Lecture 1

Introduction to Probability Distributions

Selected Topics

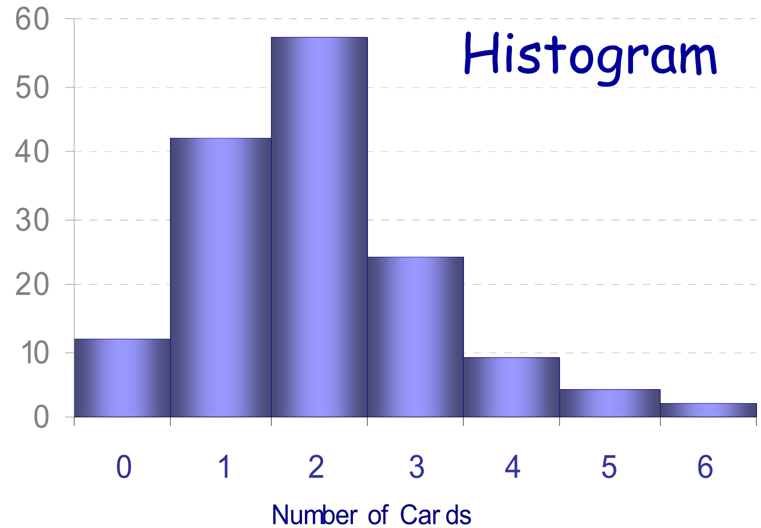
1. Random Variables and their Prob. Distributions
2. Expected Value ("Central Tendency")
3. Variance ("Spread")
4. The Normal Distribution
5. Normal Approximation to Binomial Distribution
6. Continuous Distributions

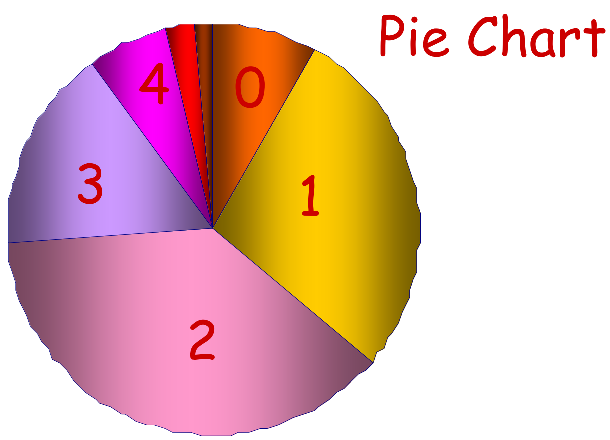
A glance into 'Representation of data'

Example: In a survey the college students were asked how many credit cards they own. The results are reported in the table:

# Cards	# Students
0	12
1	42
2	57
3	24
4	9
5	4
6	2

Total: 150





Random Variable

We can think of a random variable as a rule that assigns **a numeric value** and **a probability** to an outcome of a chance experiment[1].

- **Finite discrete** – assumes only finitely many values.

  Example: Rolling a die.

- **Infinite discrete** – assumes infinitely many values that may be arranged in a sequence.

  Example: Counting die rolls until the outcome is 6.

- **Continuous** – assumes values that make up an interval of real numbers. The probability is assigned to intervals, not individual numbers.

  Examples: Time between arrivals of two customers.

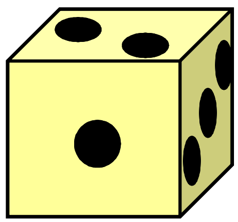
  Tomorrow's temperature at noon.

[1] Not a mathematical definition.

Prob. Distribution of a Random Variable

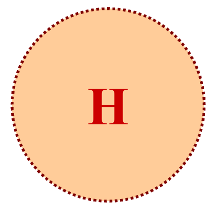
Examples: Probability distributions of:

(a) a die roll,



x	$P(D=x)$
1	$^1/_6$
2	$^1/_6$
3	$^1/_6$
4	$^1/_6$
5	$^1/_6$
6	$^1/_6$

(b) a coin toss.



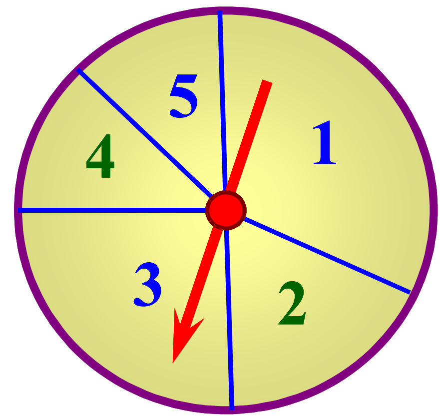
Let C denote a random variable.

C must have numerical values, so we agree on:

Tail = 0, Head = 1

x	$P(C=x)$
0	$^1/_2$
1	$^1/_2$

(c) a hand spin.



x	$P(H=x)$
1	$^1/_3$
2	$^1/_6$
3	$^1/_4$
4	$^1/_8$
5	$^1/_8$

Prob. Distribution of a Random Variable

OPIM 5603 Fall 2019

Slide 1.6

Example: Random variable X assumes (only) the values

$$-8, -3, -1, 0, 1, 4, 6$$

(hence a finite discrete random variable).

Its probability distribution is given by:

x	-8	-3	-1	0	1	4	6
$P(X=x)$	0.13	0.15	0.17	0.20	0.15	0.11	0.09

Find

(a) $P(X \le 0) = P(\{-8, -3, -1, 0\}) = 0.13 + 0.15 + 0.17 + 0.2 = 0.65$

(b) $P(-3 \le X \le 1) = P(\{-3, -1, 0, 1\}) = 0.67$

Credit Cards example revisited

Students were asked how many credit cards they own. X is the random variable representing the number of cards and the results are below.

x	#Students	$P(X=x)$
0	12	0.08
1	42	0.28
2	57	0.38
3	24	0.16
4	9	0.06
5	4	0.02666
6	2	0.01333

Total: 150

$\frac{12}{150}$

Probability Distribution

Histogram revisited

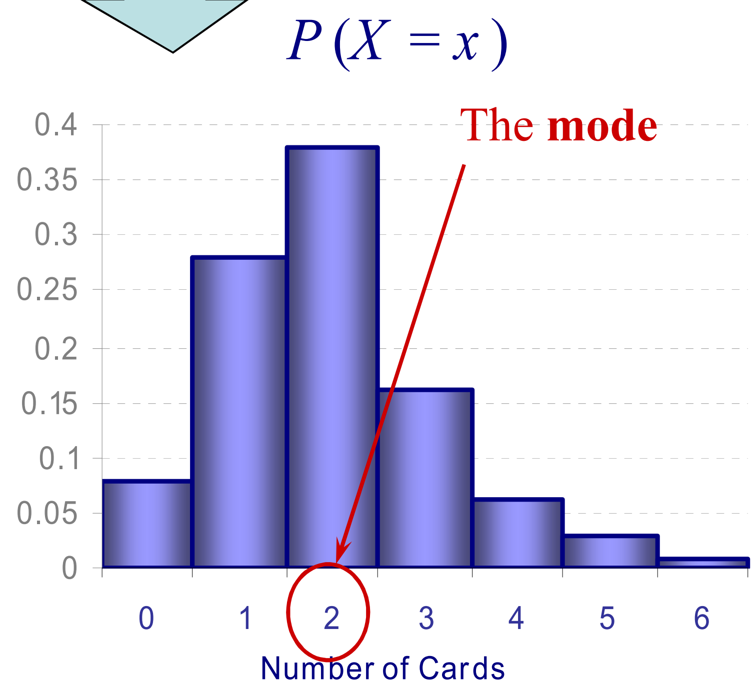
OPIM 5603 Fall 2019

Slide 1.8

A way to represent a probability distribution of a random variable graphically.

Credit card results:

x	$P(X=x)$
0	0.08
1	0.28
2	0.38
3	0.16
4	0.06
5	0.02666
6	0.01333



Mean, Median, Mode

The average (mean) of the n numbers $x_1, x_2, \ldots, x_n$ is defined as

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

The median is the middle value in a set of data that is arranged in increasing or decreasing order. For an even number of data points the median is the average of the middle two.

The mode is the most frequent number in a set of data.

Example

The quiz scores for a particular student are given below:

22, 25, 20, 18, 12, 20, 24, 20, 20, 25, 24, 25, 18

Find the average, median and mode.

Average: $\dfrac{\text{sum of entries}}{\text{number of data points}} = \dfrac{273}{13} = 21$

Median: Sort the numbers:

Middle number = 20

$12, 18, 18, \underbrace{20, 20, 20, 20}, 22, 24, 24, 25, 25, 25$

Mode (most frequent): 20 (occurs 4 times)

Expected Value of a Discrete Random Variable

Let X be a random variable that assumes the values $x_1, x_2, \ldots, x_n$ with associated probabilities $p_1, p_2, \ldots, p_n$, respectively.

Then the expected value (mean) of X, denoted by $E(X)$, is

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$

Example: Let D be the random variable recording the outcome of the single roll of a fair die. Find the expected value of D.

Solution: The probability distribution is

x	$P(D = x)$
1	$^1/_6$
2	$^1/_6$
3	$^1/_6$
4	$^1/_6$
5	$^1/_6$
6	$^1/_6$

Mean: $E(D) = x_1 p_1 + x_2 p_2 + \ldots + x_6 p_6$

$$= 1 \cdot {}^1/_6 + 2 \cdot {}^1/_6 + 3 \cdot {}^1/_6 +$$

$$+ 4 \cdot {}^1/_6 + 5 \cdot {}^1/_6 + 6 \cdot {}^1/_6$$

$$= {}^{21}/_6 = 3.5$$

Example

The quiz scores for a particular student are given below:

22, 25, 20, 18, 12, 20, 24, 20, 20, 25, 24, 25, 18

Find the expected value of the random variable S that measures this student quiz performance.

Solution: The frequency table and prob. distribution of S are given by

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

$$= 12 \cdot {}^1/_{13} + 18 \cdot {}^2/_{13} + 20 \cdot {}^4/_{13} +$$

$$+ 22 \cdot {}^1/_{13} + 24 \cdot {}^2/_{13} + 25 \cdot {}^3/_{13}$$

$$= \frac{12 + 36 + 80 + 22 + 48 + 75}{13}$$

$$= \frac{273}{13} = 21$$

x	# quizzes	$P(S=x)$
12	1	${}^1/_{13}$
18	2	${}^2/_{13}$
20	4	${}^4/_{13}$
22	1	${}^1/_{13}$
24	2	${}^2/_{13}$
25	3	${}^3/_{13}$

Recall: The average of the scores is also 21 (slide 1.10).

