Probability distribution

R script

```
# Normal distribution
dnorm(seq( -3, 3,    by = 0.1), mean = 0, sd = 1)
pnorm(seq( -3, 3,    by = 0.1), mean = 0, sd = 1)
qnorm(seq( 0.1, 0.9, by = 0.1), mean = 0, sd = 1)
rnorm(10, mean = 0, sd = 1)
```

'dnorm' gives the pdf
'pnorm' gives the cdf
'qnorm' gives the quantile function
'rnorm' generates random deviates

```
# Binomial distribution
dbinom(x, N, p)
pbinom(x, N, p)
qbinom(q, N, p)
rbinom(n, N, p)
```

```
# F distribution
df(x, df1, df2, log = FALSE)
pf(q, df1, df2, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2)
```

```
rchisq(100, df = 3) # chi square distribution
runif(x, min, max) # Uniform distribution
rnbinom(10, mu=3, theta=2) # Negative binomial distribution
rpois(x, lambda) # Poisson distribution
rgamma(100, shape = 3, scale = 2) # Gamma distribution
rweibull(100, shape = 3, scale = 2) # Weibull distribution
```

Table of content

Simple probability

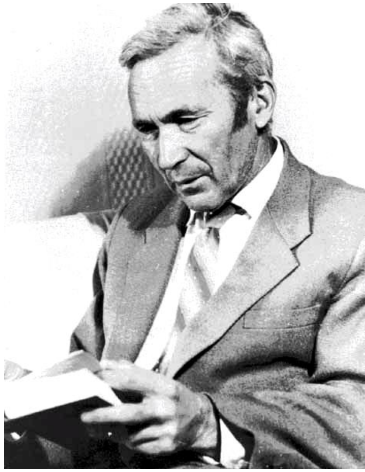
- What is the probability that a card drawn at random from a deck of cards will be an ace?

In this case there are four favorable outcomes:

(1) the ace of spades
(2) the ace of hearts
(3) the ace of clubs
(4) the ace of diamonds

The probability is 4/52

Kolmogorov's axiomization of probability theory



http://free-math.ru/history/biogr/kolmogorov.jpg

Andrey Nikolaevich **Kolmogorov** (Russian: Андре́й Никола́евич Колмого́ров) (1903 –1987) was a Soviet Russian mathematician. He made significant contributions to the mathematics of probability theory, topology, intuitionistic logic, turbulence, classical mechanics, algorithmic information theory and computational complexity.

Kolmogorov thought about the nature of probability calculations and finally realized that finding the probability of an event was exactly like finding the area of an irregular shape.

Kolmogorov was able to identify a small set of axioms upon which he could construct the entire body of probability theory. This is Kolmogorov's "axiomization of probability theory." It is taught today as the only way to view probability .

Definition of probability

- ***Experiment***: toss a coin twice
- ***Sample space***: possible outcomes of experiments

  S = {HH, HT, TH, TT}
- ***Event***: a subset of possible outcomes

  A={HH}, B={HT, TH}
- ***Probability of an event*** : an number assigned to an event Pr(A)

  Pr(A) $\geq$ 0

  Pr(S) = 1

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

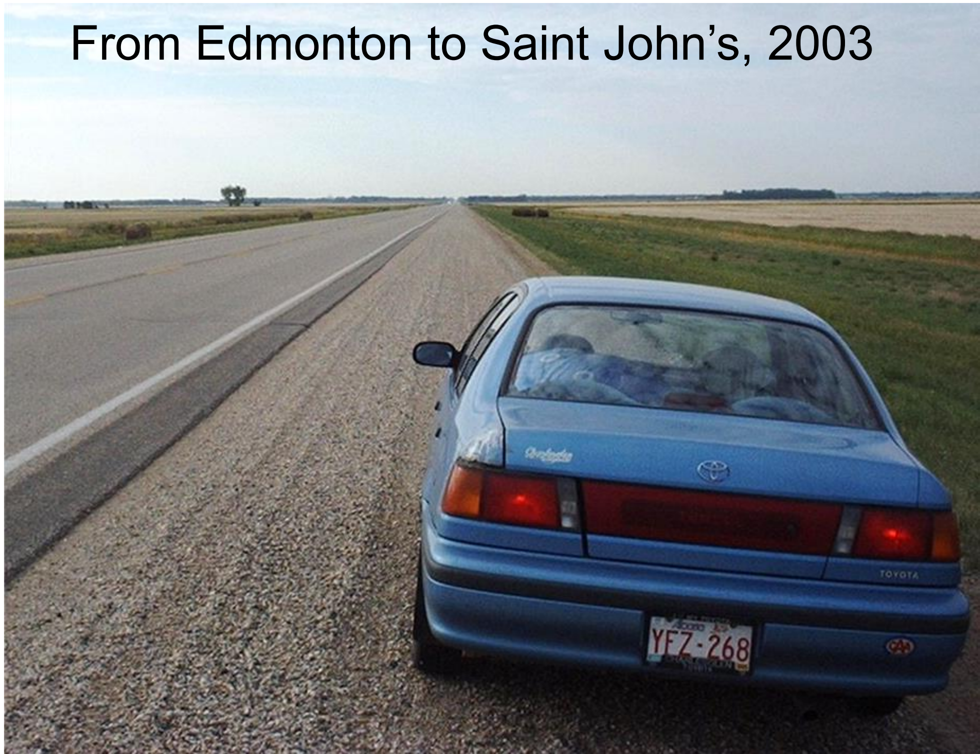
Sample space

How many different license plates are available if each plate contains a sequence of 3 letters followed by 3 digits?

letter digit

□ □ □ □ □ □ →

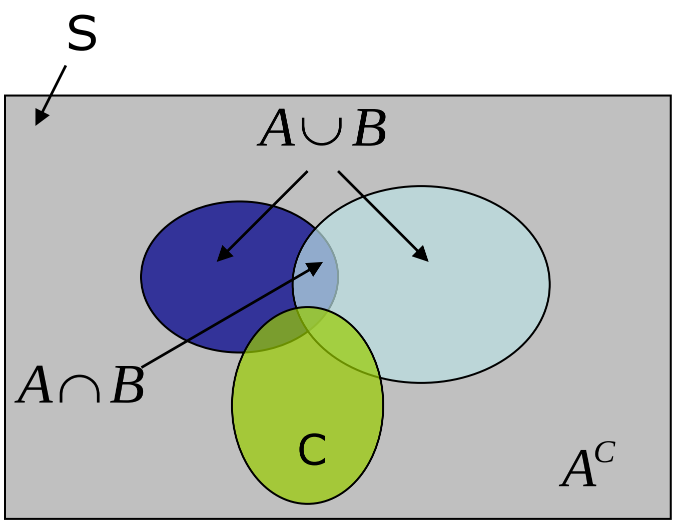
$26^3 \times 10^3$



From Edmonton to Saint John's, 2003

Operations

- Union $A \cup B$
- Intersection $A \cap B$
- Complement A^C



- Properties
  - Commutation $A \cup B = B \cup A$
  - Associativity $A \cup (B \cup C) = (A \cup B) \cup C$
  - Distribution $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - De Morgan's Rule $(A \cup B)^C = A^C \cap B^C$

Axioms and Corollaries

Axioms

$$0 \le P[A]$$

$$P[S] = 1$$

If $A \cap B = \varnothing$

$$P[A \cup B] = P[A] + P[B]$$

If A_1, A_2, … are pairwise exclusive

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

Corollaries

$$P\left[A^C\right] = 1 - P[A]$$

$$P[A] \le 1$$

$$P[\varnothing] = 0$$

$$P[A \cup B] =$$

$$P[A] + P[B] - P[A \cap B]$$

Permutations

Definition

A set of distinct elements in an ordered arrangement.

r-permutation.

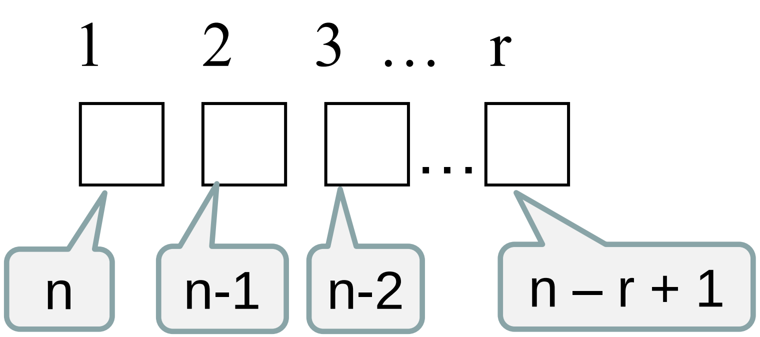
Example Let $S = \{1, 2, 3\}$.

The arrangement 3,1,2 is a permutation of S.

The arrangement 3,2 is a 2-permutation of S.

Theorem 1. The number of r-permutation of a set with n distinct elements is:

$$P(n,r) = n\cdot(n-1)\cdot(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$



Permutations

Example Suppose that a saleswoman has to visit 7 different cities. She can visit the cities in any order she wishes. How many possible orders can the saleswoman use when visiting these cities ?

$$P(n,r) = n \cdot (n-1) \cdot (n-2) \ldots (n-r+1) = \frac{n!}{(n-r)!}$$

7 factorial $7! = 5040$

Combinations

Definition An r-combination of elements of a set is an unordered selection of r elements from the set.

Example Let S be the set $\{1, 2, 3, 4\}$.
Then $\{1, 3, 4\}$ is a 3-combination from S.

Theorem 2 The number of r-combinations of a set with n elements, where n is a positive integer and r is an integer with $0 \le r \le n$, equals

$$C_r^n = C(n,r) = \binom{n}{r} = \frac{p(n,r)}{r!} = \frac{n!}{r!(n-r)!}$$

$$P(n,r) = C(n,r) \times r!$$

Combinations

Example We see that $C(4,2)=6$, since the 2-combinations of $\{a,b,c,d\}$ are the six subsets $\{a,b\}$, $\{a,c\}$, $\{a,d\}$, $\{b,c\}$, $\{b,d\}$ and $\{c,d\}$

Corollary 2. Let n and r be nonnegative integers with $r \le n$. Then $C(n,r) = C(n,n-r)$

From Theorem 2

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

Combinations

Example

How many ways are there to select 5 players from a 10-member tennis team to make a trip to a match at another school?

Sol: $C(10, 5)=252$

Example

9 faculty members in the math department and 11 in the computer science department.

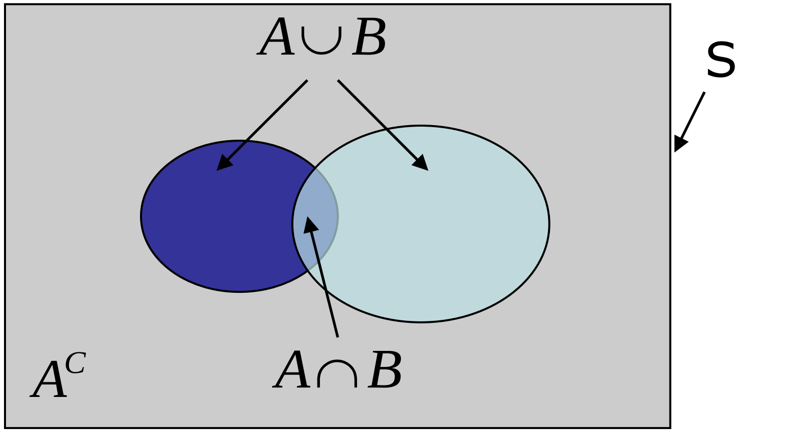
How many ways are there to select a committee if the committee is to consist of 3 faculty members from the math department and 4 from the computer science department?

Sol: $C(9, 3) \times C(11, 4)$

Conditional Probability

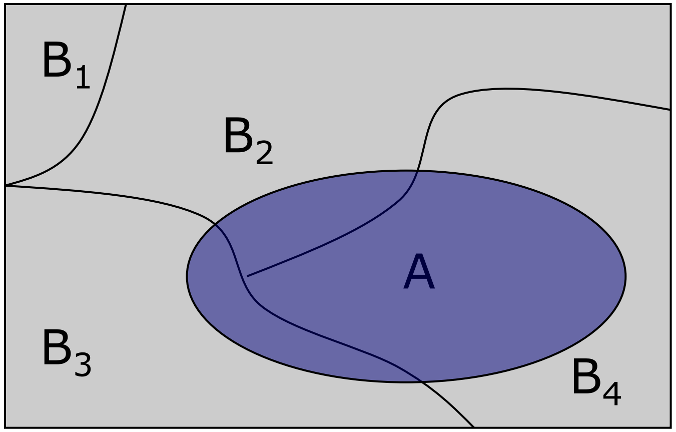
- Conditional Probability of event A given that event B has occurred

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$



- If B_1, B_2,…,B_n a ***partition*** of S, then

$$P[A] = P[A|B_1]P[B_1] + \ldots + P[A|B_j]P[B_j]$$



(Law of Total Probability)

Conditioning Probability

- If A and B are events with Pr(B) > 0, the ***conditional probability of A given B*** is

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

- Example: Drug test

	Women	Men
Success	200	19
Failure	1800	1

A = {Drug fails}

B = {Patient is a Women}

Pr(A|B) = 1800/2000

Pr(B|A) = 1800/1801

Joint Probability

- For events A and B, **joint probability** Pr(AB) stands for the probability that both events happen.

