Statistics 1B

Lecture 1. Introduction and probability review

What is “Statistics”?

There are many definitions: I will use

"A set of principles and procedures for gaining and processing quantitative evidence in order to help us make judgements and decisions"

It can include

- Design of experiments and studies
- Exploring data using graphics
- Informal interpretation of data
- Formal statistical analysis
- Clear communication of conclusions and uncertainty

It is NOT just data analysis!

In this course we shall focus on formal *statistical inference*: we assume

- we have data generated from some unknown probability model
- we aim to use the data to learn about certain properties of the underlying probability model

Idea of parametric inference

- Let X be a random variable (r.v.) taking values in $\mathcal{X}$
- Assume distribution of X belongs to a family of distributions indexed by a scalar or vector parameter θ, taking values in some parameter space Θ
- Call this a *parametric family:*

For example, we could have

- $X \sim \text{Poisson}(\mu)$, $\theta = \mu \in \Theta = (0, \infty)$
- $X \sim \text{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$.

BIG ASSUMPTION

For some results (bias, mean squared error, linear model) we do not need to specify the precise parametric family.

But generally we assume that we know which family of distributions is involved, but that the value of θ is unknown.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (iid) with the same distribution as X, so that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a simple random sample (our data).

We use the observed $\mathbf{X} = \mathbf{x}$ to make inferences about θ, such as,

(a) giving an estimate $\hat{\theta}(\mathbf{x})$ of the true value of θ (point estimation);

(b) giving an interval estimate $(\hat{\theta}_1(\mathbf{x}), (\hat{\theta}_2(\mathbf{x}))$ for θ;

(c) testing a hypothesis about θ, eg testing the hypothesis $H : \theta = 0$ means determining whether or not the data provide evidence against H.

We shall be dealing with these aspects of *statistical inference*.

Other tasks (not covered in this course) include

- Checking and selecting probability models
- Producing predictive distributions for future random variables
- Classifying units into pre-determined groups ('supervised learning')
- Finding clusters ('unsupervised learning')

Statistical inference is needed to answer questions such as:

- What are the voting intentions before an election? *[Market research, opinion polls, surveys]*
- What is the effect of obesity on life expectancy? *[Epidemiology]*
- What is the average benefit of a new cancer therapy? *Clinical trials*
- What proportion of temperature change is due to man? *Environmental statistics*
- What is the benefit of speed cameras? *Traffic studies*
- What portfolio maximises expected return? *Financial and actuarial applications*
- How confident are we the Higgs Boson exists? *Science*
- What are possible benefits and harms of genetically-modified plants? *Agricultural experiments*
- What proportion of the UK economy involves prostitution and illegal drugs? *Official statistics*
- What is the chance Liverpool will best Arsenal next week? *Sport*

Probability review

Let Ω be the *sample space* of all possible outcomes of an experiment or some other data-gathering process.

E.g when flipping two coins, $\Omega = \{HH, HT, TH, TT\}$.

'Nice' (measurable) subsets of Ω are called *events*, and $\mathcal{F}$ is the set of all events - when Ω is countable, $\mathcal{F}$ is just the power set (set of all subsets) of Ω.

A function $\mathbb{P} : \mathcal{F} \to [0,1]$ called a *probability measure* satisfies

- $\mathbb{P}(\phi) = 0$
- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$, whenever $\{A_n\}$ is a disjoint sequence of events.

A *random variable* is a (measurable) function $X : \Omega \to \mathbb{R}$.

Thus for the two coins, we might set

$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0,$

so X is simply the number of heads.

Our data are modelled by a vector $\mathbf{X} = (X_1, \dots, X_n)$ of random variables – each observation is a random variable.

The *distribution function* of a r.v. X is $F_X(x) = \mathbb{P}(X \le x)$, for all $x \in \mathbb{R}$. So F_X is

- non-decreasing,
- $0 \le F_X(x) \le 1$ for all x,
- $F_X(x) \to 1$ as $x \to \infty$,
- $F_X(x) \to 0$ as $x \to -\infty$.

A *discrete* random variable takes values only in some countable (or finite) set $\mathcal{X}$, and has a *probability mass function* (pmf) $f_X(x) = \mathbb{P}(X = x)$.

- $f_X(x)$ is zero unless x is in $\mathcal{X}$.
- $f_X(x) \ge 0$ for all x,
- $\sum_{x \in \mathcal{X}} f_X(x) = 1$
- $\mathbb{P}(X \in A) = \sum_{x \in A} f_X(x)$ for a set A.

We say X has a continuous (or, more precisely, absolutely continuous) distribution if it has a *probability density function* (pdf) f_X such that

- $\mathbb{P}(X \in A) = \int_A f_X(t)dt$ for "nice" sets A.