Example

OPIM 5603 Fall 2019

Use the table to find out the expected number of credit cards that a student will own.

x	# Students	$P(X=x)$
0	12	0.08
1	42	0.28
2	57	0.38
3	24	0.16
4	9	0.06
5	4	0.02666
6	2	0.01333

Solution: Let X be the random variable recording the number of credit cards students have. The probability distribution of X is:

The expected value:

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$

$$= 0 \cdot 0.08 + 1 \cdot 0.28 + 2 \cdot 0.38 + 3 \cdot 0.16 +$$

$$+ 4 \cdot 0.06 + 5 \cdot 0.02666 + 6 \cdot 0.01333$$

$$= 1.97333$$

Example

Your friend tosses a fair coin. If the outcome is a Head, you win \$5. Otherwise you lose \$5. What is your expected win?

Solution: Let W be the random variable recording your winnings in a single toss of a fair coin. The probability distribution of W is given by:

x	$P(W=x)$
-5	$^1/_2$
5	$^1/_2$

Expected value is: $E(W) = x_1 p_1 + x_2 p_2 = -5 \cdot {}^1/_2 + 5 \cdot {}^1/_2 = 0.$

A game in which the expected win is 0 is called a fair game.

Example

What is the expected win for a \$1 bet on red in a single roll of American roulette?

Note: The American roulette wheel has 38 numbered fields, two of which are green (**0** and **00**), 18 red and 18 black.

Solution: Let R be the random variable recording your winnings from a \$1 bet on red in a single roll of American roulette.

The probability distribution of R is given by:

x	$P(W = x)$
-1	$^{20}/_{38}$
1	$^{18}/_{38}$

Expected value is:

$$E(R) = x_1 p_1 + x_2 p_2 = -1 \cdot {}^{20}/_{38} + 1 \cdot {}^{18}/_{38} = -{}^{2}/_{38} = -{}^{1}/_{19}.$$

Expected **loss** is \$0.052632, i.e., about 5.3 cents per \$1 bet.

Variance and Standard Deviation

Variance is a measure of the spread of the data. The larger the variance, the larger the spread.

Suppose a random variable has a probability distribution

x	x_1	x_2	x_3	…	x_n
$P(X=x)$	p_1	p_2	p_3	…	p_n

and expected value $E(X) = \mu$.

The variance of a random variable X is defined by:

$$\mathrm{Var}(X) = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \ldots + p_n(x_n-\mu)^2 = E\big((X-\mu)^2\big)$$

The standard deviation of a random variable X is defined as a square root of the variance: $\sigma = \sqrt{\mathrm{Var}(X)}$.

It measures the spread of the data using *the same unit* as the data.

Example

The daily sales of *Impalas* at two *Chevrolet* dealerships are given:

Shiny Chevy Ltd.

# cars sold	7	8	9
Frequency	62	106	62

Chevy Rules Co.

# cars sold	2	3	4	5	6	7	9	10	11	23
Frequency	5	10	14	17	22	28	59	49	24	2

Find the variance and standard deviation of their daily sales.

Note: Both dealerships sold the same number of cars during 230 days: 1840.

Solution: Let S be the random variable recording the daily sales at *Shiny Chevy*.

The probability distribution of S is:

x	7	8	9
Frequency	62	106	62
$P(S=x)\approx$	0.27	0.46	0.27

Expected value $\mu = p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 \approx 0.27 \cdot 7 + 0.46 \cdot 8 + 0.27 \cdot 9 = 8$

$$\begin{aligned}\mathrm{Var}(S) &= p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + p_3(x_3-\mu)^2 \\ &\approx 0.27 \cdot (-1)^2 + 0.46 \cdot 0^2 + 0.27 \cdot 1^2 \\ &\approx 0.53913\end{aligned}$$

$$\sigma = \sqrt{\mathrm{Var}(S)} \approx 0.73426$$

Example (cont)

Let C be the random variable recording the daily sales at *Chevy Rules*. The probability distribution of C is:

x	2	3	4	5	6	7	9	10	11	23
Frequency	5	10	14	17	22	28	59	49	24	2
$P(C=x)\approx$	0.02	0.04	0.06	0.07	0.1	0.12	0.26	0.21	0.1	0.01

Expected value $\mu = p_1 \cdot x_1 + \ldots + p_{10} \cdot x_{10} = \ \ldots \ = 8$

$$\mathrm{Var}(C) = p_1(x_1-\mu)^2 + \ldots + p_{10}(x_{10}-\mu)^2 = \ \ldots \ \approx 8.01739$$

$$\sigma = \sqrt{\mathrm{Var}(C)} \approx 2.8315$$

Example (conclusion)

OPIM 5603
Fall 2019

Slide 1.19

Shiny Chevy Ltd.

x	7	8	9
Frequency	62	106	62

Chevy Rules Co.

x	2	3	4	5	6	7	9	10	11	23
Frequency	5	10	14	17	22	28	59	49	24	2

$$\mu_S = E(S) = 8$$
$$\text{Var}(S) \approx 0.53913$$
$$\sigma_S \approx 0.73426$$

$$\mu_C = E(C) = 8$$
$$\text{Var}(C) \approx 8.01739$$
$$\sigma_C \approx 2.8315$$

Conclusion: These two probability distributions have the same mean yet significantly different variances. Variance (i.e., standard deviation) measures the spread of data around its mean.

Bernoulli Random Variable

A random variable with outcomes 0 and 1 is called *Bernoulli variable* (17th century Swiss mathematician Jacob Bernoulli).

The probability of outcome 1 is denoted by p.

The probability of 0 is $q = 1 - p$ (i.e., $p + q = 1$).

x	$P(X=x)$
1	p
0	$1 - p$

Expected value of a Bernoulli variable is:

$$\mu = E(X) = x_1 p_1 + x_2 p_2 = 1 \cdot p + 0 \cdot q = p.$$

The variance of a Bernoulli variable is:

$$Var(X) = p_1(x_1 - \mu)^2 + p_2(x_2 - \mu)^2 = p \cdot (1 - p)^2 + q \cdot (0 - p)^2$$

$$= pq^2 + qp^2 = pq(q + p) = pq.$$

Bernoulli variable models a *biased coin toss* experiment.

Independent repetitions of this experiment are called *Binomial Trials*.

Binomial (Bernoulli) Trials

Binomial Trials have the properties:

1. Number of trials in the experiment is fixed,
2. The only outcomes are **success** and **failure**,
3. In each trial the **success** probability is the same, and
4. The trials are independent of each other.

In a binomial trial in which the probability of **success** in any trial is p, the probability of exactly k **successes** in n independent trials is given by

$$C(n,k)\, p^k\, (1-p)^{n-k}$$

where $C(n,k) = \binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Probability Distribution

k	$P(X=k)$
0	q^n
1	$C(n,1)\, p\, q^{n-1}$
2	$C(n,2)\, p^2\, q^{n-2}$
3	$C(n,3)\, p^3\, q^{n-3}$
…	…
$n-1$	$C(n,n-1)\, p^{n-1}\, q$
n	p^n

Why combinatorial coefficient $C(n,k)$?

Take for example $n = 5$ and $k = 2$: in five repeated independent experiments we want to list all outcomes that have exactly two successes.

Here's one: $S\ S\ F\ F\ F$ Others are obtained by shuffling S's and F's:

$S\ F\ S\ F\ F$

$S\ F\ F\ S\ F$

$S\ F\ F\ F\ S$

$F\ S\ S\ F\ F$

$F\ S\ F\ S\ F$

$F\ S\ F\ F\ S$

$F\ F\ S\ S\ F$

$F\ F\ S\ F\ S$

$F\ F\ F\ S\ S$

How many are there?

$$C(5,2) = \frac{5!}{2! \cdot (5-2)!} = \frac{5 \cdot 4 \cdot \cancel{3 \cdot 2 \cdot 1}}{2 \cdot 1 \cdot \cancel{3 \cdot 2 \cdot 1}} = 10$$

Mean, Variance, and Standard Deviation

of a Binomial Random Variable X

If X is a binomial random variable associated with a binomial experiment consisting of n trials with probability of **success** p and probability of **failure** q, then the mean, variance, and standard deviation of X are

$$\mu = E(X) = np \qquad \mathrm{Var}(X) = npq \qquad \sigma_X = \sqrt{npq}$$

Example: Five cards are drawn, with replacement, from a standard 52-card deck. If drawing a club is considered a success, find the mean, variance, and standard deviation of the number of successes X.

$$p = \frac{1}{4},\ q = 1 - \frac{1}{4} = \frac{3}{4} \qquad \mu = np = 5\left(\frac{1}{4}\right) = 1.25$$

$$\mathrm{Var}(X) = npq = 5\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = 0.9375 \qquad \sigma_X = \sqrt{npq} = \sqrt{0.9375} \approx 0.968$$

Example

If the probability of a student successfully passing the class (D or better) is 0.82, find the probability that given 8 students

(a) all 8 pass: $C(8,8) \cdot 0.82^8 \cdot 0.18^0 \approx 0.2044$

(b) none pass: $C(8,0) \cdot 0.82^0 \cdot 0.18^8 \approx 0.0000011$

(c) at least 6 pass. Means: 6, **or** 7, **or** 8 'successes':

$$C(8,6) \cdot 0.82^6 \cdot 0.18^2 +$$
$$+\ C(8,7) \cdot 0.82^7 \cdot 0.18^1 +$$
$$+\ C(8,8) \cdot 0.82^8 \cdot 0.18^0$$
$$\approx 0.2758 + 0.3590 + 0.2044 = 0.8392$$

Example revisited

If the probability of a student successfully passing the class (D or better) is 0.82, find the probability that given **800** students at least 650 pass.