Independence

- Two events ***A and B are independent*** in case

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- A set of events $\{A_i\}$ is independent in case

$$\Pr(A) = \prod_i \Pr(A_i)$$

Conditional independence

- Example: There are three events: A, B, C

A: Tom is late;

B: John is late;

C: No games (e.g. FIFA World Cup)

  - $P(A|C) = 1/10$, $P(B|C) = 1/20$; $P(AB|C) = 1/200$
  - $P(A) = 1/4$, $P(B) = 1/5$; $P(A,B) = 1/8$
  - A, B are conditionally independent given C

- A and B are not independent $\neq$ A and B are not conditionally independent
- Conditional Independence (A is independent of B given C)
  - $P(A|C) \times P(B|C) = P(AB|C)$
  - $P(A|C) = P(A|(B,C))$

The Monty Hall problem

Marilyn vos Savant (August 11, 1946 -) is an American magazine (Parade) columnist, author, and lecturer. She is listed in the Guinness Book of World Records under "Highest IQ (228)".

Since 1986 she has written "Ask Marilyn", a Sunday column in Parade magazine in which she solves puzzles and answers questions from readers on a variety of subjects. On the 9 September 1990 column, she gave a question:

- Suppose you're on a game show, and you're given the choice of three doors. Behind the doors are a car and two goats respectively. If you select a door with a car, you win.

- You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat.

- He says to you, "Do you want to switch (picking door #2)?"

Distribution of random variables

Types of Random Variables

A *discrete* random variable is an random variable whose possible values either constitute a finite set or else can listed in an infinite sequence.

A random variable is *continuous* if its set of possible values consists of an entire interval on a number line.

Probability function

Probability function describes probabilities of the values of a random variable:

$$p(x) = p(X = x)$$ for discrete variable

$$p(a < x \le b) = \int_a^b f(x)dx$$ for continuous variable

Probability function can be conveniently depicted by a bar graph or histogram.

Distribution function is a cumulative probability function:

$$F(x) = p(X \le x) = \sum_{t \le x} p(t) = \int_{x_0}^{x} f(t)dt$$

Binomial experiment

An experiment for which the following four conditions are satisfied is called a *binomial experiment*.

1. The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.
2. The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success (S) or failure (F).
3. The trials are independent.
4. The probability of success is constant from trial to trial: denoted by p.

Binomial random variable

Given a binomial experiment consisting of n trials, the *binomial random variable X* associated with this experiment is defined as

$$X = \text{the number of success among } n \text{ trials}$$

```
n = 100 # number of trials
p = 0.5 # probability of success
N.EXP = 5

trial = numeric(n)
BN = numeric(N.EXP)

trial.1 = sample(c(1, 0), 1, p=c(p, 1-p))
```

```
for (j in 1:N.EXP) {
    for (i in 1:n){
      trial[i] = sample(c(1, 0), 1, p=c(p, 1-p))
    }
    X = sum(trial)
    BN[j] = X
}


BN
hist(BN)
mean(BN)
var(BN)
```

Examples of binomial distribution

- If you toss a coin 4 times, what's the probability that you'll get 3 heads?

- If you draw a card 5 times (with replacement), what's the probability that you'll get one ace?

- If you generate words randomly, what's the probability that you'll have two *the'*s in the first 10 words?

Probability mass function (PMF)

Because the PMF of a binomial random variable X depends on the two parameters n and p, we denote the pmf by $b(x; n, p)$.

Computation of a binomial PMF

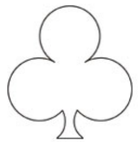
$$b(x;\, n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{(n-x)} & x = 0,1,2,\ldots,n \\ 0 & otherwise \end{cases}$$

$$\binom{n}{x} = C(n, x) = \frac{n!}{x!(n-x)!}$$

```
# R function
p.x <- function(n=10, p=1/6, x)
       {gamma(n+1)/gamma(x+1)/gamma(n-x+1)
        *p^x*(1-p)^(n-x)}
p.x(x=5)
```

Example

A card is drawn from a standard 52-card deck. If drawing a club is considered a success, find the probability of



a. exactly one success in 4 draws (with replacement).

$p = ¼; q = 1 – ¼ = ¾$

$$\binom{4}{1}\left(\frac{1}{4}\right)^{1}\left(\frac{3}{4}\right)^{3} \approx 0.422$$

b. no successes in 5 draws (with replacement).

$$\binom{5}{0}\left(\frac{1}{4}\right)^{0}\left(\frac{3}{4}\right)^{5} \approx 0.237$$

CDF for a binomial distribution

For X ~ Bin(n, p), the CDF will be denoted by

$$P(X \le x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

$$x = 0, 1, 2, \ldots n$$

Example

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

a. all 8 pass. $\binom{8}{8}(0.82)^8(0.18)^0 \approx 0.2044$

b. none pass. $\binom{8}{0}(0.82)^0(0.18)^8 \approx 0.0000011$

c. at least 6 pass.

$$\binom{8}{6}(0.82)^6(0.18)^2 + \binom{8}{7}(0.82)^7(0.18)^1 + \binom{8}{8}(0.82)^8(0.18)^0$$

$$\approx 0.2758 + 0.3590 + 0.2044 \quad = 0.8392$$

Mean and variance

For X ~ Bin(n, p), then

$E(X) = np$,

$V(X) = np(1 – p) = npq$,

$\sigma_X = \sqrt{npq}$

(where $q = 1 – p$).

$$\mu = n \cdot p$$

$$\sigma^2 = n \cdot p \cdot q$$

$$\sigma = \sqrt{n \cdot p \cdot q}$$


Example

5 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 *X* (where *X* is the number of successes).

$$p = 1/4; \quad q = 1 - 1/4 = 3/4$$

$$\mu = np = 5\left(\frac{1}{4}\right) = 1.25$$

$$V(X) = npq = 5\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = 0.9375$$

$$\sigma_X = \sqrt{npq} = \sqrt{0.9375} \approx 0.968$$

Example

Randomly guessing on a 100 question multiple-choice test, where each question has 4 possible answers, the mean and variance of the score:

$$\mu = 100 \cdot \tfrac{1}{4} = 25$$

$$\sigma^2 = 100 \cdot \tfrac{1}{4} \cdot \tfrac{3}{4} = 18.8$$

$$\sigma = \sqrt{100 \cdot \tfrac{1}{4} \cdot \tfrac{3}{4}} = 4.3$$

The shape of the binomial distribution

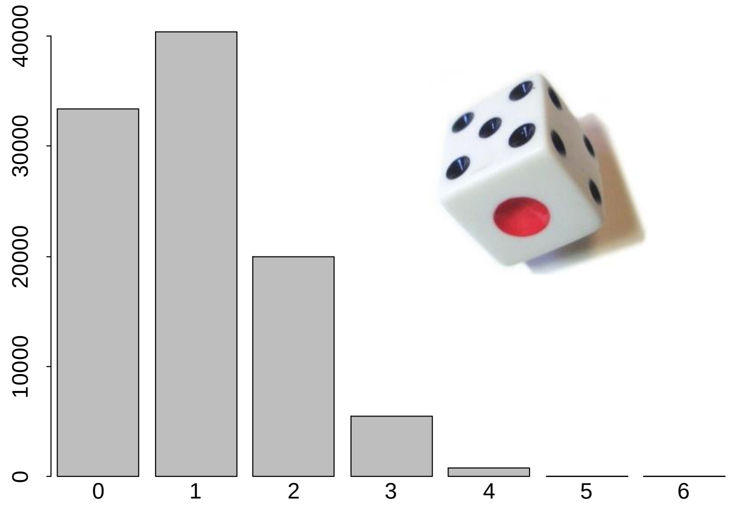
- Shape is determined by values of n and p
  - Only truly symmetric if $p = 0.5$
  - Approaches normal distribution if n is large, unless p is far away from 0.5

To draw a binomial distribution

n = 6; number of dice rolled

p = 1/6; probability of rolling a ‘5’ on any die

x = [0 1 2 3 4 5 6]; number of ‘5’s out of 6



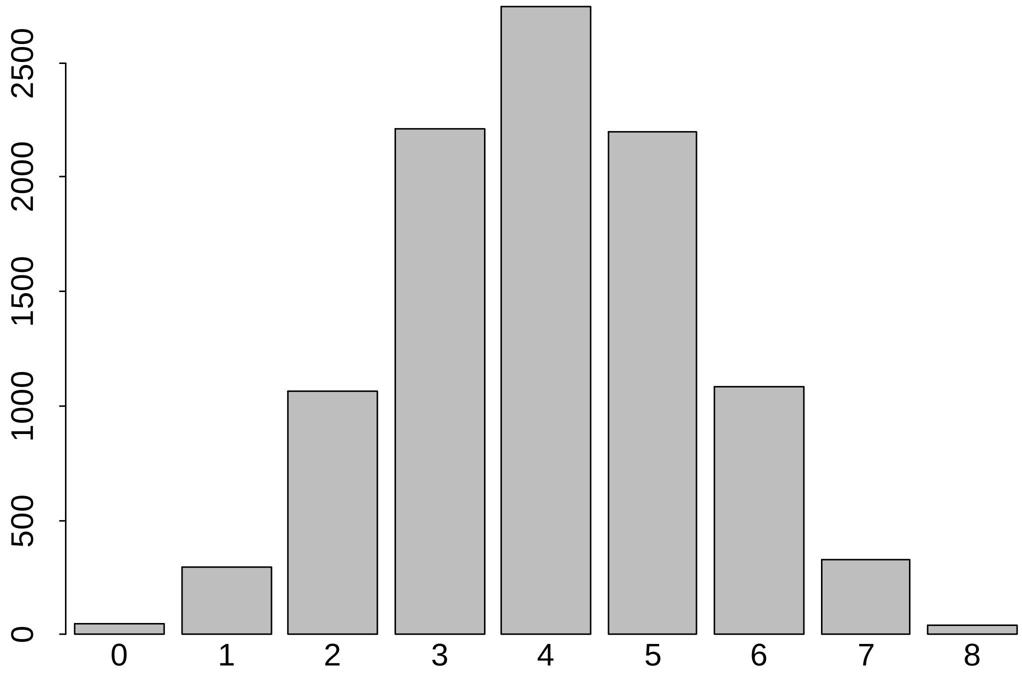

```
barplot(table(rbinom(100000, 6, 1/6)))
```

To draw a binomial distribution

n = 8; number of puppies in litter

p = 1/2; probability of any pup being male

x = [0 1 2 3 4 5 6 7 8]; number of males out of 8



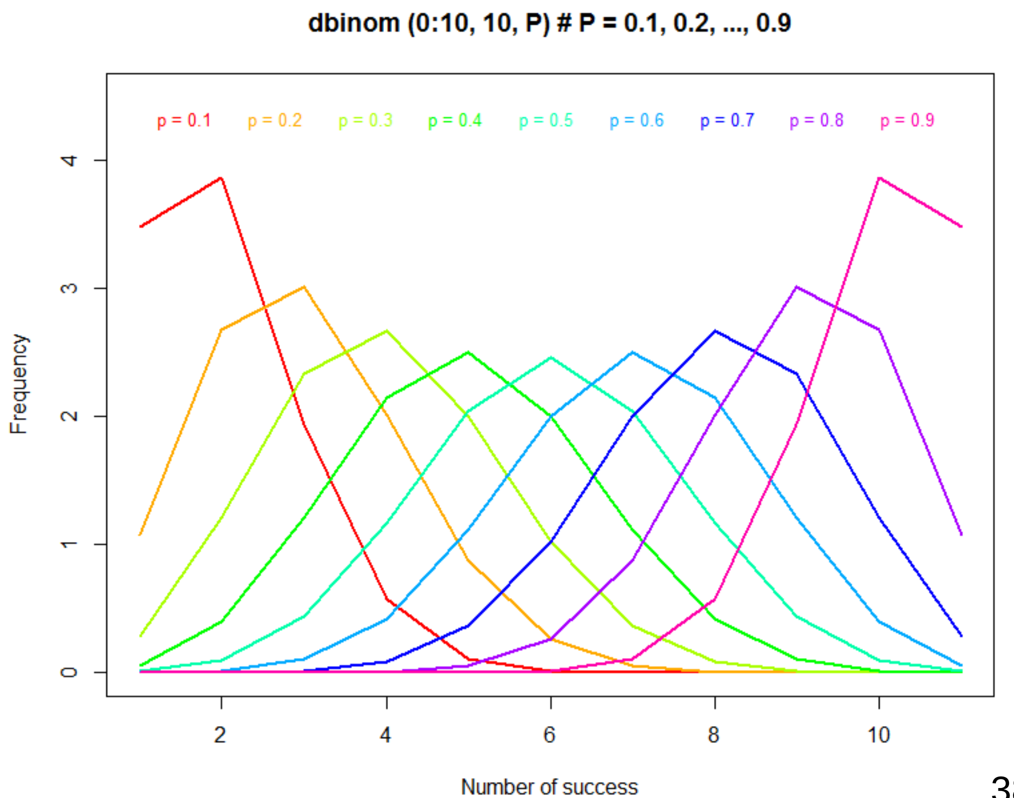

```
barplot(table(rbinom(10000, 8, 0.5)))
```

Binomial distributions at a series of P
(n=10)

```
binomial.PDF = dbinom(0:10, 10, 0.5)
plot(binomial.PDF * 10, type = 'l', ylim = c(0, 4.5),
         xlab = "Number of success", ylab = "Frequency",
         main = paste('dbinom (0:10, 10, P) # P = 0.1, 0.2, ..., 0.9', sep = " "))
```

```
for (i in seq(0.1, 0.9, by = 0.1)) {
  binomial.PDF = dbinom(0:10, 10, i)
  lines(binomial.PDF * 10, type = 'l',
        col = rainbow(9)[i*10], lwd=2)

  legend(-0.3 + 11 * i, 4.5,
          paste('p =', i, sep = " "),
          text.col = rainbow(9)[i * 10],
          box.lty = 0, cex = 0.8)
}
```



Assumptions

- Dichotomous (fall into only two categories)
- Mutually exclusive of the two categories
- Independent between trials
- Randomly selected to be such as “head” or “tail”

Poisson distribution

The distribution was derived by the French mathematician Siméon Poisson in 1837.