Thus

- $\int_{-\infty}^{\infty} f_X(t)dt = 1$
- $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$

[Notation note: There will be inconsistent use of a subscript in mass, density and distributions functions to denote the r.v. Also f will sometimes be p.]

Expectation and variance

If X is discrete, the *expectation* of X is

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x\mathbb{P}(X = x)$$

(exists when $\sum |x|\mathbb{P}(X = x) < \infty$).

If X is continuous, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

(exists when $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$).

$\mathbb{E}(X)$ is also called the expected value or the mean of X.

If $g : \mathbb{R} \to \mathbb{R}$ then

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x \in \mathcal{X}} g(x)\mathbb{P}(X = x) & \text{if } X \text{ is discrete} \\ \int g(x) f_X(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

The *variance* of X is $var(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$.

Independence

The random variables $X_1, \dots, X_n$ are *independent* if for all $x_1, \dots, x_n$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n).$$

If the independent random variables $X_1, \dots, X_n$ have pdf's or pmf's $f_{X_1}, \dots, f_{X_n}$, then the random vector $\mathbf{X} = (X_1, \dots, X_n)$ has pdf or pmf

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_i f_{X_i}(x_i).$$

Random variables that are independent and that all have the same distribution (and hence the same mean and variance) are called *independent and identically distributed (iid) random variables*.

Maxima of iid random variables

Let $X_1, \ldots, X_n$ be iid r.v.'s, and $Y = \max(X_1, \ldots, X_n)$.

Then

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\max(X_1, \ldots, X_n) \leq y)$$
$$= \mathbb{P}(X_1 \leq y, \ldots, X_n \leq y) = \mathbb{P}(X_i \leq y)^n = [F_X(y)]^n$$

The density for Y can then be obtained by differentiation (if continuous), or differencing (if discrete).

Can do similar analysis for minima of iid r.v.'s.

Sums and linear transformations of random variables

For any random variables,

$$
\begin{aligned}
\mathbb{E}(X_1 + \cdots + X_n) &= \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) \\
\mathbb{E}(a_1 X_1 + b_1) &= a_1 \mathbb{E}(X_1) + b_1 \\
\mathbb{E}(a_1 X_1 + \cdots + a_n X_n) &= a_1 \mathbb{E}(X_1) + \cdots + a_n \mathbb{E}(X_n) \\
\mathrm{var}(a_1 X_1 + b_1) &= a_1^2 \mathrm{var}(X_1)
\end{aligned}
$$

For independent random variables,

$$\mathbb{E}(X_1 \times \ldots \times X_n) = \mathbb{E}(X_1) \times \ldots \times \mathbb{E}(X_n),$$

$$\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n),$$

and

$$\mathrm{var}(a_1 X_1 + \cdots + a_n X_n) = a_1^2 \mathrm{var}(X_1) + \cdots + a_n^2 \mathrm{var}(X_n).$$

Standardised statistics

Suppose $X_1, \ldots, X_n$ are iid with $\mathbb{E}(X_1) = \mu$ and $\mathrm{var}(X_1) = \sigma^2$.
Write their sum as

$$S_n = \sum_{i=1}^{n} X_i$$

From preceding slide, $\mathbb{E}(S_n) = n\mu$ and $\mathrm{var}(S_n) = n\sigma^2$.
Let $\bar{X}_n = S_n/n$ be the sample mean.
Then $\mathbb{E}(\bar{X}_n) = \mu$ and $\mathrm{var}(\bar{X}_n) = \sigma^2/n$.
Let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

Then $\mathbb{E}(Z_n) = 0$ and $\mathrm{var}(Z_n) = 1$.
Z_n is known as a *standardised statistic*.

Moment generating functions

The *moment generating function* for a r.v. X is

$$M_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \sum_{x \in \mathcal{X}} e^{tx} \mathbb{P}(X = x) & \text{if } X \text{ is discrete} \\ \int e^{tx} f_X(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

provided M exists for t in a neighbourhood of 0.

Can use this to obtain moments of X, since

$$\mathbb{E}(X^n) = M_X^{(n)}(0),$$

i.e. nth derivative of M evaluated at $t = 0$.

Under broad conditions, $M_X(t) = M_Y(t)$ implies $F_X = F_Y$.

Mgf's are useful for proving distributions of sums of r.v.'s since, if $X_1, ..., X_n$ are iid, $M_{S_n}(t) = M_X^n(t)$.

Example: sum of Poissons

If $X_i \sim \text{Poisson}(\mu)$, then

$$M_{X_i}(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} e^{-\mu} \mu^x / x! = e^{-\mu(1-e^t)} \sum_{x=0}^{\infty} e^{-\mu e^t} (\mu e^t)^x / x! = e^{-\mu(1-e^t)}.$$

And so $M_{S_n}(t) = e^{-n\mu(1-e^t)}$, which we immediately recognise as the mgf of a Poisson($n\mu$) distribution.