Means: 650, 651, 652, ..., 799, or 800 ‘successes’:

$$
\begin{aligned}
& C(800,650) \cdot 0.82^{650} \cdot 0.18^{150} + \\
&+ C(800,651) \cdot 0.82^{651} \cdot 0.18^{149} + \\
&+ C(800,652) \cdot 0.82^{652} \cdot 0.18^{148} + \\
& \quad \cdots \\
&+ C(800,799) \cdot 0.82^{799} \cdot 0.18^{1} + \\
&+ C(800,800) \cdot 0.82^{800} \cdot 0.18^{0}
\end{aligned}
$$

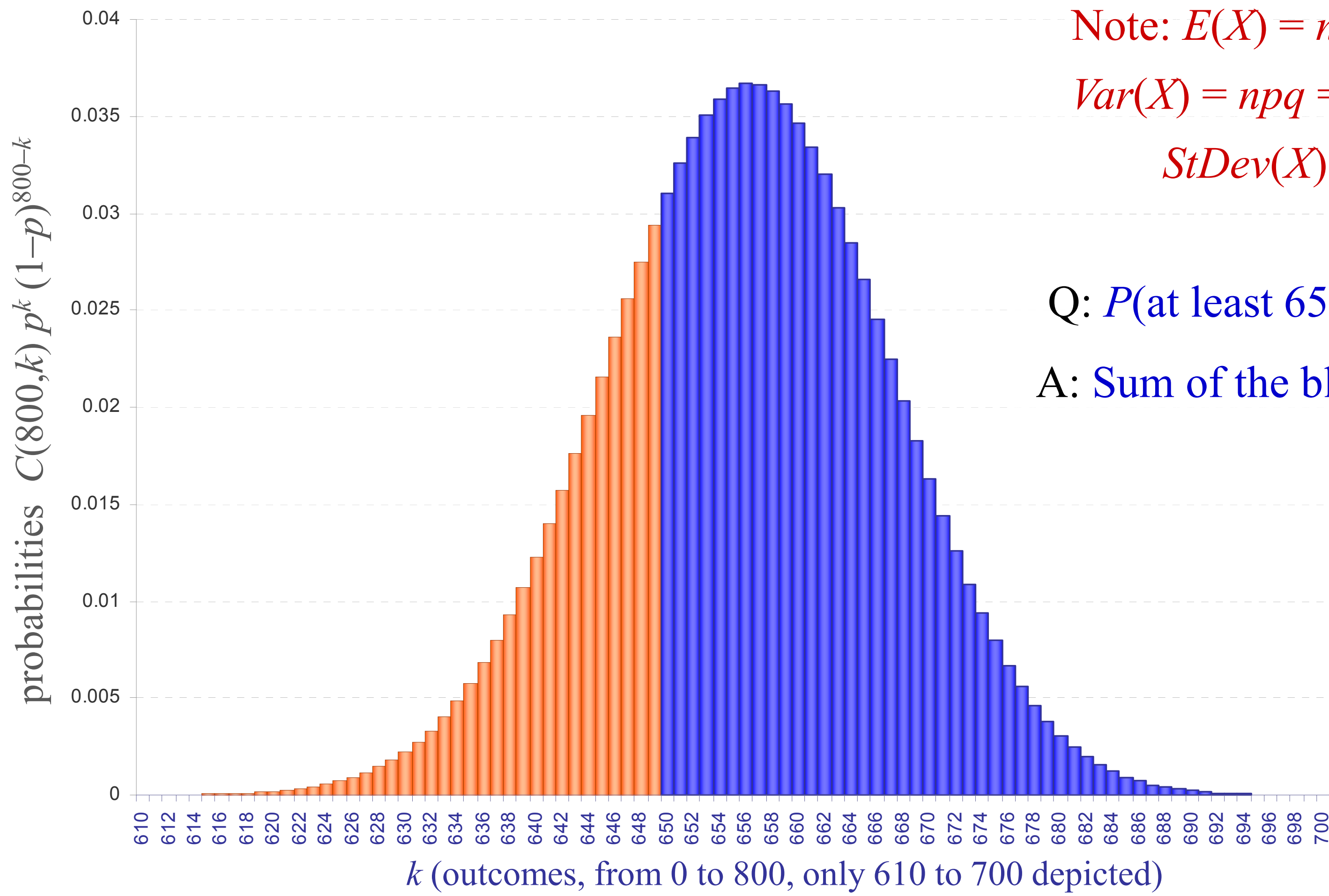
Pretty cumbersome computation! But easy with R (≈ 0.72722).

Histogram of Binomial distribution

OPIM 5603 Fall 2019

Slide 1.26

If we compute the probability distribution (table) for the binomial random variable $X \sim B(n,p)$, with $n = 800$ and $p = 0.82$ (from previous example) and visualize the resulting values with a histogram, we get:



Note: $E(X) = np = 656$

$Var(X) = npq = 118.08$

$StDev(X) \approx 10.87$

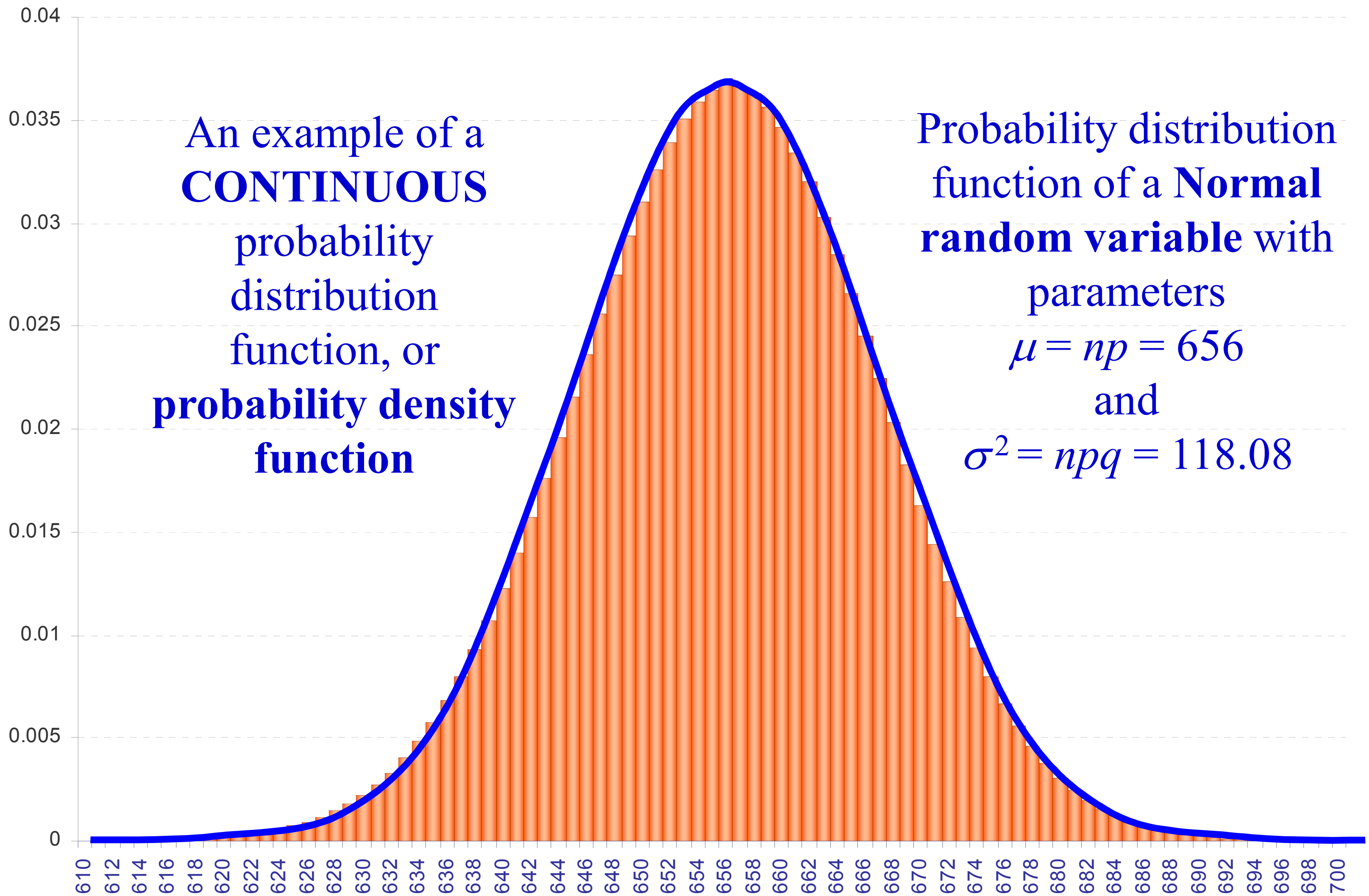
Q: P(at least 650 pass)?

A: Sum of the blue bars.

OPIM 5603 Fall 2019

Slide 1.27

Binomial histogram approximation



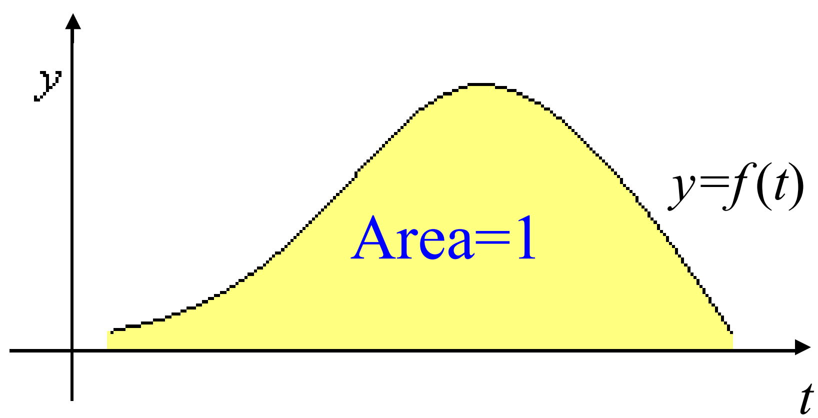
What would happen if we increased the number of trials: 800, 8000, 80000,…?

Probability Density Function

A probability density function f defines a continuous probability distribution and coincides with the interval of values taken on by the random variable associated with an experiment.

A pdf must satisfy:

- $f(t) \geq 0$ for all t in $(-\infty, +\infty)$, and
- the area of the region between the graph of f and the t-axis is equal to 1.