The first application was the description of the number of deaths by horse kicking in the Prussian army.

http://www.ats.ucla.edu/STAT/stata/seminars/count_presentation/sdpoisson.jpg

Poisson distribution

The Poisson distribution is used to model the number of events occurring within a given time interval.

The formula for the Poisson probability mass function is

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

λ is the shape parameter which indicates the average number of events in the given time interval.

Poisson distribution – rare occurrences

Some events are rather rare - such as car accidents.

Still, over a period of time, we can see rare events.

Poisson distribution – rare occurrences

Other phenomena that often follow a Poisson distribution are:

- death of infants
- the number of misprints in a book
- the number of customers arriving

Poisson distribution mean and variance

If X has a Poisson distribution with parameter λ, then

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

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

Example- arrivals at a bus-stop

Arrivals at a bus-stop follow a Poisson distribution with an average of 4.5 every quarter of an hour.

Calculate the probability of fewer than 3 arrivals in a quarter of an hour.

Example - arrivals at a bus-stop

The probabilities of 0 up to 2 arrivals can be calculated directly from the formula

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$ with λ =4.5

$$p(0) = \frac{e^{-4.5}4.5^{0}}{0!}$$ So p(0) = 0.01111

Similarly p(1)=0.04999 and p(2)=0.11248

So the probability of fewer than 3 arrivals is 0.01111+ 0.04999 + 0.11248 =0.17358

Example of Poisson distribution - wars by year

- Number of wars beginning by year for years 1482-1939. Table of frequency counts and proportions (458 years):

Wars	*Frequency*	Proportion
0	242	0.5284
1	148	0.3231
2	49	0.1070
3	15	0.0328
4	4	0.0087
More	0	0

```r
# R script
x1 = rep(0, 242)
x2 = rep(1, 148)
x3 = rep(2, 49)
x4 = rep(3, 15)
x5 = rep(4, 4)
x = c(x1, x2, x3, x4, x5)
sum(x) # 307
mean(x) # 0.67
var(x) # 0.738
```

Total wars: $0\times(242) + 1\times(148) + 2\times(49) + 3\times(15) + 4\times(4) = 307$

Average wars per year: 307 wars / 458 years = 0.67 wars/year

variance = 0.738

Using as approximation

- Since mean of empirical (observed) distribution is 0.67, use that as mean for Poisson distribution (that is, set λ = 0.67)
  - $p(0) = (e^{-\lambda}\lambda^0)/0! = e^{-0.67} = 0.5117$
  - $p(1) = (e^{-\lambda}\lambda^1)/1! = e^{-0.67}(0.67) = 0.3428$
  - $p(2) = (e^{-\lambda}\lambda^2)/2! = e^{-0.67}(0.67)^2/2 = 0.1149$
  - $p(3) = (e^{-\lambda}\lambda^3)/3! = e^{-0.67}(0.67)^3/6 = 0.0257$
  - $p(4) = (e^{-\lambda}\lambda^4)/4! = e^{-0.67}(0.67)^4/24 = 0.0043$
  - $P(Y \geq 5) = 1 - P(Y \leq 4) =$

    $1-.5117-.3428-.1149-.0257-.0043=0.0006$

Comparison of observed and model probabilities

- In EXCEL, the function =POISSON(y,λ,*FALSE*) returns $p(y) = e^{-\lambda}\lambda^{y}/y!$

Wars	*Frequency*	Proportion	Model
0	242	0.5284	0.5117
1	148	0.3231	0.3428
2	49	0.1070	0.1149
3	15	0.0328	0.0257
4	4	0.0087	0.0043
More	0	0	0.0006

The model provides a good fit to the observed data.

```
# R script
options(digits = 3)
dpois(0:4, 0.67)
```

Example – products quality

- A production line produces 600 parts per hour with an average of 5 defective parts an hour.

- If you test every part that comes off the line in 15 minutes, what are your chances of finding no defective parts (and incorrectly concluding that your process is perfect)?

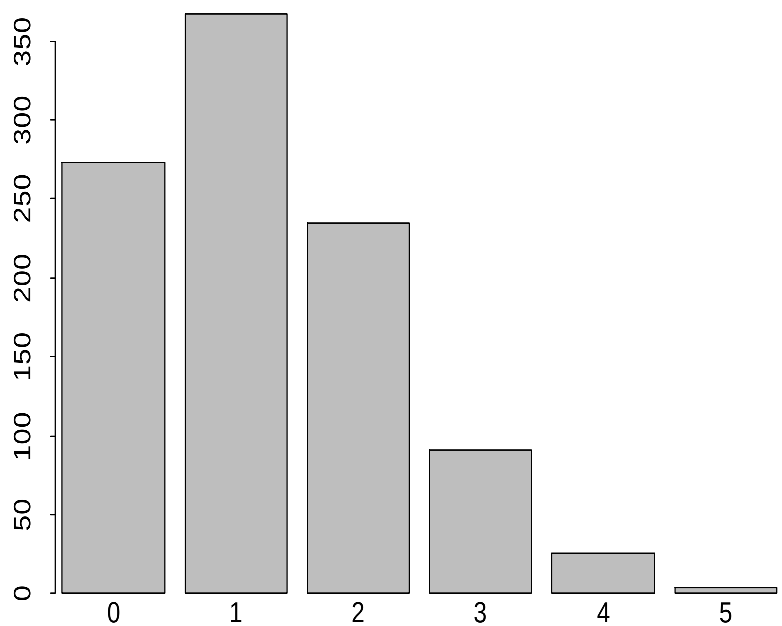
Example – products quality

- A production line produces 600 parts per hour with an average of 5 defective parts an hour. If you test every part that comes off the line in 15 minutes, what are your chances of finding no defective parts (and incorrectly concluding that your process is perfect)?
- λ = (5 parts/hour) $\times$ (0.25 hours observed) = 1.25 parts
- x = 0
- $p(0) = e^{-1.25}(1.25)^0 / 0! = e^{-1.25} = 0.297$

  or about 29 %

To draw a Poisson distribution

- lambda = 1.25; average defects in 15 min
- x = [0 1 2 3 4 5]; number observed
- y = pdf;




```
barplot(table(rpois(1000,1.25)))
```

Example – products quality

- A production line produces 600 parts per hour with an average of 5 defective parts an hour. If you test every part that comes off the line in 15 minutes, what are your chances of finding no defective parts (and incorrectly concluding that your process is perfect)?

- Why not the binomial distribution?

  n = 600 / 4 = 150 ------ large

  p = 5 / 600 = 0.008 ------ small

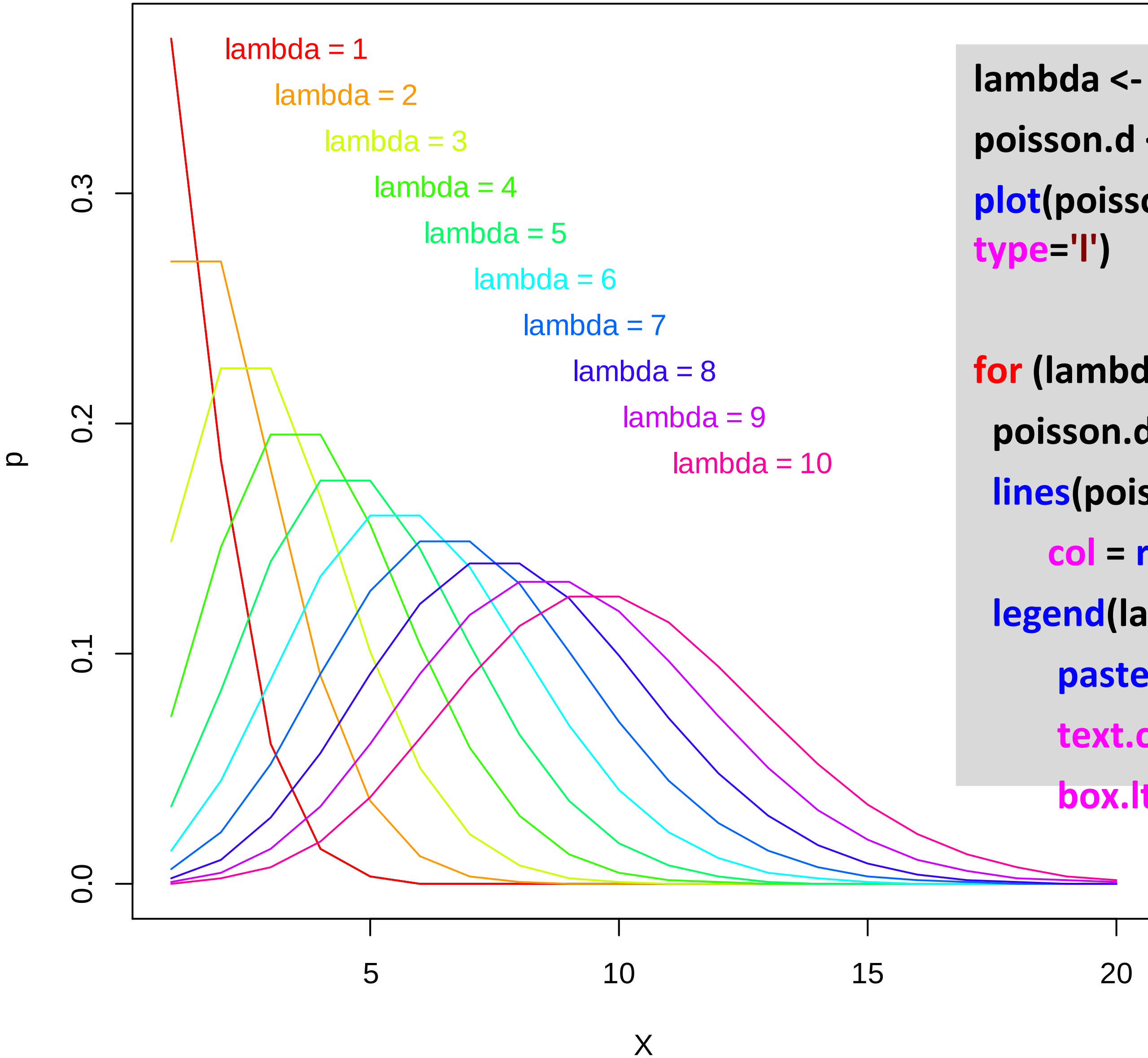
  No needs to calculate 150!

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

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

$$\binom{n}{x} = \frac{p(n,x)}{x!} = \frac{n!}{x!(n-x)!}$$

A collection of graphs for different values of λ




```
lambda <- 1
poisson.d <- dpois(seq(1,20), lambda)
plot(poisson.d, xlab = 'X', ylab='p‘,
type='l')

for (lambda in 1:10){
  poisson.d <- dpois(seq(1,20), lambda)
  lines(poisson.d, type = 'l',
        col = rainbow(10)[lambda])
  legend(lambda, 0.4-lambda/50,
         paste('lambda =', lambda, sep=' '),
         text.col = rainbow(10)[lambda],
         box.lty = 0, cex = 0.8)}
```

Properties of Poisson distrbution

The mean and variance are both equal to λ.

The sum of independent Poisson variables is a further Poisson variable with mean equal to the sum of the individual means.

The Poisson distribution provides an approximation for the Binomial distribution.

The Poisson distribution as a limit

Suppose that in the binomial pdf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\lambda > 0$.

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

Example: death by horse-kick

Binomial situation, $n = 200$, $p = 0.075$

Calculate the probability of fewer than 5 successes.

Advantage: no need to know n and p; estimate the parameter λ from data

X= Number of deaths	frequencies
0	109
1	65
2	22
3	3
4	1
Total year	200
Mean	0.61 death/year

200 yearly reports of death by horse-kick from 10 cavalry corps over a period of 20 years in 19th century by Prussian officials.

