Y(t)$$

https://online.stat.psu.edu/stat414/lesson/26/26.1

The standard normal distribution

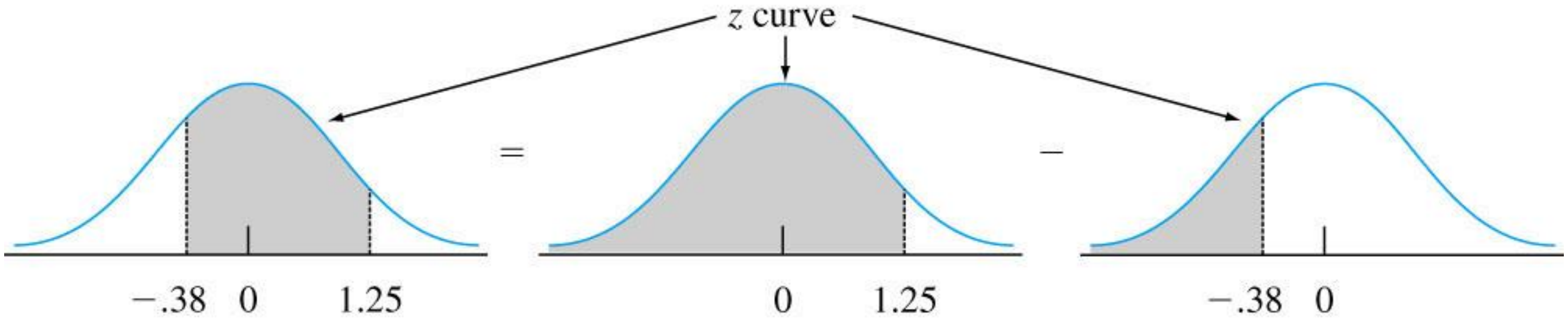
- $N(0,1)$ is called the standard normal distribution. A standard normal random variable will be denoted by Z. The cdf of Z will be denoted by $\Phi(z) = P(Z < z)$.
- Appendix Table A.3 (reproduced on the inside front cover) can be used to obtain $\Phi(z)$ and the (100p)th percentile of $N(0,1)$.
- Example : Find $P(-0.38 \le Z \le 1.25)$ and the 99th percentile of the standard normal distribution

$$P(-0.38 \le Z \le 1.25) = 0.8944 - 0.3520 = 0.5424$$

99th percentile : $P(Z \le 2.326) = 0.99$

```
> pnorm(1.25)-pnorm(-0.38)
> qnorm(0.99)
```

- Notation : z_α denotes the value for which α of the area under the standard normal density curve lies to the right of z_α. That is z_α is the [100(1- α)]th percentile of $N(0,1)$. For example, $z_{0.05} = 1.645$



0.8944

0.3520

Fig. 4-16, p. 147

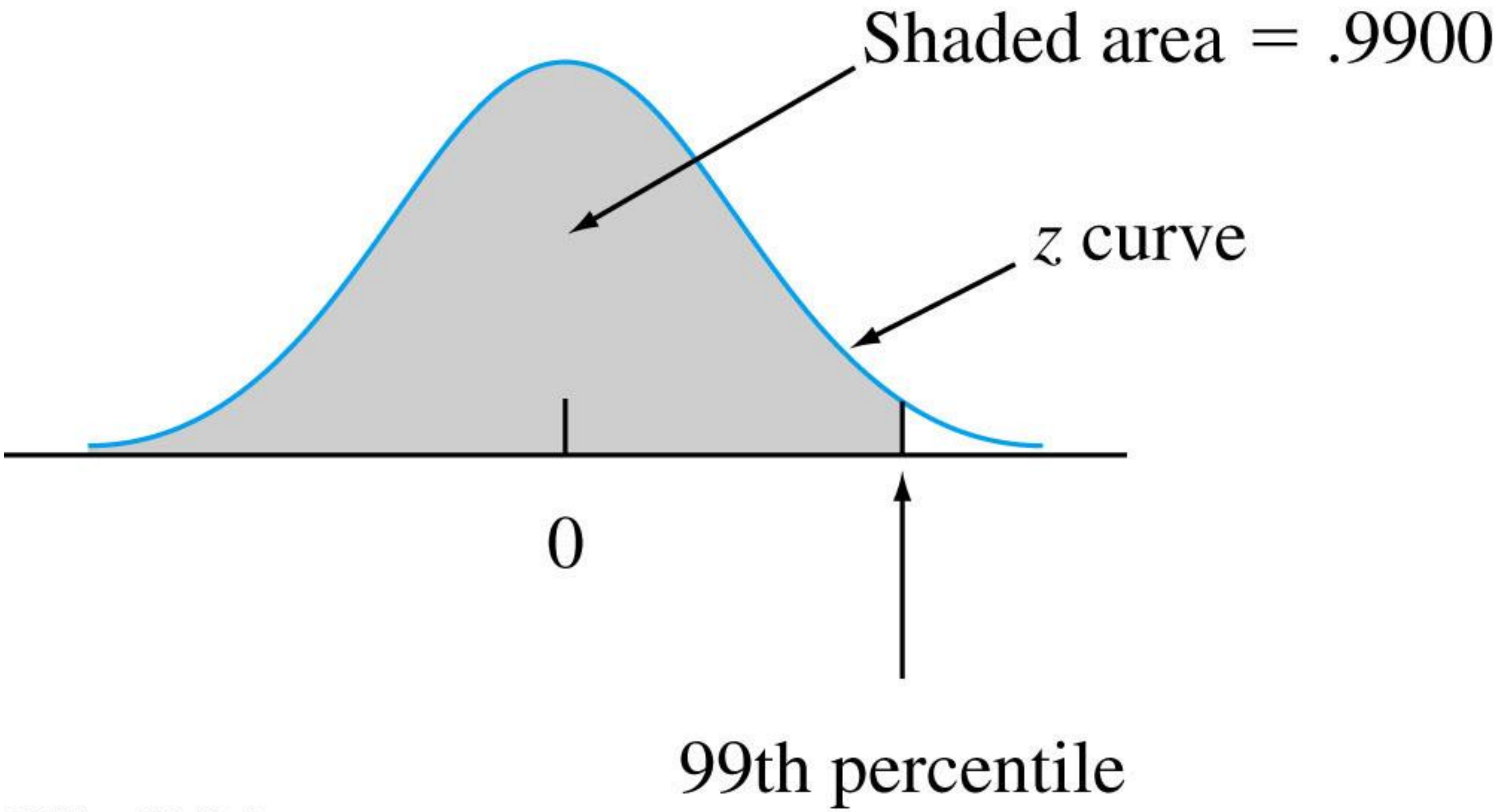


Fig. 4-17, p. 147

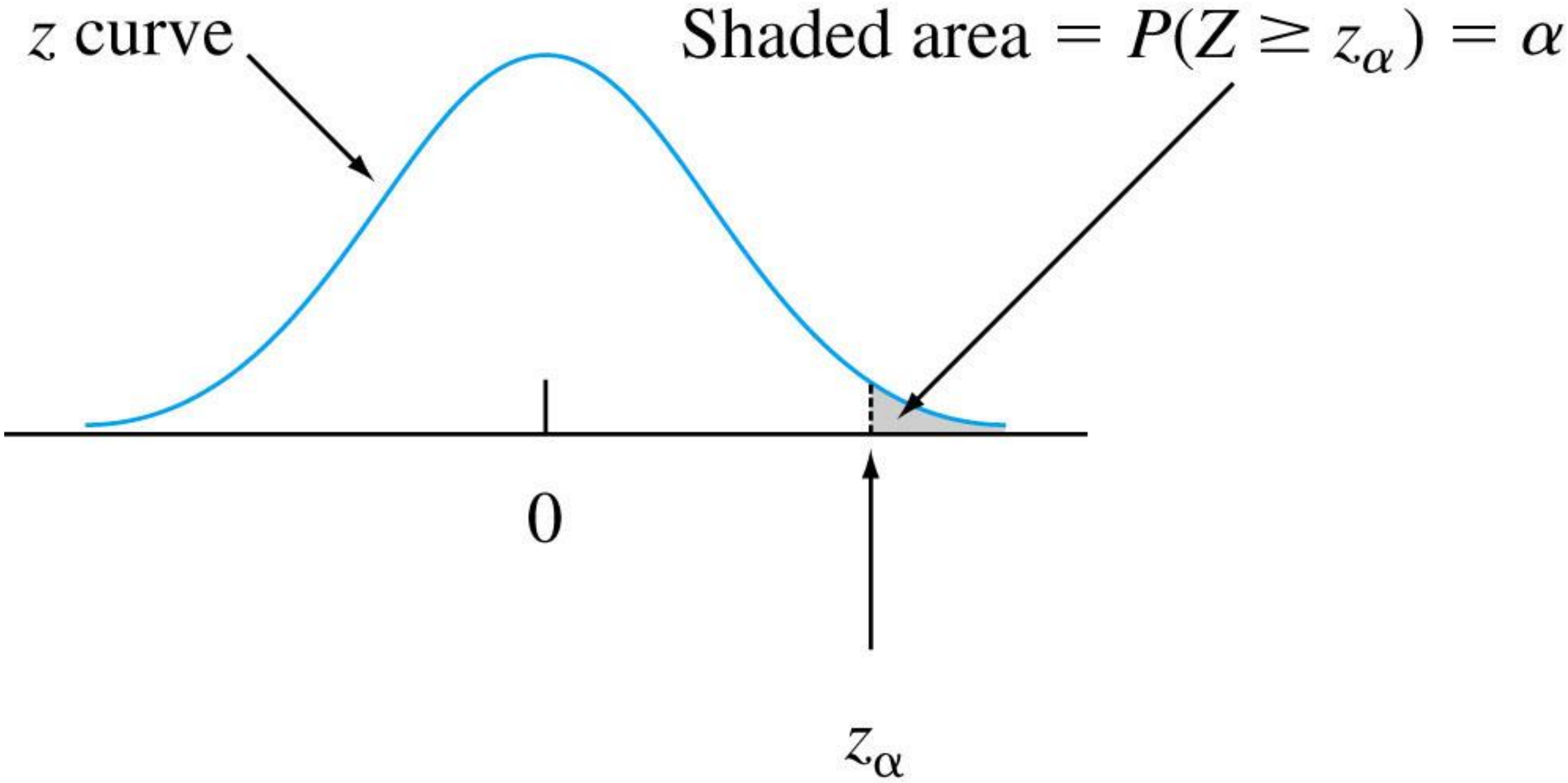


Fig. 4-19, p. 148

The standard normal distribution

Table 4.1 Standard Normal Percentiles and Critical Values

Percentile	90	95	97.5	99	99.5	99.9	99.95
α(tail area)	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
z_α=100(1-α)th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

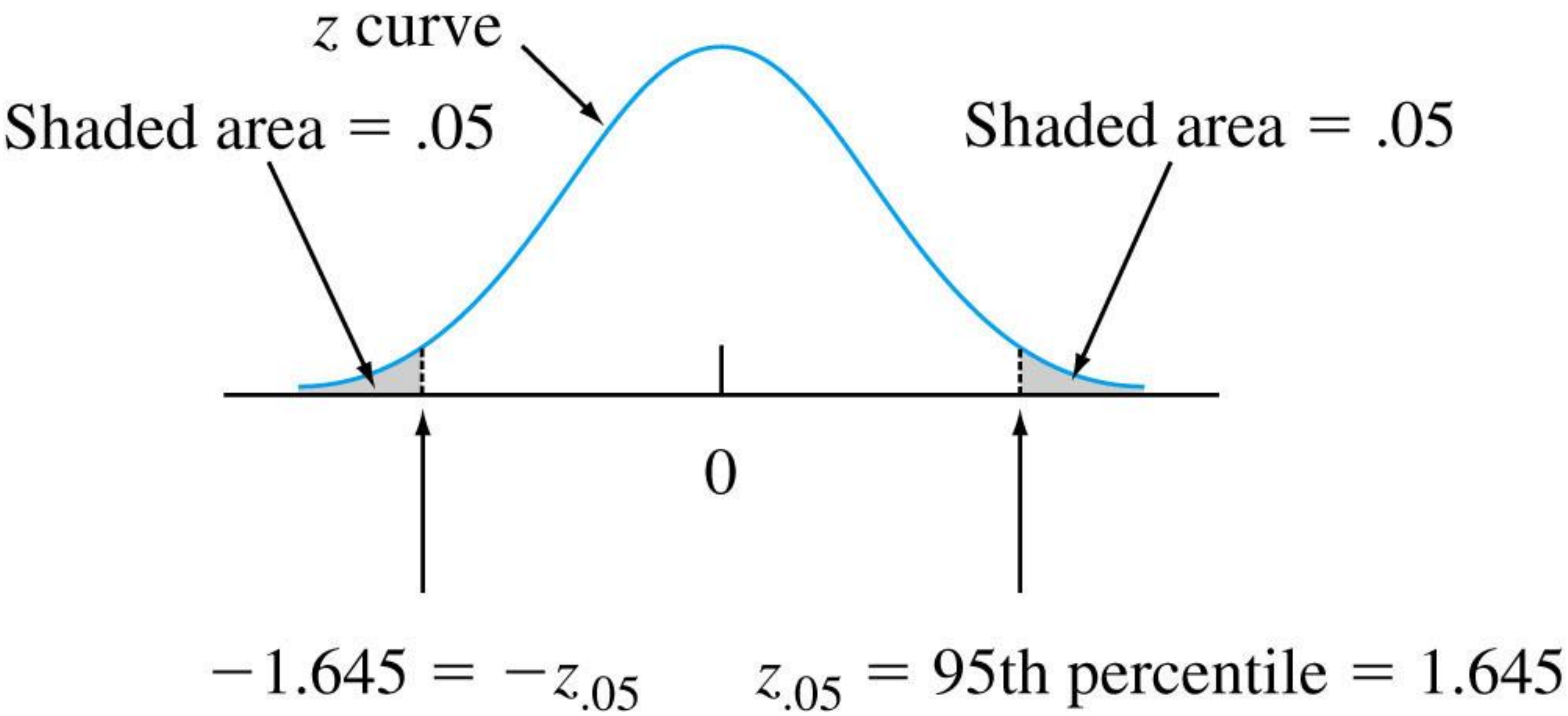
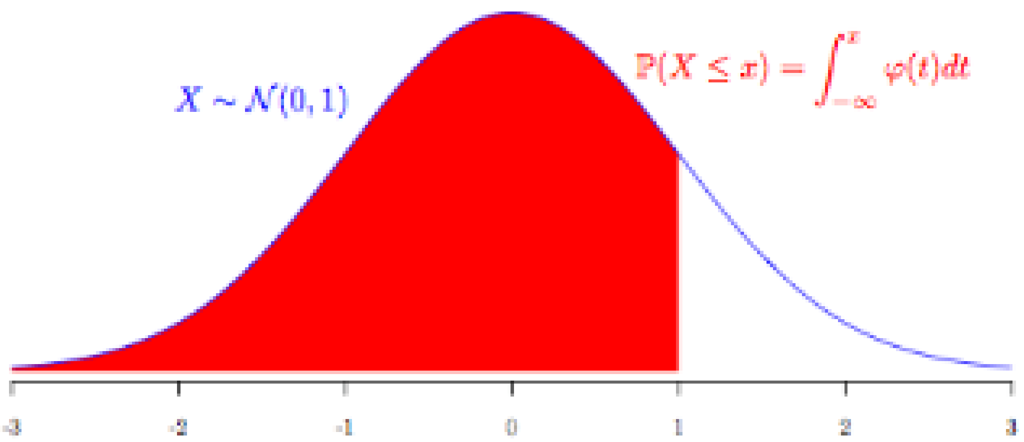

z curve
Shaded area = .05
Shaded area = .05
0
$-1.645 = -z_{.05}$
$z_{.05}$ = 95th percentile = 1.645

Fig. 4-20, p. 149

Normal Distribution Table



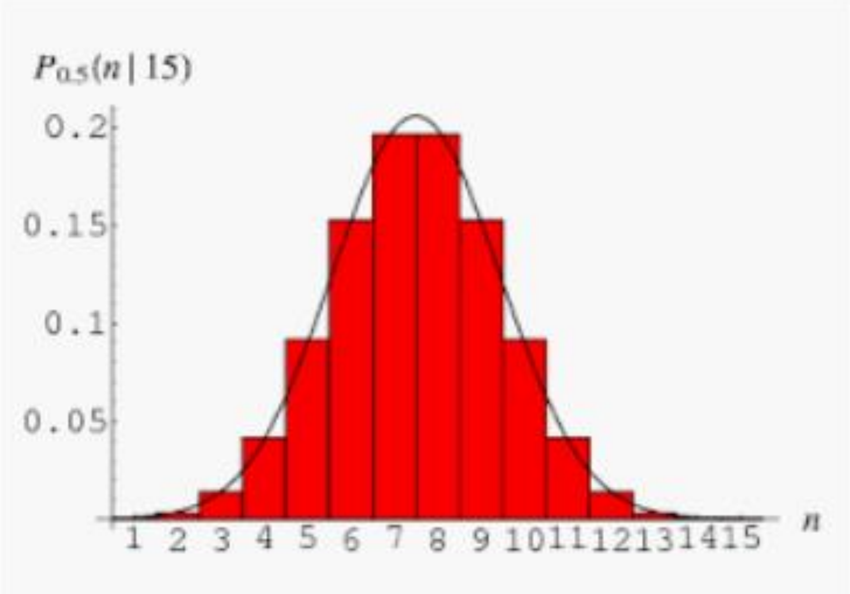
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Numerical integration of Normal distribution $\int_{-1}^{1} f(x)dx$

divide (-1, 1) into 10000 subintervals and sum the area of the 10000 rectangles

```
x <- seq(-1, 1, length=10000)
sum((1/sqrt(2*pi)*exp(-x^2/2))*1/5000)
```



Find

- $P(Z \leq 0)$
- $P(Z < 1.5)$
- $P(Z \geq 1.5)$
- $P(Z < -1.5)$
- $P(Z \leq 3)$
- $P(Z > 2)$

$> \ pnorm(0)$

$[1]0.5$

$> \ pnorm(1.5)$

$[1]0.9331928$

$> \ 1 - pnorm(1.5)$

$[1]0.0668072$

$> \ pnorm(-1.5)$

$[1]0.0668072$

$> \ pnorm(3)$

$[1]0.9986501$

$> \ 1 - pnorm(2)$

$[1]0.02275013$

Nonstandard normal distribution

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$ has a standard normal distribution. Thus

$$\mathrm{P}(\mathrm{a} \leq \mathrm{X} \leq b) = \mathrm{P}\left(\frac{a-\mu}{\sigma} \leq \mathrm{Z} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Example 4.16: The reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled as $X \sim N(1.25, 0.46^2)$. Then

$$\mathrm{P}(1 \leq \mathrm{X} \leq 1.75) = \mathrm{P}\left(\frac{1-1.25}{0.46} \leq \mathrm{Z} \leq \frac{1.75-1.25}{0.46}\right)$$

$$= \mathrm{P}(-0.54 \leq \mathrm{Z} \leq 1.09)$$=0.862-0.295=0.5675

```
> diff(pnorm(c(-0.543,1.087)))
[1] 0.5679167
> diff(pnorm(c(1,1.75), mean=1.25, sd=0.46))
[1] 0.5680717
```

Percentiles of normal distribution

The $(100\mathrm{p})th$ percentile for $N(\mu, \sigma^2)$ can be computed as

$$\mu + [\text{100pth percentile for N}(0,1)] \cdot \sigma$$

Example 4.16 (cont.) : $X \sim N(1.25, 0.46^2)$.

The 99th percentile of X is

$$1.25 + 2.326 \cdot 0.46 = 2.31996$$

```
> qnorm(0.99, mean=1.25, sd=0.46)
[1] 2.32012
> qnorm(0.99)
[1] 2.326348
```

Normal approximation to binomial

Suppose $X \sim Bin(n,p)$. Provided that $\mathrm{np} \geq 10$ and $\mathrm{n}(1-\mathrm{p}) \geq 10$, X has approximately a normal distribution with $\mu = \mathrm{np}$ and $\sigma = \sqrt{\mathrm{np}(1-\mathrm{p})}$ and

$$\text{Binomial } \mathrm{P}(\mathrm{X} \leq x) \approx \text{Nomal } \mathrm{P}(\mathrm{X} \leq x + 0.5) = \Phi\left(\frac{x+0.5-np}{\sqrt{\mathrm{np}(1-\mathrm{p})}}\right) \qquad x = 0, 1, \cdots, n$$

Example 4.20: Suppose 25% of all licensed drivers do not have insurance.

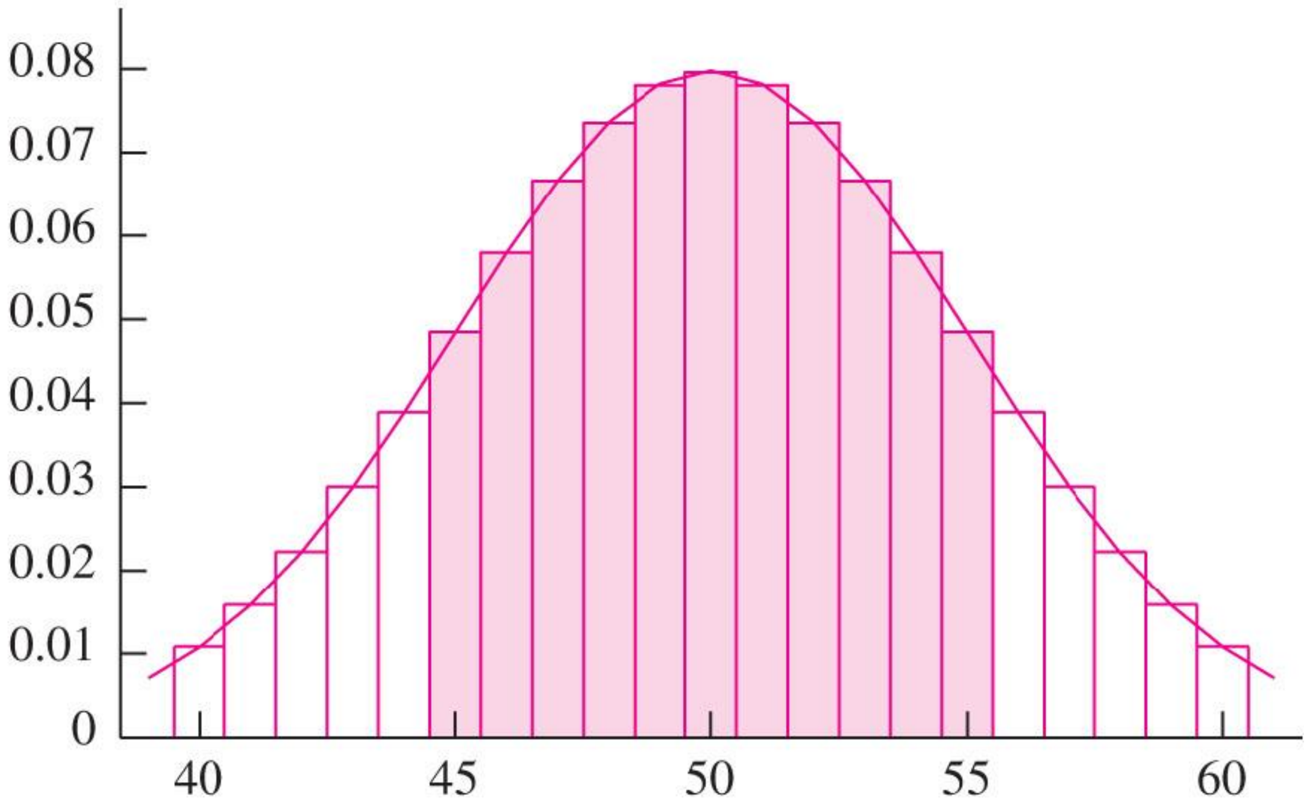
Let X be the number of uninsured drivers in a random sample of size 50.

Then $X \sim Bin(50, 0.25)$,

$$\mathrm{P}(5 \leq \mathrm{X} \leq 15) = P(X \leq 15) - P(X \leq 4)$$

$$\approx \Phi\left(\frac{15.5-12.5}{3.06}\right) - \Phi\left(\frac{4.5-12.5}{3.06}\right) = 0.8320$$

```
> diff(pnorm(c(4.5,15.5), 12.5, sqrt(50*0.25*0.75)))
[1] 0.8319162
> sum(dbinom(5:15, 50, 0.25))
[1] 0.8348084
```



Binomial : P(45$\leq X \leq$ 55) $\leftrightarrow$ Normal : P(44.5 $\leq X \leq$ 55.5)

The Gamma distribution

For $\alpha > 0$, the Gamma function $\Gamma(\alpha)$ is defined by

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

Property : $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$, $\Gamma(n) = (n-1)!, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

A continuous random variable X is said to have a gamma distribution if the pdf of X is

$$\mathrm{f}(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & otherwise \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. The standard gamma distribution has $\beta = 1$.