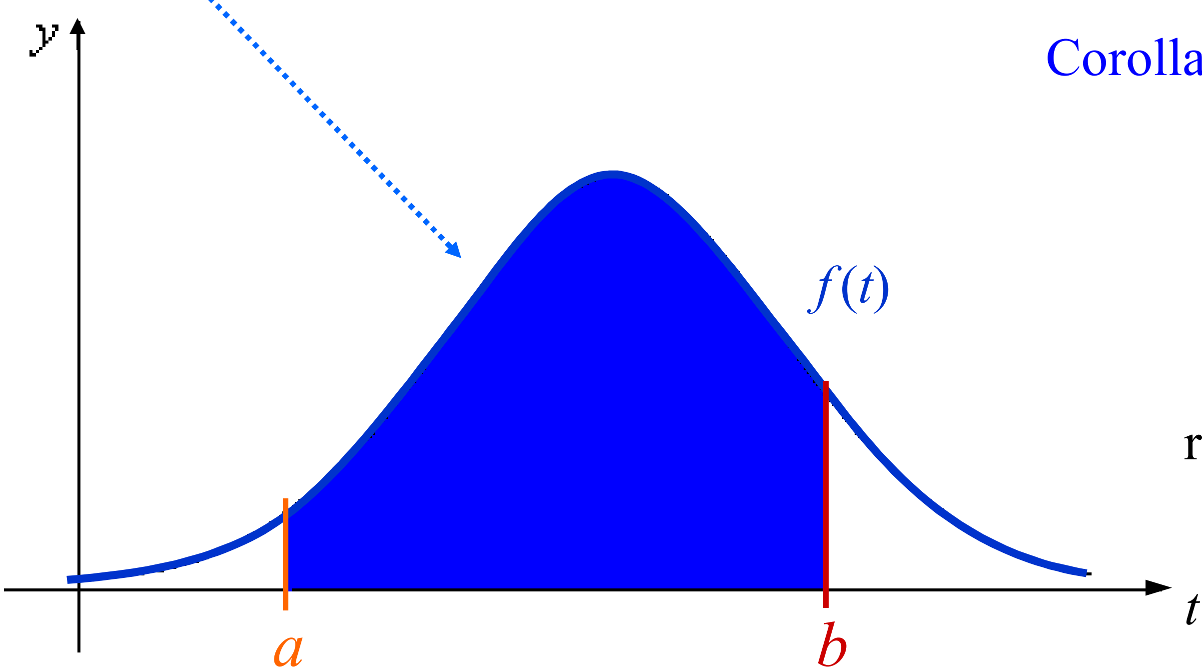
Probability Density Function

OPIM 5603 Fall 2019

Slide 1.29

If X is a random variable with pdf f (X is a *continuous* random variable) then $P(a < X \le b)$ is given by the area of the shaded region.

Note: $P(X = c) = 0$, for any c.



Corollary: $P(a < X \le b) =$

$= P(a < X < b) =$

$= P(a \le X < b) =$

$= P(a \le X \le b)$,

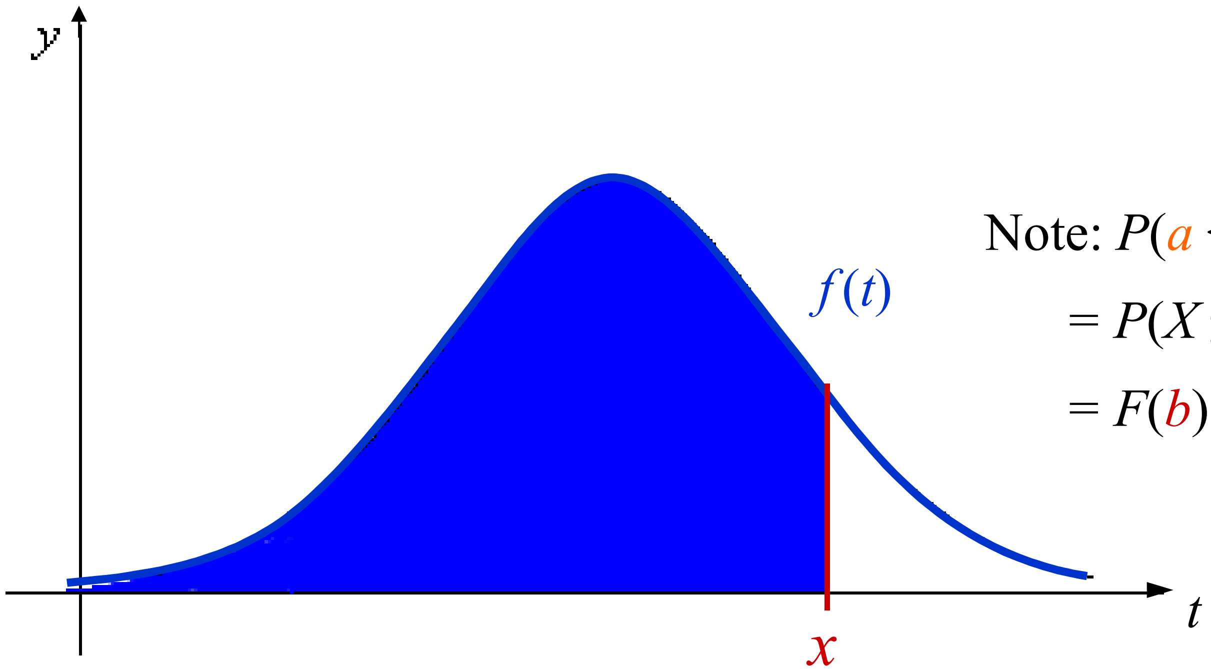
i.e., for continuous r.v.'s the probability of having value in an interval is the same regardless whether end-point(s) are included or not.

Basic calculus: $P(a < X \le b) = \int_a^b f(t)\,dt$

Cumulative Distribution Function

A cumulative distribution function (CDF) F associated with a probability density function f is 'defined' by

$$F(x) = \text{area under } f \text{ over the interval } (-\infty, x].$$



Note: $P(a < X \leq b) =$

$= P(X \leq b) - P(X \leq a) =$

$= F(b) - F(a)$

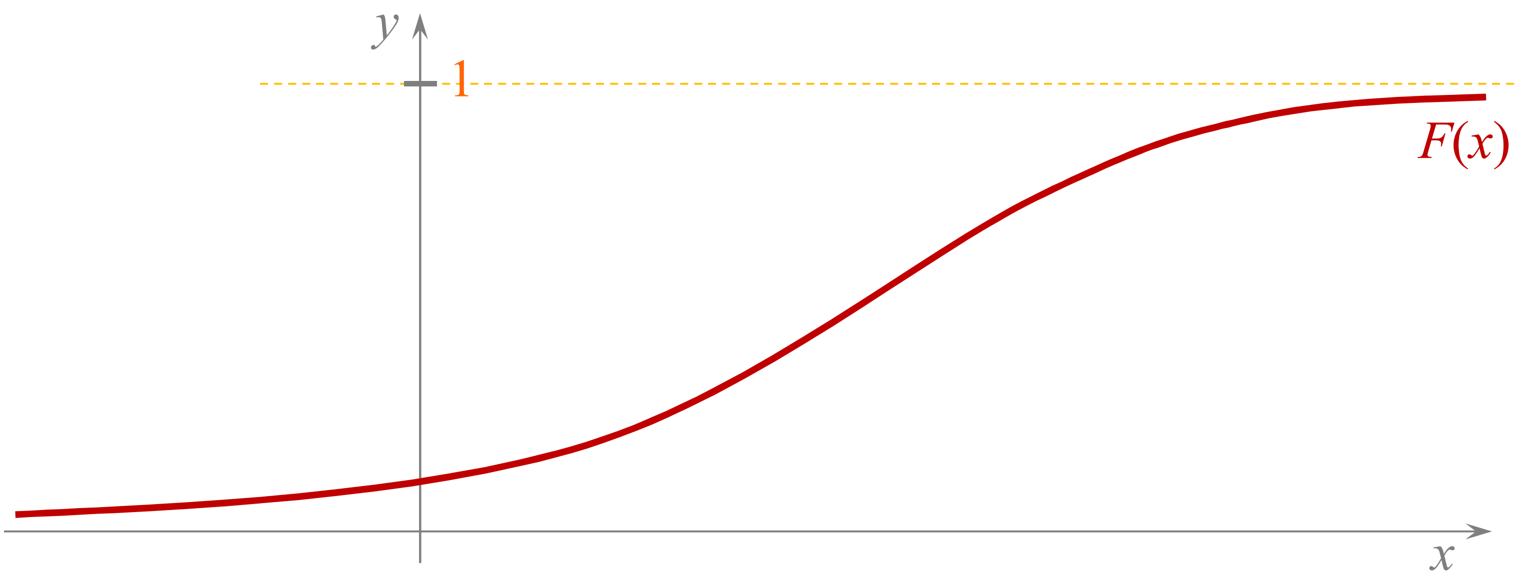
Given a random variable X with a pdf f, we have

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

Properties of CDF

A **CDF** must satisfy:

- $F(x) \geq 0$ for all x in $(-\infty, +\infty)$,
- F is increasing* on $(-\infty, +\infty)$,
- $\lim F(x) = 0$, as $x \to -\infty$, and
- $\lim F(x) = 1$, as $x \to +\infty$.



(*) F needs not be strictly increasing, i.e., it can be constant on some intervals.

Normal Distribution

OPIM 5603 Fall 2019

Normal (Gaussian) distributions are a class of continuous probability density functions. Normal distributions are described by real parameters μ and $\sigma^2 > 0$.

Many real-world phenomena can be accurately modeled by assuming normal distribution $N(\mu, \sigma^2)$ with properly chosen parameters.

The probability density function of $N(\mu, \sigma^2)$:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

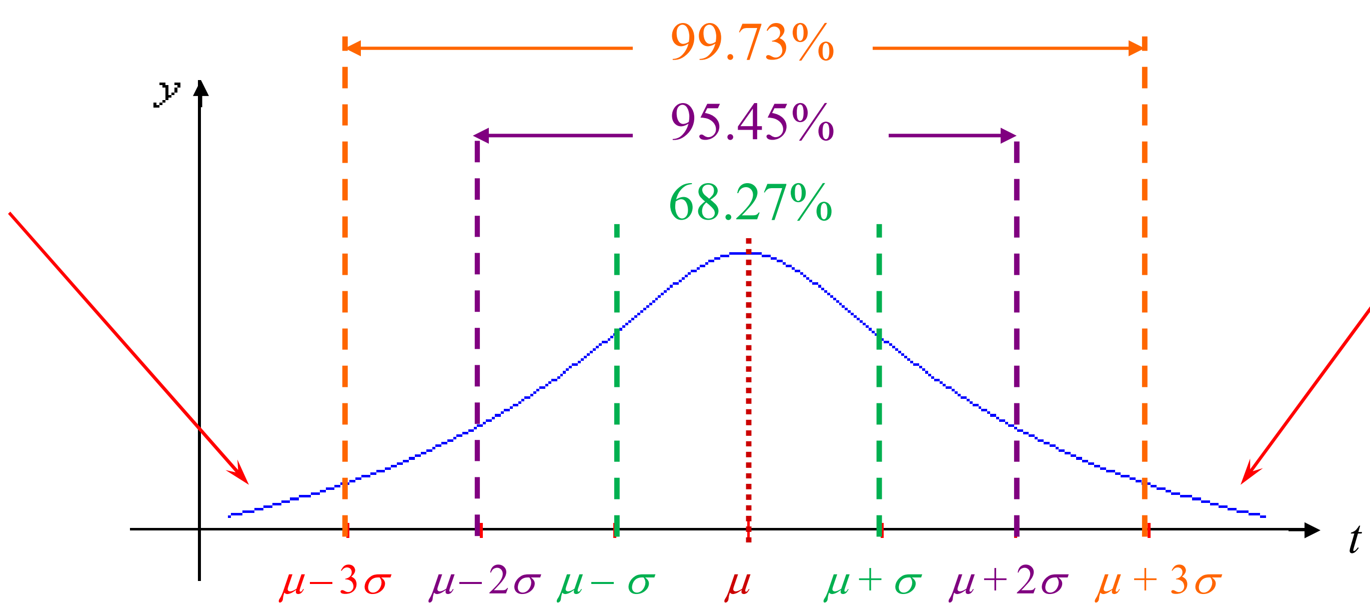
For $\mu = 0$ and $\sigma^2 = 1$ we have the standard normal pdf:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

The graphs of these pdf's are called normal or bell curves.