Observed vs. expected

x	Data frequencies	Poisson probability	Expected frequencies
0	109	.5435	108.7
1	65	.3315	66.3
2	22	.101	20.2
3	3	.0205	4.1
4	1	.003	0.6
	200		

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

Negative binomial distribution

The number of successes in a sequence of Bernoulli trials before a specified (non-random) number of failures (denoted r) occur.

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in a success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so P(S on trial i) = p for i = 1, 2, 3, …
4. The experiment continues until a total of *r* failures have been observed, where *r* is a specified positive integer.
5. The number of successes has a negative binomial distribution.

Probability mass function (PMF) of a negative binomial distribution

The PMF of the negative binomial random variable X with parameters r = number of failures and $p = P(\text{success})$ is

$$nb(x; r, p) = \binom{x+r-1}{x} p^x (1-p)^r$$

$x = 0, 1, 2, \ldots$

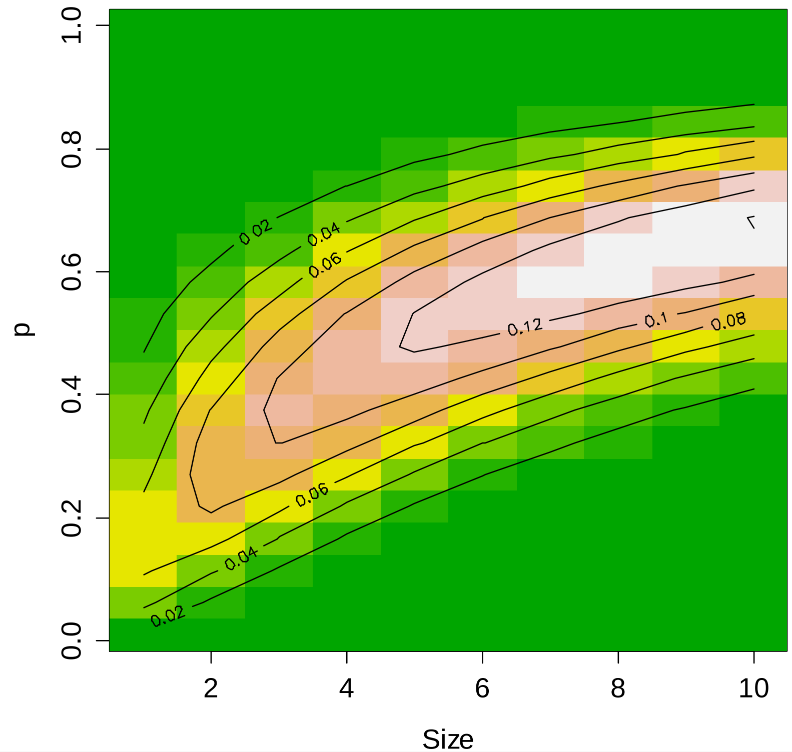
$$nb(x; r, p) = \binom{x + r - 1}{x} p^x (1 - p)^r$$

Negative binomial distribution mean and variance

$$E(X) = \frac{rp}{1-p} \qquad V(X) = \frac{rp}{(1-p)^2}$$

r = number of failures and $p = P$(success)

Negative binomial distribution: effects of parameters



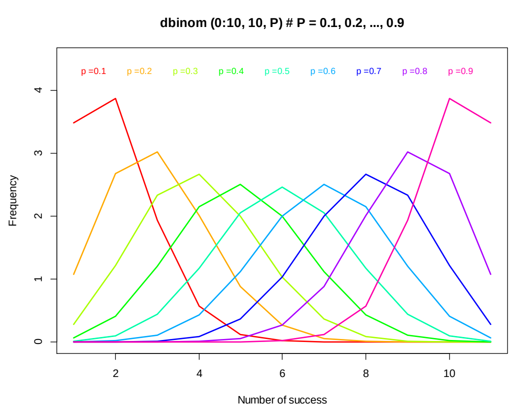

```
size  <- seq(1, 10, len = 10); prob = seq(0.01, 1, len = 20)
p.nb <- matrix(apply(expand.grid(size, prob), 1,
      function(z) dnbinom(5, size = z[1], prob = z[2])),   # x=5
      nrow = length(size), ncol = length(prob))
colnames(p.nb) <- round(prob, 2); rownames(p.nb) <- size
image(size, prob, p.nb, col = terrain.colors(12), xlab = 'Size', ylab = 'p')
contour(size, prob, p.nb, add = TRUE)
```

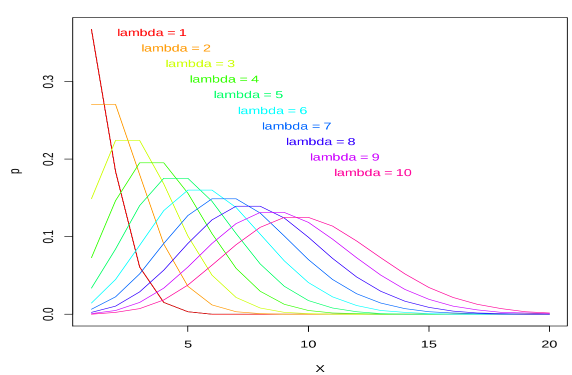
The three discrete distributions

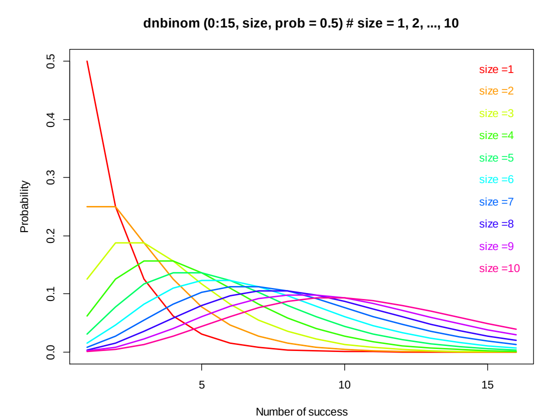
Binomial $E(X) = np \quad V(X) = np(1-p)$

Poisson $E(X) = V(X) = \lambda$

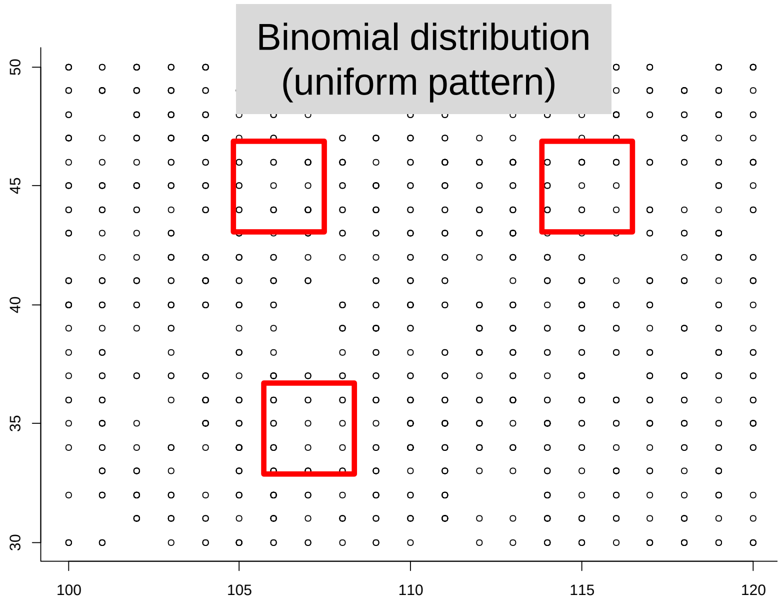
Negative binomial $E(X) = \dfrac{rp}{1-p} \qquad V(X) = \dfrac{rp}{(1-p)^2}$

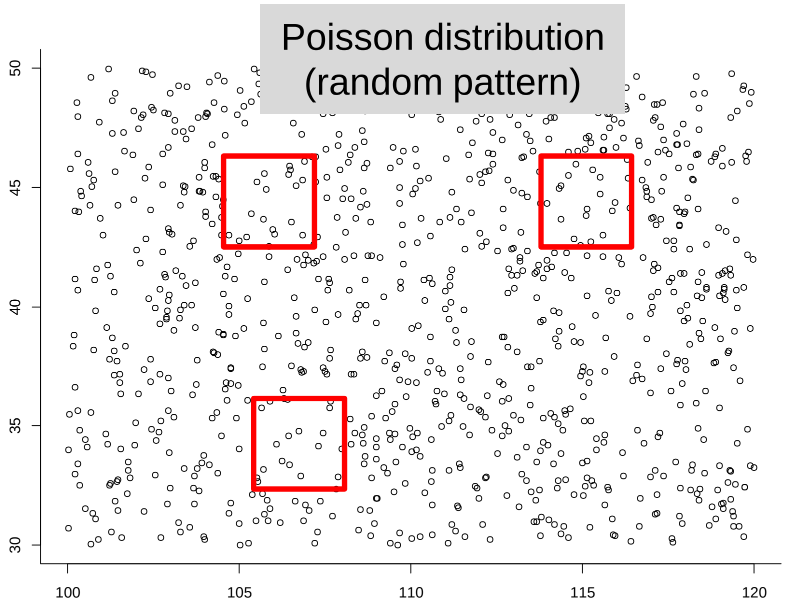


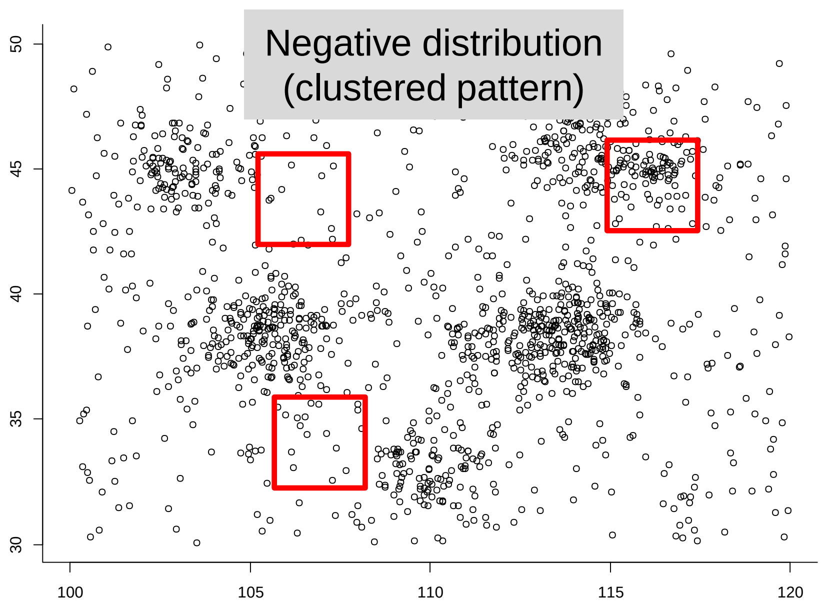




Spatial point patterns







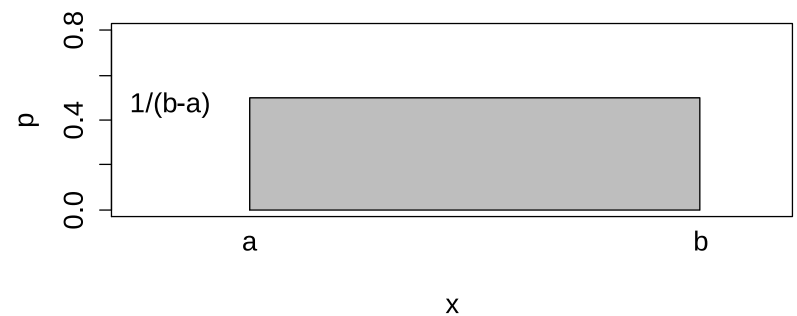
Some import continuous distributions

- The uniform distribution
- The normal distribution
- The chi-square distribution
- The F distribution
- Exponential distribution
- Gamma distribution
- Weibull distribution

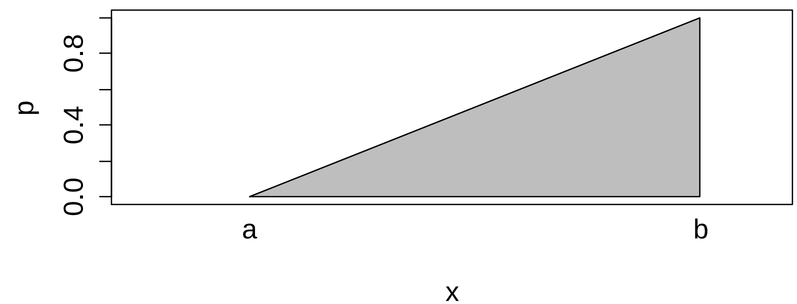
Uniform distribution

Let X denote the outcome when a point is selected at random from an interval $[a,b]$, $-\infty < a < b < \infty$.