$E(X) = \alpha\beta$ and $V(X) = \alpha\beta^2$

The Gamma function**

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

- $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$

Let $dv = e^{-x}dx$ and $u = x^{\alpha-1}$. Then $v = -e^{-x}, du = (\alpha - 1)\ x^{\alpha-2}dx$

$$\int x^{\alpha-1}e^{-x}dx = \int udv$$

$$= uv - \int vdu$$

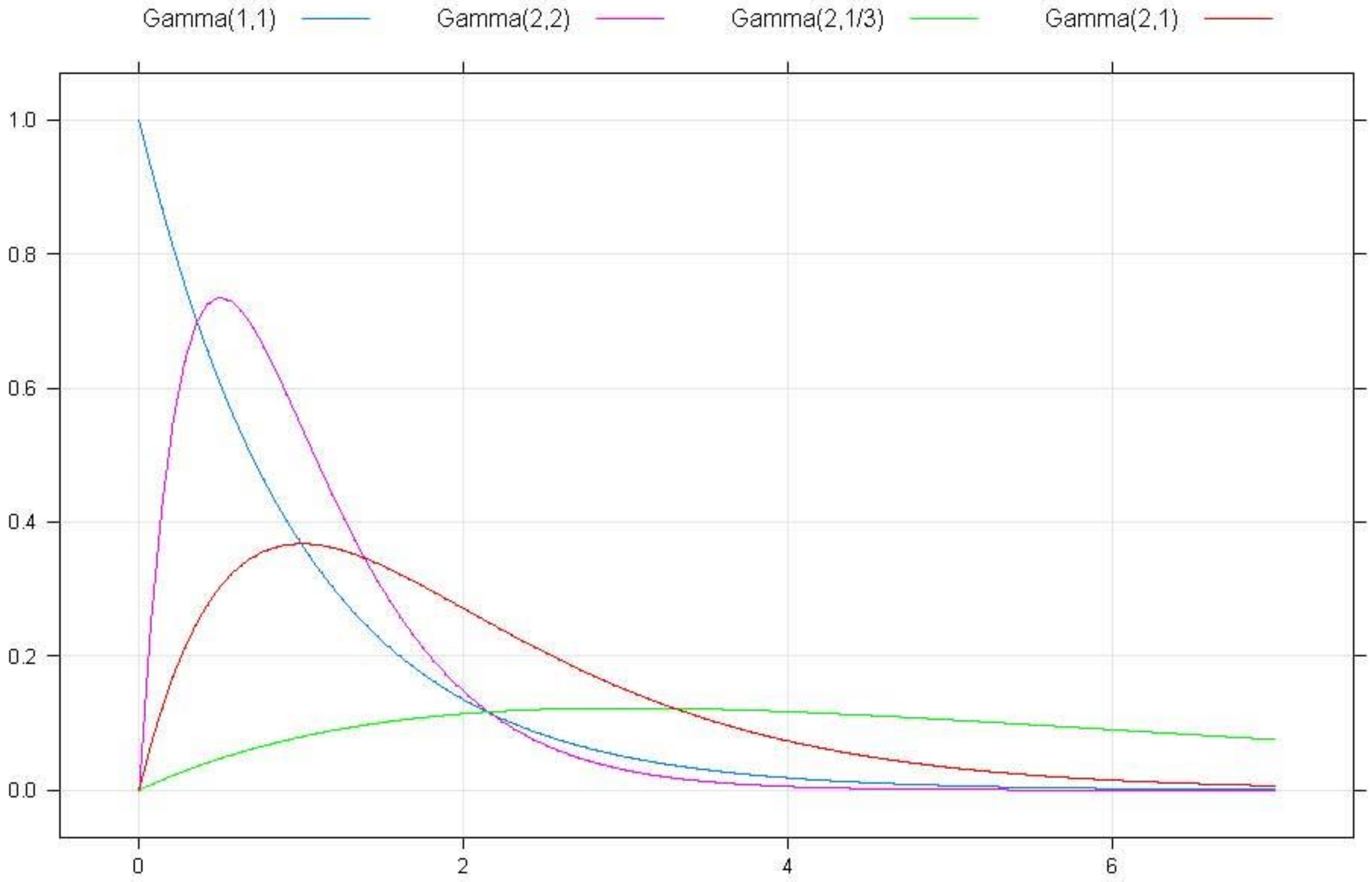
$$= -x^{\alpha-1}e^{-x} + (\alpha - 1)\int x^{\alpha-2}e^{-x}\, dx$$

As the bounds go to $[0, \infty)$, we find that $-x^{\alpha-1}e^{-x} \to 0$, hence

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

$$\Gamma(1) = \int_0^{\infty} x^{1-1}\, e^{-x}dx\ = \int_0^{\infty} e^{-x}\, dx\ = [-e^{-x}]_0^{\infty} = 1$$

Gamma density curves



Probabilities from the gamma distribution : Example 4.21

Suppose the survival time X in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has a gamma distribution with $\alpha = 8$ and $\beta = 15$.

What is the probability that a mouse survives between 60 and 120 weeks?

$$\mathrm{P}(60 \leq \mathrm{X} \leq 120) = P(X \leq 120) - P(X \leq 60) = 0.496$$

```
> diff(pgamma(c(60, 120), shape=8, scale=15))
[1] 0.4959056
```

By Equation Calculator :

$$\int_{60}^{120} \frac{1}{15^8 \Gamma(8)} x^{8-1} e^{-x/15} dx = \frac{2499.36409}{7!} = 0.4959056$$

The exponential distribution

A gamma distribution with $\alpha = 1$ and $\beta = \frac{1}{\lambda}$ is called an exponential distribution with parameter λ. The exponential distribution pdf is

$$\mathrm{f}(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

IF X is an exponential random variable with parameter λ, then

$$E(X) = \frac{1}{\lambda}(= \alpha\beta) \text{ and } V(X) = \frac{1}{\lambda^2}\ (= \alpha\beta^2)$$

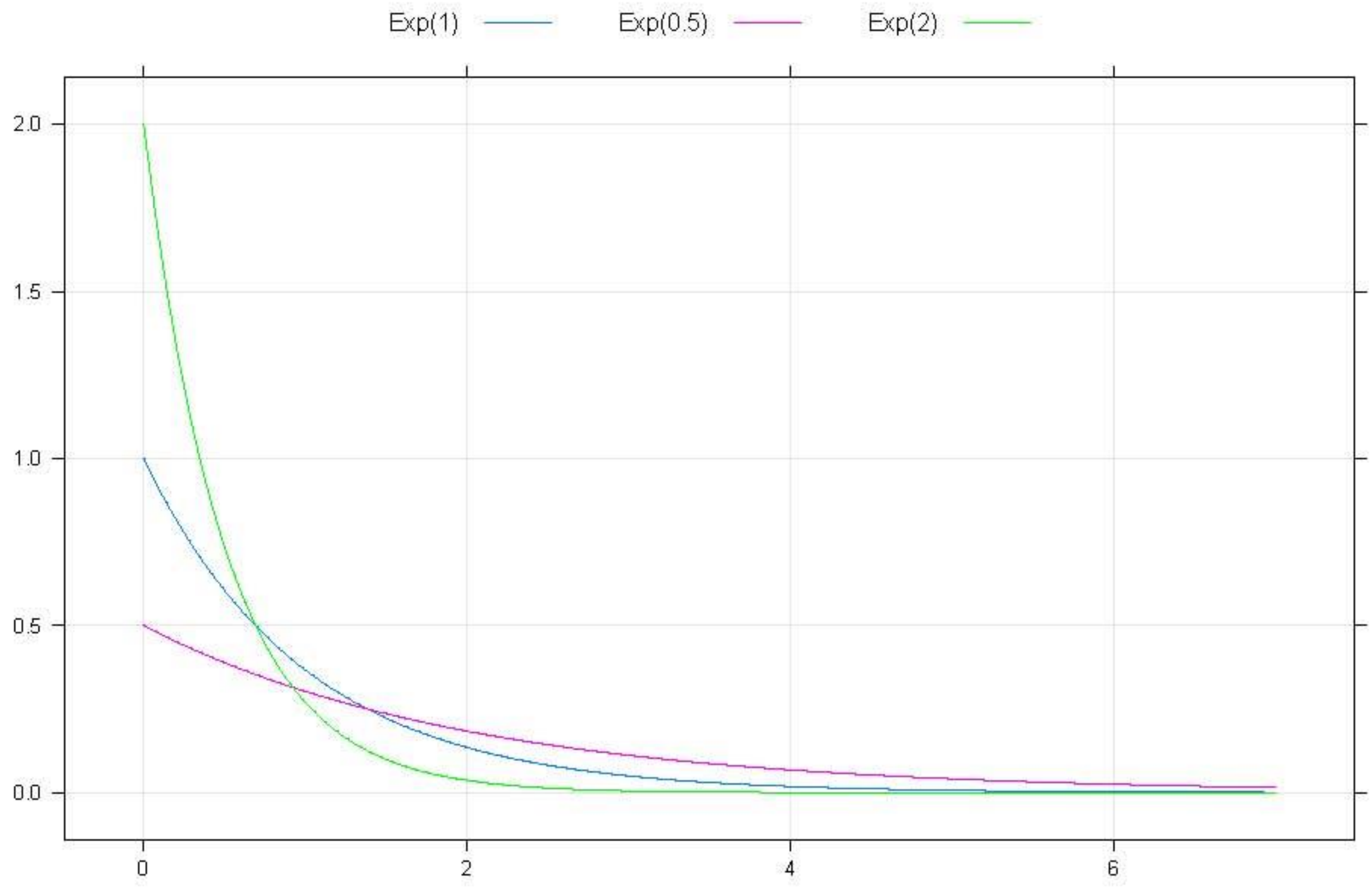
The exponential pdf is easily integrated to obtain the cdf.

$$\mathrm{F}(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$$\mathrm{F}(x;\lambda) = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$$

Exp

Exponential density curves



Expected value of an exponential random variable

Let X be a continuous random variable with an exponential density function with parameter λ.

Integrating by parts with $\mathrm{u} = \lambda x$ and $\mathrm{dv} = e^{-\lambda x} dx$ so that

$\mathrm{du} = \lambda dx$ and $\mathrm{v} = -\frac{1}{\lambda} e^{-\lambda x}$, we find

$$E(X) = \int_0^\infty x f(x) dx$$

$$= \int_0^\infty x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x}\right]\Big|_0^\infty - \int_0^\infty -e^{-\lambda x} dx = \int_0^\infty e^{-\lambda x} dx$$

$$= \left[-\frac{1}{\lambda} e^{-\lambda x}\right]\Big|_0^\infty = \frac{1}{\lambda}$$

Variance of an exponential random variable

Integrating by parts with $\mathrm{u} = \lambda x^2$ and $\mathrm{dv} = e^{-\lambda x} dx$ so that

$\mathrm{du} = 2\lambda x dx$ and $\mathrm{v} = -\frac{1}{\lambda} e^{-\lambda x}$, we find

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$= \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x}\right]\Big|_0^{\infty} - \int_0^{\infty} -2x e^{-\lambda x} dx$$

$$= \int_0^{\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

So,

$$V(X) = \frac{2}{\lambda^2} - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

The Moment Generating Function of the Exponential Distribution**

$$M_X(t) = \int_0^\infty e^{tx}\,\lambda e^{-\lambda x} dx$$

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

$$= \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)x}\right]_0^\infty = -\frac{\lambda}{t-\lambda} = \frac{\lambda}{\lambda-t} = \lambda(\lambda-t)^{-1} \quad \text{for} \quad t < \lambda$$

$$M_X'(0) = \lambda(-1)(\lambda-t)^{-2}(-1)|_{t=0} = \lambda(\lambda-t)^{-2}|_{t=0} = \frac{1}{\lambda} \qquad E(X) = \frac{1}{\lambda}$$

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

$$M_X''(0) = \lambda(-2)(\lambda-t)^{-3}(-1)|_{t=0} = 2\lambda(\lambda-t)^{-3}|_{t=0} = \frac{2}{\lambda^2}$$

$$\mathrm{V(X)} = \mathrm{E}(X^2) - [E(X)]^2$$

$$= M_X''(0) - [M_X'(0)]^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

The Moment Generating Function of the Gamma Distribution**

$$M_X(t) = \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx$$

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

Let's use the change of variable technique with

$$\mathrm{y} = x\left(\frac{1}{\beta} - t\right) \rightarrow x = \frac{\beta}{1-\beta t} y \text{ and } \mathrm{d}x = \frac{\beta}{1-\beta t} dy$$

$$M_X(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-y} \left[\frac{\beta}{1-\beta t} y\right]^{\alpha-1} \frac{\beta}{1-\beta t} dy$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} \left[\frac{\beta}{1-\beta t}\right]^\alpha dy$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{1-\beta t}\right)^\alpha \int_0^\infty y^{\alpha-1} e^{-y} dy$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{1-\beta t}\right)^\alpha \Gamma(\alpha) = \frac{1}{(1-\beta t)^\alpha} \qquad \text{for } t < \frac{1}{\beta}$$

Example 4.22

Suppose that calls are received at a 24-hour “suicide hotline” according to a Poisson process with rate $\alpha = 0.5$ call per day. $\left(P_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}\right)$

Then the number of days X between successive calls has an exponential distribution with parameter value 0.5.

So the probability that more than 2 days elapse between calls is

$$\mathrm{P(X > 2) = 1 - P(X \leq 2) = 1 - F(2; 0.5) = e^{-(0.5)(2)} = 0.36}$$

The expected time between successive calls is 1/0.5=2 days.

```
> 1-pexp(2, 0.5)
[1] 0.3678794
```

The Chi-squared distribution

Let ν be a positive integer.

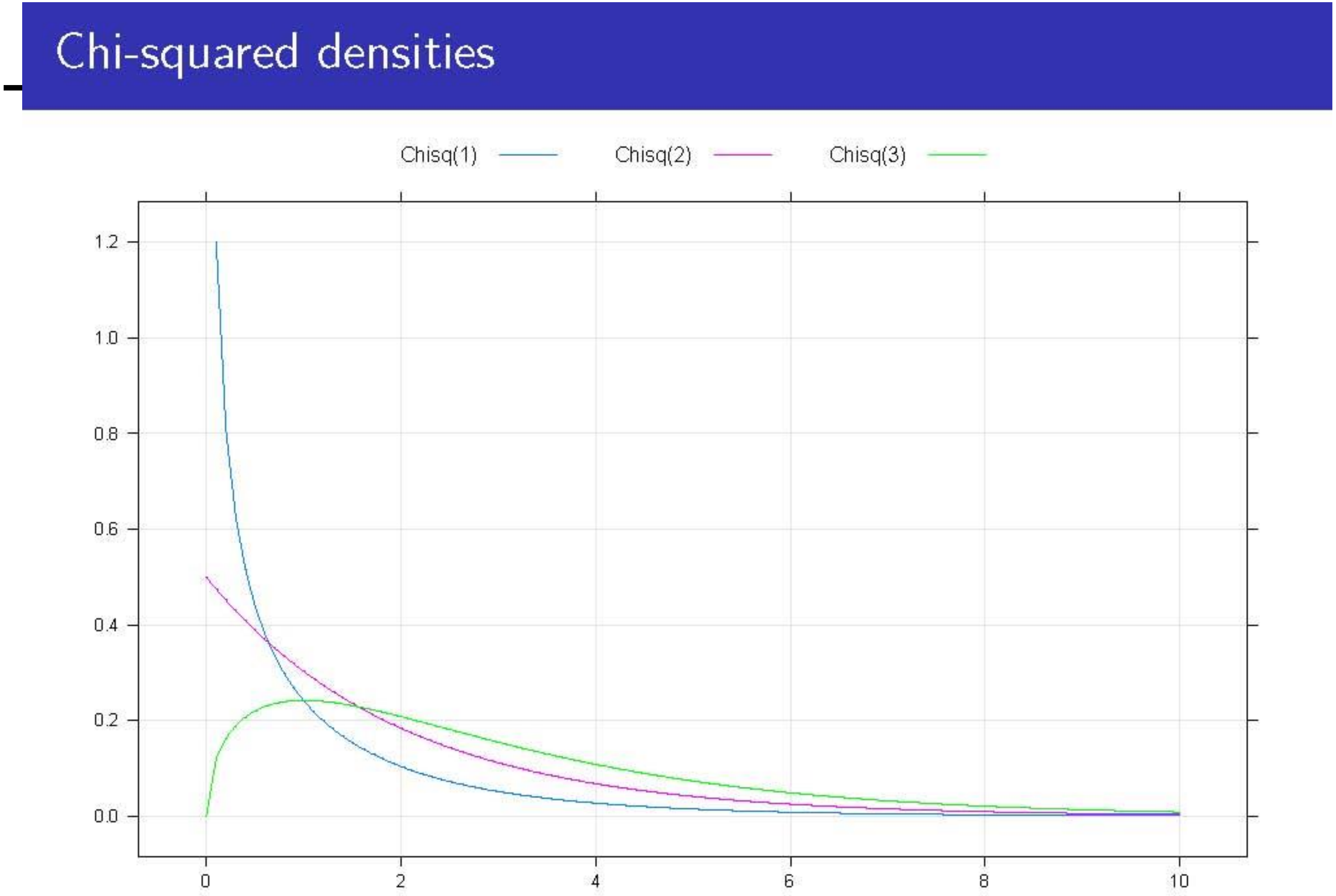
A random variable X is said to have a chi-squared distribution with parameter ν if the pdf of X is the gamma density with $\alpha = \nu/2$ and $\beta = 2$.

The pdf of a chi-squared r.v. $\chi^2(\nu)$ is thus

$$\mathrm{f}(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\left(\frac{\nu}{2}\right)-1} e^{-x/2} & x \geq 0 \\ 0 & otherwise \end{cases}$$

The parameter ν is called the number of degrees of freedom (df) of X.

Chi-



Chi-squared densities
Chisq(1)
Chisq(2)
Chisq(3)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10

If $Z_1, \cdots, Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = \sum_{i=1}^{k} Z_i^2$$

is distributed according to the chi-squared distribution with k degrees of freedom.

This is usually denoted as

$$Q \sim \chi^2(k) \quad \text{or} \quad Q \sim \chi_k^2$$

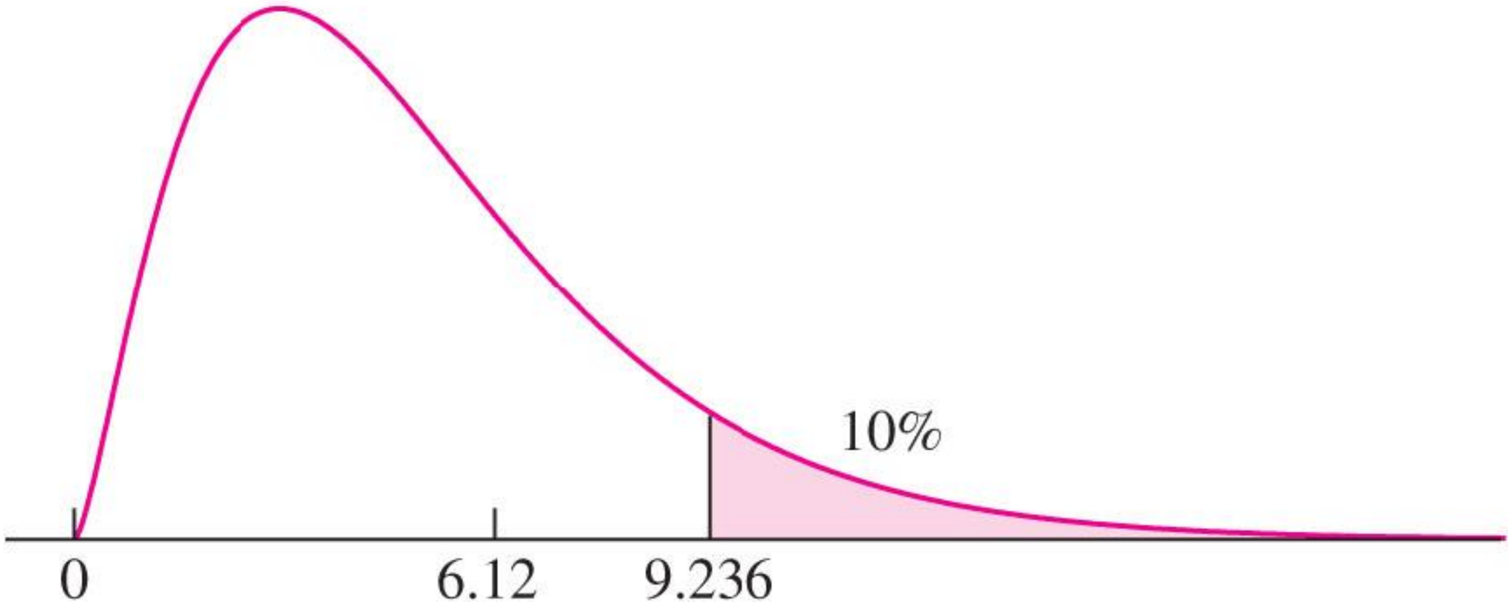
Table 6.1

TABLE 6.1 Observed and expected values for 600 rolls of a die

Category	Observed	Expected
1	115	100
2	97	100
3	91	100
4	101	100
5	110	100
6	86	100
Total	600	600

$$\chi^2 = \frac{(115-100)^2}{100} + \frac{(97-100)^2}{100} + \cdots + \frac{(86-100)^2}{100} = 6.12$$

Figure 6.10



Since $P - \text{value} = P\left(\chi_5^2 > 6.12\right) = 0.2947$, we cannot reject the H_0

> 1- pchisq(6.12, 5) #0.2947

Normal probability plot

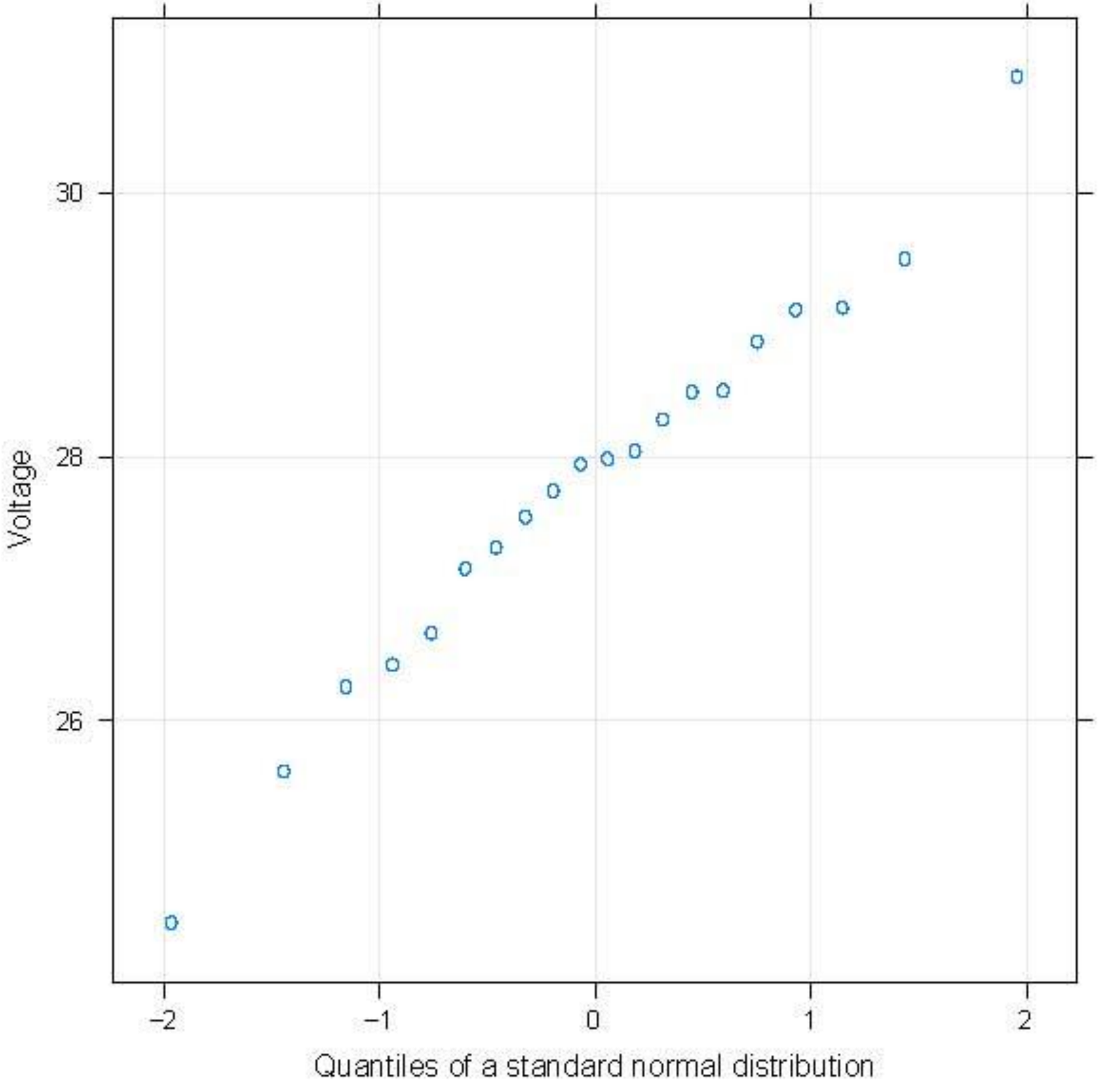
Probability plots can be used to check if a sample came from a particular distribution.