Normal Curve Properties

1. The area under the curve is 1.
2. The peak is at $t = \mu$, and the curve is symmetric with respect to the vertical line $t = \mu$.
3. The curve lies above and approaches the t-axis.
4. 68.27% of the area lies within $(\mu - \sigma, \mu + \sigma)$,
   95.45% within $(\mu - 2\sigma, \mu + 2\sigma)$,
   99.73% within $(\mu - 3\sigma, \mu + 3\sigma)$.

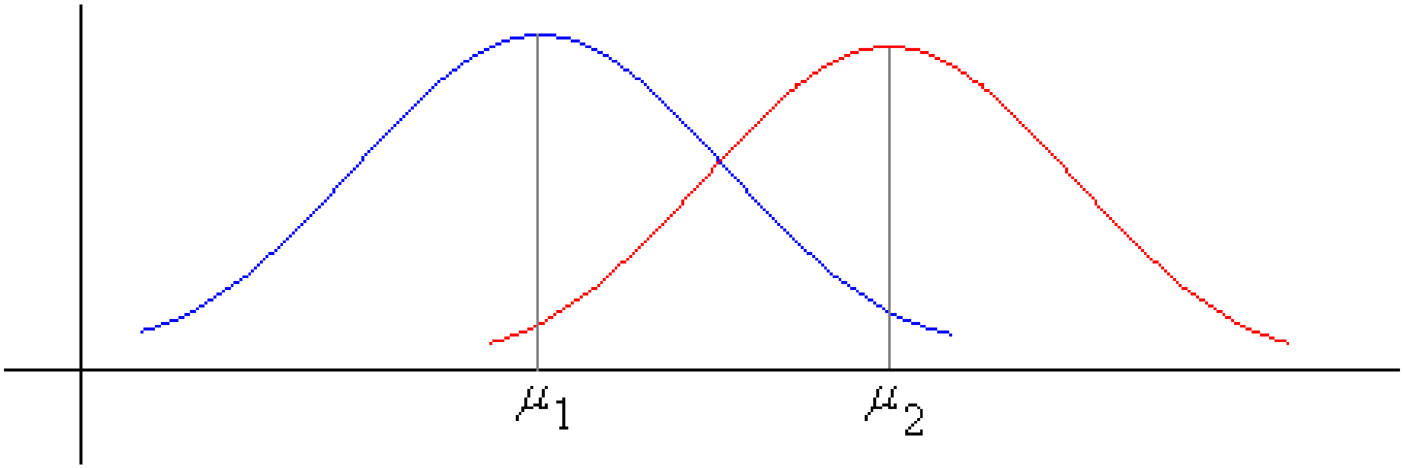


Normal Curves

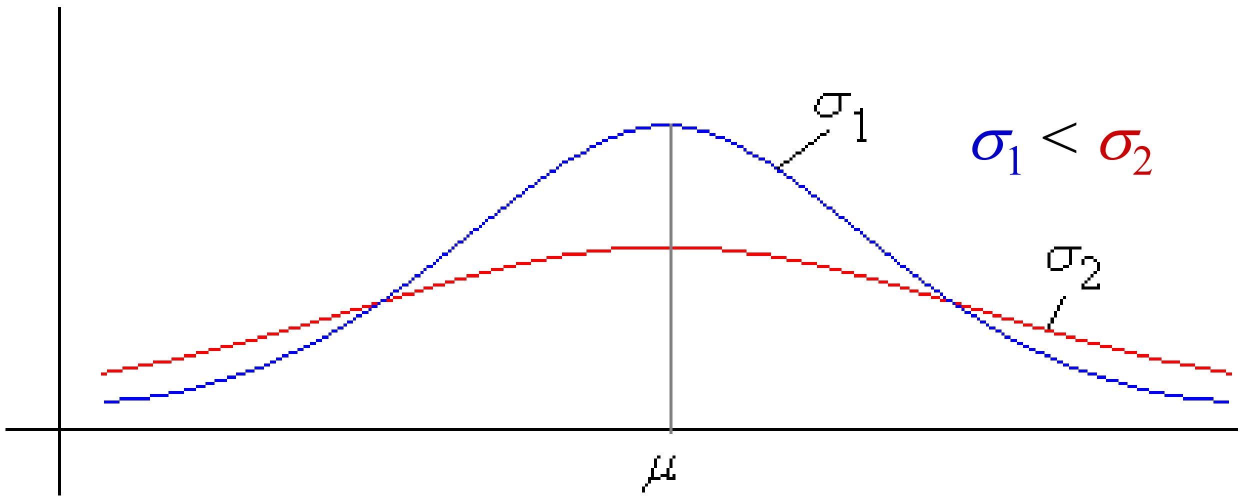
OPIM **5603** Fall 2019

Slide 1.34

Normal curves with same σ and different μ's.



Normal curves with same μ but different σ's'.

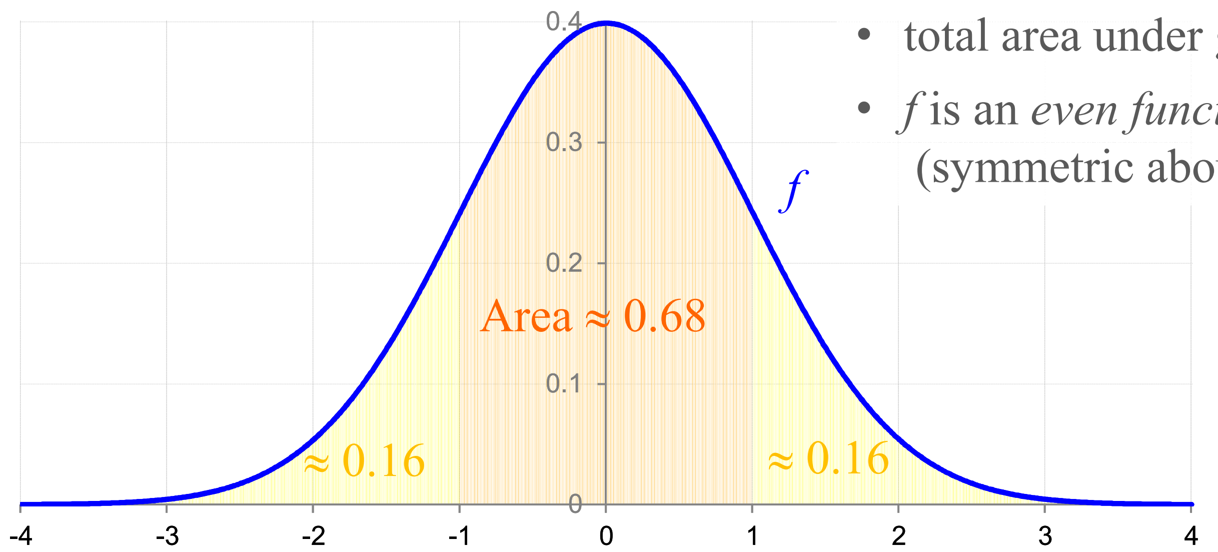


Standard Normal Distribution

Typically denoted by Z: $\mu = 0$ and $\sigma^2 = 1$.

pdf: $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} =$ dnorm(t) *R function*

- $f(t) > 0$ for all t in $(-\infty, +\infty)$,
- total area under graph of f is 1,
- f is an *even function* (symmetric about the y-axis)



CDF: $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt =$ pnorm(x) *R function*

Note: $F(-1) \approx 0.16$

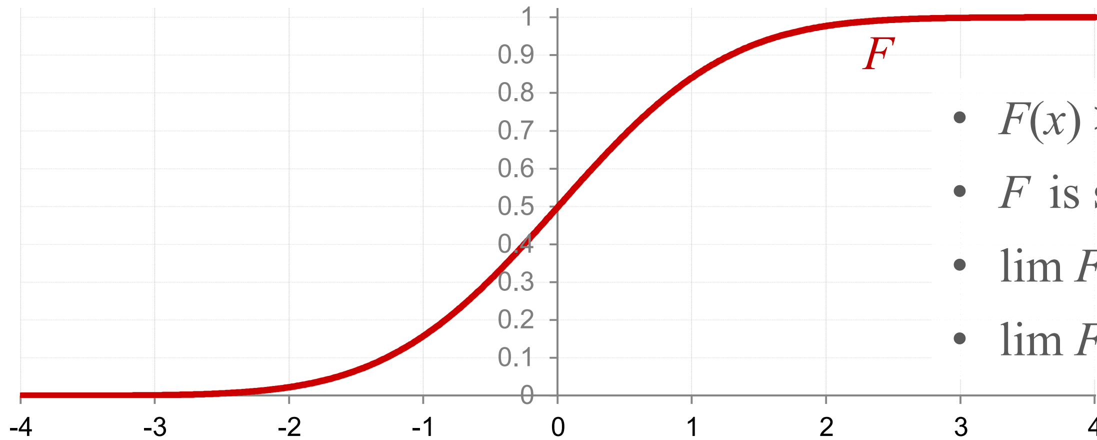
$F(1) \approx 0.16 + 0.68$

$F(0) = 0.5$

Standard Normal Distribution cont.

OPIM
5603
Fall
2019

CDF: $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \; = \text{pnorm}(x)$



- $F(x) > 0$ for all x in $(-\infty, +\infty)$,
- F is strictly increasing,
- $\lim F(x) = 0$ as $x \to -\infty$,
- $\lim F(x) = 1$ as $x \to +\infty$.

"Quantile Function"

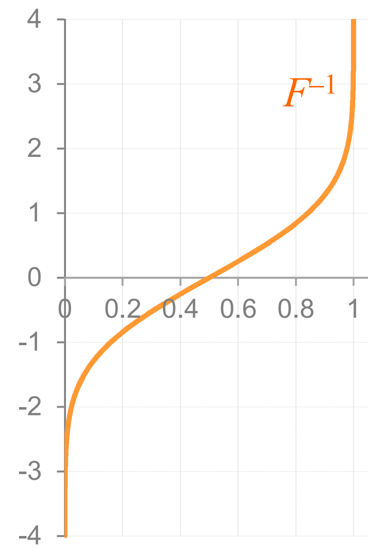
R function

CDF inverse: $F^{-1}(y) = \text{qnorm}(y)$

$$x = F^{-1}(y) \text{ for } y \text{ in } (0,1)$$

if and only if

$$F(x) = y$$



Slide
1.36

Normal distribution in *R*

R provides *density* (pdf), *distribution function* (CDF), *quantile function* (CDF^{-1}) and *random number generator* for the normal distribution with parameters μ (*mean*) and σ (*sd*).