The PDF is $f(x;a,b) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$



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



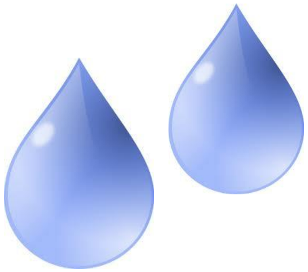
Mean $\mu = E(x) = (a+b)/2$

Variance $= \sigma^2 = Var(x) = E(x-\mu)^2 = (b-a)^2/12$

```
par(mfrow = c(2,1), mar = c(4,4,2,2))
plot(c(1,1,3,3), c(0,0.5,0.5,0), type='l', xlim=c(0,4), ylim=c(0,1), xlab='x', ylab='p', cex=2, xaxt='n')
mtext(c('a','b'), side = 1,at = c(1,3)); legend(0.3, 0.6, '1/(b-a)', box.lty = 0, cex = 1, bty='n')
polygon(c(1,1, 3,3), c(0,0.5,0.5,0), col='grey')

plot(c(1,3),c(0,1), type='l', xlim=c(0,4), ylim=c(0,1), xlab='x', ylab='p', cex=2, xaxt='n')
polygon(c(1,3,3), c(0,0,1), col='grey'); mtext(c('a','b'), side=1, at = c(1,3))
```

Normal distribution



- Measure 5 ml of liquid and weigh it
- Repeat this process many times
- You won’t get the same answer every time, but if you make a lot of measurements, a histogram of your measurements will approach the appearance of a normal distribution.

Normal distribution

- Hold a ping pong ball over a target on the floor, drop it, and record the distance between where it fell and the center of the target
- Repeat this process many times
- You won't get the same distance every time, but if you make a lot of measurements, the histogram of your measurements will approach a normal distribution.

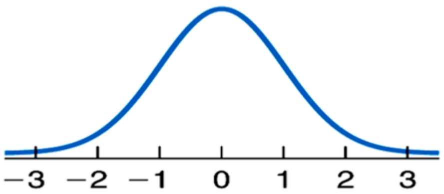
Normal distribution

- Any situation in which the exact value of a continuous variable is altered randomly from trial to trial.
- There is a random error for each trial.
- Note: If your measurement is biased (e.g., there is a steady wind blowing the ping pong ball), then your measurements can be normally distributed around some value other than the true value or target.

Normal or Gaussian distribution

- Start off with something simple, like this:

$$e^{-x^2}$$

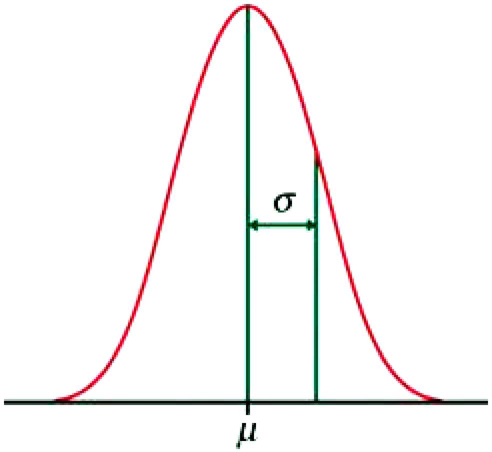


That's symmetric around the y-axis, the frequency slides to 0 as you go off to infinity, either positive or negative.

Normal or Gaussian distribution

- Well, x's average can be something other than 0: it can be any μ

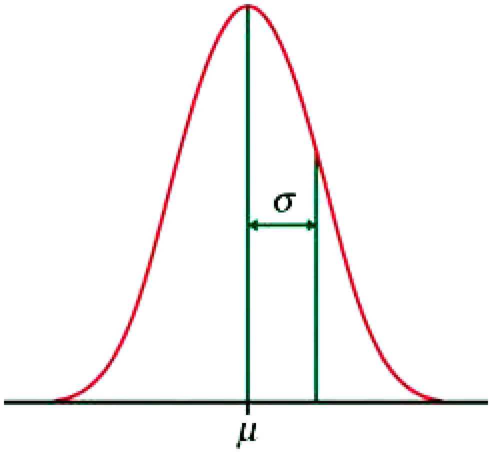
$$e^{-(x-\mu)^2}$$



Normal or Gaussian distribution

And its variance (σ^2) can be other than 1

$$e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



Normalization

It adds up (integrates, really) to 1, we have to divide by a normalizing factor:

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

How to get μ and σ?

- To draw a normal distribution (and integrate to find the area under it), you must know μ and σ

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

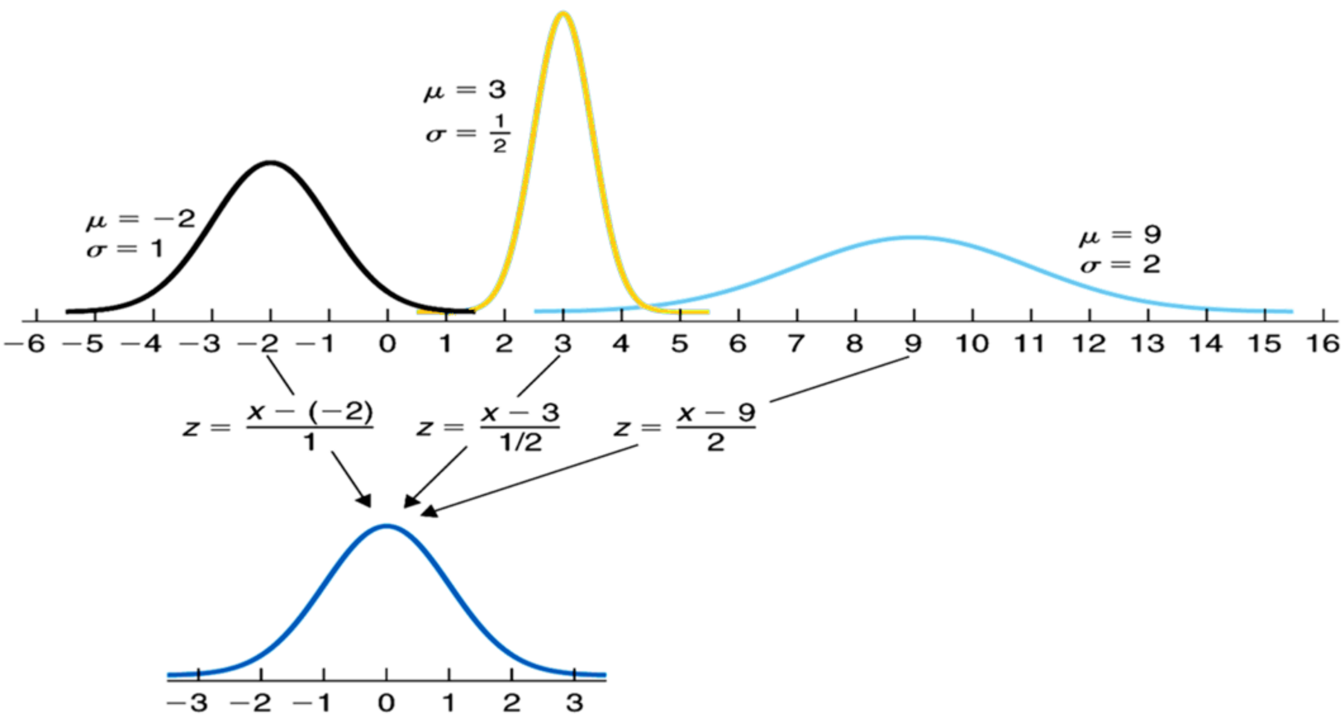
  - If you made an infinite number of measurements, their mean would be μ and their standard deviation would be σ
  - In practice, you have a finite number of measurements with mean x and standard deviation s

- For now, μ and σ will be given
  - Later we'll use x and s to estimate μ and σ

The standard normal distribution

- It is tedious to integrate a new normal distribution for every single measurement, so use a "standard normal distribution" with tabulated areas.

- Convert your measurement x to a standard score



The standard normal distribution

- A random variable X has a normal (or Gaussian) distribution with mean μ and standard deviation σ, if and only if

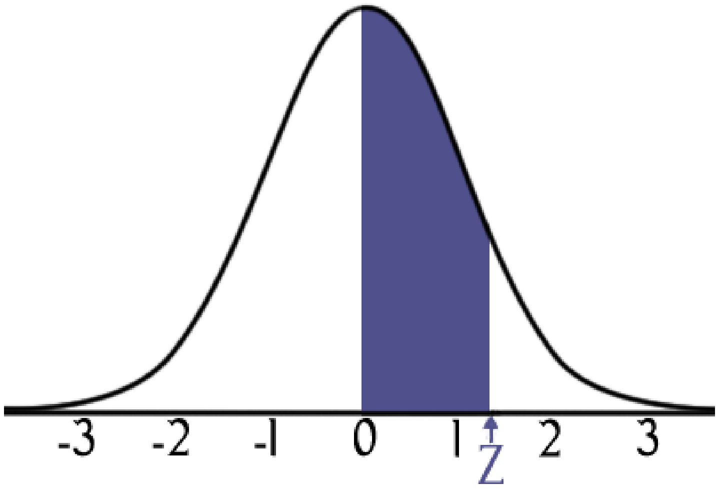
$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

- we write X ~ N(μ, σ^2) and say that X has a normal distribution

- For N(0,1), μ = 0 and σ = 1

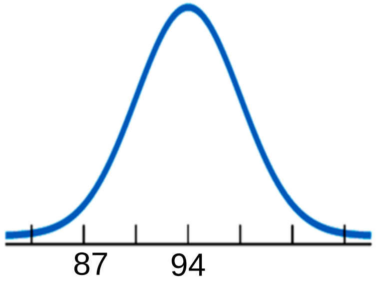
How to use the normal distribution?

- Use the area UNDER the normal distribution
- For example, the area under the curve between x=a and x=b is the probability that your next measurement of x will fall between a and b



Example

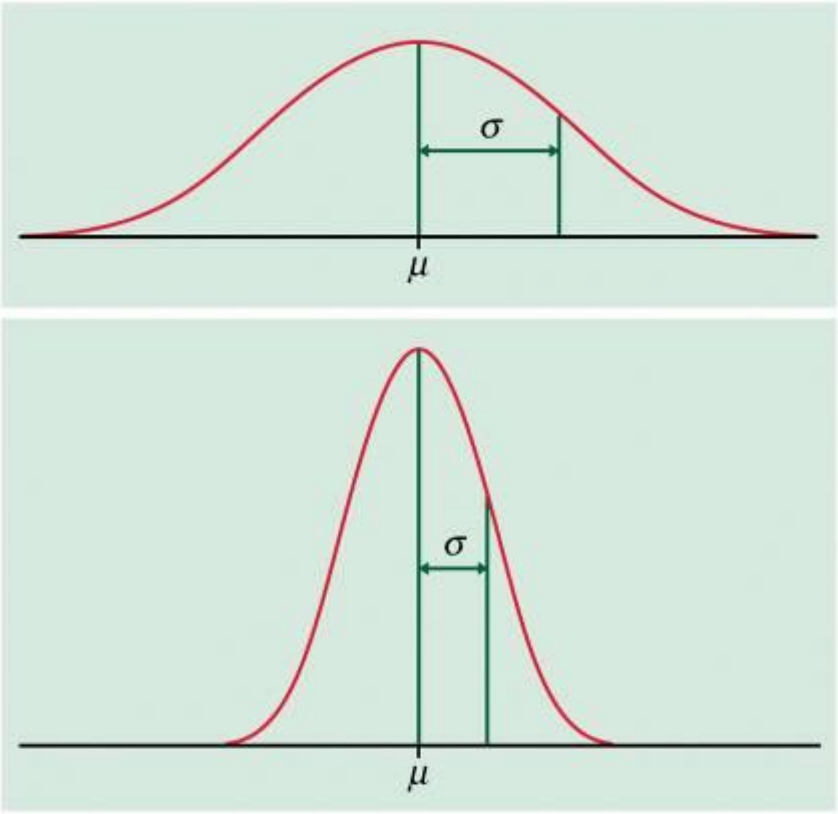
- Historical data shows that the temperature of a particular pipe in a normally-operating continuous production line is (94 $\pm$ 5)°C ($\pm 1\sigma$). You glance at the control display and see that T = 87 °C. How abnormal is this measurement?



- z = (87 – 94)/5 = -1.4
- From the table of z distribution, -1.4 gives an area of 0.0808.
- In other words, when the line is operating normally, you would expect to see even lower temperatures about 8 % of the time.
- This measurement alone should not worry you.