The normal probability plot can be used to check normality.

Example 4.30: 20 observations on dielectric breakdown voltage of a piece of epoxy resin.

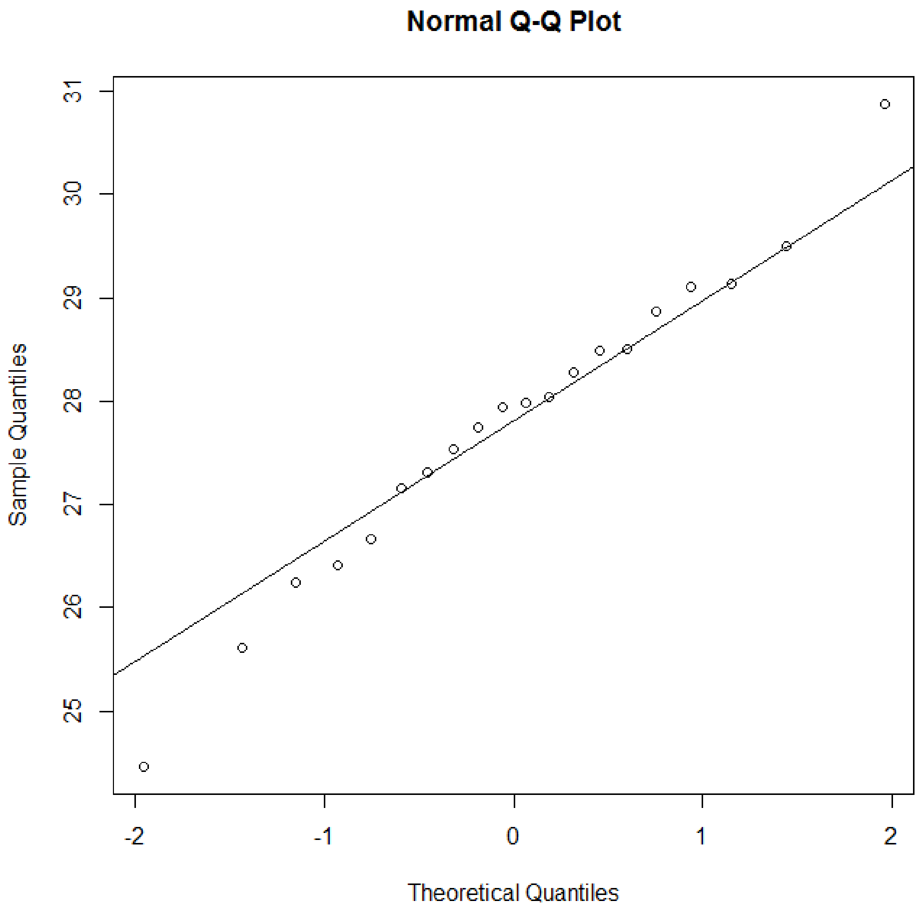
```
> library(Devore7)
> str(xmp04.30)
'data.frame':   20 obs. of  2 variables:
 $ Voltage     : num  24.5 25.6 26.2 26.4 26.7 ...
 $ z.percentile: num  -1.96 -1.44 -1.15 -0.93 -0.76 -0.6 -0.45 -0.32 -0.19 -0.06 ...
> qqmath(~Voltage, xmp04.30, aspect=1, type=c("g", "p", "r"))
```

Normal probability plot



Normal probability plot

```
library(Devore7)
str(xmp04.30)
qqmath(~Voltage, xmp04.30, aspect=1,
type=c("g", "p"))

# or you can use following command
qqnorm(xmp04.30$Voltage)
qqline(xmp04.30$Voltage)
```



How to Draw a Normal Probability Plot

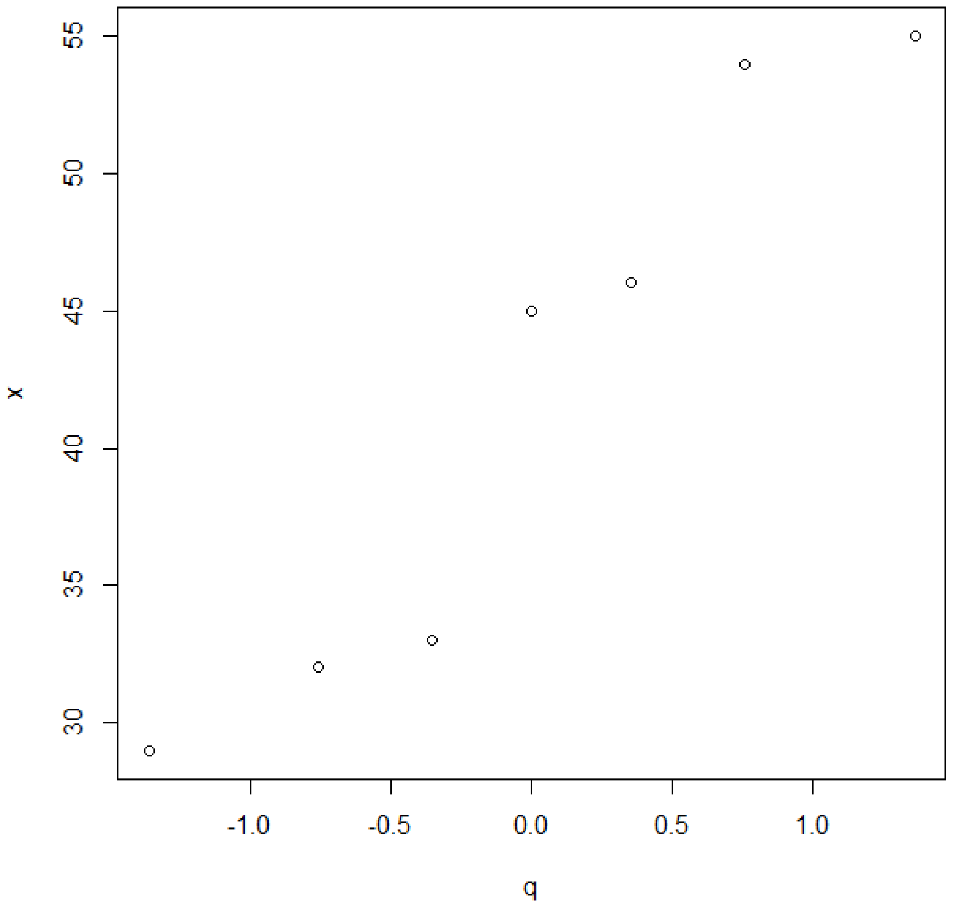
- Arrange your x-values in ascending order.
- Calculate f_i = (i-0.375)/(n+0.25), where i is the position of the data value in the ordered list and n is the number of observations.
- Find the z-score for each f_i
- Plot your z-score on the horizontal axis and the corresponding x values on the vertical axis.
- Ex) We have 7 values : 29, 32, 33, 45, 46, 5

	29	32	33	45	46		
f	0.0862	0.224	0.362	0.5	0.638	0.776	0.914
P(Z<z)=f	-1.364	-0.758	-0.353	0	0.353	0.758	1.364

1	2	3	4	5
0.1	0.3	0.5	0.7	0.9

How to Draw a Normal Probability Plot

```
> x <- c(29, 32, 33, 45, 46, 54, 55)
> qqnorm(x)
> qqline(x)

> f <- (1:length(x) - 0.375)/(length(x)+0.2
> q <- qnorm(f)
> plot(q, x)
```



How to Draw a Normal Probability Plot

- Arrange your x-values in ascending order.
- Calculate f_i = (i-0.375)/(n+0.25), where i is the position of the data value in the ordered list and n is the number of observations.
- Find the z-score for each f_i
- Plot your z-score on the horizontal axis and the corresponding x-values on the vertical axis.
- Ex) We have 7 values : 29, 32, 33, 45, 46, 5

	29	32	33	45	46		
f	0.0862	0.224	0.362	0.5	0.638	0.776	0.914
P(Z<z)=f	-1.364	-0.758	-0.353	0	0.353	0.758	1.364

1	**2**	**3**	**4**	**5**
0.1	0.3	0.5	0.7	0.9

EasyFit

- http://www.download3000.com/download-easyfit-count-reg-9662.html
- *Download*

  http://www.mathwave.com/downloads.html

Free Downloads

The free trial versions are provided for your use during a trial period, allowing you to evaluate the functionality of the products and determine whether they meet your needs.

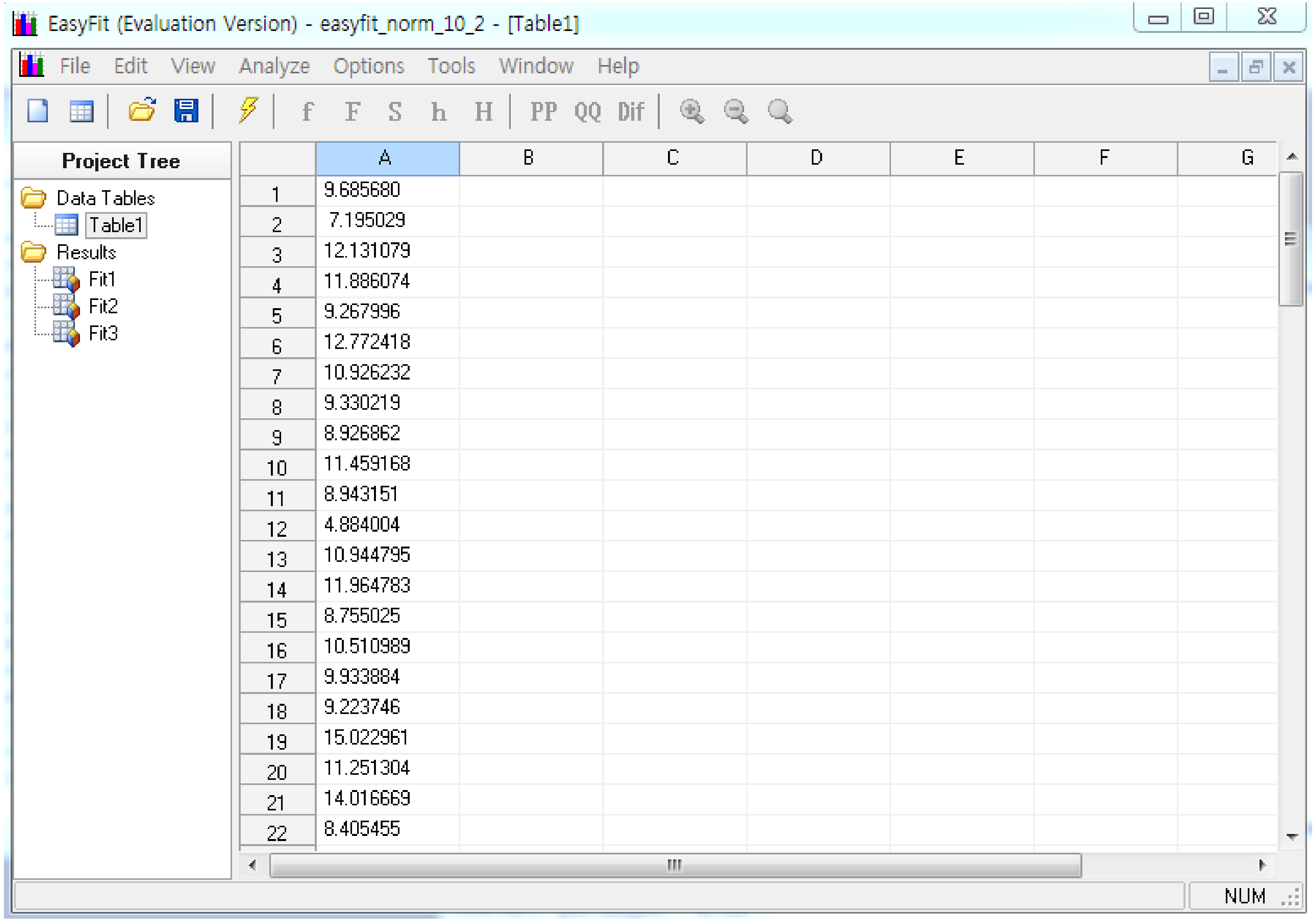
EasyFit (Standard & Professional)	**Simulation & Probabilistic Analysis SDK** (Software Development Kit)
Download	Download
Version: 5.6 File size: 3.6 MB	Version: 1.2 File size: 1.7 MB

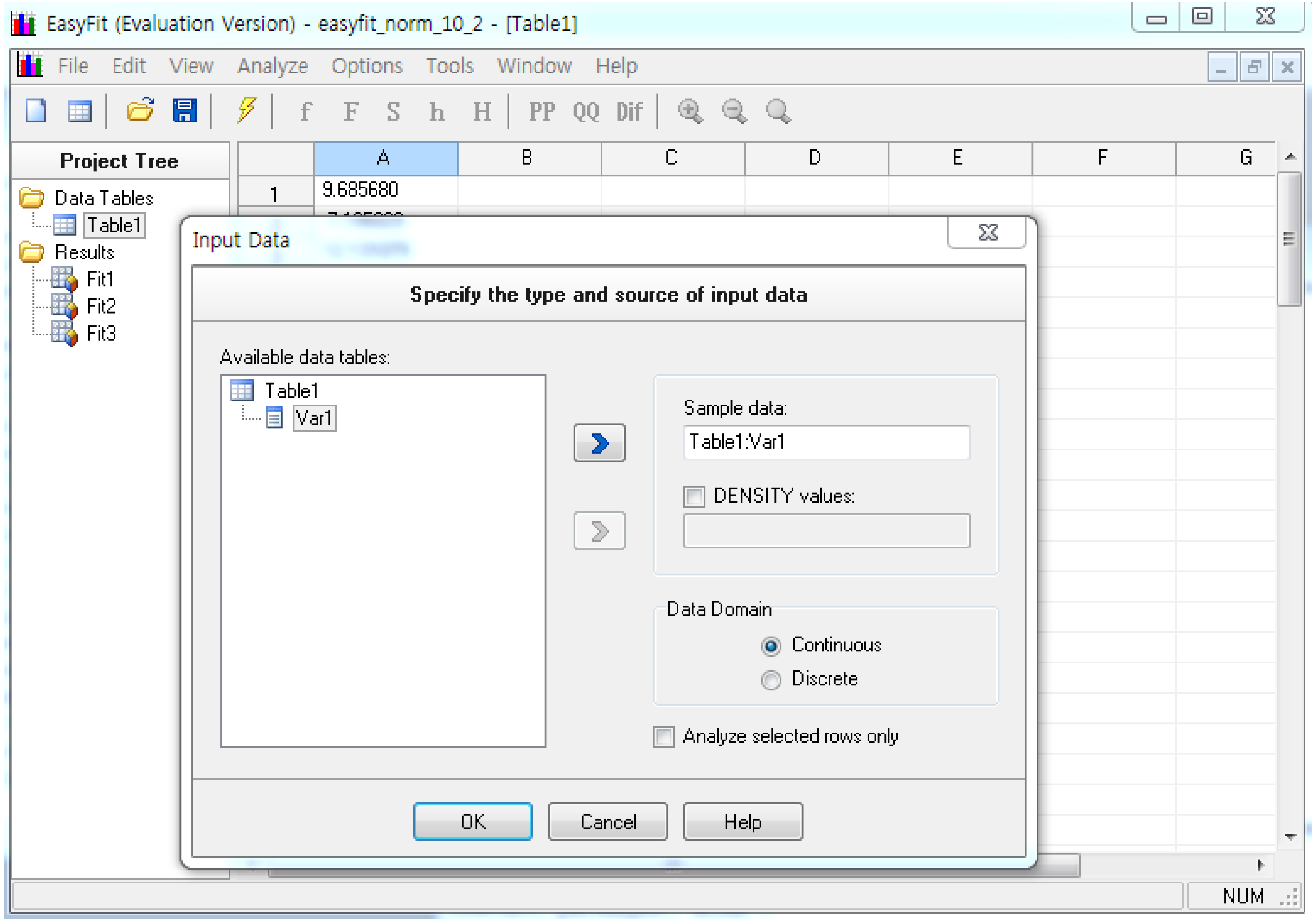
System Requirements:
Windows 8/7/Vista/XP/2000, Windows Server 2003/2008
20 MB available hard disk space
128 MB available RAM

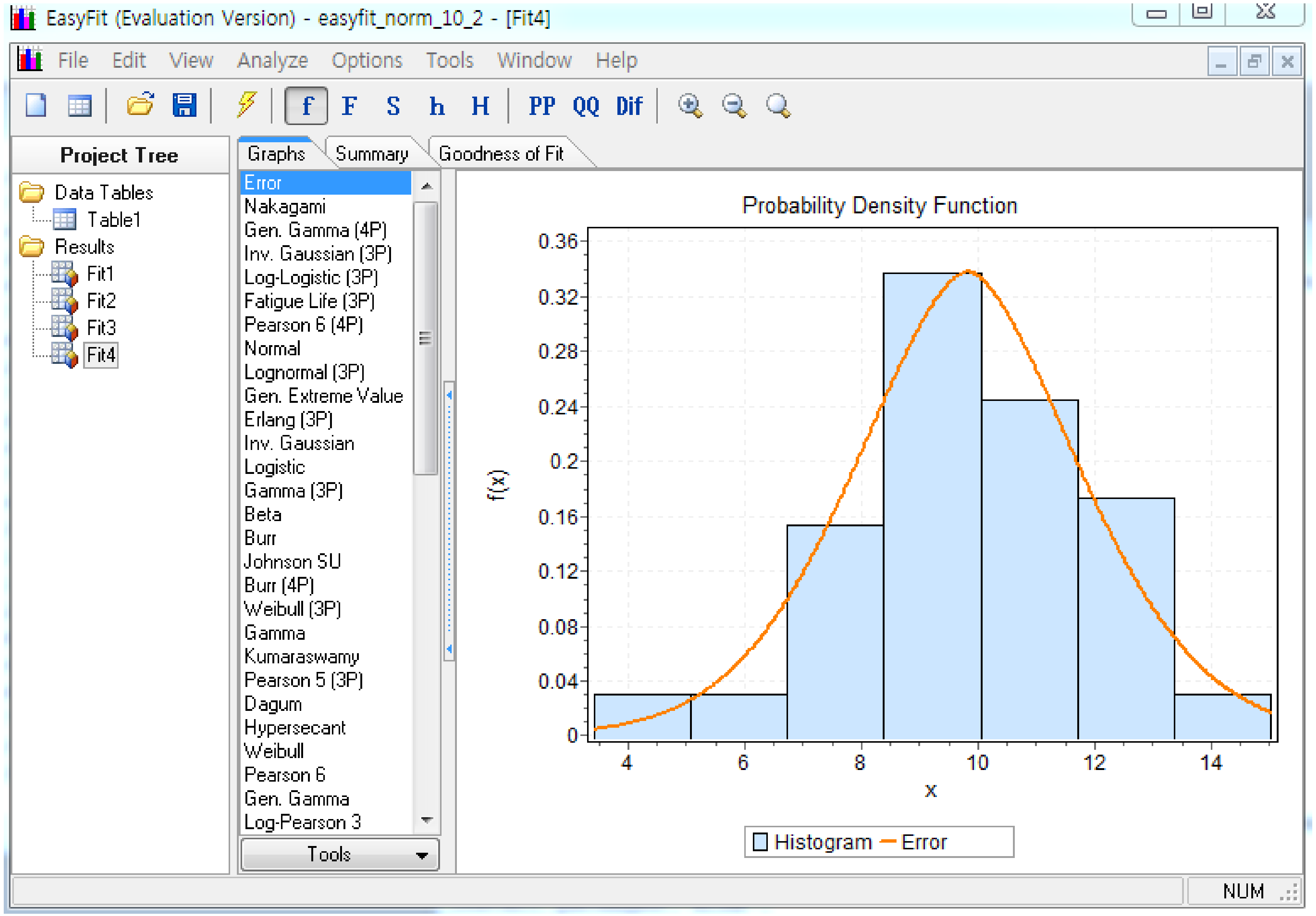
Note: EasyFitXL add-in (a part of EasyFit Professional) is currently compatible with 32-bit versions of Excel (support for 64-bit Excel coming soon)

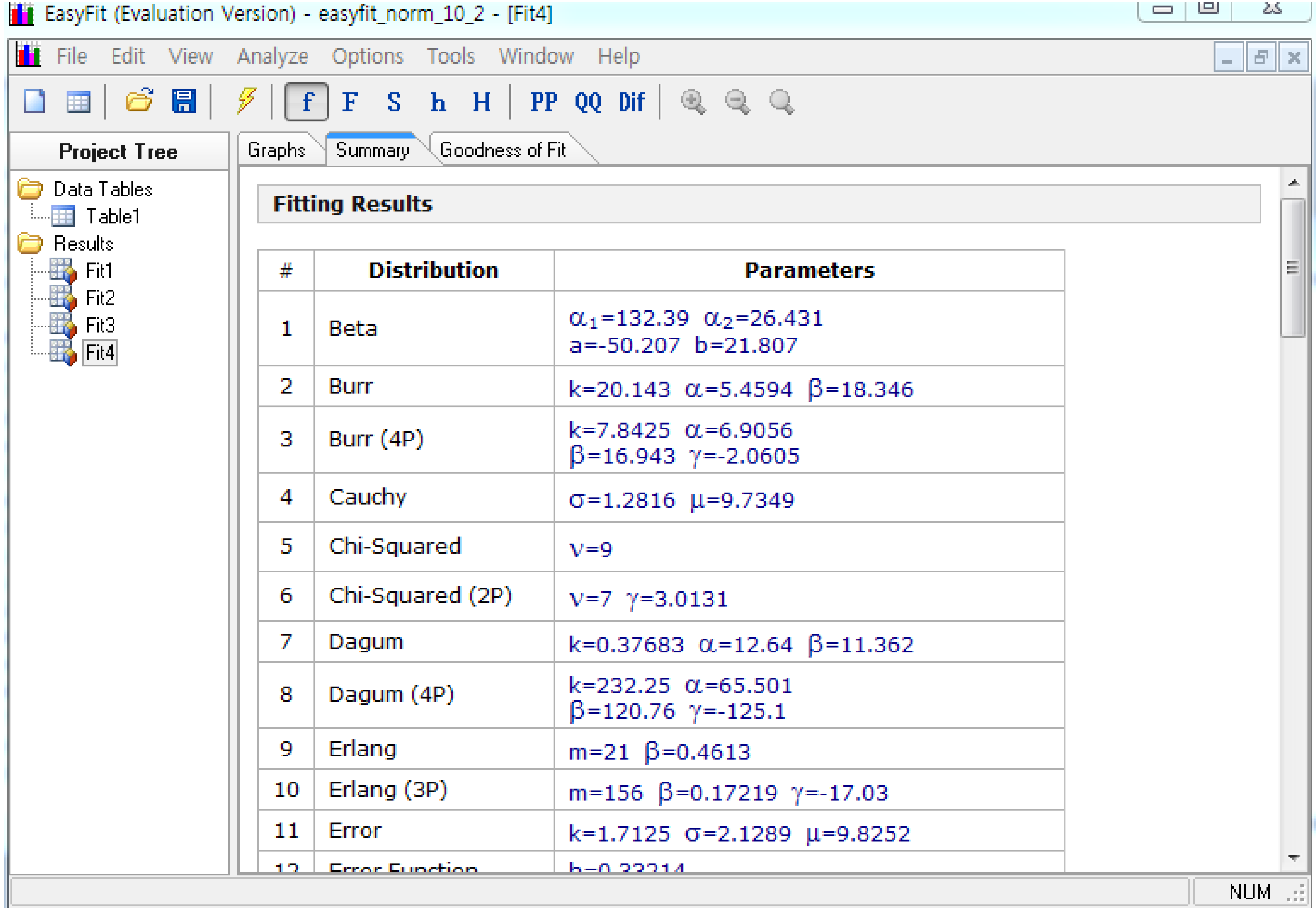
Airport Simulation

- Is it necessary to build new runway to serve the increasing traveling customers after 10 years?
- We can analyze by computer simulation.
- In order to reproduce the operation of the runway inside the computer, we need to know the inter-arrival time distribution of the airplanes, the landing time distributions of the airplanes.
- We measure the landing times of the airplane and want to know the distribution type that fits well to the measured data to generate the random values that depict the landing times in the simulation.
- So we need to find the appropriate probability distribution for the given data.