Usage:

dnorm(*t*, *mean*=0, *sd*=1, log=FALSE) pdf

pnorm(*x*, *mean*=0, *sd*=1, lower.tail=TRUE, log.p=FALSE) CDF

qnorm(*p*, *mean*=0, *sd*=1, lower.tail=TRUE, log.p=FALSE) CDF^{-1}

rnorm(*n*, *mean*=0, *sd*=1) random number generator

Using the same naming convention, *R* provides these functions for many other common parametric distributions:

d______pdf q______quantile

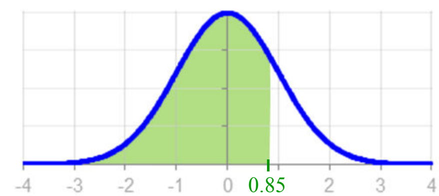
p______CDF r______ random number generator

Simple examples

Example: Let Z be the standard normal variable. Find:

(a) $P(Z < 0.85)$ = (area to the left of 0.85)

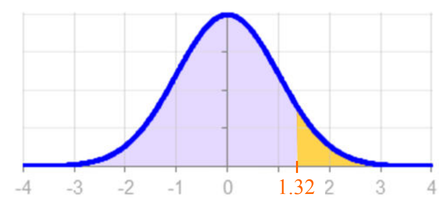
$$= F(0.85) = \text{pnorm}(0.85) \approx 0.8023$$



(b) $P(Z > 1.32)$ = (area to the right of 1.32)

$$= 1 - \text{(area to the left of 1.32)}$$

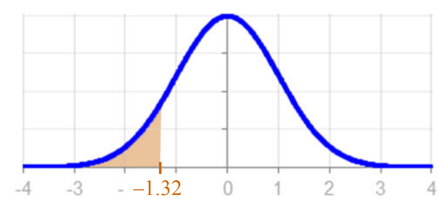
$$= 1 - F(1.32) = 1 - \text{pnorm}(1.32) \approx 0.0934$$



Alternatively, using the fact that pdf f is an even function, the area to the right of 1.32 is the same as the area to the left of –1.32:

$$P(Z > 1.32) = P(Z < -1.32)$$

$$= \text{pnorm}(-1.32) \approx 0.0934$$



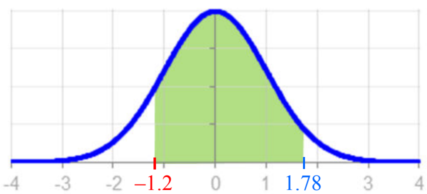
Furthermore: pnorm(1.32, lower.tail=FALSE) ≈ 0.0934

OPIM 5603 Fall 2019

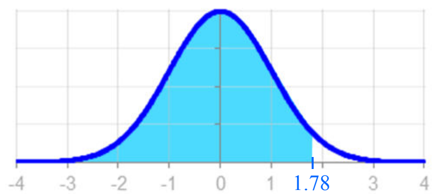
Slide 1.39

Examples cont.

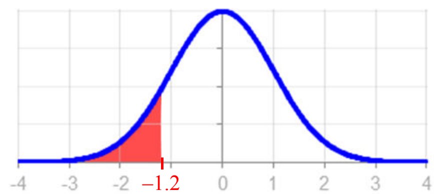
(c) $P(-1.2 < Z < 1.78)$ = (area to the right of –1.2 and to the left of 1.78)



=



–



= (area left of 1.78) minus (area left of –1.2)

$= F(1.78) - F(-1.2)$

= pnorm(1.78) – pnorm(–1.2)

$\approx 0.9625 - 0.1151 = 0.8474$

Expected Value and Variance

Definitions: A random variable is *continuous* if it has the pdf.
(if so, the pdf is *the derivative* of its CDF)

Given a continuous random variable X with a pdf f, we have

$$\text{Expected Value} \quad EX = \int_{-\infty}^{\infty} t\, f(t)\, dt,$$

$$\text{Variance} \quad VarX = E(X - EX)^2.$$

Both values are real numbers (variance is non-negative).

Properties: Given a random variable X and a real number c:

(1) $E(X + c) = EX + c$ (3) $Var(X + c) = VarX$

(2) $E(cX) = cEX$ (4) $Var(cX) = c^2\, VarX$

Note: (3) and (4) follow from (1) and (2), respectively.

Furthermore: for any random variables X and Y, (5) $E(X + Y) = EX + EY$

if X and Y are *independent**, (6) $Var(X + Y) = VarX + VarY$

* *will be introduced in the next chapter.*

Additional properties of N(μ, σ^2)

Given $X \sim N(\mu, \sigma^2)$ (X has a $N(\mu, \sigma^2)$ distribution),

(1) $EX = \int_{-\infty}^{\infty} t\, f(t)\, dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \text{(substitutions, …)} = \mu.$

Note: This justifies naming the parameter μ the *mean*.

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

Hence the parameter σ^2 is called the *variance* (and σ is the *standard deviation*).

(3) For any real numbers $a \neq 0$ and b, $Y = aX + b$ is a normal random variable.

Furthermore, $Y \sim N(a\mu+b, (a\sigma)^2)$.

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

Examples

Example: Suppose $X \sim N(3, 2)$. What is the distribution of $2X - 5$?

Solution: $2X - 5$ is a normal random variable, by (3) on slide 1.41.

Notice that $\mu = 3$, $\sigma^2 = 2$, and with notation used in (3), $a = 2$ and $b = -5$.

Hence $2X - 5 \sim N(a\mu+b, a^2\sigma^2) \sim N(2 \cdot 3 + (-5), 2^2 \cdot 2) \sim N(1,8)$.

Note: $E(2X - 5) = E(2X) - 5 = 2E(X) - 5 = 2 \cdot 3 - 5 = 1$ (by (1), (2) on 1.40)

$Var(2X - 5) = Var(2X) = 2^2 Var(X) = 4 \cdot 2 = 8$ (by (3), (4) on 1.40)

Example: A particular rash has shown up at an elementary school. It has been determined that the length of time that the rash will last is normally distributed with $\mu = 6$ days and $\sigma = 1.5$ days. Find the probability that for a student selected at random, the rash will last for less than 3 days.

$$P(X < 3) = P(X - 6 < 3 - 6) = P\left(\underbrace{\boxed{\frac{X-6}{1.5}}}_{\text{Standard Normal}} < \underbrace{\boxed{\frac{3-6}{1.5}}}_{-2}\right) = \text{pnorm}(-2) \approx 0.02275.$$

Alternatively: $P(X < 3) = P(N(6,1.5^2) < 3) = \text{pnorm}(3,6,1.5) \approx 0.02275.$

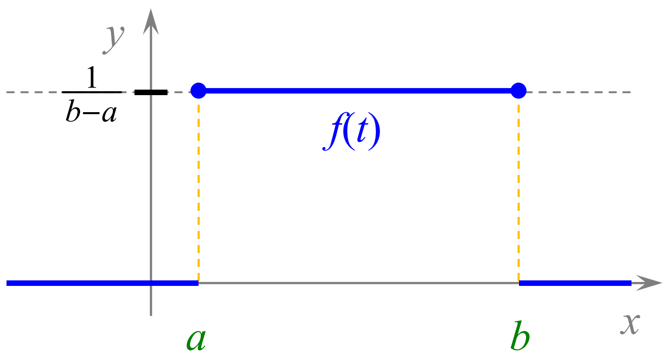
Uniform Distribution

Uniform distribution is the simplest of continuous distributions.

It has two parameters: lower bound a and upper bound b, and it is usually denoted as $U(a,b)$ or $Unif(a,b)$.

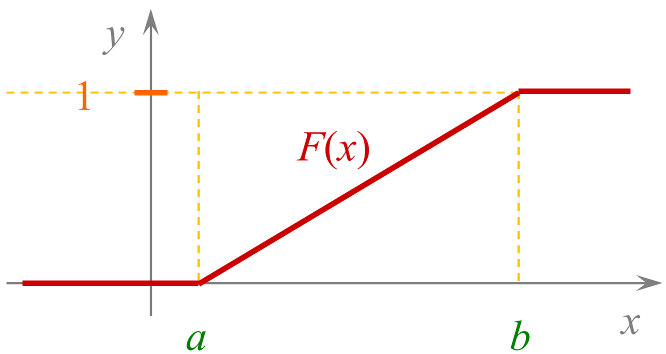
Probability density function:

$$f(t) = \begin{cases} \frac{1}{b-a}, & a \le t \le b \\ 0, & \text{otherwise} \end{cases}$$



Cumulative distribution function:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1 & x > b \end{cases}$$



Expected value = Median = $\frac{a+b}{2}$ Variance = $\frac{(b-a)^2}{12}$

Examples: Properties (1) – (6) on 1.40

Example: Suppose that $X \sim Unif(1,7)$. Compute:

(a) $E(X+3)$ = prop.(1) $= E(X) + 3 = \frac{1+7}{2} + 3 = 7.$

(b) $E(4X)$ = prop.(2) $= 4E(X) = 4 \cdot 4 = 16.$

(c) $Var(X+2)$ = prop.(3) $= Var(X) = \frac{(7-1)^2}{12} = 3.$

(d) $Var(5X)$ = prop.(4) $= 5^2 \cdot Var(X) = 25 \cdot 3 = 75.$

(e) $Var(9-2X)$ = prop.(3) $= Var(-2X)$ = prop.(4) $= (-2)^2 \cdot Var(X) = 4 \cdot 3 = 12.$

Example: Suppose that $X \sim N(3,2)$ and $Y \sim Unif(5,8)$. Compute $E(X+Y)$.