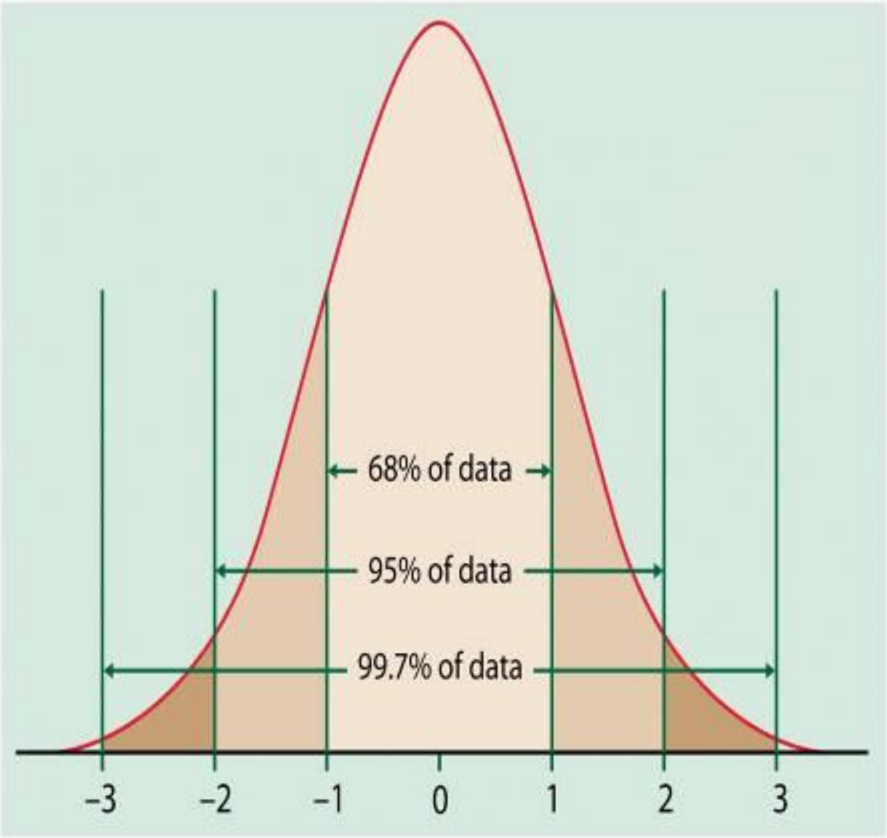
Characteristics

- A normal distribution is bell-shaped and symmetric.
- The distribution is determined by the mean mu (μ), and the standard deviation sigma (σ).
- The mean mu controls the center and sigma controls the spread.



68.26 -95.44-99.74 Rule

- For any normal curve with mean mu and standard deviation sigma:
- 68 percent of the observations fall within one standard deviation sigma of the mean.
- 95 percent of observation fall within 2 standard deviations.
- 99.7 percent of observations fall within 3 standard deviations of the mean.



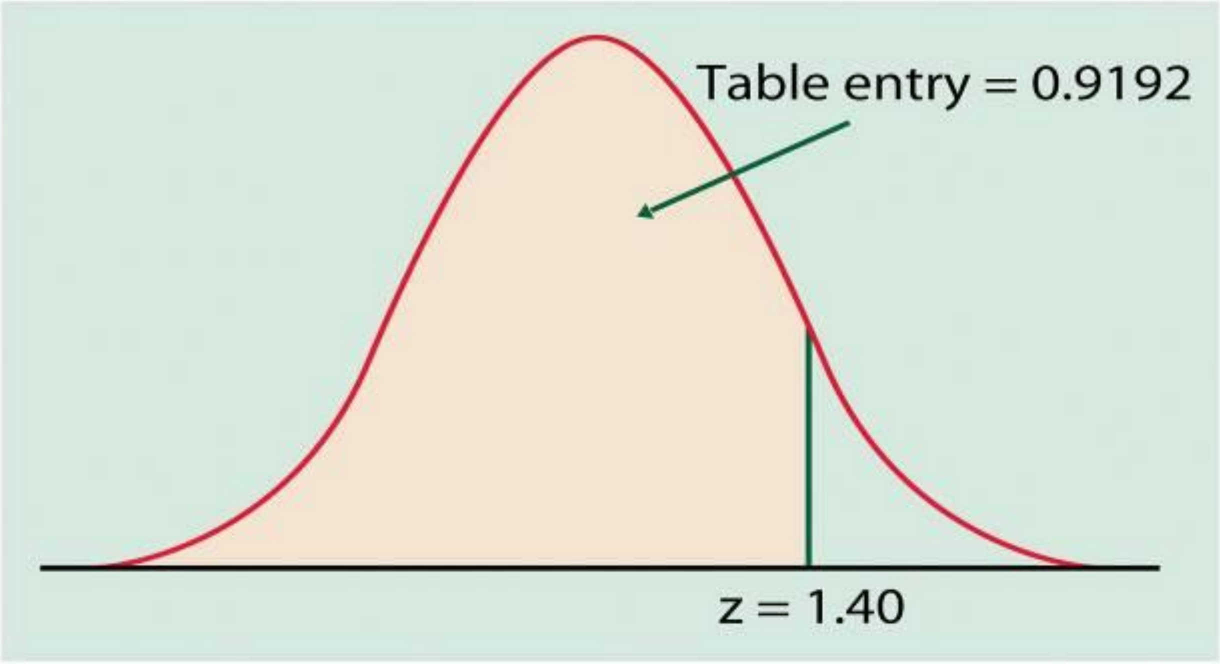
Example questions

- If mu=30 and sigma=4, what are the values (a, b) around 30 such that 95 percent of the observations fall between these values?

- If mu=40 and sigma=5, what are the bounds (a, b) such that 99.7 percent of the values fall between these values?

Normal table usage

- What proportion of standard normal distribution values Z are less than 1.40? That is, P(Z < 1.40) = ?
- 0.9192 `pnorm(1.4)`



Example

- Jim gets 680 on SAT math exam. Mean on this exam is 500 and SD is 100.
- Jim's standardized score is:

$$Z = \frac{680 - 500}{100} = \frac{180}{100} = 1.80$$

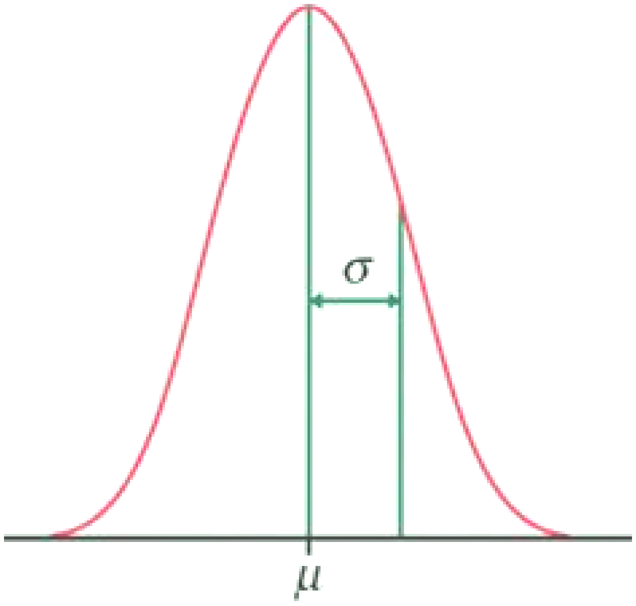
Example continued

- Ben got 27 on ACT math. Mean is 18 with SD of 6.
- Ben's Z-Score is:

$$Z = \frac{27 - 18}{6} = \frac{9}{6} = 1.50$$

- Jim did better !

The standard normal curve

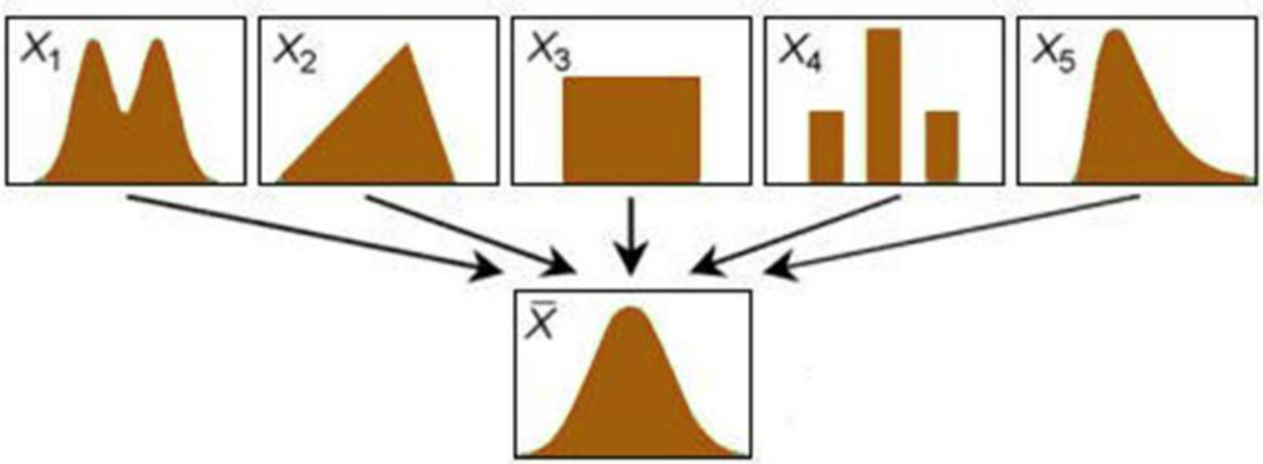


- Skewness = 0
- Kurtosis = 3

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

Central limit theorem

- Sampling distribution of means becomes normal as N increases, regardless of shape of original distribution.

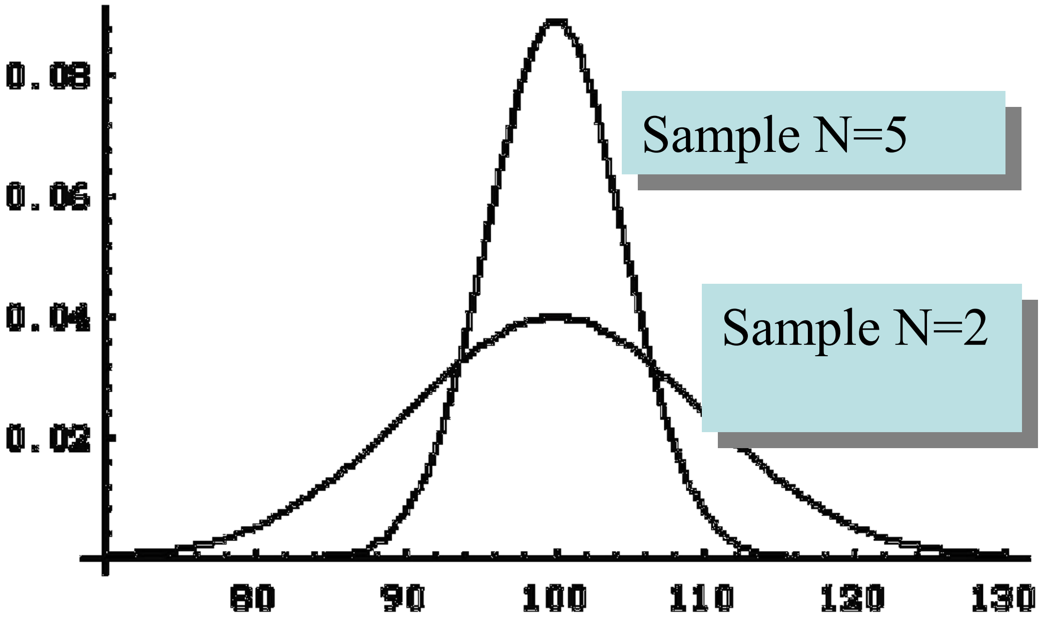


Central limit theorem - lots of means

- This distribution of survey results would follow a normal distribution
  - Mean = μ
  - standard deviation = $\sigma/(N)^{1/2}$

Standard deviation of the sampling distribution

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$



Underlying Population
Mean = 100
SD = 10

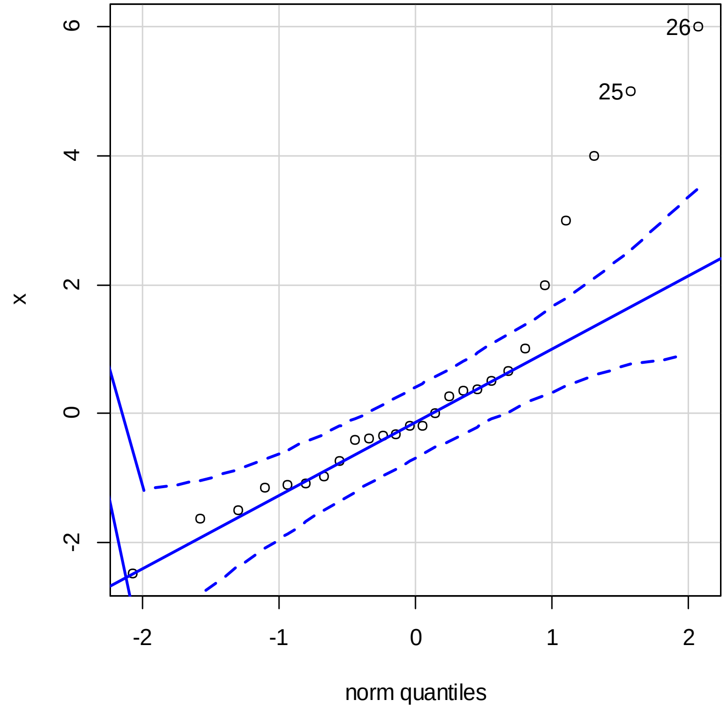
This is a distribution of means of samples of size N

How to check the normality of a vector

```
x = c(rnorm(20), 1:6)
shapiro.test(x)
```

Shapiro-Wilk normality test
data: x
W = 0.85, p-value = 0.002

```
library(car)
qqPlot(x)
```