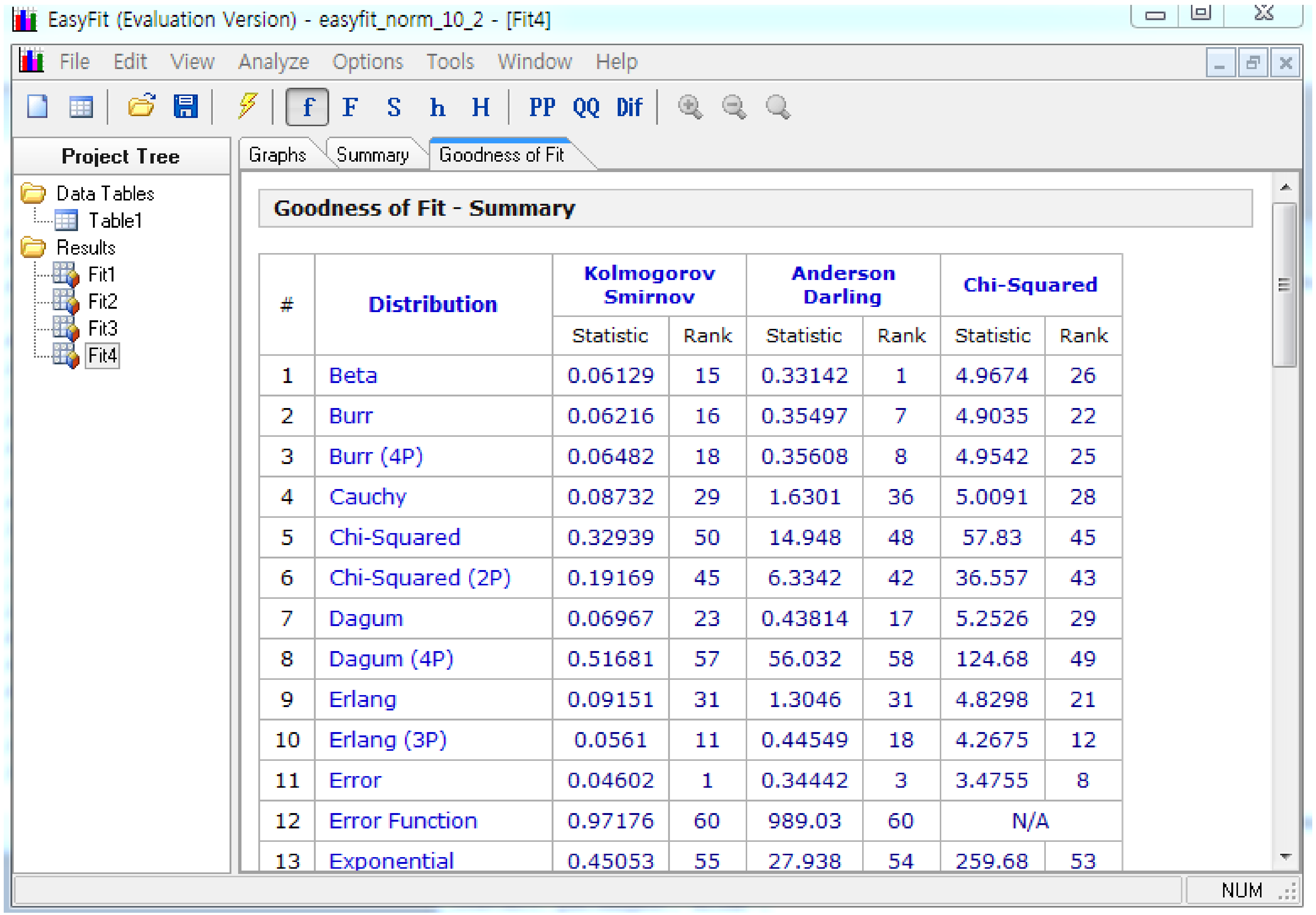

EasyFit (Evaluation Version) - easyfit_norm_10_2 - [Table1]
File Edit View Analyze Options Tools Window Help
f F S h H PP QQ Dif
Project Tree
Data Tables
Table1
Results
Fit1
Fit2
Fit3
A B C D E F G
1 9.685680
2 7.195029
3 12.131079
4 11.886074
5 9.267996
6 12.772418
7 10.926232
8 9.330219
9 8.926862
10 11.459168
11 8.943151
12 4.884004
13 10.944795
14 11.964783
15 8.755025
16 10.510989
17 9.933884
18 9.223746
19 15.022961
20 11.251304
21 14.016669
22 8.405455
NUM


EasyFit (Evaluation Version) - easyfit_norm_10_2 - [Table1]
File Edit View Analyze Options Tools Window Help
Project Tree
Data Tables
Table1
Results
Fit1
Fit2
Fit3
9.685680
Input Data
Specify the type and source of input data
Available data tables:
Table1
Var1
Sample data:
Table1:Var1
DENSITY values:
Data Domain
Continuous
Discrete
Analyze selected rows only
OK
Cancel
Help
NUM


EasyFit (Evaluation Version) - easyfit_norm_10_2 - [Fit4]
File Edit View Analyze Options Tools Window Help
f F S h H PP QQ Dif
Project Tree
Data Tables
Table1
Results
Fit1
Fit2
Fit3
Fit4
Graphs
Summary
Goodness of Fit
Error
Nakagami
Gen. Gamma (4P)
Inv. Gaussian (3P)
Log-Logistic (3P)
Fatigue Life (3P)
Pearson 6 (4P)
Normal
Lognormal (3P)
Gen. Extreme Value
Erlang (3P)
Inv. Gaussian
Logistic
Gamma (3P)
Beta
Burr
Johnson SU
Burr (4P)
Weibull (3P)
Gamma
Kumaraswamy
Pearson 5 (3P)
Dagum
Hypersecant
Weibull
Pearson 6
Gen. Gamma
Log-Pearson 3
Tools
Probability Density Function
f(x)
0.36
0.32
0.28
0.24
0.2
0.16
0.12
0.08
0.04
0
4
6
8
10
12
14
x
Histogram
Error
NUM



#	Distribution	Parameters
1	Beta	α_1=132.39 α_2=26.431 a=-50.207 b=21.807
2	Burr	k=20.143 α=5.4594 β=18.346
3	Burr (4P)	k=7.8425 α=6.9056 β=16.943 γ=-2.0605
4	Cauchy	σ=1.2816 μ=9.7349
5	Chi-Squared	ν=9
6	Chi-Squared (2P)	ν=7 γ=3.0131
7	Dagum	k=0.37683 α=12.64 β=11.362
8	Dagum (4P)	k=232.25 α=65.501 β=120.76 γ=-125.1
9	Erlang	m=21 β=0.4613
10	Erlang (3P)	m=156 β=0.17219 γ=-17.03
11	Error	k=1.7125 σ=2.1289 μ=9.8252
12	Error Function	h=0.33214



Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Beta	0.06129	15	0.33142	1	4.9674	26
2	Burr	0.06216	16	0.35497	7	4.9035	22
3	Burr (4P)	0.06482	18	0.35608	8	4.9542	25
4	Cauchy	0.08732	29	1.6301	36	5.0091	28
5	Chi-Squared	0.32939	50	14.948	48	57.83	45
6	Chi-Squared (2P)	0.19169	45	6.3342	42	36.557	43
7	Dagum	0.06967	23	0.43814	17	5.2526	29
8	Dagum (4P)	0.51681	57	56.032	58	124.68	49
9	Erlang	0.09151	31	1.3046	31	4.8298	21
10	Erlang (3P)	0.0561	11	0.44549	18	4.2675	12
11	Error	0.04602	1	0.34442	3	3.4755	8
12	Error Function	0.97176	60	989.03	60	N/A	
13	Exponential	0.45053	55	27.938	54	259.68	53

Fitting distribution in R

- #https://stats.stackexchange.com/questions/132652/how-to-determine-which-distribution-fits-my-data-best

mydata <- scan("C:₩₩mydata.txt")

mydata

install.packages("fitdistrplus")

library(fitdistrplus)

fitdist(mydata, "weibull")

ks.test(mydata, "pweibull", scale=43.2474500, shape=6.4632971)

ks.test(mydata, "pnorm", mean=mean(mydata), sd=sd(mydata))

- The p-values are 0.8669 for the Weibull distribution, and 0.5522 for the normal distribution. Thus I can assume that my data follows a Weibull as well as a normal distribution.

Fitting distribution in R

```
> fit_w  <- fitdist(mydata, "weibull")
> fit_n <- fitdist(mydata, "norm")
> gofstat(list(fit_w, fit_n), fitnames = c("weibull", "norm"))
Goodness-of-fit statistics
                                 weibull       norm
Kolmogorov-Smirnov statistic  0.06863193 0.09013127
Cramer-von Mises statistic    0.05673634 0.09587878
Anderson-Darling statistic    0.38619340 0.59723294

Goodness-of-fit criteria
                                weibull     norm
Akaike's Information Criterion 519.8537 523.3079
Bayesian Information Criterion 524.5151 527.9694
```

- the AIC of the Weibull fit is lower compared to the normal fit: the Weibull maybe looks a bit better

- http://www.di.fc.ul.pt/~jpn/r/distributions/fitting.html

```
data("groundbeef", package = "fitdistrplus")
my_data <- groundbeef$serving

fit_w  <- fitdist(my_data, "weibull")
fit_g  <- fitdist(my_data, "gamma")
fit_ln <- fitdist(my_data, "lnorm")
```

```
> gofstat(list(fit_w, fit_g, fit_ln), fitnames = c("weibull", "gamma", "lnorm"))
```

```
Goodness-of-fit statistics
                              weibull     gamma     lnorm
Kolmogorov-Smirnov statistic 0.1396646 0.1281246 0.1493090
Cramer-von Mises statistic   0.6840994 0.6934112 0.8277358
Anderson-Darling statistic   3.5736460 3.5660192 4.5436542

Goodness-of-fit criteria
                               weibull    gamma    lnorm
Akaike's Information Criterion 2514.449 2511.250 2526.639
Bayesian Information Criterion 2521.524 2518.325 2533.713
```

- Gamma distribution is the preferred one among the candidates.
- https://cran.r-project.org/web/packages/fitdistrplus/vignettes/paper2JSS.pdf

- https://www.sciencedirect.com/topics/medicine-and-dentistry/akaike-information-criterion
- A good model is the one that has minimum AIC among all the other models.
- A lower AIC or BIC value indicates a better fit.

$$AIC = -2 * \ln(L) + 2 * k$$

$$BIC = -2 * \ln(L) + 2 * \ln(N) * k$$

- Where L is the value of the likelihood, N is the number of recorded measurements, and k is the number of estimated parameters.

- http://www.di.fc.ul.pt/~jpn/r/distributions/fitting.html

```
data("endosulfan", package = "fitdistrplus")
my_data <- endosulfan$ATV

install.packages("actuar")
library(actuar)

fit_ln <- fitdist(my_data, "lnorm")
fit_ll <- fitdist(my_data, "llogis", start = list(shape = 1, scale = 500)) # logistic distribution
fit_P  <- fitdist(my_data, "pareto", start = list(shape = 1, scale = 500))
fit_B  <- fitdist(my_data, "burr",   start = list(shape1 = 0.3, shape2 = 1, rate = 1))
```

> gofstat(list(fit_ln, fit_ll, fit_P, fit_B), fitnames = c("lnorm", "llogis", "Pareto", "Burr"))

```
Goodness-of-fit statistics
                                 lnorm    llogis     Pareto      Burr
Kolmogorov-Smirnov statistic 0.1493090 0.1397238  0.3009686 0.1424432
Cramer-von Mises statistic   0.8277358 0.8059024  6.1003998 0.6986946
Anderson-Darling statistic   4.5436542 4.4812669 31.9420338 3.5852937

Goodness-of-fit criteria
                                 lnorm   llogis   Pareto     Burr
Akaike's Information Criterion 2526.639 2529.063 2696.027 2514.190
Bayesian Information Criterion 2533.713 2536.138 2703.101 2524.802
```

- It seems that the Burr distribution seems the preferred one among the candidates.

Test if data follows normal distribution

```
> #A package called nortest (it should be downloaded from CRAN website) allows to
perform 5
> #different normality test:
> library(nortest) ## package loading
> #test a data that follows normal distribution
> x.norm<-rnorm(n=200,m=10,sd=2)
> library(nortest)
> sf.test(x.norm)#Shapiro-Francia test
> ad.test(x.norm)#Anderson-Darling21 test:
> cvm.test(x.norm)#Cramer-Von Mises test
> lillie.test(x.norm)#Lilliefors22 test:
> pearson.test(x.norm)#Pearson's chi-square test:

> #test a data that follows poisson distribution
> x.poi<-rpois(n=200,lambda=2.5)
> sf.test(x.poi)
> sf.test(aov.ex1$residuals)
```

Chapter 5

Chapter 5 – Joint Probability Distributions

Outline

① Jointly Distributed Random Variables

② Expected Values, covariance and correlation

③ Statistics and their distributions

④ Distribution of the Sample Mean

⑤ The Distribution of Linear Combinations

Joint Probability Mass Function

Let X and Y be two discrete rv's defined on S, the sample space for an experiment.

Their joint probability mass function is

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

The marginal probability mass functions of X and Y are

$$p_X(x) = \sum_y p(x, y) \text{ and } p_Y(y) = \sum_x p(x, y)$$

$$p_X(100) = \sum_y p(100, y) = p(100, 0) + p(100, 100) + p(100, 200) = 0.2 + 0.1 + 0.2 = 0.5$$

		y		
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

Two Discrete Random Variables – Example 5.1

A large insurance agency services a number of customers who purchase both a homeowner's policy and an automobile policy.

A deductible amount :

For an automobile policy : \$100 and \$250

For a homeowner's policy : 0, \$100, and \$200.

Suppose an individual is selected at random from the agency's files. Let

$X =$ the deductible amount on the automobile policy

$Y =$ the deductible amount on the homeowner's policy

Possible (X, Y) : $(100{,}0)$, $(100{,}100)$, $(100{,}200)$, $(250{,}0)$, $(250{,}100)$, and $(250{,}200)$

The joint pmf specifies the probability associated with each one of these pairs, with any other pair having probability zero.

Two Discrete Random Variables - Examples

Suppose the joint pmf is given in the joint probability table: p(x,y)

		y		
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

Then

$$\mathrm{p}(100{,}100) = P(X = 100 \text{ and } Y = 100)$$

$$= P(\$100 \text{ deductible on both policy})$$

$$= 0.10$$

The probability $P(Y \geq 100)$ is computed by summing probabilities of all (x, y) pairs for which $y \geq 100$

$$P(Y \geq 100) = p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) = 0.75$$

Two Discrete Random Variables - Examples

The possible X values are $x = 100$ and $x = 250$, so computing row totals in the joint probability table yields

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = 0.50$$

and

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = 0.50$$

The marginal pmf of X is then

$$p_X(x) = \begin{cases} 0.5 & if\ x = 100, 200 \\ 0 & otherwise \end{cases}$$

and the marginal pmf of Y is then

$$p_Y(y) = \begin{cases} 0.25 & if\ y = 0, 100 \\ 0.50 & if\ y = 200 \\ 0 & otherwise \end{cases}$$

Two Discrete Random Variables - Examples

y				
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- The expected value of a function $h(x, y)$, denoted by $\mathrm{E}[h(X, Y)]$, is given by

$$\mathrm{E}[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot p(x,y) \text{ cf : } \mathrm{E}[h(X)] = \sum_x h(x) \cdot p(x)$$

$$\mathrm{E}[X] = \sum_x \sum_y x \cdot p(x,y)$$

$$= 100 \cdot p(100,0) + 100 \cdot p(100,100) + 100 \cdot p(100,200)$$

$$+250 \cdot p(250,0) + 250 \cdot p(250,100) + 250 \cdot p(250,200)$$

$$= 100 \cdot 0.2 + 100 \cdot 0.1 + 100 \cdot 0.2 + 250 \cdot 0.05 + 250 \cdot 0.15 + 250 \cdot 0.3$$

$$= 20 + 10 + 20 + 12.5 + 37.5 + 75 = 175$$

$$\mathrm{E}[X] = \sum_x x\, p_X(x) = 100 \cdot p_X(100) + 250 \cdot p_X(250) \qquad (p_X(100) = p_X(250) = 0.5)$$

$$= 100 \cdot 0.5 + 250 \cdot 0.5 = 175$$

Joint Probability Density Function

If X and Y are two continuous r.v.'s then $f(x,y)$ is their joint density function if

$$P[(X,Y) \in A] = \int \int_A f(x,y) dx dy$$

cf) $P(a \le x \le b) = \int_a^b f(x) dx$

In particular, if A is the two-dimensional rectangle $\{(x,y): a \le x \le b, c \le y \le d\}$, then

$$P[(X,Y) \in A] = P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x,y) dx dy$$

The marginal probability density functions of X and Y are

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

Example of Joint Probability Density : Example 5.3

A bank operates both a drive-up facility and a walk-up window.

Let X = the proportion of time that the drive-up facility is in use,

Y = the proportion of time that walk-up window is in use.

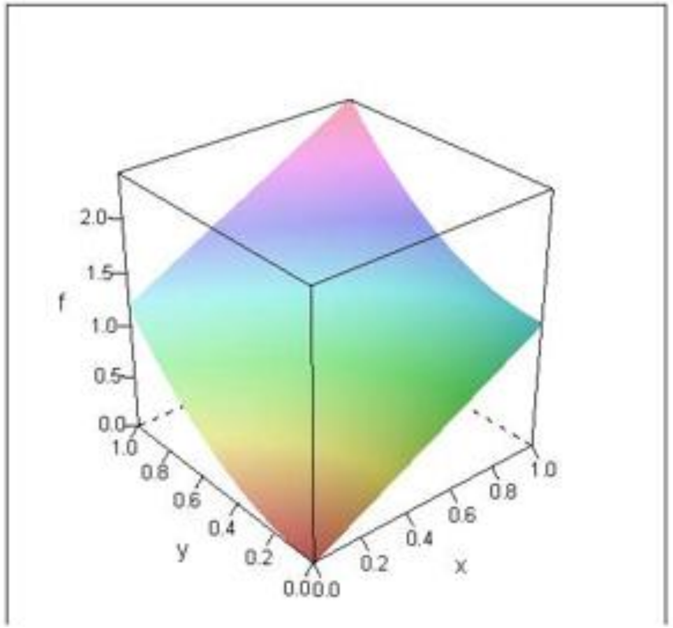
Suppose the joint pdf of (x, y) is given by

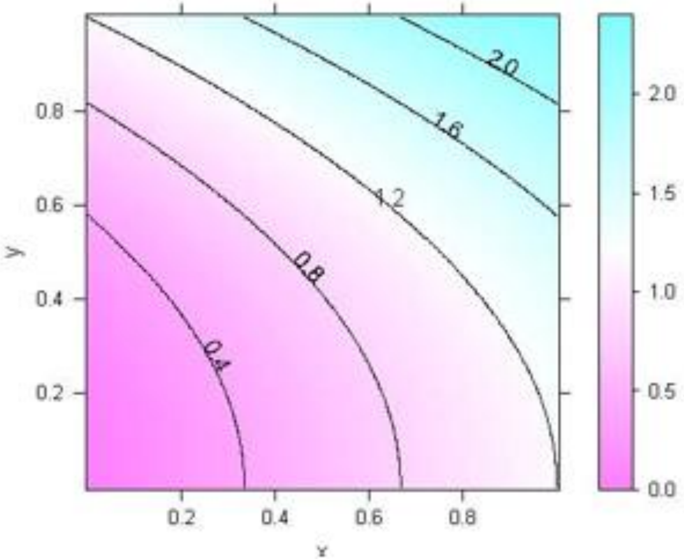
$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

We can see that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^1\int_0^1 \frac{6}{5}(x+y^2)dxdy = \int_0^1 \frac{6}{5}\left[\frac{x^2}{2} + xy^2\bigg|\begin{matrix} x=1 \\ x=0 \end{matrix}\right]dy$$

$$= \int_0^1 \left(\frac{3}{5} + \frac{6}{5}y^2\right)dy = \left(\frac{3}{5}y + \frac{2}{5}y^3\right)\bigg|\begin{matrix} y=1 \\ y=0 \end{matrix} = 1 \qquad \left(\int x^a dx = \frac{x^{a+1}}{a+1} + c\right)$$





Contour line : shows the same height

Example of Joint Probability Density : Example 5.3

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)dxdy$$

$$= \int_0^{1/4} \frac{6}{5}\left[\frac{x^2}{2} + xy^2 \Big|\begin{matrix} x = 1/4 \\ x = 0 \end{matrix}\right]dy$$

$$= \int_0^{1/4} \frac{6}{5}\left(\frac{1}{32} + \frac{1}{4}y^2\right)dy = \frac{6}{5}\left(\frac{1}{32}y + \frac{1}{12}y^3\right)\Big|\begin{matrix} y = 1/4 \\ y = 0 \end{matrix}$$

$$= \frac{6}{5}\left(\frac{1}{32} \times \frac{1}{4} + \frac{1}{12} \times \frac{1}{64}\right) = \frac{7}{640} = 0.0109$$

The marginal pdf of X

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{6}{5}(x+y^2)dy = \left(\frac{6}{5}xy + \frac{2}{5}y^3\right)\Big|_{y=0}^{y=1}$$

$$= \frac{6}{5}x + \frac{2}{5} \qquad \text{for } 0 \le x \le 1$$

$$= 0 \qquad \text{otherwise}$$

The marginal pdf of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 \frac{6}{5}(x+y^2)dx = \left(\frac{3}{5}x^2 + \frac{6}{5}xy^2\right)\Big|_{x=0}^{x=1}$$

$$= \frac{6}{5}y^2 + \frac{3}{5} \qquad \text{for } 0 \le y \le 1$$

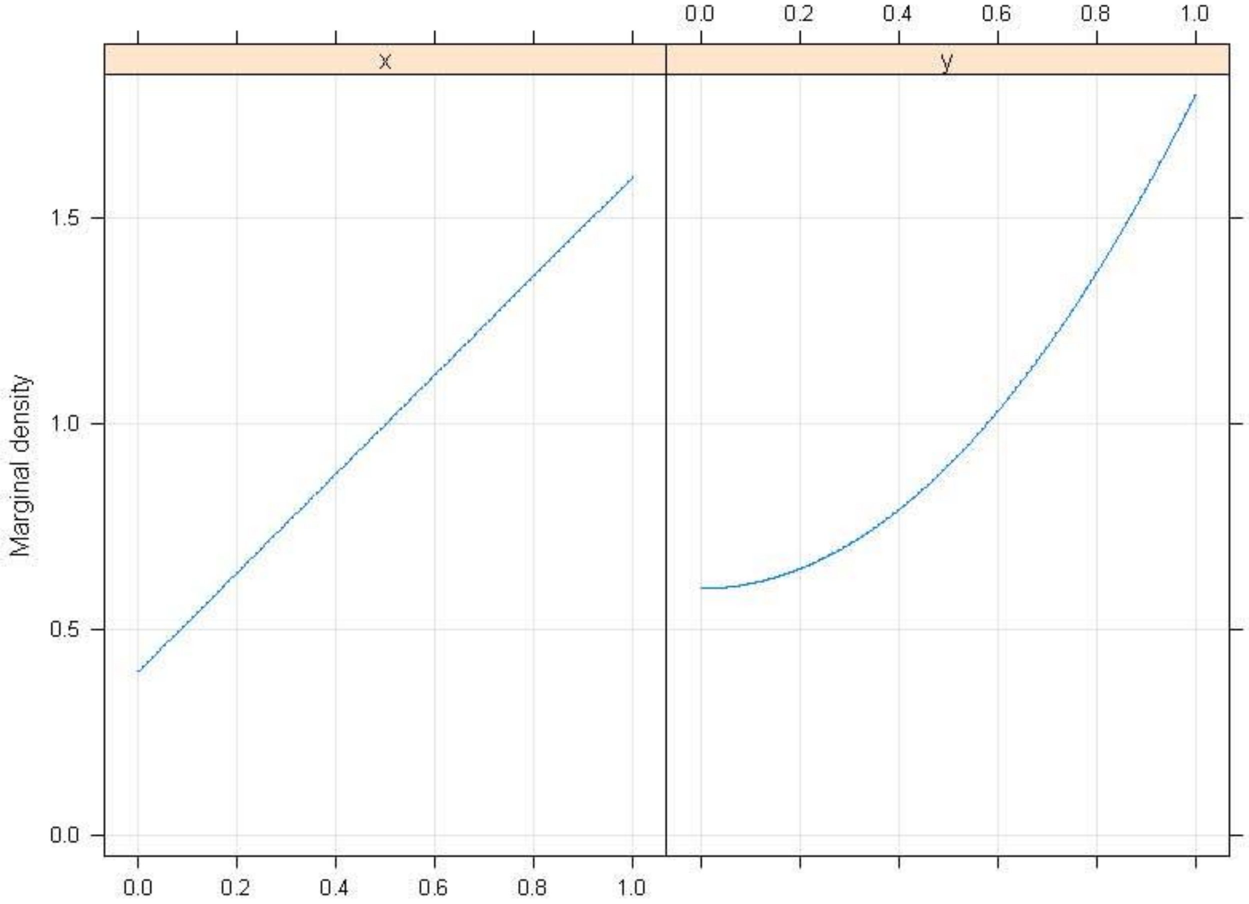
$$= 0 \qquad \text{otherwise}$$

Then $\qquad P\left[\frac{1}{4} \le Y \le \frac{3}{4}\right] = \int_{1/4}^{3/4} \int_0^1 \frac{6}{5}(x+y^2)dxdy$

or $\qquad P\left[\frac{1}{4} \le Y \le \frac{3}{4}\right] = \int_{1/4}^{3/4} f_Y(y)dy = \int_{1/4}^{3/4} \left(\frac{6}{5}y^2 + \frac{3}{5}\right)dy = \left(\frac{2}{5}y^3 + \frac{3}{5}y\right)\Big|_{y=1/4}^{y=3/4}$

$$= \left(\frac{2}{5}\times\frac{27}{64} + \frac{3}{5}\times\frac{3}{4}\right) - \left(\frac{2}{5}\times\frac{1}{64} + \frac{3}{5}\times\frac{1}{4}\right) = \frac{37}{80} = 0.4625$$

Marginal densities in example



Independent Random Variables

- Discrete random variables X and Y are said to be independent if

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

- Continuous random variables X and Y are said to be independent if

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

- If these conditions don't hold then X and Y are said to be dependent.