Solution: $E(X+Y)$ = prop.(5) $= E(X) + E(Y) = 3 + \frac{5+8}{2} = 3 + 6.5 = 9.5$

Example: Assume that X and Y above are independent. Compute:

(a) $Var(X+Y)$ = prop.(6) $= Var(X) + Var(Y) = 2 + \frac{(8-5)^2}{12} = 2 + 0.75 = 2.75$

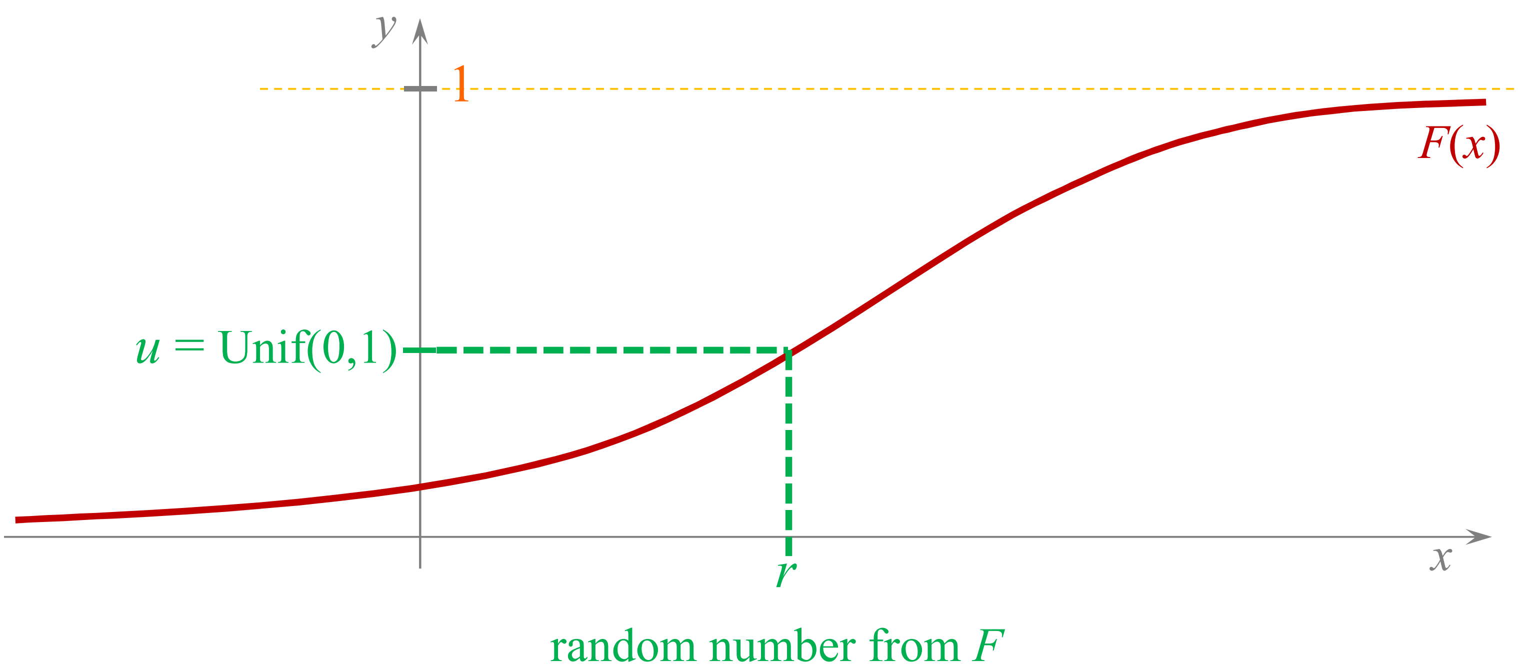
(b) $Var(2X-4Y)$ = prop.(6) $= Var(2X) + Var(-4Y)$ ($2X$ and $-4Y$ are also independent)

= prop.(4) $= 2^2 \cdot Var(X) + (-4)^2 \cdot Var(Y) = 4 \cdot 2 + 16 \cdot 0.75 = 8 + 12 = 20$

Generating random numbers from a CDF

Given a CDF F one can easily generate random numbers from this distribution, assuming:

- F is an ***invertible*** function, and
- Uniform random number generator is available.



Hence $r = F^{-1}(u)$, where u is a uniform random number in [0,1].

Student *t* distribution(s)

Student t distribution with n degrees of freedom is a continuous distribution with a probability density function

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in (-\infty, \infty).$$

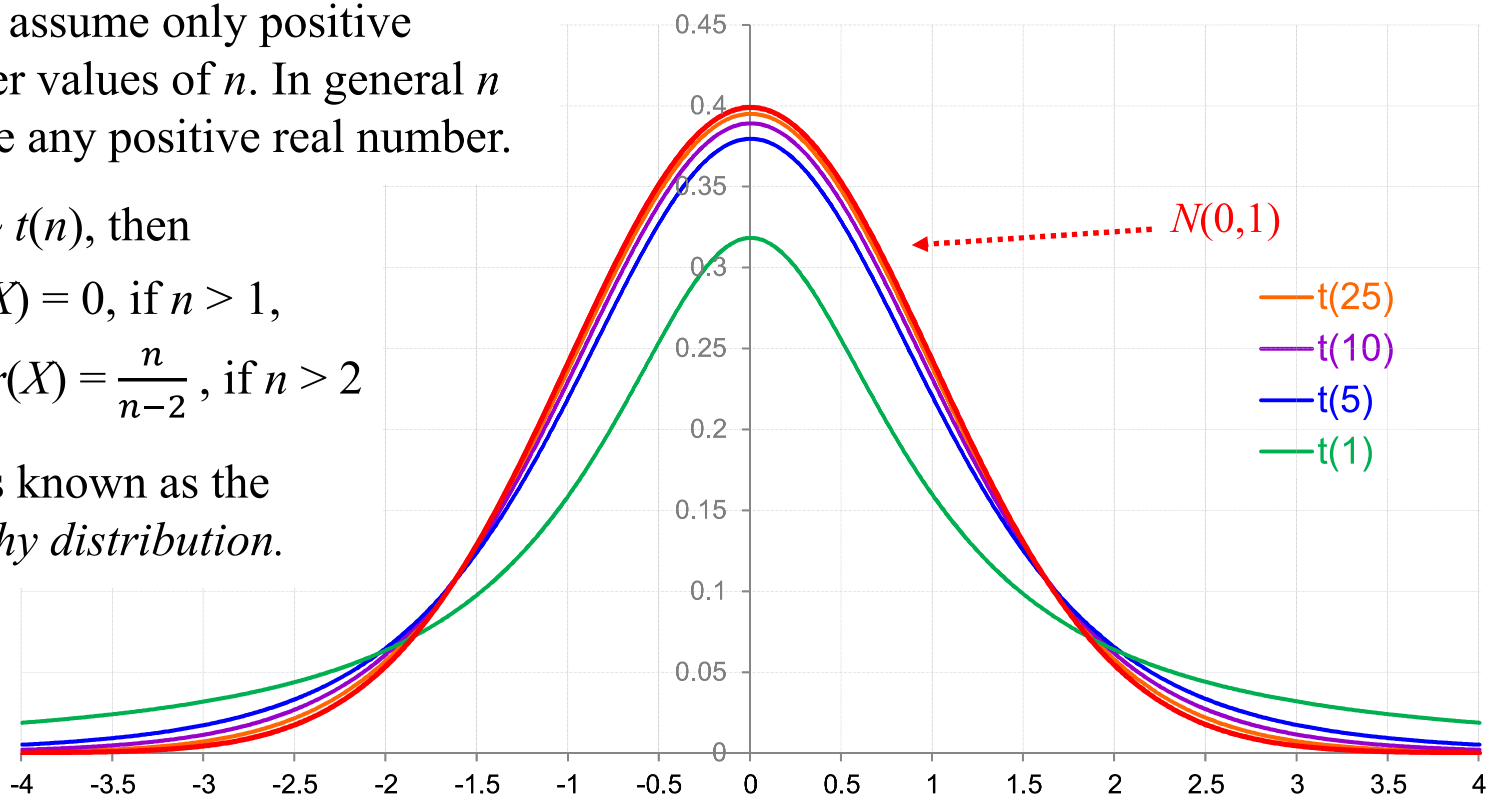
We'll assume only positive integer values of n. In general n can be any positive real number.

If $X \sim t(n)$, then

$E(X) = 0$, if $n > 1$,

$Var(X) = \dfrac{n}{n-2}$, if $n > 2$

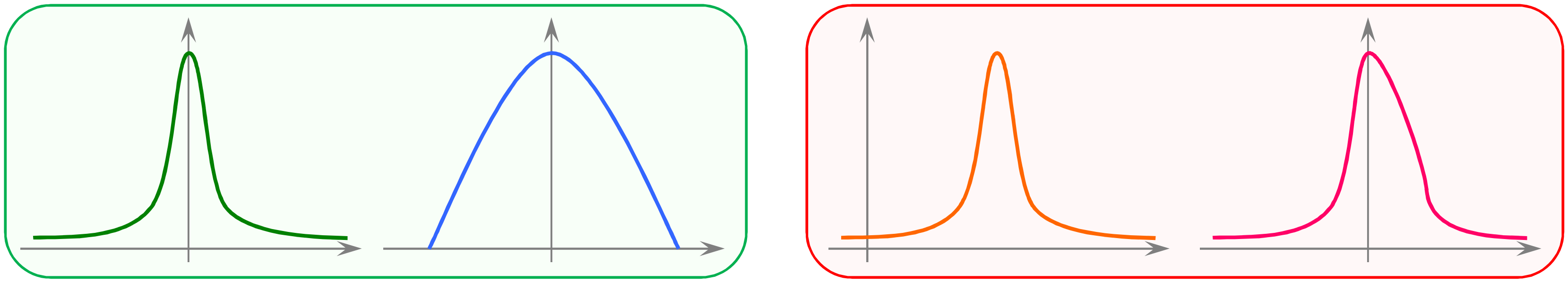
$t(1)$ is known as the *Cauchy distribution.*



Symmetric distributions

$N(0,1)$ and t-distributions are *symmetric*.

Random variable X is symmetric if X and $-X$ have the same distribution. Equivalently, its pdf must be symmetric around the y-axis, i.e., must be an even function.

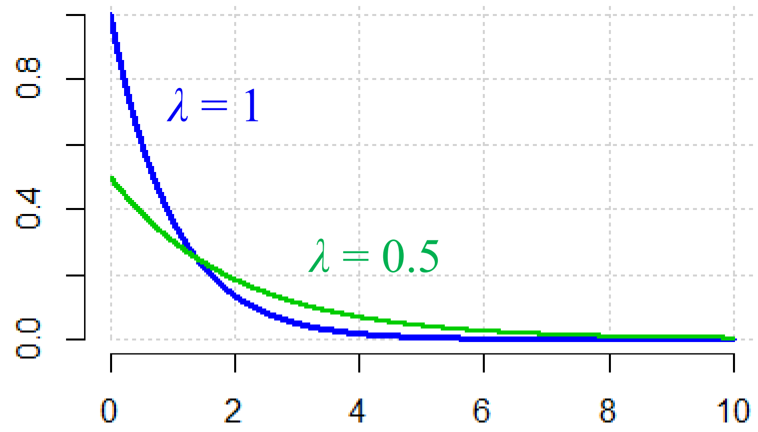


Exponential Distribution

Exponential distribution has only one parameter: *rate* $\lambda > 0$.