Chi-square distribution

$$z = \frac{(X - \overline{X})}{SD};\ z = \frac{(X - \mu)}{\sigma}$$ z score

$$z^2 = \frac{(y - \mu)^2}{\sigma^2}$$ z score squared

$$z^2 = \chi^2_{(1)}$$ Make it Greek

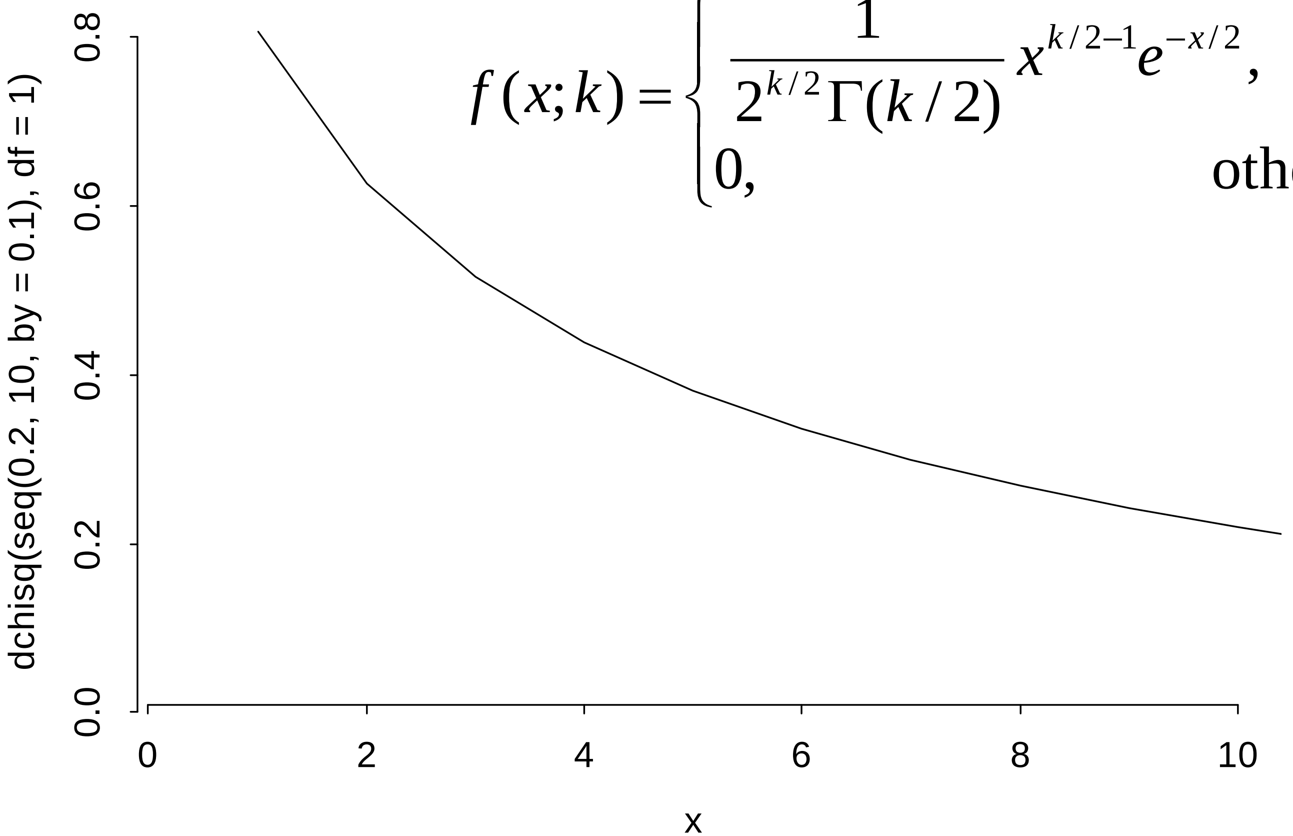
What would its sampling distribution look like?

Minimum value is zero.

Maximum value is infinite.

Most values are between zero and 1;
most around zero.

PDF of Chi-square distribution



$$f(x;k)=\begin{cases}\dfrac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}, & x\geq 0;\\ 0, & \text{otherwise.}\end{cases}$$

```
plot(dchisq(seq(0.1, 10, by=0.1), df = 1), type='l',
     xlab = 'x', xlim = c(0,10))
```

Chi-square distribution is additive

What if we took 2 values of z^2 at random and added them?

$$z_1^2 = \frac{(y_1-\mu)^2}{\sigma^2};\ z_2^2 = \frac{(y_2-\mu)^2}{\sigma^2} \qquad \chi^2_{(2)} = \frac{(y_1-\mu)^2}{\sigma^2} + \frac{(y_2-\mu)^2}{\sigma^2} = z_1^2 + z_2^2$$

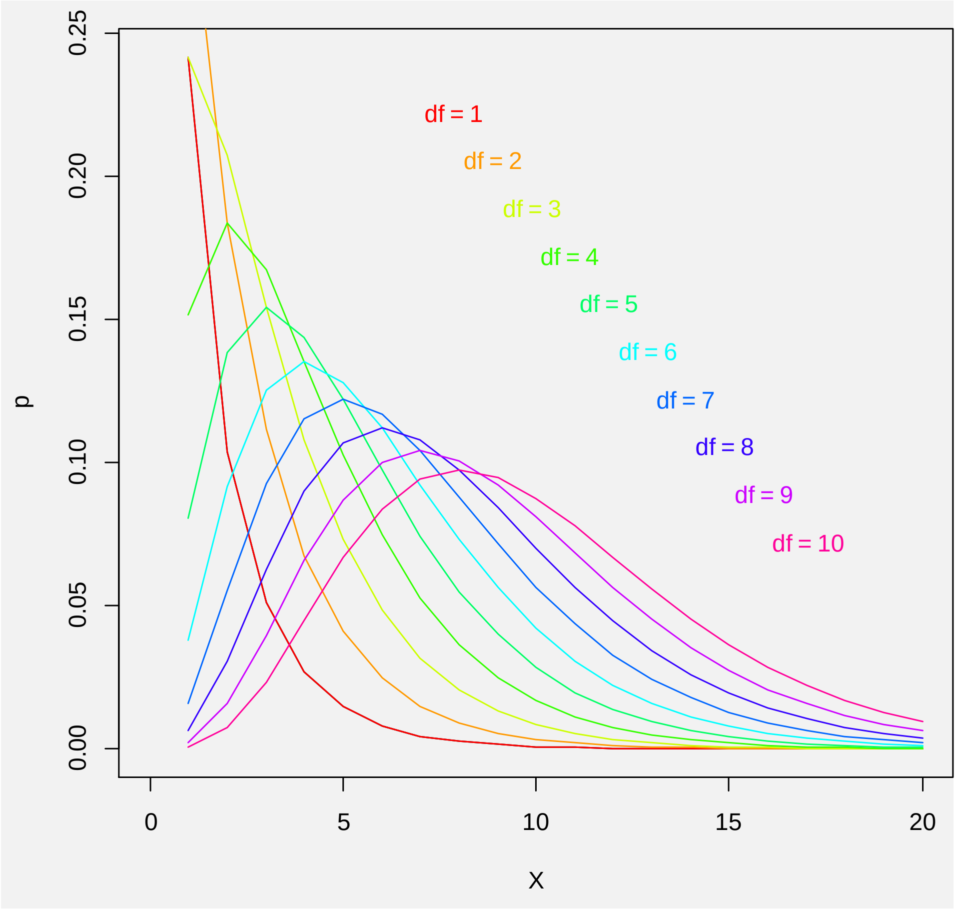
Same minimum and maximum as before, but now average should be a bit bigger.

Chi-square is the distribution of a sum of squares. Each squared value (usually deviation) is taken from the unit normal: N(0,1). The shape of the chi-square distribution depends on the number of squared values that are added together.

$$\chi^2_{(v_1+v_2)} = \chi^2_{(v_1)} + \chi^2_{(v_2)}$$

Chi-square distribution (one parameter: df)

The distribution of chi-square depends on 1 parameter, its degrees of freedom (*df* or *v*). As *df* gets large, curve is less positively skewed, more normal.



Chi-square distribution (mean and variance)

- The expected value of chi-square is *df*.
  - The mean of the chi-square distribution is its degrees of freedom.
- The expected variance of the distribution is 2*df*.
  - If the variance is 2*df*, the standard deviation must be sqrt(2*df*).

A case of Chi-square distribution: sample variance

$$s^2 = \frac{\sum (y - \bar{y})^2}{N - 1}$$

Sample estimate of population variance (unbiased).

$$\chi^2_{(N-1)} = \frac{(N-1)s^2}{\sigma^2}$$

Multiply variance estimate by N-1 to get sum of squares. Divide by population variance to normalize. Result is a random variable distributed as chi-square with (N-1) *df*.

We can use info about the sampling distribution of the variance estimate to find confidence intervals and conduct statistical tests.

F distribution

- The F distribution is the ratio of two variance estimates of normal distributions:

$$F = \frac{s_1^2}{s_2^2} = \frac{est.\sigma_1^2}{est.\sigma_2^2}$$

- Also the ratio of two chi-squares, each divided by its degrees of freedom:

$$F = \frac{\chi^2_{(v_1)} / v_1}{\chi^2_{(v_2}) / v_2}$$

In most applications, v_2 will be larger than v_1, and v_2 will be larger than 2. In such a case, the mean of the F distribution (expected value) is $v_2 /(v_2 -2)$.

F distribution: parameters and PDF

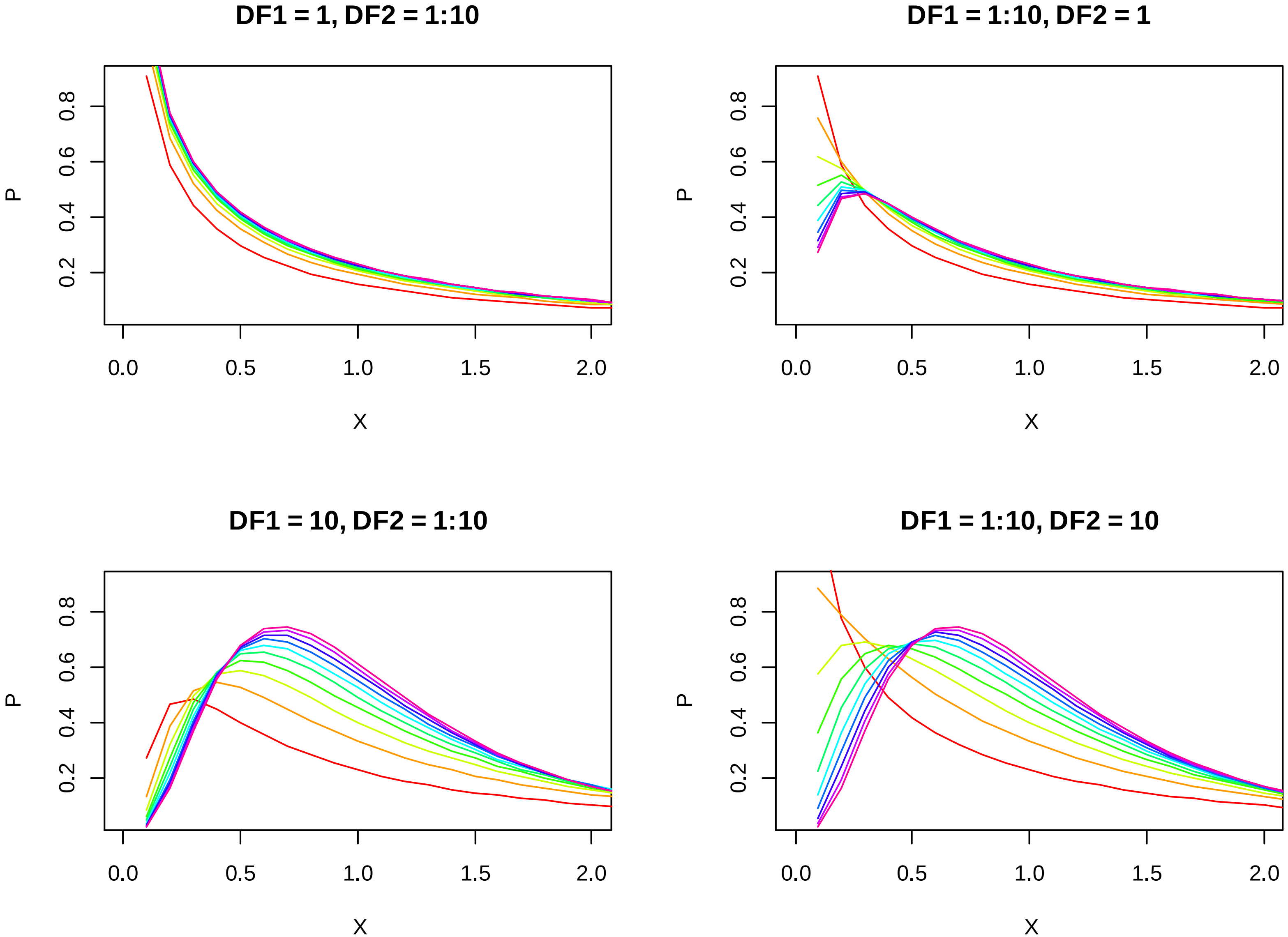
- *F* depends on two parameters: v_1 and v_2 (df_1 and df_2). The shape of *F* changes with these. Range is 0 to infinity. Shaped a bit like chi-square.
- *F* tables show critical values for *df* in the numerator and *df* in the denominator.
- *F* tables are 1-tailed; can figure 2-tailed if you need to (but you usually don't).