Two Continuous Random Variables - Examples

In the insurance situation

$$\mathrm{p}(100{,}100) = 0.10 \neq (0.5)(0.25) = p_X(100) \cdot p_Y(100)$$

so X and Y are not independent.

Independence of two random variables is most useful when the description of the experiment under study suggests that X and Y have no effect on one another.

Then once the marginal pmfs or pdfs have been specified, the joint pmf or pdf is simply the product of the two marginal functions. It follows that

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b)P(c \leq Y \leq d)$$

proof) $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)dxdy = \int_a^b \int_c^d f_X(x) \cdot f_Y(y)dxdy$

$$= \int_a^b f_X(x)dx \int_c^d f_Y(y)dy$$

Two Discrete Random Variables - Examples

Suppose the joint pmf is given in the joint probability table: p(x,y)

		y		
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

Then

$$\mathrm{p}(100{,}100) = P(X = 100 \text{ and } Y = 100)$$

$$= P(\$100 \text{ deductible on both policy})$$

$$= 0.10$$

The probability $P(Y \geq 100)$ is computed by summing probabilities of all (x, y) pairs for which $y \geq 100$

$$P(Y \geq 100) = p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) = 0.75$$

Exercise 5.1 #19

You have two lightbulbs for a particular lamp.

Let $X =$ the lifetime of the first bulb

$Y =$ the lifetime of the second bulb (both in 1000s of hours).

Suppose that X and Y are independent and that each has an exponential distribution with parameter $\lambda = 1$.

① What is the joint pdf of X and Y?

② What is the probability that each bulb last at most 1000 hours (i.e. $X \leq 1$ and $Y \leq 1$)?

③ What is the probability that the total lifetime of the two bulbs is at most 2?

[Hint: Draw a picture of the region $A = \{(x, y): x \geq 0, y \geq 0, x + y \leq 2\}$ before integrating.]

④ What is the probability that the total lifetime of the two bulbs is larger than

- The joint pdf is

$$f(x, y) = f_X(x) \cdot f_Y(y) \qquad\qquad f(x) = \lambda e^{-\lambda x} \quad x > 0$$

$$= \begin{cases} e^{-x} \cdot e^{-y} = e^{-x-y} & x > 0, y > 0 \\ 0 & otherwise \end{cases}$$

- $P(X \leq 1, Y \leq 1) = P(X \leq 1) \cdot P(Y \leq 1)$

$$= \int_0^1 e^{-x} dx \cdot \int_0^1 e^{-y} dy = (-e^{-x})\Big|_{x=0}^{x=1} \cdot (-e^{-y})\Big|_{y=0}^{y=1}$$

$$= (1 - e^{-1}) \cdot (1 - e^{-1}) = 0.3996$$

$$\left(\int e^{ax} = \frac{1}{a} e^{ax}, \qquad \frac{d}{dx} e^{f(x)} = e^{f(x)} f'(x), \qquad \int f'(x) e^{f(x)} dx = e^{f(x)} \right)$$

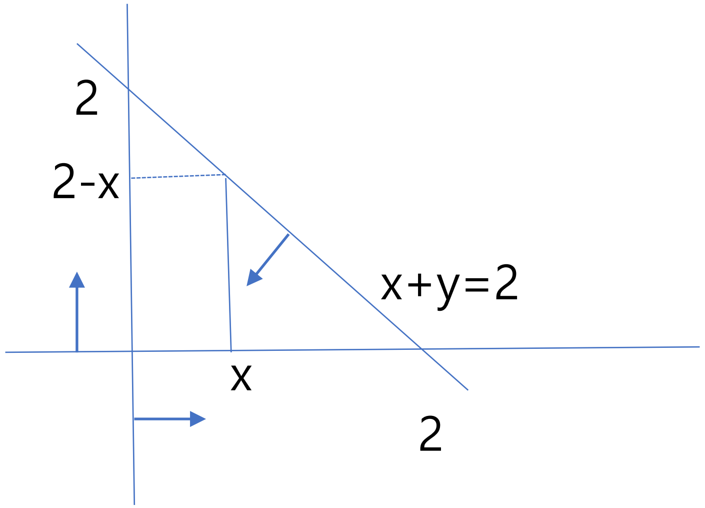
$$P(x+y\le 2) = \int_0^2\int_0^{2-x} e^{-x-y}\,dydx$$

$$= \int_0^2 \left(-e^{-x-y}\Big|_{y=0}^{y=2-x}\right)dx = \int_0^2 [-e^{-x-(2-x)} - (-e^{-x})]\,dx$$

$$= \int_0^2 (-e^{-2} + e^{-x})\,dx$$

$$= (-e^{-2}x - e^{-x})\Big|_{x=0}^{x=2} = (-2e^{-2} - e^{-2}) - (-1)$$

$$= 1 - 2e^{-2} - e^{-2} = 0.594$$



$$P(1 < x + y < 2) = P(x + y < 2) - P(x + y < 1)$$

$$P(x + y < 1) = \int_0^1 \int_0^{1-x} e^{-x-y} dydx$$

$$= \int_0^1 (-e^{-x-y} \Big|_{y=0}^{y=1-x}) dx = \int_0^1 (-e^{-1} + e^{-x})\, dx$$

$$= (-e^{-1}x - e^{-x}) \Big|_{x=0}^{x=1} = (-e^{-1} - e^{-1}) - (-1)$$

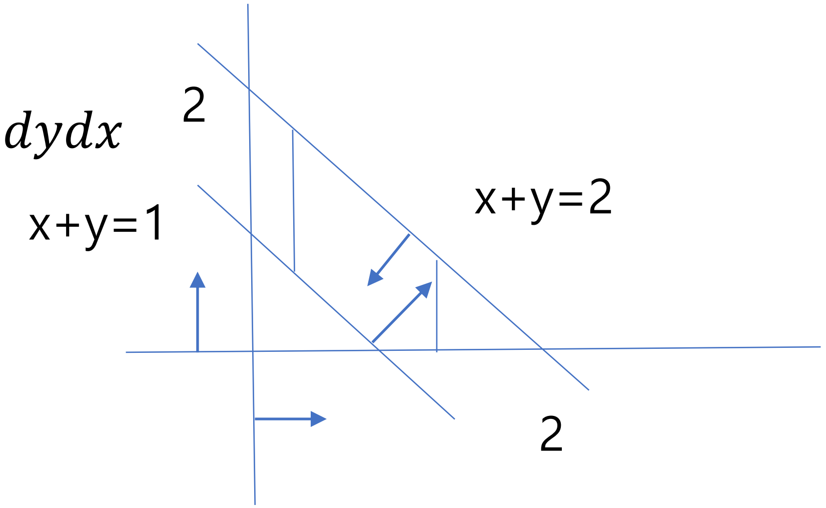
$$= 1 - 2e^{-1} = 0.264$$

$$P(1 < x + y < 2) = 0.594 - 0.264 = 0.33$$

or

$$P(1 < x + y < 2)$$

$$= \int_0^1 \int_{1-x}^{2-x} e^{-x-y} dydx + \int_1^2 \int_0^{2-x} e^{-x-y} dydx$$



Conditional distributions

- For continuous random variables X and Y with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$, the conditional probability density of Y, given $X = x$ is

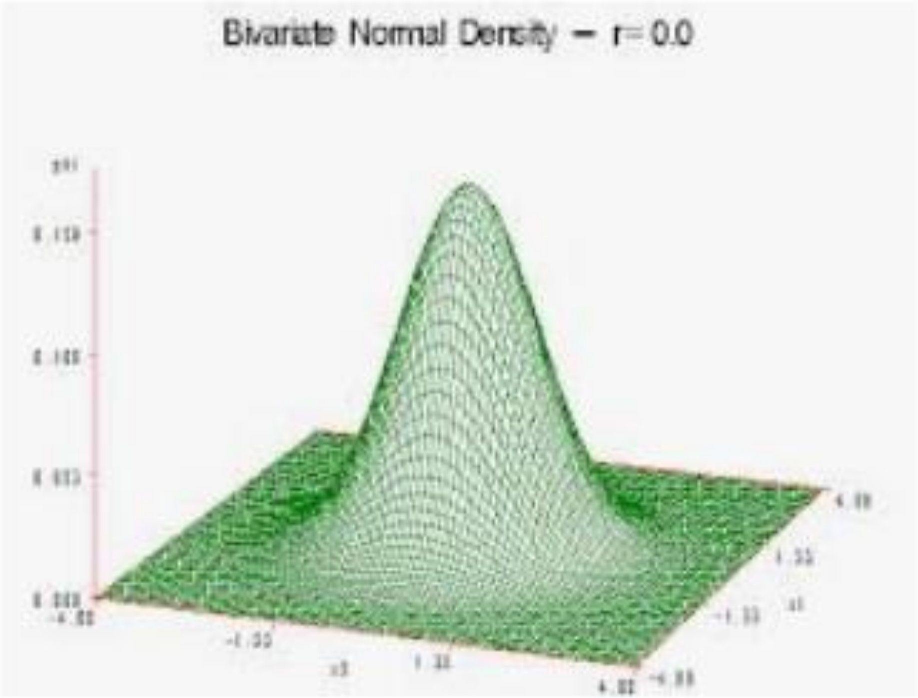
$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \qquad -\infty \leq y \leq \infty \qquad \text{cf) } \mathrm{P}(A|B) = \frac{P(A \cap B)}{P(B)}$$

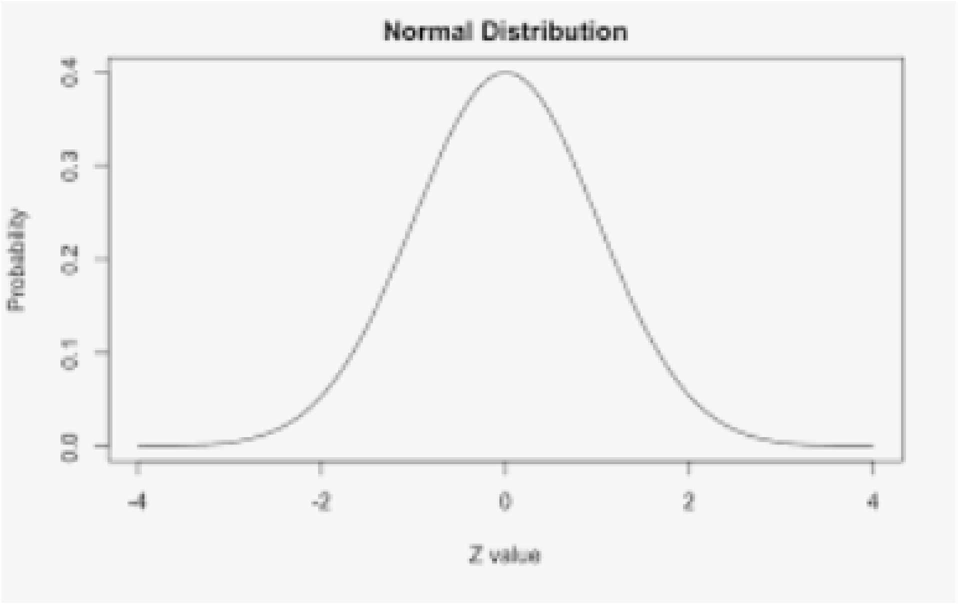
provided that $f_X(x) > 0$

- For discrete random variables X and Y with joint pmf $p(x, y)$ and marginal pmfs $p_X(x)$ and $p_Y(y)$, the conditional pmf of Y, given $X = x$ is

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}$$

provided that $p_X(x) > 0$





Cross Section at
X=x

Two Discrete Random Variables - Examples

Suppose the joint pmf is given in the joint probability table: p(x,y)

The marginal pmf of X is then

		y		
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

$$p_X(x) = \begin{cases} 0.5 & if\ x = 100, 250 \\ 0 & otherwise \end{cases}$$

and the marginal pmf of Y is then

$$p_Y(y) = \begin{cases} 0.25 & if\ y = 0, 100 \\ 0.50 & if\ y = 200 \\ 0 & otherwise \end{cases}$$

$$P(Y = 0|X = 100) = \frac{P(X=100, Y=0)}{P(X=100)} = \frac{0.2}{0.5} = 0.4$$

$$P(Y = 100|X = 100) = \frac{P(X=100, Y=100)}{P(X=100)} = \frac{0.1}{0.5} = 0.2$$

$$P(Y = 200|X = 100) = \frac{P(X=100, Y=200)}{P(X=100)} = \frac{0.2}{0.5} = 0.4$$

Example 5.12 (Reconsider Example 5.4)

$X =$ proportion of time that a bank's drive-up facility is busy $f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & otherwise \end{cases}$

$Y =$ proportion of time that a bank's walk-up window is busy $f_X(x) = \frac{6}{5}x + \frac{2}{5}$ for $0 \le x \le 1$

The conditional pdf of Y given that $X = 0.8$ is

$$f_{Y|X}(y|0.8) = \frac{f(0.8,y)}{f_X(0.8)} = \frac{1.2\left(0.8+y^2\right)}{1.2(0.8)+0.4} = \frac{1}{34}(24+30y^2) \quad 0 \le y \le 1$$

The probability that the walk-up facility is busy at most half the time given that $X = 0.8$ is

$$P(Y \le 0.5|X = 0.8) = \int_{-\infty}^{0.5} f_{Y|X}(y|0.8)dy$$

$$= \int_0^{0.5} \frac{1}{34}(24+30y^2)dy = \left(\frac{24}{34}y + \frac{10}{34}y^3\right)\Big|_{y=0}^{y=0.5} = 0.390$$

Using the marginal pdf of Y :

$$P(Y \le 0.5) = \int_{-\infty}^{0.5} f_Y(y)dy = \int_0^{0.5}\left(\frac{6}{5}y^2 + \frac{3}{5}\right)dy = \left(\frac{2}{5}y^3 + \frac{3}{5}y\right)\Big|_{y=0}^{y=0.5}$$

Example 5.12 (Reconsider Example 5.4)

The expected proportion of time that the walk-up facility is busy given that $X = 0.8$ is

$$E(Y|X = 0.8) = \int_0^1 y \cdot f_{Y|X}(y|0.8)dy$$

$$= \frac{1}{34}\int_0^1 y \cdot (24 + 30y^2)dy = \frac{1}{34}\left(12y^2 + \frac{30}{4}y^4\right)\Big|_{y=0}^{y=1} = 0.574$$

Using the marginal pdf of Y :

$$E(Y) = \int_0^1 y \cdot f_Y(y)dy = \int_0^1 y \cdot \left(\frac{6}{5}y^2 + \frac{3}{5}\right)dy = \left(\frac{6}{20}y^4 + \frac{3}{10}y^2\right)\Big|_{y=0}^{y=1}$$

$$= \frac{6}{20} + \frac{3}{10} = 0.6$$

Expected value

- The expected value of a function $h(x, y)$, denoted by $\mathrm{E}[h(X, Y)]$, is given by

$$\mathrm{E}[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & discrete \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dxdy & continuous \end{cases} \quad (\mathrm{E}[h(X)] = \sum_x h(x) \cdot p(x))$$

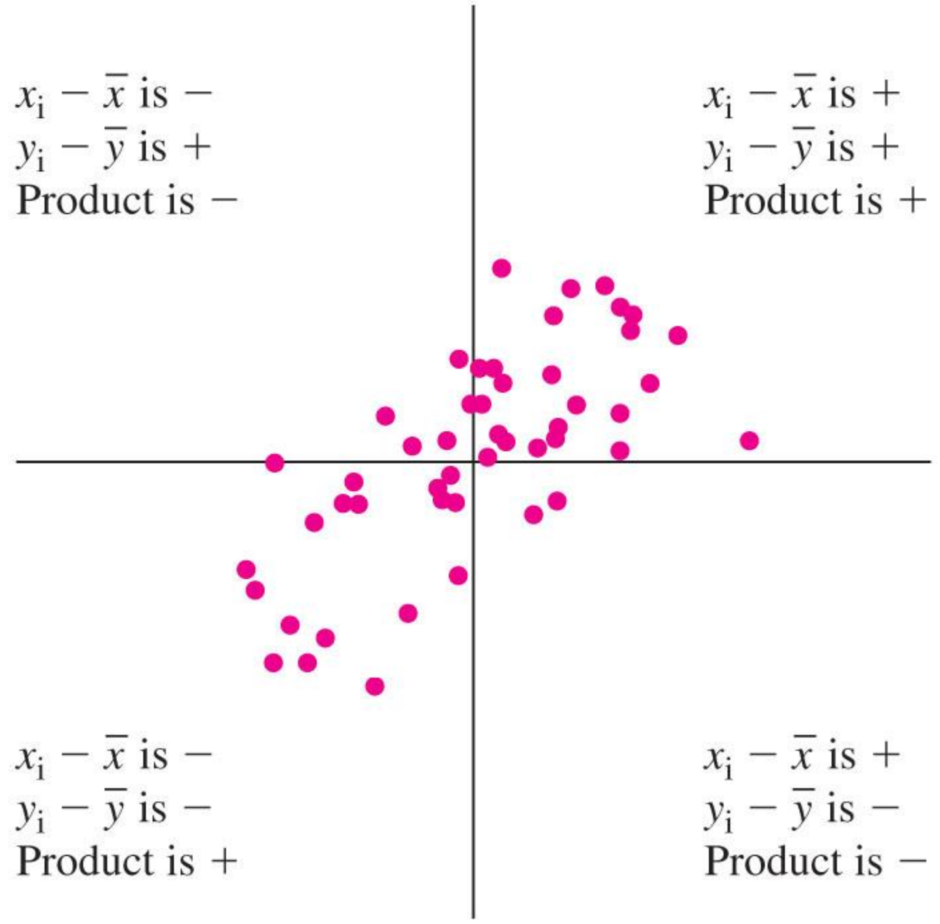
- The covariance between X and Y is

$$Cov(X, Y) = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \qquad V(X) = \mathrm{E}[(X - \mu)^2]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) \cdot p(x, y) & discrete \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y)\, dxdy & continuous \end{cases}$$

- Sometimes it is more convenient to evaluate

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



$x_i - \bar{x}$ is −
$y_i - \bar{y}$ is +
Product is −
$x_i - \bar{x}$ is +
$y_i - \bar{y}$ is +
Product is +
$x_i - \bar{x}$ is −
$y_i - \bar{y}$ is −
Product is +
$x_i - \bar{x}$ is +
$y_i - \bar{y}$ is −
Product is −

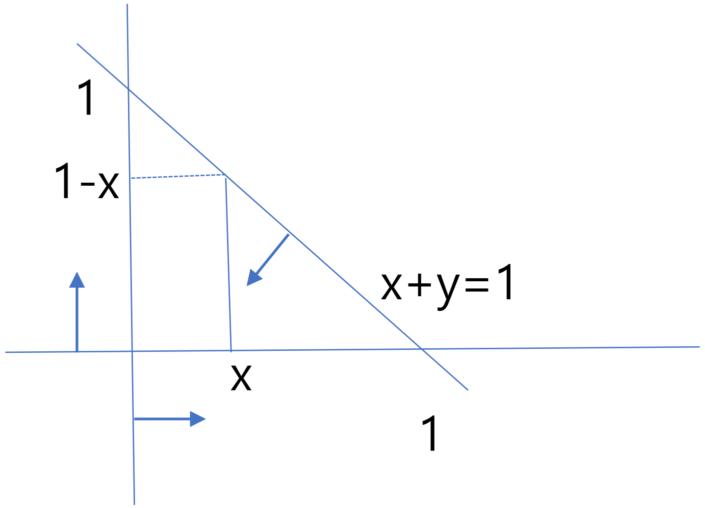
Expected Values – Example 5.14

The joint pdf of the amount X of almonds and amount Y of cashews in a 1-lb can of nuts was

$$f(x,y) = \begin{cases} 24xy & 0 \le x \le 1,\ 0 \le y \le 1,\ \ x+y \le 1 \\ 0 & otherwise \end{cases}$$

If 1 lb of almonds costs the company \$1.00, 1 lb of cashews costs \$1.50, and 1lb of peanuts costs \$0.50, then the cost of the contents of a can is

$$h(X,Y) = (1)X + (1.5)Y + (0.5)(1 - X - Y) = 0.5 + 0.5X + Y$$



Expected Values – Example 5.14

The expected total cost is

$$\mathrm{E}[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\cdot f(x,y)\,dxdy$$

$$= \int_0^1\int_0^{1-x}(0.5+0.5x+y)24xy\,dxdy$$

$$=24\int_0^1\int_0^{1-x}(0.5xy+0.5x^2y+xy^2)\ dxdy$$

$$= 24\int_0^1 \left(\frac{1}{4}xy^2+\frac{1}{4}x^2y^2+\frac{xy^3}{3}\right)\Big|_{y=0}^{y=1-x} dx$$

$$= 24\int_0^1\left(\frac{1}{4}x(1-x)^2+\frac{1}{4}x^2(1-x)^2+\frac{x(1-x)^3}{3}\right)dx$$

$$= 24\int_0^1\left(\frac{1}{4}x-\frac{1}{2}x^2+\frac{1}{4}x^3+\frac{1}{4}x^2-\frac{1}{2}x^3+\frac{1}{4}x^4+\frac{x}{3}-x^2+x^3-\frac{1}{3}x^4\right)dx$$

$$= 24\int_0^1\left(\frac{7}{12}x-\frac{5}{4}x^2+\frac{3}{4}x^3-\frac{1}{12}x^4\right)dx \quad = 24\left(\frac{7}{24}x^2-\frac{5x^3}{12}+\frac{3}{16}x^4-\frac{1}{60}x^5\right)\Big|_{x=0}^{x=1}$$

$$= 24\left(\frac{7}{24}-\frac{5}{12}+\frac{3}{16}-\frac{1}{60}\right) = 1.1$$

Expected Values – Example 5.16

The joint and marginal pdf's of X = amount of almonds and Y = amount of cashews in a 1-lb can of nuts was

$$f(x,y) = \begin{cases} 24xy & 0 \le x \le 1,\ 0 \le y \le 1,\ x+y \le 1 \\ 0 & otherwise \end{cases}$$

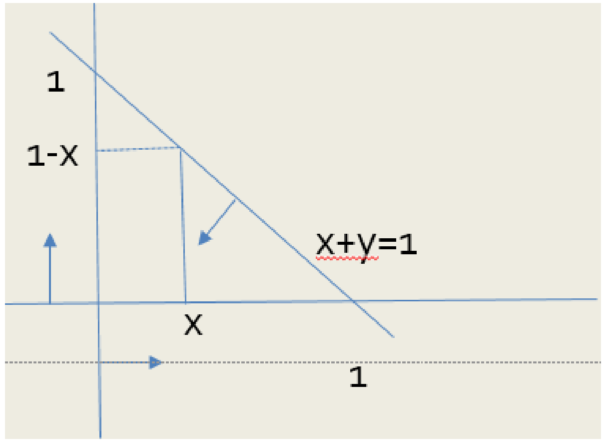
$$f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^{1-x} 24xydy = (12xy^2)\Big|_{y=0}^{y=1-x}$$

$$= 12x(1-x)^2$$

with $f_Y(y)$ obtained by replacing x by y in $f_X(x)$.

$$f_Y(y) = \begin{cases} 12y(1-y)^2 & 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$