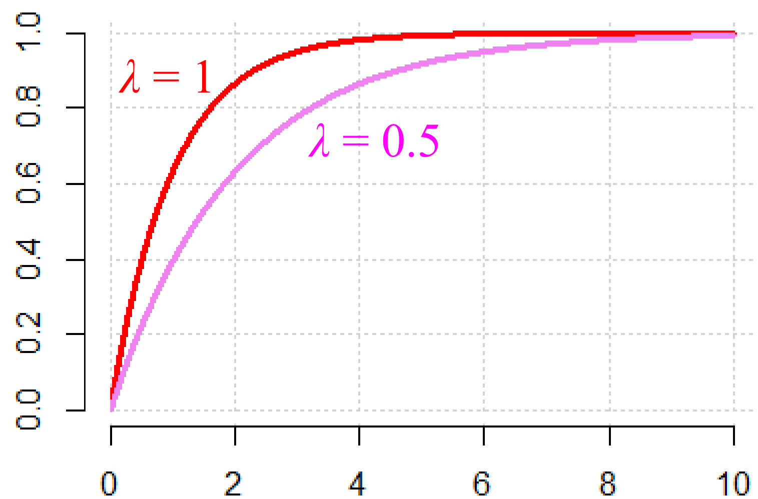
Probability density function:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



Cumulative distribution function:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



Expected value $= \frac{1}{\lambda}$ Median $= \frac{\ln(2)}{\lambda}$ Variance $= \frac{1}{\lambda^2}$

Poisson Distribution

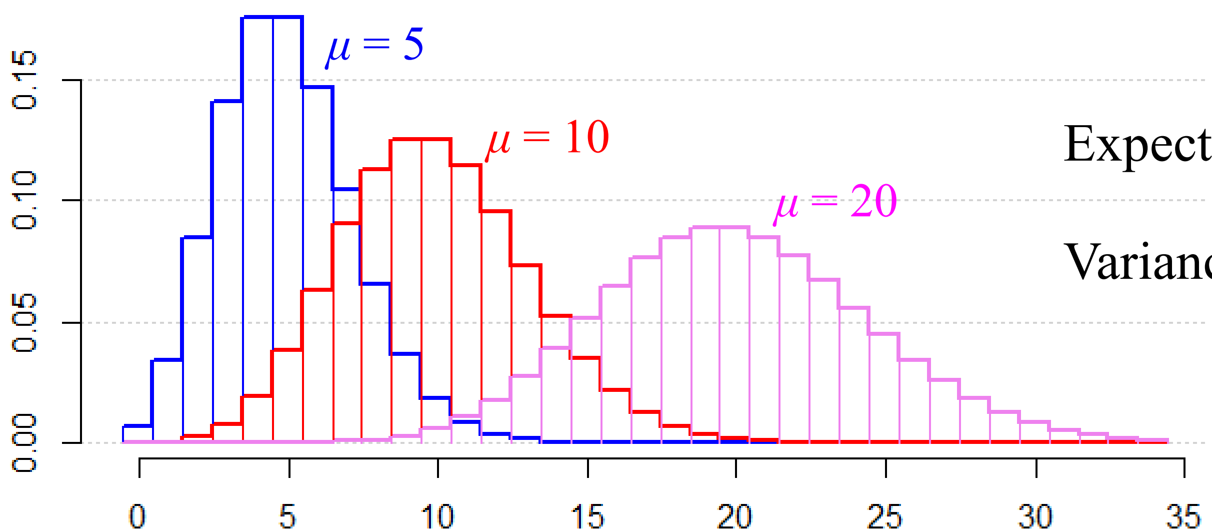
OPIM 5603 Fall 2019

Slide 1.49

Poisson distribution is an infinite discrete distribution with a parameter $\mu > 0$.

Probability distribution: $P(N=k) = e^{-\mu}\dfrac{\mu^k}{k!},\ k = 0,1,2,\ldots$

Alternatively: $N = Pois(\mu) \sim \begin{pmatrix} 0 & 1 & 2 & 3 & \ldots \\ e^{-\mu} & \mu e^{-\mu} & \frac{\mu^2}{2}e^{-\mu} & \frac{\mu^3}{6}e^{-\mu} & \ldots \end{pmatrix}$



Expected value $= \mu$

Variance $= \mu$

Exponential and Poisson distribution play essential roles in the *Poisson Arrival Process*.

All these distributions ?!

Here are couple of examples we will dissect with *R*:

Example 1: *BirthWeight.csv* contains the birth weights (in *kg*) of approx. 65 thousand newborn babies from the northeastern U.S. in 2018. How is the data distributed?

Example 2: The rate at which the low energy (around 1 *GeV*) cosmic ray particles arrive at the top of the atmosphere is about one per square centimeter per second. *Rays.csv* contains the *inter-arrival times** of low-energy cosmic ray particles hitting a space station plate of area 1 *sqcm* during the period of roughly 36 hours. How is the data distributed?

* *inter-arrival times* are the differences between the consecutive arrival times.

Poisson Arrival Process

Arrival Process is a sequence of random variables $0 < A_1 < A_2 < \ldots$ for which *inter-arrival times* random variables $T_k = A_k - A_{k-1}$ (for $k = 1, 2, \ldots$ with $A_0 = 0$)

1) are positive,
2) have the same distribution (*identically distributed*), and
3) are *independent* (more on this shortly).

Notice that two or more arrivals cannot happen at exactly the same instant.

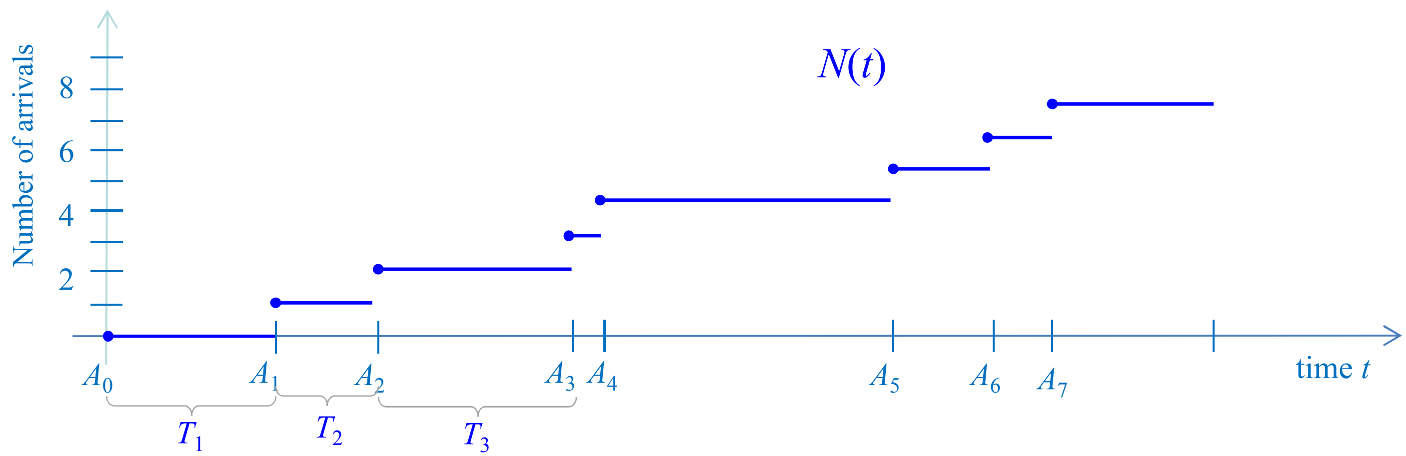


Illustration of arrival process: $N(t)$ counts the number of arrivals by time $t \geq 0$.

Poisson Arrival Process is an arrival process whose inter-arrival times have $Exp(\lambda)$ distribution. λ is called the Poisson Arrival Process *rate*.

The rate λ at which arrivals occur is constant: it cannot be higher in some intervals and lower in other intervals.

Poisson Arrival Process (cont)

Theorem: Number of arrivals of a Poisson process with rate λ in time interval of length Δt has $Pois(\lambda \cdot \Delta t)$ distribution.

Note: Number of arrivals of a Poisson process depends only on the ***length of the interval***, not on the interval itself (endpoints are not relevant).

Example: Arrival of customers at a county diner on Saturday afternoons is modeled as a *Poisson Arrival* process with a rate of 2.2 customers per minute. Let X be a random variable recording the number of arrivals between 12:45 PM and 1:05 PM.

(a) Is this variable finite discrete, infinite discrete, or continuous?

Solution: Number of customer arrivals is a non-negative integer. In principle it cannot be bounded: if you argue (for the sake of argument) that it is bounded by the population of the town, state, U.S., or the world, one could bring over the extra-terrestrials, parallel universes folks, etc., to exceed your bound however large it may be. Thus this is (modeled as) an infinite discrete random variable.

Example (cont.)

Example: Arrival of customers at a county diner on Saturday afternoons is modeled as a *Poisson Arrival* process with a rate of 2.2 customers per minute. Let X be a random variable recording the number of arrivals between 12:45 PM and 1:05 PM.

(b) What is the distribution of X?

Solution: The rate of customer arrivals per minute is $\lambda = 2.2$. Since the customer arrival is a *Poisson Process*, the number of arrivals in time interval of length Δt minutes is a *Poisson random variable* with parameter $\mu = \lambda \cdot \Delta t$.

In this case $\lambda = 2.2$ and $\Delta t = 20$, thus $\mu = 2.2 \cdot 20 = 44$. All told, $X \sim Pois(44)$.

(c) What is the expected value of X?

Solution: $X \sim Pois(44)$ (part (b)), so $E(X) = E(Pois(44)) = 44$.

(d) What is the probability that at most 50 people (i.e., 50 or less) will arrive between 12:45 PM and 1:05 PM?

Solution: By (b), P(at most 50 arrivals between 12:45 and 1:05 PM)

$= P(X \leq 50) =$ `ppois(50,44)` ≈ 0.836891

Exponential Distribution is *Memoryless*

OPIM 5603 Fall 2019

Optional

For any positive s, t: $P(T > t + s | T > s) = P(T > t)$

Note: Conditional probability $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$$P(T > t + s | T > s) = \frac{P((T > t + s) \cap (T > s))}{P(T > s)}$$

$$= \frac{P(T > t + s)}{P(T > s)} = \frac{1 - F(t + s)}{1 - F(s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}$$

$$= \frac{e^{-\lambda t} e^{-\lambda s}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P(T > t)$$

Consequence: Knowledge that arrival did not occur before time s does not yield any information about probability of its arrival between s and $s + t$; this probability depends only on t.