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} {d_2}^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x B(\frac{d_1}{2}, \frac{d_2}{2})} = \frac{1}{B(\frac{d_1}{2}, \frac{d_2}{2})} (\frac{d_1}{d_2})^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} (1 + \frac{d_1}{d_2} x)^{-\frac{d_1+d_2}{2}}$$

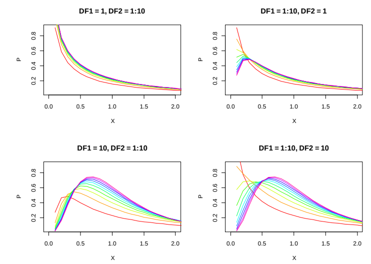
F distribution: mean, mode, range, SD, CV

Mean	$\dfrac{\nu_2}{(\nu_2 - 2)}$	$\nu_2 > 2$
Mode	$\dfrac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}$	$\nu_1 > 2$
Range	0 to positive infinity	
Standard Deviation	$\sqrt{\dfrac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}}$	$\nu_2 > 4$
Coefficient of Variation	$\sqrt{\dfrac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)}}$	$\nu_2 > 4$
Skewness	$\dfrac{(2\nu_1 + \nu_2 - 2)\sqrt{8(\nu_2 - 4)}}{\sqrt{\nu_1}(\nu_2 - 6)\sqrt{(\nu_1 + \nu_2 - 2)}}$	$\nu_2 > 6$

F distribution PDF shapes



R code for plotting *F* distribution




```
X = seq(0.1, 3, length=30); Y = df(X, 1,1)
par(mfrow=c(2,2))

plot(X, Y, type='n', xlab = 'X', ylab = 'P', xlim=c(0,2), main="DF1=1, DF2=1:10")
for (i in 1:10)  lines(X, df(X, 1, i), col=rainbow(10)[i])

plot(X, Y, type='n', xlab = 'X', ylab = 'P', xlim=c(0,2), main="DF1=1:10, DF2=1")
for (i in 1:10) lines(X, df(X, i, 1), col=rainbow(10)[i])

plot(X, Y, type='n', xlab = 'X', ylab = 'P', xlim=c(0,2), main="DF1=10, DF2=1:10")
for (i in 1:10) lines(X, df(X, 10, i), col=rainbow(10)[i])

plot(X, Y, type='n', xlab = 'X', ylab = 'P', xlim=c(0,2), main="DF1=1:10, DF2=10")
for (i in 1:10)  lines(X, df(X, i, 10), col=rainbow(10)[i])
```

Exponential distribution

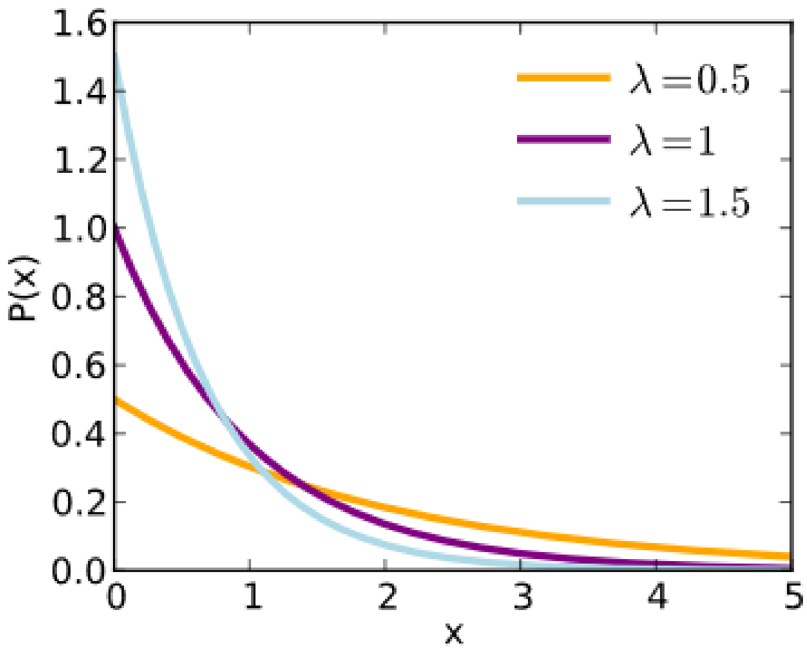
A with one parameter often used for describing the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

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

Mean: $E(x) = 1/\lambda$

Variance: $var(x) = 1/(\lambda^2)$

Mode: 0



Gamma distribution

A continuous probability distribution with two parameters.

There are three different parametrizations in common use:

- 1. With a shape parameter k and a scale parameter θ (common in econometrics).
- 2. With a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called a rate parameter (common in Bayesian statistics).
- 3. With a shape parameter k and a mean parameter $\mu = k/\beta$.

In each of these three forms, both parameters are positive real numbers.

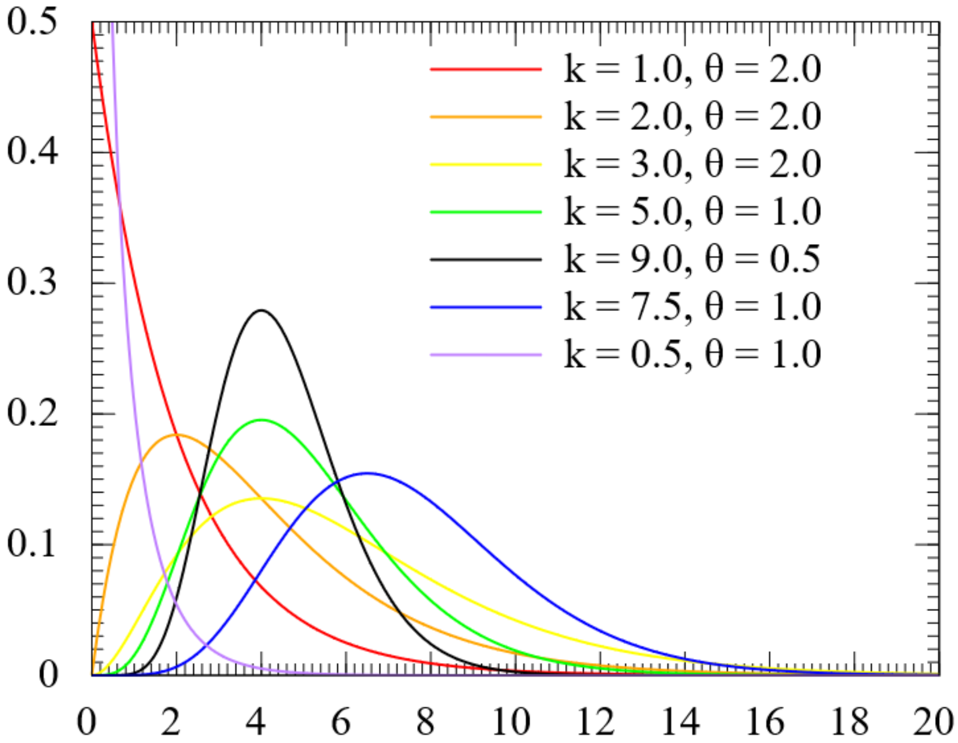
The exponential distribution and chi-squared distribution are special cases of the gamma distribution.

$$\Gamma(\alpha = 1, \beta = 1/\lambda) = \exp(\lambda)$$

$$\Gamma(\alpha = n/2, \beta = 1/2) = X^2(n)$$

Gamma distribution: PDF and parameters

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$



Mean: E(x) = α/β

Variance: var(x) = $\alpha/(\beta^2)$

Mode: $(\alpha-1)/\beta$

Applications

- The gamma distribution has been used to model the size of insurance claims and rainfalls. This means that aggregate insurance claims and the amount of rainfall accumulated in a reservoir are modelled by a gamma process. The gamma distribution is also used to model errors in multi-level Poisson regression models, because the combination of the Poisson distribution and a gamma distribution is a negative binomial distribution.

- In wireless communication, the gamma distribution is used to model the multi-path fading of signal power.

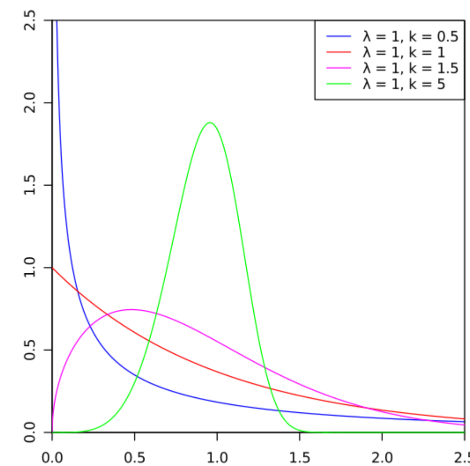
- In neuroscience, the gamma distribution is often used to describe the distribution of inter-spike intervals.
- In bacterial gene expression, the copy number of a constitutively expressed protein often follows the gamma distribution, where the scale and shape parameter are, respectively, the mean number of bursts per cell cycle and the mean number of protein molecules produced by a single mRNA during its lifetime.

- The gamma distribution is widely used as a conjugate prior in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution.

Weibull distribution

A continuous probability distribution with two parameters, named after Swedish mathematician Waloddi Weibull (1951). It was first identified by Fréchet (1927) and first applied by Rosin & Rammler (1933) to describe a particle size distribution.

$$f(x;\lambda,\kappa)=\begin{cases}\frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1}e^{-(x/\lambda)^{\kappa}}, & x\geq 0\\ 0, & x<0\end{cases}$$



where k > 0 is the shape parameter and λ > 0 is the scale parameter.

Fréchet, Maurice (1927), "Sur la loi de probabilité de l'écart maximum", Annales de la Société Polonaise de Mathematique, Cracovie 6: 93–116.
Rosin, P.; Rammler, E. (1933), "The Laws Governing the Fineness of Powdered Coal", Journal of the Institute of Fuel 7: 29–36.
Weibull, W. (1951), "A statistical distribution function of wide applicability" (PDF), J. Appl. Mech.-Trans. ASME 18 (3): 293–297.

Gamma distribution vs. Weibull distribution

- Both the Gamma and Weibull distributions can be seen as generalisations of the exponential distribution.

$$f_{\Gamma} = \frac{\beta^{\kappa}}{\Gamma(\kappa)} x^{\kappa-1} e^{-\frac{x}{\theta}} \qquad f_W = \frac{\kappa}{\lambda^{\kappa}} x^{\kappa-1} e^{-(\frac{x}{\lambda})^{\kappa}}$$

- The pdf for the Weibull distribution drops off much more quickly (for k>1) or slowly (for k<1) than the Gamma distribution. In the case where k=1, they both reduce to the exponential distribution.

- If we look at the exponential distribution as describing the waiting time of a Poisson process (the time we have to wait until an event happens, if that event is equally likely to occur in any time interval), then the $\Gamma(k, \theta)$ distribution describes the time we have to wait for k independent events to occur; the Weibull distribution describes the time we have to wait for one event to occur, if that event becomes more or less likely with time. Here the k parameter describes how quickly the probability ramps up.

R script

```
# Normal distribution
dnorm(seq( -3, 3,     by = 0.1), mean = 0, sd = 1)    'dnorm' gives the pdf
pnorm(seq( -3, 3,     by = 0.1), mean = 0, sd = 1)    'pnorm' gives the cdf
qnorm(seq( 0.1, 0.9, by = 0.1), mean = 0, sd = 1)    'qnorm' gives the quantile function
rnorm(10, mean = 0, sd = 1)                           'rnorm‘ generates random deviates
```

```
dbinom(x, N, p) # Binomial distribution
pbinom(x, N, p)
qbinom(q, N, p)
rbinom(n, N, p)
```

```
# F distribution
df(x, df1, df2, log = FALSE)
pf(q, df1, df2, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2)
```

```
rchisq(100, df = 3) # chi square distribution
runif(x, min, max) # Uniform distribution
rnbinom(10, mu=3, theta=2) # Negative binomial distribution
rpois(x, lambda) # Poisson distribution
rgamma(100, shape = 3, scale = 2) # Gamma distribution
rweibull(100, shape = 3, scale = 2) # Weibull distribution
```

Assignment

General objectives: Fit your data with a normal distribution, plot the 95% confident interval.

You have a number of specific objectives in this assignment. I will provide you with the programs that you will use.

For the report, you will provide a brief introduction to the data set, and print out the data at the section **Data**, print programs (R code) in the section **Method**. List model output in your **Results** section, and address further important points in the **Discussion** section.

R codes for doing your assignment

R script

```
obs = c(7,5,6,6,7,5,3,4,5,8,2,4,5,6,7,6,4,5,9,3,6,4)

hist(obs, freq = F)
mean = mean(obs)
SD    = sd(obs)

x = seq(min(obs), max(obs), by = .1)
norm = dnorm(x, mean = mean, sd = SD)
lines(x, norm, type = 'l', col = 'red')
abline(v=c(qnorm(.025, mean, SD), qnorm(.975, mean, SD)), col = 'brown')
abline(v=c(mean - 2*SD, mean + 2*SD), col = ‘blue')
```