Expected Values – Example 5.16

It is verified that $\mu_X = \mu_Y = \frac{2}{5}$

$$\mu_X = \mathrm{E}(X) = \int_0^1 x \cdot 12x(1-x)^2 dx = 12\int_0^1 (x^2 - 2x^3 + x^4) dx$$

$$= 12\left(\frac{x^3}{3} - \frac{2}{4}x^4 + \frac{1}{5}x^5\right)\Big|\begin{matrix} y=1 \\ y=0 \end{matrix} = 12\left(\frac{1}{3} - \frac{2}{4} + \frac{1}{5}\right) = 0.4$$

$$\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot 24xy\, dxdy$$

$$= \int_0^1 \int_0^{1-x} xy \cdot 24xy\, dxdy = 24\int_0^1 \int_0^{1-x} (x^2y^2)\ dxdy$$

$$= 24\int_0^1 \left(\frac{x^2y^3}{3}\right)\Big|\begin{matrix} y = 1-x \\ y=0 \end{matrix} dx = 24\int_0^1 \frac{1}{3}x^2(1-x)^3\, dx$$

$$= 8\int_0^1 (x^2 - 3x^3 + 3x^4 - x^5) dx$$

$$= 8\left(\frac{x^3}{3} - \frac{3}{4}x^4 + \frac{3}{5}x^5 - \frac{1}{6}x^5\right)\Big|\begin{matrix} x=1 \\ x=0 \end{matrix} = 8\left(\frac{1}{3} - \frac{3}{4} + \frac{3}{5} - \frac{1}{6}\right) = \frac{2}{15} = 0.1333$$

Cov(X, Y)$=\frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = -\frac{2}{75}$

Two Discrete Random Variables – Example 5.15

y				
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- The expected value of a function $h(x, y)$, denoted by $\mathrm{E}[h(X, Y)]$, is given by

$$\mathrm{E}[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot p(x,y)$$

$$\mathrm{E}[XY] = \sum_x \sum_y xy \cdot p(x,y)$$

$$= 100 \cdot 0 \cdot p(100,0) + 100 \cdot 100 \cdot p(100,100) + 100 \cdot 200 \cdot p(100,200)$$

$$+250 \cdot 0 \cdot p(250,0) + 250 \cdot 100 \cdot p(250,100) + 250 \cdot 200 \cdot p(250,200)$$

$$= 10000 \cdot 0.1 + 20000 \cdot 0.2 + 25000 \cdot 0.15 + 50000 \cdot 0.3$$

$$= 23750$$

$$\mathrm{E}[X] = \sum_x x \cdot P_X(x) = 100 \cdot 0.5 + 250 \cdot 0.5 = 175$$

$$\mathrm{E}[Y] = \sum_y y \cdot P_Y(y) = 0 \cdot 0.25 + 100 \cdot 0.25 + 200 \cdot 0.5 = 125$$

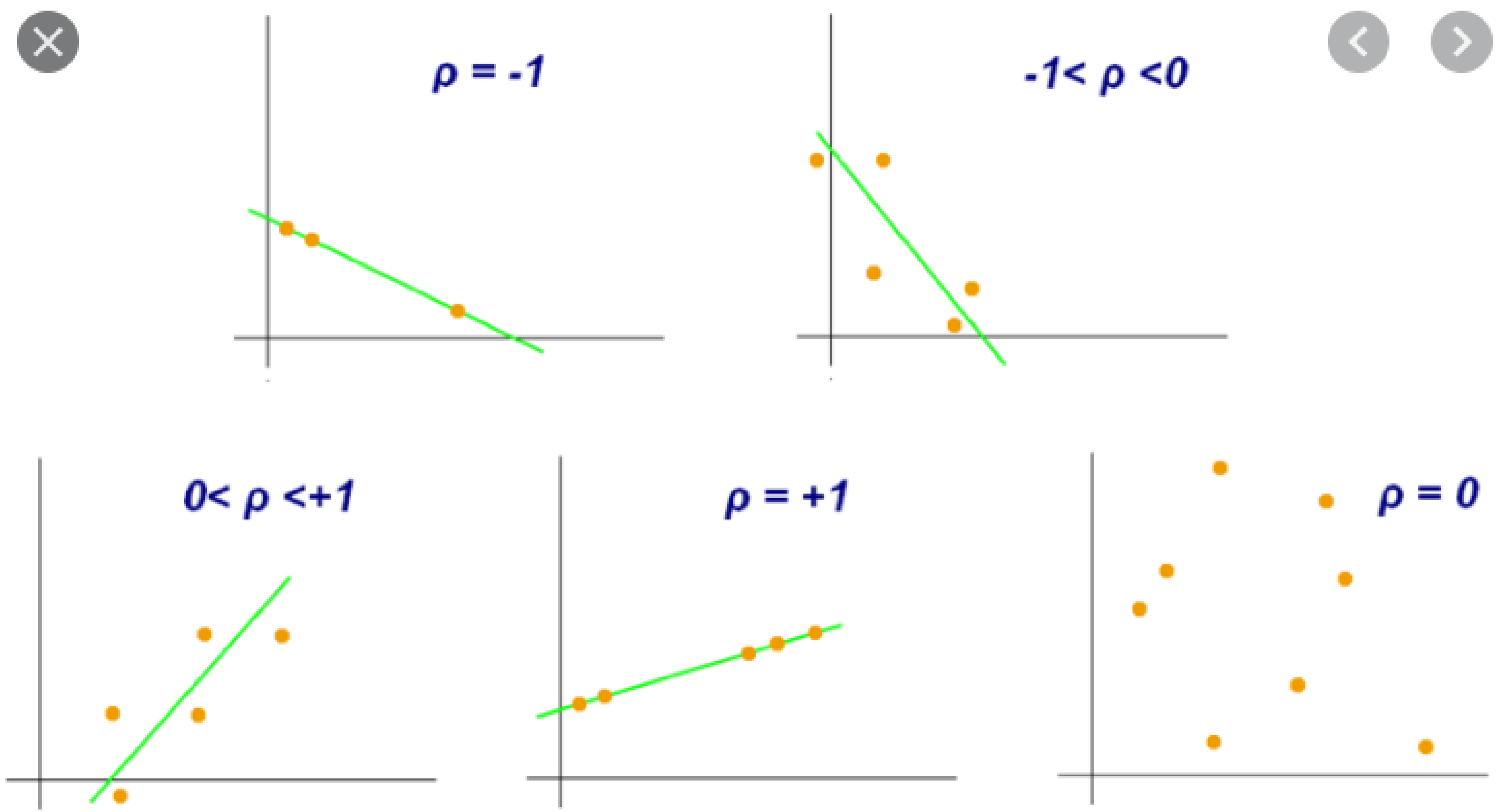
$$\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = 23750 - 175 \cdot 125 = 1875$$

Correlation

- The correlation coefficient of X and Y, denoted by $\rho_{X,Y}$ or simply ρ, is defined as

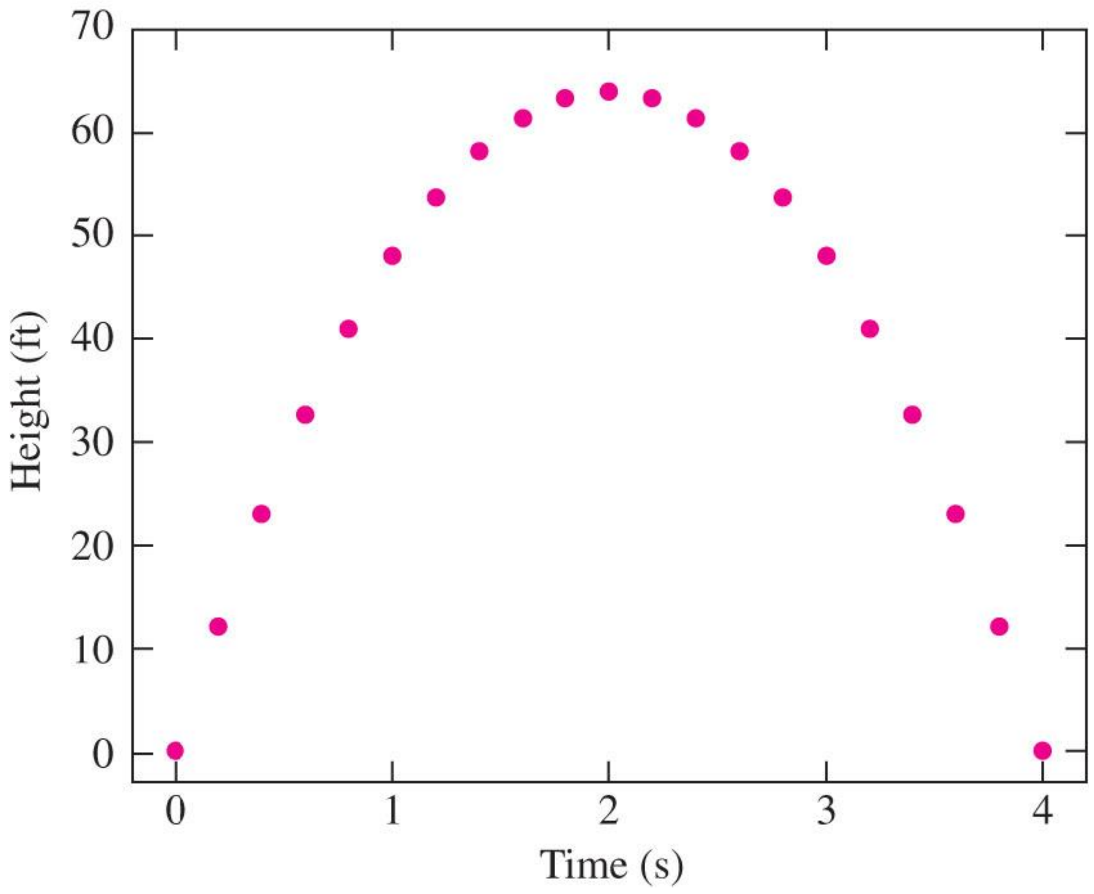
$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

- For any two r.v.'s X and Y, $-1 \le \rho_{X,Y} \le 1$
- IF X and Y are independent, then $\rho = 0$. However, $\rho = 0$ does not imply that X and Y are independent.
- $\rho = -1$ or $\rho = 1$ if and only if $Y = aX + b$ for some numbers a and b.


ρ = -1
-1< ρ <0
0< ρ <+1
ρ = +1
ρ = 0

Correlation coefficient : 0

An object is fired upward from the ground.
The relationship between the height of a free falling object and the time.



Example 5.17

		y		
		0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

$$p_X(x) = \begin{cases} 0.5 & if\ x = 100, 250 \\ 0 & otherwise \end{cases}$$

$$\mathrm{E}[X] = \sum_x x\, p_X(x) = 100 \cdot p_X(100) + 250 \cdot p_X(250)$$

$$= 100 \cdot 0.5 + 250 \cdot 0.5 = 175$$

$$\mathrm{V}[X] = \mathrm{E}[(x - \mathrm{E}[X])^2] = \sum_x (x - \mathrm{E}[X])^2 p_X(x) = (100 - 175)^2 \cdot 0.5 + (250 - 175)^2 \cdot 0.5 = 5625$$

$$p_Y(y) = \begin{cases} 0.25 & if\ y = 0, 100 \\ 0.50 & if\ y = 200 \\ 0 & otherwise \end{cases}$$

$$\mathrm{E}[Y] = \sum_y y\, p_Y(y) = 0 \cdot 0.25 + 100 \cdot 0.25 + 200 \cdot 0.5 = \mathit{125}$$

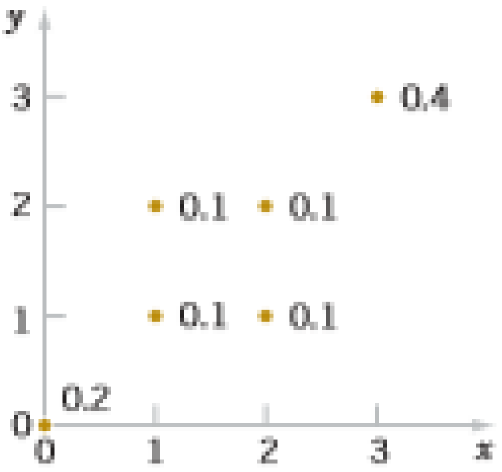
$$\mathrm{V}[Y] = \sum_y (y - \mathrm{E}[Y])^2 p_Y(y) = (0 - 125)^2 \cdot 0.25 + (100 - 125)^2 \cdot 0.25 + (200 - 125)^2 \cdot 0.5$$

$$= 6875$$

$$\mathrm{Cov(X, Y)} = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = 23750 - 175 \cdot 125 = 1875$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{1875}{\sqrt{5625}\sqrt{6875}} = 0.3015$$

Example



- $\mathrm{E}[XY] \;= \sum_x \sum_y xy \cdot p(x,y)$

$$= 0 \cdot 0 \cdot 0.2 + 1 \cdot 1 \cdot 0.1 + 1 \cdot 2 \cdot 0.1$$

$$+2 \cdot 1 \cdot 0.1 + 2 \cdot 2 \cdot 0.1 + 3 \cdot 3 \cdot 0.4 = 4.5$$

$$\mathrm{E}[X] \;= \sum_x x \;\; p_X(x) = 0 \cdot 0.2 + 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.4 = 1.8$$

$$\mathrm{V}[X] \;= \sum_x (x - \mathrm{E}[X])^2 p_X(x) = (0 - 1.8)^2 \cdot 0.2 + (1 - 1.8)^2 \cdot 0.2$$

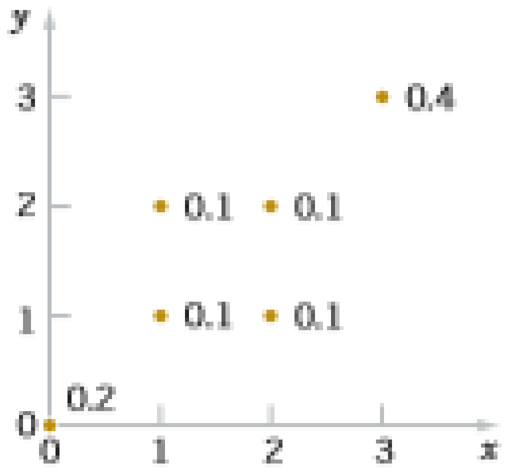
$$+(2 - 1.8)^2 \cdot 0.2 + (3 - 1.8)^2 \cdot 0.4 = 1.36$$

X and Y have the same marginal pdf. $\mathrm{E}[Y] = 1.8, \mathrm{V}[Y] \;= 1.36$

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y}) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = 4.5 - 1.8 \cdot 1.8 = 1.26$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = = \frac{1.26}{\sqrt{1.36}\sqrt{1.36}} = 0.926$$

Example



- $\mathrm{E}[XY] = \sum_x \sum_y xy \cdot p(x,y)$

$$= 0 \cdot 0 \cdot 0.2 + 1 \cdot 1 \cdot 0.1 + 1 \cdot 2 \cdot 0.1$$

$$+ 2 \cdot 1 \cdot 0.1 + 2 \cdot 2 \cdot 0.1 + 3 \cdot 3 \cdot 0.4 = 4.5$$

$$\mathrm{E}[X] = \sum_x p_X(x) = 0 \cdot 0.2 + 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.4 = 1.8$$

$$\mathrm{V}[X] = \sum_x (x - \mathrm{E}[X])^2 p_X(x) = (0-1.8)^2 \cdot 0.2 + (1-1.8)^2 \cdot 0.2$$

$$+ (2-1.8)^2 \cdot 0.2 + (3-1.8)^2 \cdot 0.4 = 1.36$$

X and Y have the same marginal pdf. $\mathrm{E}[Y] = 1.8, \mathrm{V}[Y] = 1.36$

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y}) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[X] = 4.5 - 1.8 \cdot 1.8 = 1.26$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = = \frac{1.26}{\sqrt{1.36}\sqrt{1.36}} = 0.926$$

- $f(x,y) = \begin{cases} 24xy & 0 \le x \le 1,\ 0 \le y \le 1,\ \ x+y \le 1 \\ 0 & otherwise \end{cases}$ $\quad f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$

$$\mathrm{E}(X) = \int_0^1 x \cdot 12x(1-x)^2 dx = 0.4$$

$$\mathrm{E}(\,X^2) = \int_0^1 \ x^2 \cdot 12x(1-x)^2 dx = \int_0^1 \ x^2 \cdot 12x(1-2x+x^2)dx = 12\int_0^1 (x^3 - 2x^4 + x^5)dx$$

$$= 12\left(\frac{x^4}{4} - \frac{2}{5}x^5 + \frac{1}{6}x^6\right)\Big|_{x=0}^{x=1} = 12\left(\frac{1}{4} - \frac{2}{5} + \frac{1}{6}\right) = 0.2$$

$$\mathrm{V}(X) = \mathrm{E}(\,X^2) - \mathrm{E}(X)^2 = 0.2 - 0.16 = 0.04$$

By symmetry, $\mathrm{V}(Y) = 0.04$.

$$\mathrm{Cov(X,Y)} = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = -\frac{2}{75}$$

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{-2/75}{\sqrt{0.04}\sqrt{0.04}} = -0.667$$

Definition of a statistic

- A statistic is any value that can be calculated from sample data. (ex, $\bar{X}, S^2, \ldots$)
- Because a statistic is calculated from the sample data, the statistic itself is a random variable.
- Once the data are obtained, we can evaluate the observed value of the statistic.
- To evaluate the distribution of a statistic, we must consider not only the sample we did observe but also the possibility of other samples that we could have observed.

Random samples

- Evaluating the distribution of a statistic calculated from a sample can be very difficult.
- Frequently we make the simplifying assumption that our data constitute a random sample $X_1, X_2, \cdots, X_n$ from a distribution.
- This means that

① The X_is are independent

② All the X_is have the same probability distribution

Simulation experiments

- R provides functions for generating random samples from many different families of distributions.
- Given a sample we can evaluate a statistic of interest, e.g. $\bar{X}$, the sample mean, then repeat this process for a large number of samples.
- The values of the statistic calculated from these samples allow us to examine the distribution of the statistic. We can examine density plots or histograms, normal probability plots, etc. to see the form of the distribution.
- By changing settings like the sample size we can examine how the distribution of the statistic changes.

Steps in a simulation experiment

① Choose the distribution family and the values of parameters.

② Choose the statistic of interest, e.g., the sample mean $\bar{X}$ (function mean) or the sample standard deviation (function sd)

③ Choose the sample size n (usually a small number like $n = 10$ or $n = 50$).

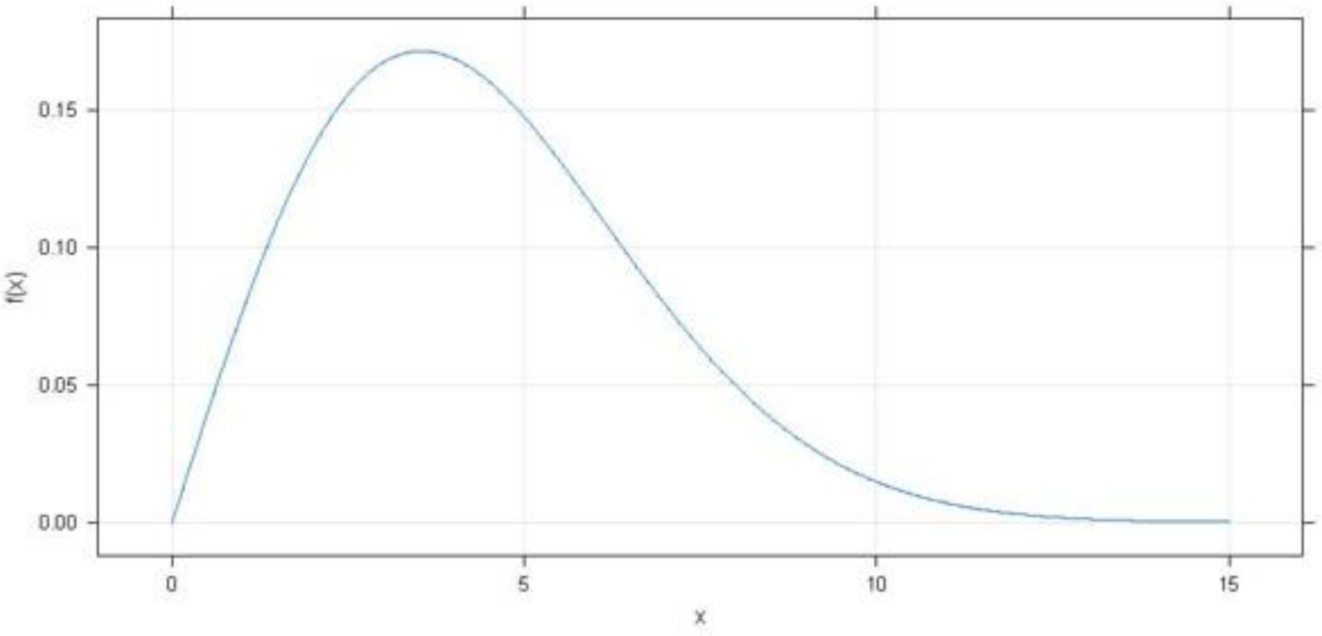
④ Choose the number of replications k (usually a very large number like $k = 10000$ or $k = 50000$).

- The larger the value of k, the better the simulated distribution will approximate the actual distribution of the statistic.
- Large values of k also mean that the simulation takes longer to run.

Simulating a sample mean from a Weibull Distribution

Consider the Weibull distribution with parameters $\alpha = 2$ (the shape parameter) and $\beta = 5$ (the scale parameter) shown below

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} & x \geq 0 \\ 0 & otherwise \end{cases}$$



Simulating a sample mean from a Weibull Distribution (cont'd)

To evaluate $k = 50000$ replications of the sample mean of samples of size $n = 10$ from this distribution we use the code

```
>k=50000

>n=10

>mns = numeric(k)

>for(i in 1:k) mns[i]=mean(rweibull(n, scale=2, shape=5))
```

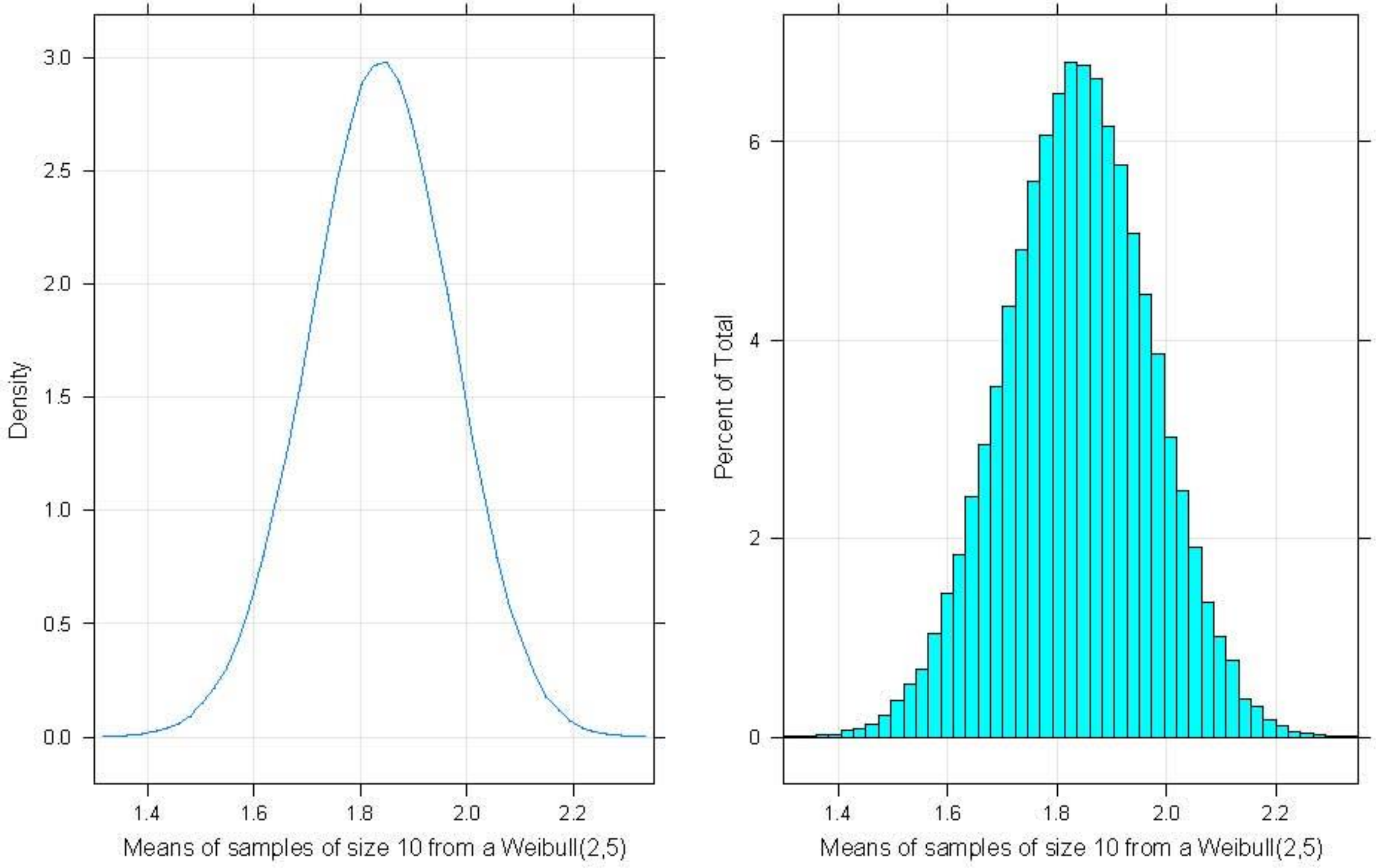
Things to note about the code

- We assign values of $k = 50000$ and $n = 10$ to those names so we can easily rerun the simulation with different settings.
- We assign a numeric vector of size k to the name mns to hold the results.
- Functions that generate a random sample from a particular distribution have names that start with "r"

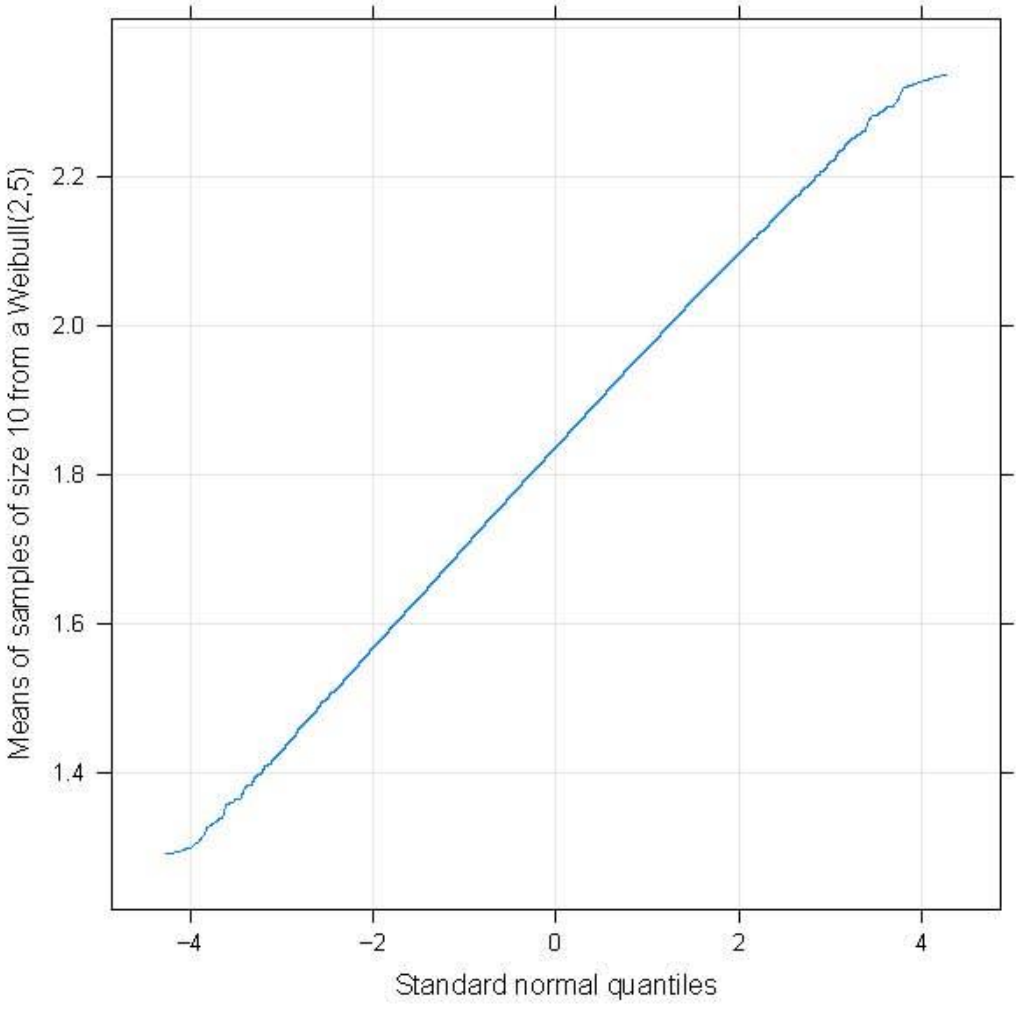
Characteristics of the simulated values

```
> summary(mns)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.290   1.747   1.838   1.836   1.927   2.338 

> sd(mns)

[1] 0.133065
> densityplot(~mns, plot.points = FALSE)
> histogram(~mns, nint = 50)
> qqmath(~mns, aspect = 1, type = c("g", "l"))
```

Density plot and histogram



Normal probability plot of sample means

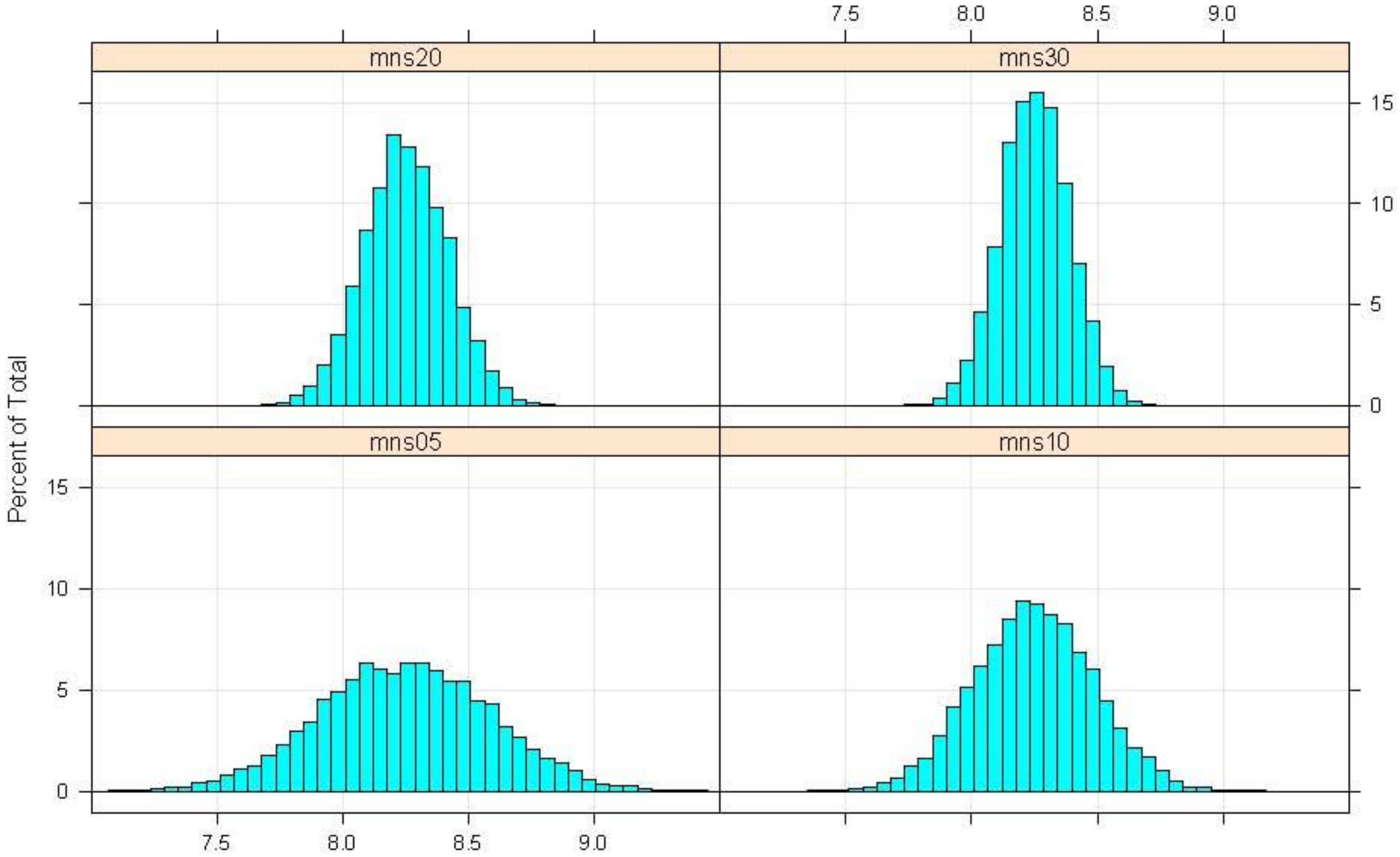


Multiple sample sizes : example 5.22

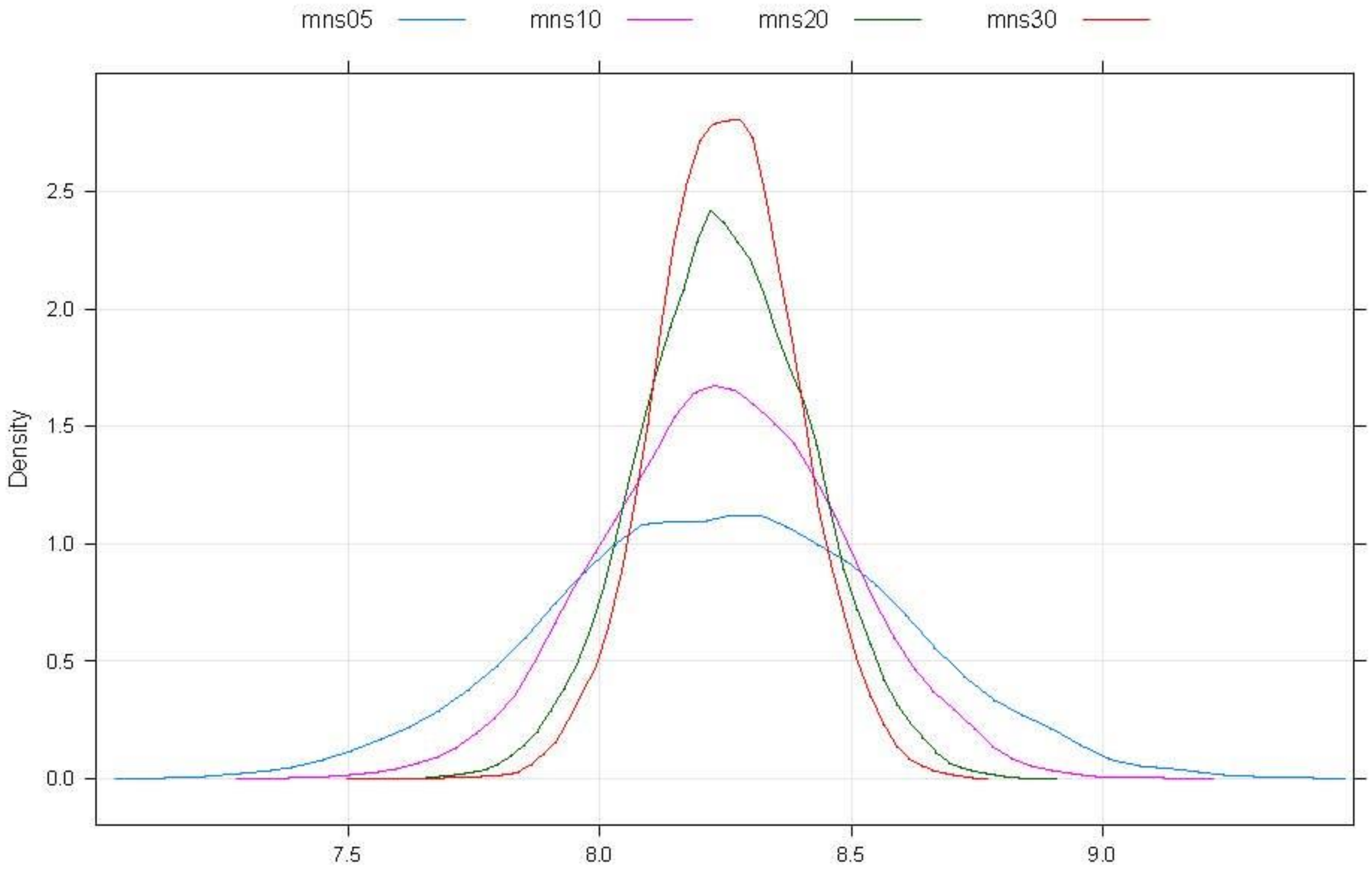
The distributions of the means of samples of size 5, 10, 20 and 30 from a normal distribution, $N(8.25, (0.75)^2)$ are simulated.

```
>k=10000
>mns05 = mns10 = mns20 = mns30 = numeric(k)
>for(i in 1:k) {
+   mns05[i]=mean(rnorm(5, mean=8.25, sd=0.75))
+   mns10[i]=mean(rnorm(10, mean=8.25, sd=0.75))
+   mns20[i]=mean(rnorm(20, mean=8.25, sd=0.75))
+   mns30[i]=mean(rnorm(30, mean=8.25, sd=0.75))
+}
>library(Devore7)
>histogram(~mns05 + mns10 + mns20 + mns30, nint = 50)
```

Histograms of means of different sizes of samples

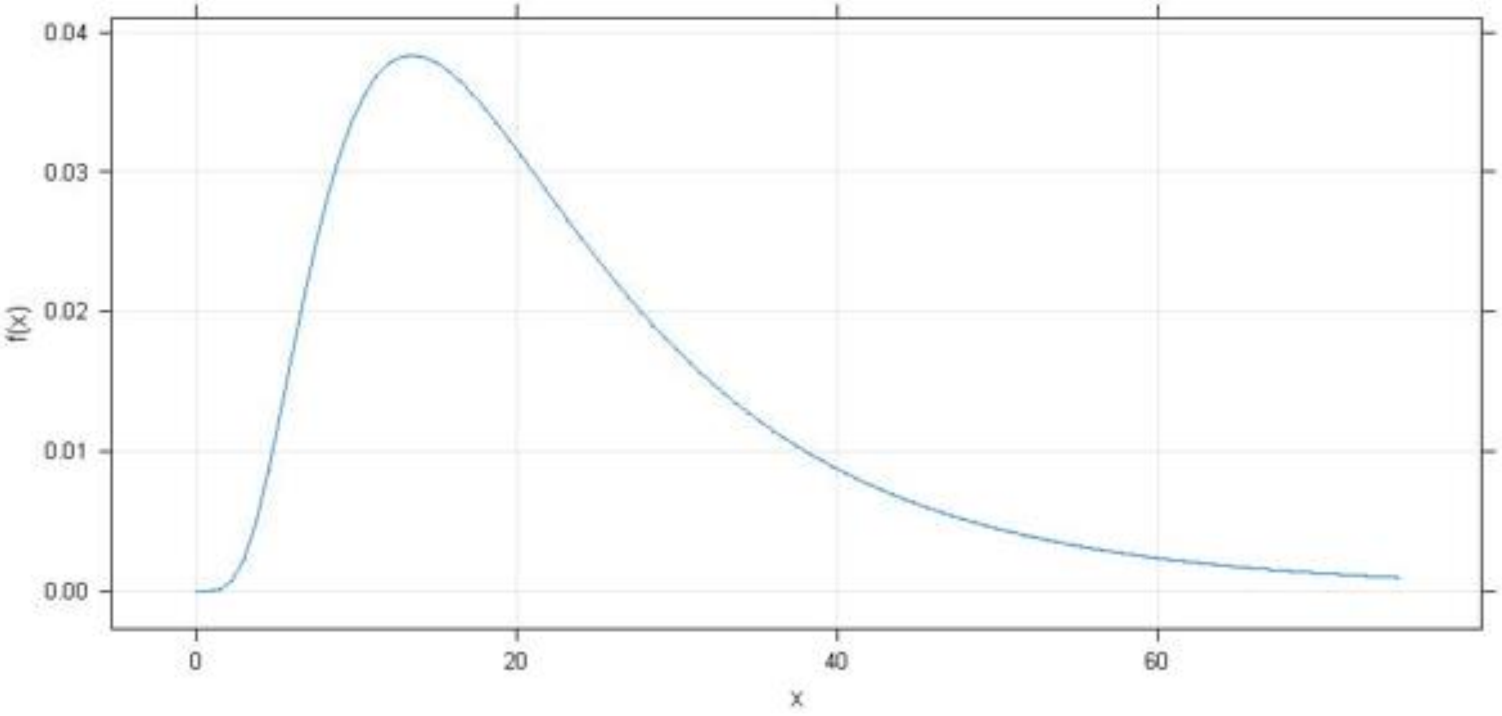


Densities of means of different sizes of samples



Example 5.23 – Simulating from a skewed distribution

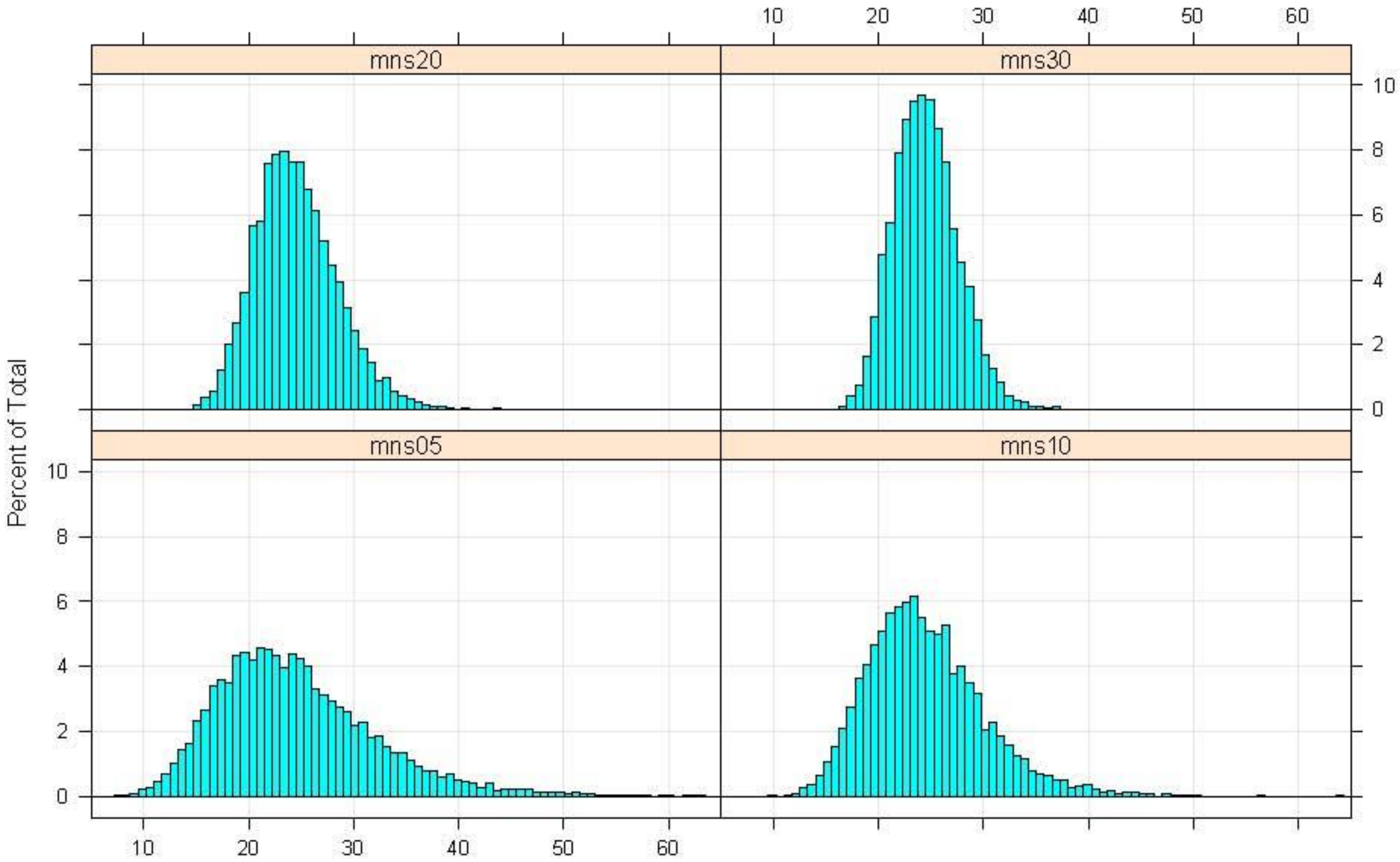
The means of samples of different sizes from a log-normal distribution with $E[\ln X] = 3$ and $Var[\ln X] = 0.4$ are simulated.



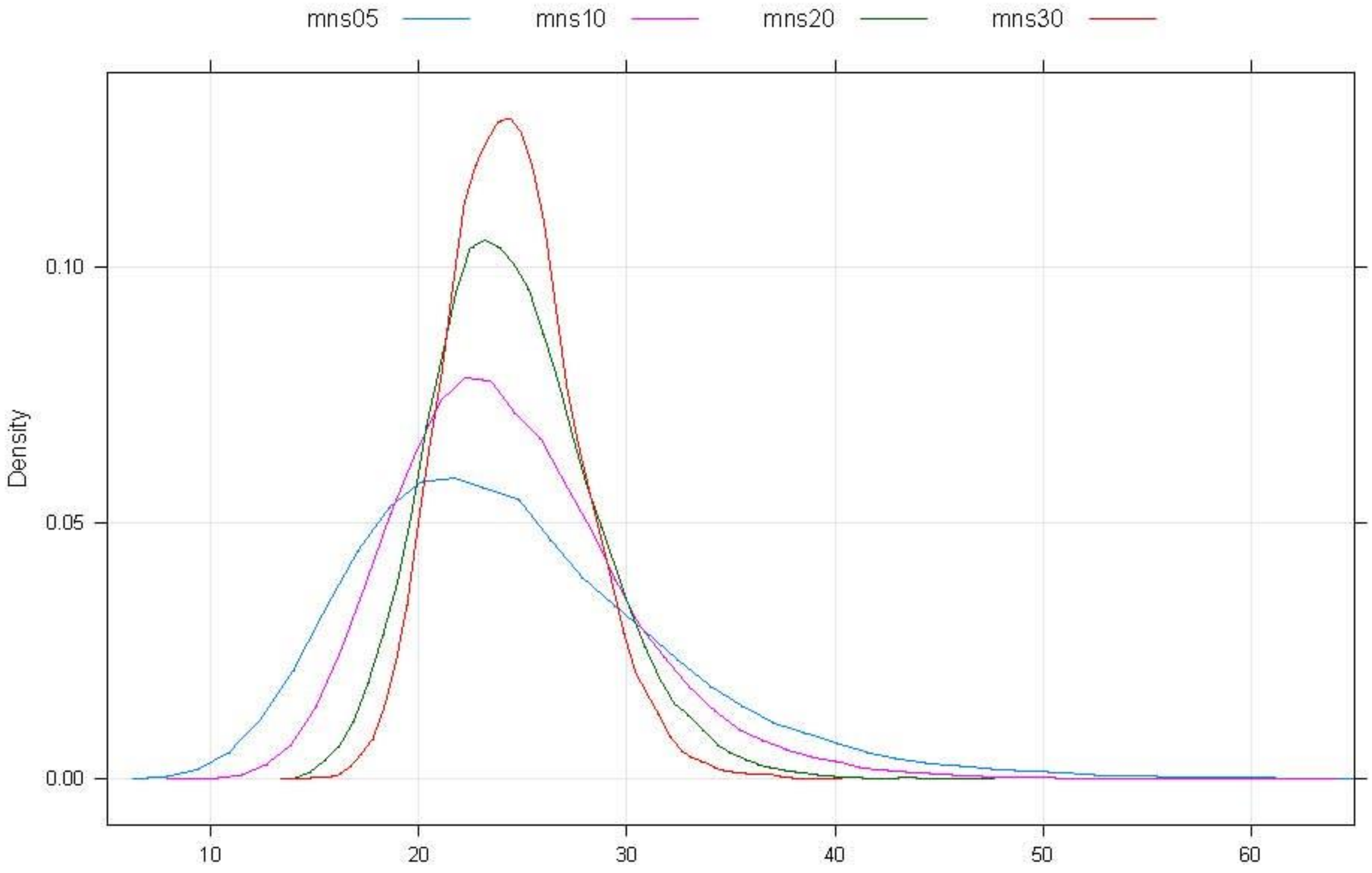
Simulation

```
> for (i in 1:k) {
+     mns05[i] = mean(rlnorm(5, meanlog = 3, sdlog = sqrt(0.4)))
+     mns10[i] = mean(rlnorm(10, meanlog = 3, sdlog = sqrt(0.4)))
+     mns20[i] = mean(rlnorm(20, meanlog = 3, sdlog = sqrt(0.4)))
+     mns30[i] = mean(rlnorm(30, meanlog = 3, sdlog = sqrt(0.4)))
+ }
> histogram(~mns05 + mns10 + mns20 + mns30, nint = 50)
> densityplot(~mns05 + mns10 + mns20 + mns30, plot.points = F)
```

Histograms of means from a log-normal distribution



Densities of means from a log-normal distribution



Properties of sample mean and sample sum

- Let X_1, $X_2, \cdots, X_n$ have mean values $\mu_1, \mu_2, \cdots, \mu_n$, respectively, and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively.

1. Whether or not the X_i's are independent,

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$
$$= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

2. IF X_1, $X_2, \cdots, X_n$ are independent,

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n)$$
$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

3. For any X_1, $X_2, \cdots, X_n$

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)$$

Variance of Linear Combinations
- https://online.stat.psu.edu/stat414/lesson/24/24.3

Now for the proof for the variance. Starting with the definition of the variance of Y, we have:

$$\sigma_Y^2 = Var(Y) = E[(Y - \mu_Y)^2]$$

Now, substituting what we know about Y and the mean of Y Y, we have:

$$\sigma_Y^2 = E\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)^2\right]$$

Because the summation signs have the same index ($i = 1$ to n), we can replace the two summation signs with one summation sign:

$$\sigma_Y^2 = E\left[\left(\sum_{i=1}^{n} (a_i X_i - a_i \mu_i)\right)^2\right]$$

And, we can factor out the constants a_i:

$$\sigma_Y^2 = E\left[\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right]$$

Variance of Linear Combinations
- https://online.stat.psu.edu/stat414/lesson/24/24.3

Now, let's rewrite the squared term as the product of two terms. In doing so, use an index of i on the first summation sign, and an index of j on the second summation sign:

$$\sigma_Y^2 = E\left[\left(\sum_{i=1}^{n} a_i(X_i - \mu_i)\right)\left(\sum_{j=1}^{n} a_j(X_j - \mu_j)\right)\right]$$

Now, let's pull the summation signs together:

$$\sigma_Y^2 = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right]$$

Then, by the linear operator property of expectation, we can distribute the expectation:

$$\sigma_Y^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E\left[(X_i - \mu_i)(X_j - \mu_j)\right]$$

Variance of Linear Combinations

- https://online.stat.psu.edu/stat414/lesson/24/24.3

Now, let's rewrite the variance of Y by evaluating each of the terms from $i = 1$ to n and $j = 1$ to n. In doing so, recognize that when $i = j$, the expectation term is the variance of X_i, and when $i \neq j$, the expectation term is the covariance between X_i and X_j, which by the assumed independence, is 0:

$$
\begin{aligned}
\sigma_Y^2 = {} & a_1 a_1 \underbrace{E\left[(X_1-\mu_1)(X_1-\mu_1)\right]}_{Var(X_1)} + a_1 a_2 \underbrace{E\left[(X_1-\mu_1)(X_2-\mu_2)\right]}_{0} + \cdots \\
& + a_1 a_n \underbrace{E\left[(X_1-\mu_1)(X_n-\mu_n)\right]}_{0} + a_2 a_1 \underbrace{E\left[(X_2-\mu_2)(X_1-\mu_1)\right]}_{0} + \cdots \\
& + a_2 a_2 \underbrace{E\left[(X_2-\mu_2)(X_2-\mu_2)\right]}_{Var(X_2)} + \cdots + a_n a_n \underbrace{E\left[(X_n-\mu_n)(X_n-\mu_n)\right]}_{Var(X_n)}
\end{aligned}
$$

Properties of sample mean and sample sum

- Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with mean μ and standard deviation σ. Then

① $E[\bar{X}] = \mu_{\bar{X}} = \mu$

$$E[\bar{X}] = E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = E\left[\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n\right]$$

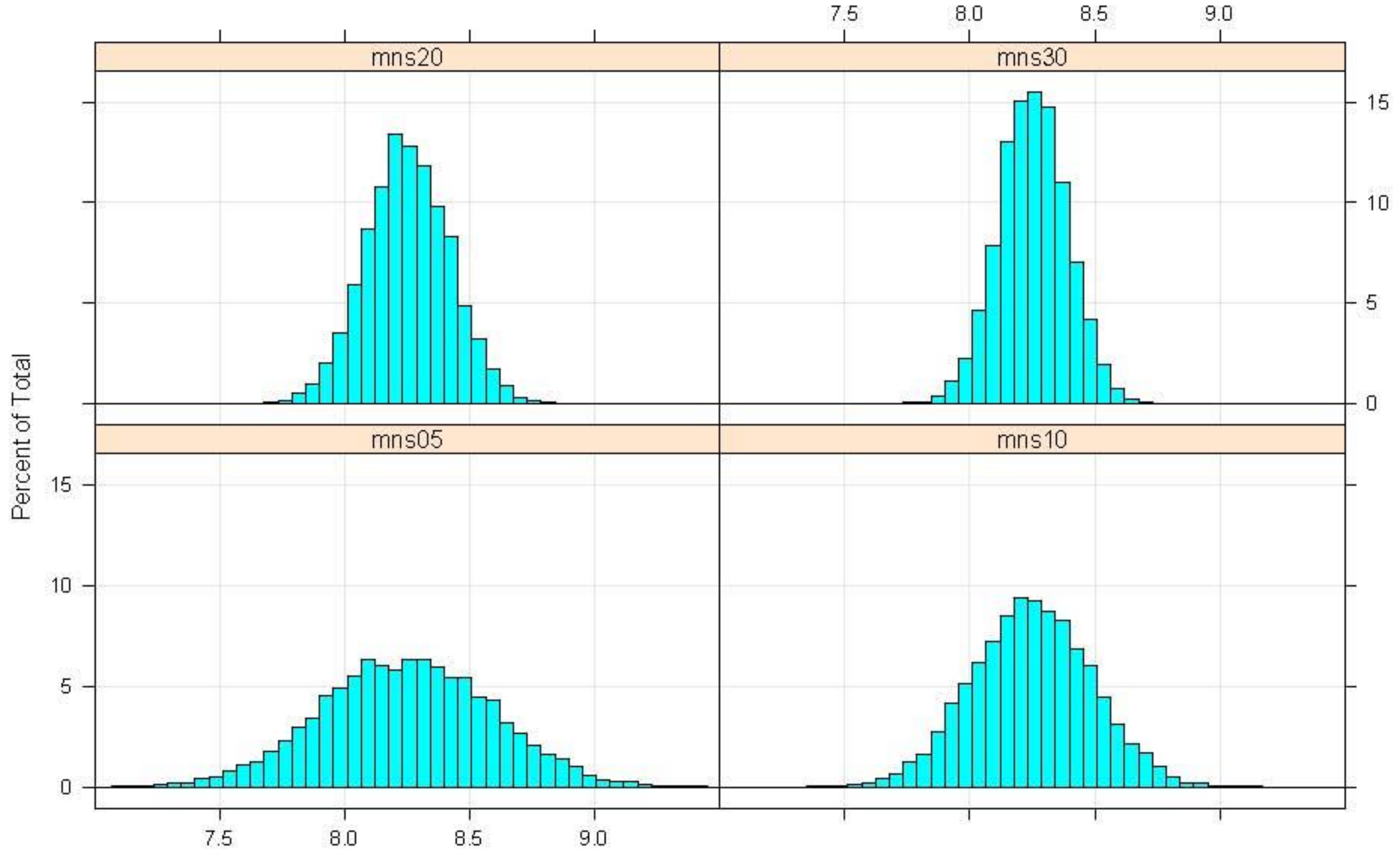
$$= [\frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \cdots + \frac{1}{n}E(X_n)] = \frac{1}{n}n\mu = \mu$$

② $V[\bar{X}] = \sigma_{\bar{X}}^2 = \sigma^2/n$ and $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

$$V[\bar{X}] = V\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = V\left[\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n\right]$$

$$= \frac{1}{n^2}[V(X_1) + V(X_2) + \cdots + V(X_n)] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Histograms of means of different sizes of samples



Properties of sample mean and sample sum

- Let $T_n = X_1 + X_2 + \cdots + X_n$ be the sample total. Then

① $E[T_n] = n\mu$

$$E[T_n] = E[X_1 + X_2 + \cdots + X_n] = E(X_1) + E(X_2) + \cdots + E(X_n) = n\mu$$

② $V[T_n] = n\sigma^2$ and $\sigma_{T_n} = \sqrt{n}\sigma$

$$V[T_n] = V[X_1 + X_2 + \cdots + X_n] = [V(X_1) + V(X_2) + \cdots + V(X_n)] = n\sigma^2$$

- If the original distribution of the X_is is normal, then the distribution of $\bar{X}$ and T_n is also normal.

Sum of the Normal Distributions

- Let $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N\left(\mu_y, \sigma_y^2\right)$ be two independently distributed normal variables. Then their sum is also a normally distributed random variable:

$$X + Y \sim N\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$$

- To prove this we need only to invoke the result that, in the case of independence, the moment generating function of the sum is the product of the moment generating functions of its elements.

$$M_{X+Y}(t) = M_X(t) M_Y(t) = e^{\mu_x t + \frac{1}{2}\sigma_x^2 t^2} e^{\mu_y t + \frac{1}{2}\sigma_y^2 t^2}$$

$$= e^{(\mu_x + \mu_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2}$$

which is recognized as the moment generating function of a normal distribution.

Moment Generating Function of the Normal Distribution

- https://www.le.ac.uk/users/dsgp1/COURSES/MATHSTAT/6normgf.pdf

$$M_X(t) = E(e^{Xt}) = \int e^{xt} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx = e^{\mu\, t+\frac{1}{2}\sigma^2 t^2}$$

Define

$$z = \frac{x-\mu}{\sigma}, \text{which implies } x = z\sigma + \mu$$

Using the change of variable technique, we get

$$M_x(t) = e^{\mu\, t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \left|\frac{dx}{dz}\right| dz$$

$$= e^{\mu\, t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \left|\frac{dx}{dz}\right| dz \quad (dx = \sigma dz, \frac{dx}{dz} = \sigma)$$

$$= e^{\mu\, t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Moment Generating Function of the Normal Distribution

$$M_x(t) = e^{\mu\, t} \int \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}z^2 + z\sigma t\right\}} dz$$

$$= e^{\mu\, t} \int \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}(z^2 - 2z\sigma t + \sigma^2 t^2 - \sigma^2 t^2)\right\}} dz$$

$$= e^{\mu\, t} \int \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}(z - \sigma t)^2 + \frac{1}{2}\sigma^2 t^2\right\}} dz$$

$$= e^{\mu\, t} e^{\frac{1}{2}\sigma^2 t^2} \int \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}(z - \sigma t)^2\right\}} dz$$

$$= e^{\mu\, t} e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu\, t + \frac{1}{2}\sigma^2 t^2}$$

The final equality follows from the fact that the expression under the integral is the $N(z; \mu = \sigma t, \sigma^2 = 1)$ which integrates to unity

Properties of sample mean and sample sum : Example 5.29

A gas station sells three grades of gasoline: regular, extra, and super.

Prices : \$3.00, \$3.20, and \$3.40 per gallon.

Let X_1, X_2, and X_3 denote the amounts of these grades sold on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, and $\sigma_3 = 50$.

The revenue from sales is $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$

$$E[Y] = 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620$$

$$V[Y] = (3.0)^2\sigma_1^2 + (3.2)^2\sigma_2^2 + (3.4)^2\sigma_3^2 = 184436$$

$$\sigma_Y = \sqrt{184436} = \$429.46$$

Central Limit Theorem

- Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with mean μ and variance σ^2.
- Then for n sufficiently large, $\bar{X}$ has approximately a normal distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma_{\bar{X}}^2 = \sigma^2/n$
- Another way of phrasing this is :

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1) \text{ as } n \to \infty$$

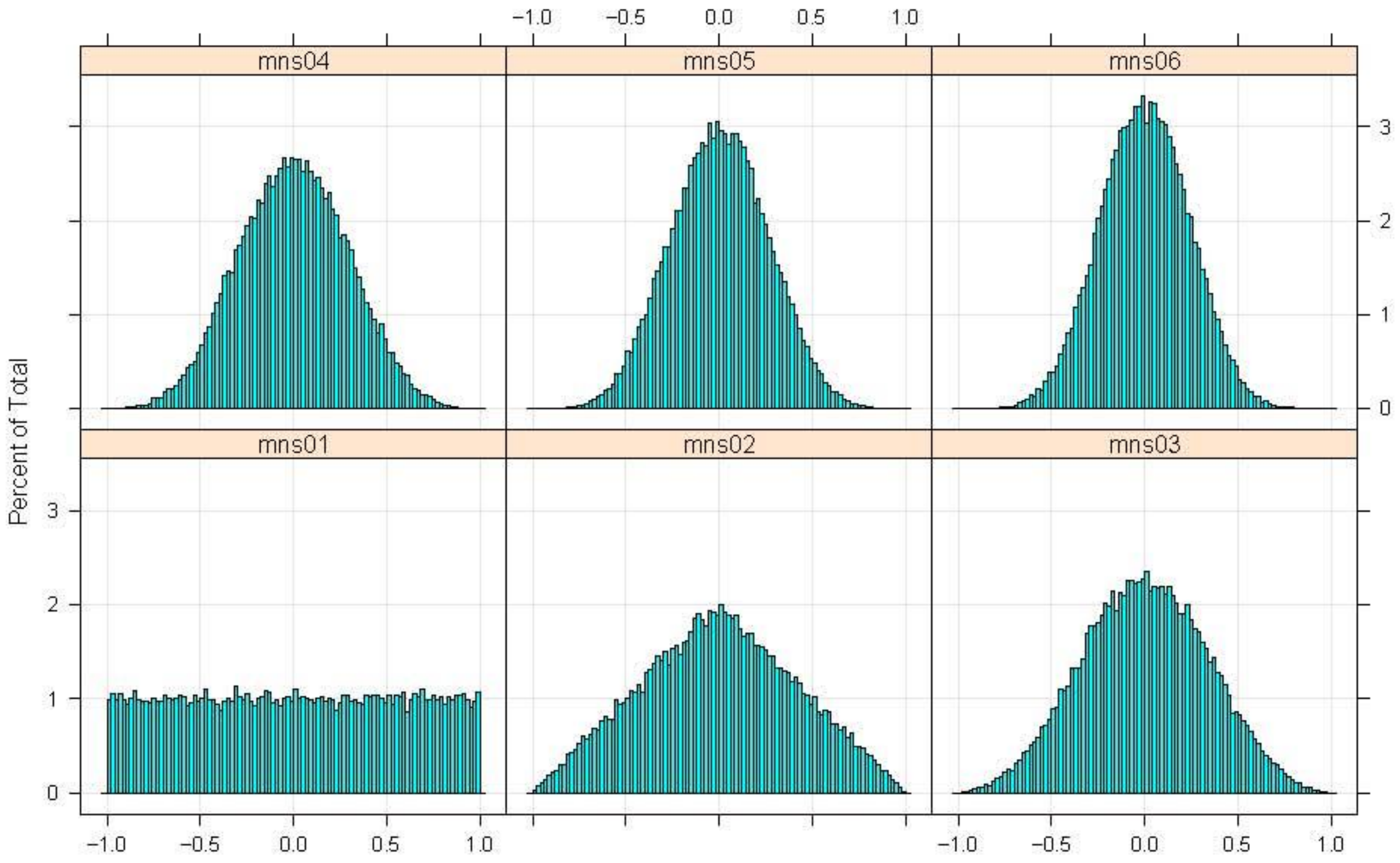
- The larger the value of n, the better the approximation.
- For distributions that are continuous and reasonably close to being symmetric, the convergence to the normal distribution is good even for small values of n.
- Proof of the central limit theorem :

Convergence of means from U[-1, 1] to a normal shape

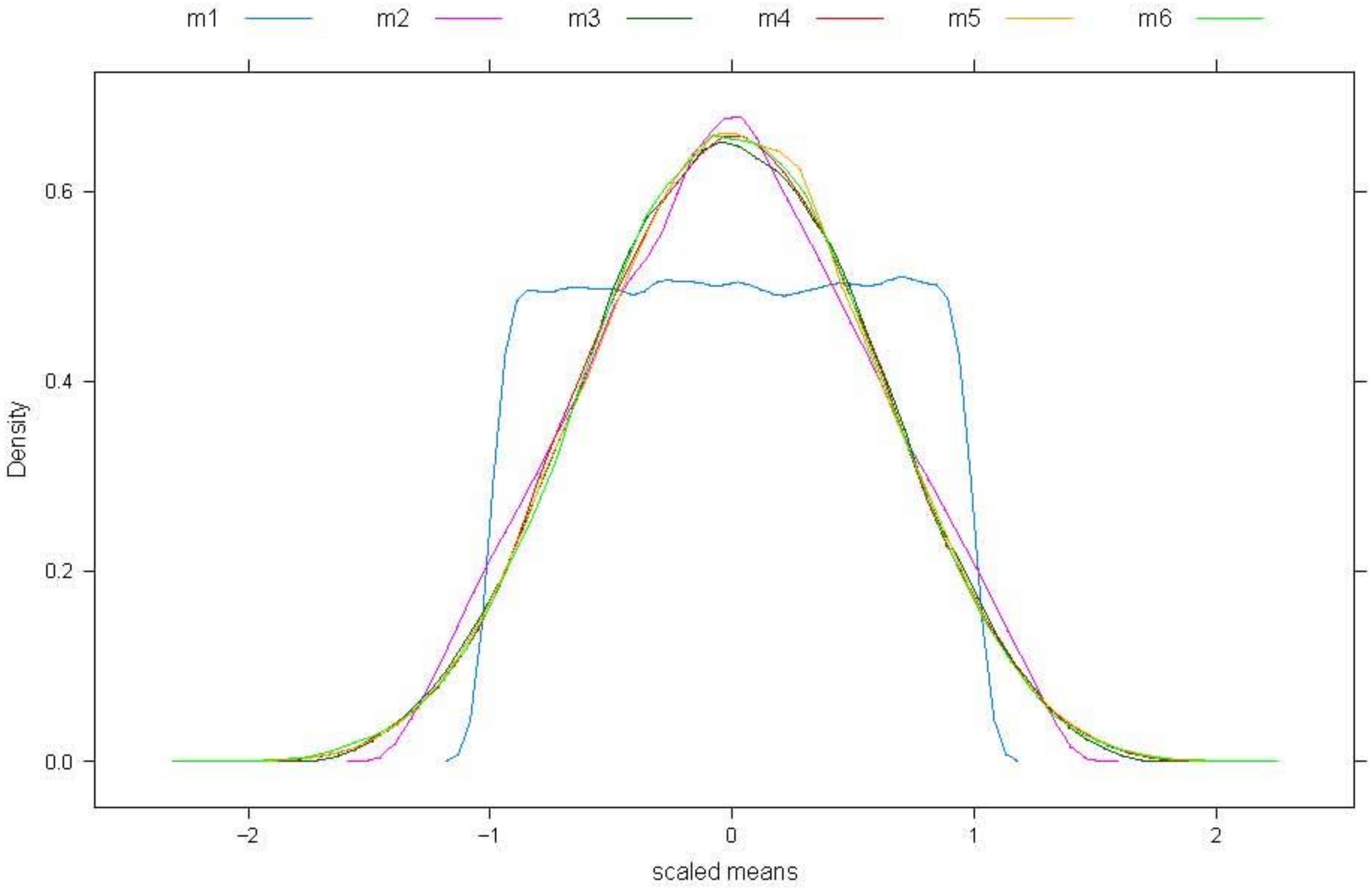
- The uniform distribution on the interval [-1, 1] has a mean of 0 and a variance of 1/3.
- We simulate 50000 replications from the original distribution, mns01, from the distribution of the means of samples of sizes 2, 3, 4, 5, and 6.
- Histograms of the means of the samples will show convergence to a normal shape and decreasing variance.
- If we multiply the means of samples of size n by $\sqrt{n}$ we can put them all on the same scale to see the convergence to a normal shape.

```
# Simulating a sample mean from a Uniform
> k=10000
> mns1 = mns2 = mns3 = mns4 = mns5 = mns6 = numeric(k)
> for(i in 1:k) {
+ mns1[i]=mean(runif(1, -1, 1))
+ mns2[i]=mean(runif(2, -1, 1))
+ mns3[i]=mean(runif(3, -1, 1))
+ mns4[i]=mean(runif(4, -1, 1))
+  mns5[i]=mean(runif(5, -1, 1))
+  mns6[i]=mean(runif(6, -1, 1))
+ }

> histogram(~mns1 + mns2 + mns3 + mns4 + mns5 + mns6, nint = 50)
> densityplot(~mns1 + mns2 + mns3 + mns4 + mns5 + mns6, plot.points = F)
```

Histograms of raw means of samples from U[-1,1].



Densities of scaled means of samples from U[-1,1].



Linear Combinations and their means

- Given a collection of n random variables $X_1, X_2, \cdots, X_n$ and n numerical constants $a_1, a_2, \cdots, a_n$, the random variable

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

  is called a linear combination of the X_is.

- Whether or not the X_is are independent,

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n]$$

- https://online.stat.psu.edu/stat414/lesson/24/24.3

Variances of Linear Combinations

- If X_1, $X_2, \cdots, X_n$ are independent with variances σ_1^2, σ_2^2, $\cdots$, σ_n^2 then

$$V[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1^2V[X_1] + a_2^2V[X_2] + \cdots + a_n^2V[X_n]$$
$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

- In general

$$V[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j \mathrm{Cov}(X_i,\ X_j)$$

- https://online.stat.psu.edu/stat414/lesson/24/24.3

$$E(Y) = E(a_1X_1 + \ a_2X_2 + \cdots + a_nX_n)$$
$$= a_1E(X_1) + a_2E(X_2) + \ \cdots + a_nE(X_n)$$

$$V(Y) = E[(Y - \mu_Y)^2]$$
$$= E[(\textstyle\sum_{i=1}^n a_iX_i - \sum_{i=1}^n a_i\mu_i)^2] \qquad (\mu_i = E(X_i))$$
$$= E[(\textstyle\sum_{i=1}^n (a_iX_i - a_i\mu_i))^2]$$
$$= E[(\textstyle\sum_{i=1}^n a_i(X_i - \mu_i)^2]$$
$$= E\left[(\textstyle\sum_{i=1}^n a_i(X_i - \mu_i))\left(\sum_{j=1}^n a_j(X_j - \mu_j)\right)\right]$$
$$= E\left[\textstyle\sum_{i=1}^n \sum_{j=1}^n a_ia_j\,(X_i - \mu_i)\left(X_j - \mu_j\right)\right]$$
$$= \textstyle\sum_{i=1}^n \sum_{j=1}^n a_ia_j\,E[(X_i - \mu_i)\left(X_j - \mu_j\right)]$$
$$= \textstyle\sum_{i=1}^n \sum_{j=1}^n a_i\,a_j\mathrm{Cov}(X_i,\ X_j)$$

$$
\begin{aligned}
V(Y) = {} & a_1 a_1 E[(X_1-\mu_1)(X_1-\mu_1)] + a_1 a_2 E[(X_1-\mu_1)(X_2-\mu_2)] + \cdots \\
& + a_1 a_n E[(X_1-\mu_1)(X_n-\mu_n)] \\
& + a_2 a_1 E[(X_2-\mu_2)(X_1-\mu_1)] + a_2 a_2 E[(X_2-\mu_2)(X_2-\mu_2)] + \cdots \\
& + a_2 a_n E[(X_2-\mu_2)(X_n-\mu_n)] + \cdots \\
& + a_n a_1 E[(X_n-\mu_n)(X_1-\mu_1)] + a_n a_2 E[(X_n-\mu_n)(X_2-\mu_2)] + \cdots \\
& + a_n a_n E[(X_n-\mu_n)(X_n-\mu_n)]
\end{aligned}
$$

If $X_1, X_2, \cdots, X_n$ are independent, covariance terms become 0 :

$$V(Y) = a_1^2 V[X_1] + a_2^2 V[X_2] + \cdots + a_n^2 V[X_n]$$

The difference between random variables

- A common special case of a linear combination is the difference of random variables $Y = X_1 - X_2$.

  That is, $n = 2, a_1 = 1, a_2 = -1$.

- $E[Y] = \mu_1 - \mu_2$.
- If X_1 and X_2 are independent then the variance of the difference is

$$V[Y] = a_1^2 V[X_1] + a_2^2 V[X_2] = (1)^2 V[X_1] + (-1)^2 V[X_2]$$
$$= V[X_1] + V[X_2]$$

  That is, the variance of the difference is the sum of the variances.

- Remember that "Variances add." in the sense that even when you take the difference of independent random variables, their variances add.
- But, $\sigma_Y = \sqrt{\sigma_1^2 + \sigma_2^2} \neq \sigma_1 + \sigma_2$

The Case of Normal Random Variables : Example 5.31

When the X_is are independent and normally distributed, any linear combination will also be normally distributed.

The total avenue from the sale of the three grades of gasoline on a particular day was

$Y = 3.0X_1 + 3.2X_2 + 3.4X_3$ and we calculated $\mu_Y = 5620$ and $\sigma_Y = 429.46$.

If X_is are normally distributed, the probability that revenue exceeds 4500 is

$$P(Y > 4500) = P\left(Z > \frac{4500-5620}{429.46}\right) = P(Z > -2.61) = 1 - \Phi(-2.61) = 0.9955$$

Properties of sample mean and sample sum : Example 5.29

A gas station sells three grades of gasoline: regular, extra, and super.

Prices : \$3.00, \$3.20, and \$3.40 per gallon.

Let X_1, X_2, and X_3 denote the amounts of these grades sold on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, and $\sigma_3 = 50$.

The revenue from sales is $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$

$$E[Y] = 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620$$

$$V[Y] = (3.0)^2\sigma_1^2 + (3.2)^2\sigma_2^2 + (3.4)^2\sigma_3^2 = 184436$$

$$\sigma_Y = \sqrt{184436} = \$429.46$$