So sum of iid Poissons is Poisson. □

Convergence

The Weak Law of Large Numbers (WLLN) states that for all $\epsilon > 0$,

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \to 0 \text{ as } n \to \infty.$$

The Strong Law of Large Numbers (SLLN) says that

$$\mathbb{P}\left(\bar{X}_n \to \mu\right) = 1.$$

The Central Limit Theorem tells us that

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \text{ is approximately N}(0,1) \text{ for large } n \text{ .}$$

Conditioning

Let X and Y be discrete random variables with joint pmf

$$p_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y).$$

Then the *marginal pmf* of Y is

$$p_Y(y)=\mathbb{P}(Y=y)=\sum_x p_{X,Y}(x,y).$$

The *conditional pmf* of X given $Y=y$ is

$$p_{X|Y}(x\mid y)=\mathbb{P}(X=x\mid Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\frac{p_{X,Y}(x,y)}{p_Y(y)},$$

if $p_Y(y)\neq 0$ (and is defined to be zero if $p_Y(y)=0$)).

Conditioning

In the continuous case, suppose that X and Y have joint pdf $f_{X,Y}(x,y)$, so that for example

$$\mathbb{P}(X \leq x_1, Y \leq y_1) = \int_{-\infty}^{y_1} \int_{-\infty}^{x_1} f_{X,Y}(x,y) dx dy.$$

Then the *marginal pdf* of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx.$$

The *conditional pdf* of X given $Y = y$ is

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

if $f_Y(y) \neq 0$ (and is defined to be zero if $f_Y(y) = 0$).

The conditional expectation of X given $Y = y$ is

$$\mathbb{E}(X \mid Y = y) = \begin{cases} \sum x f_{X|Y}(x \mid y) & \text{pmf} \\ \int x f_{X|Y}(x \mid y) dx & \text{pdf.} \end{cases}$$

Thus $\mathbb{E}(X \mid Y = y)$ is a function of y, and $\mathbb{E}(X \mid Y)$ is a function of Y and hence a r.v..

The conditional expectation formula says

$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}(X \mid Y)\right].$$

Proof [discrete case]:

$$\mathbb{E}\left[\mathbb{E}(X \mid Y)\right] = \sum_{\mathcal{Y}} \left[\sum_{\mathcal{X}} x \ f_{X|Y}(x \mid y)\right] f_Y(y) = \sum_{\mathcal{X}} \sum_{\mathcal{Y}} x \ f_{X,Y}(x, y)$$

$$= \sum_{\mathcal{X}} x \left[\sum_{\mathcal{Y}} f_{Y|X}(y \mid x)\right] f_X(x) = \sum_{\mathcal{X}} x \ f_X(x). \square$$

The conditional variance of X given $Y = y$ is defined by

$$\mathrm{var}(X \mid Y=y) = \mathbb{E}\Big[\big(X - \mathbb{E}(X \mid Y=y)\big)^2 \mid Y = y\Big],$$

and this is equal to $\mathbb{E}(X^2 \mid Y=y) - \big(\mathbb{E}(X \mid Y=y)\big)^2$.

We also have the conditional variance formula:

$$\mathrm{var}(X) = \mathbb{E}[\mathrm{var}(X \mid Y)] + \mathrm{var}[\mathbb{E}(X \mid Y)].$$

Proof:

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \mathbb{E}\big[\mathbb{E}(X^2 \mid Y)\big] - \Big[\mathbb{E}\big[\mathbb{E}(X \mid Y)\big]\Big]^2 \\
&= \mathbb{E}\Big[\mathbb{E}(X^2 \mid Y) - \big[\mathbb{E}(X \mid Y)\big]^2\Big] + \mathbb{E}\Big[\big[\mathbb{E}(X \mid Y)\big]^2\Big] - \Big[\mathbb{E}\big[\mathbb{E}(X \mid Y)\big]\Big]^2 \\
&= \mathbb{E}\big[\mathrm{var}(X \mid Y)\big] + \mathrm{var}\big[\mathbb{E}(X \mid Y)\big].
\end{aligned}$$

Change of variable (illustrated in 2-d)

Let the joint density of random variables (X, Y) be $f_{X,Y}(x, y)$.

Consider a 1-1 (bijective) differentiable transformation to random variables $(U(X, Y), V(X, Y))$, with inverse $(X(U, V), Y(U, V))$.

Then the joint density of (U, V) is given by

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))|J|,$$

where J is the Jacobian

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Some important discrete distributions: Binomial

X has a **binomial** distribution with parameters n and p ($n \in \mathbb{N}$, $0 \leq p \leq 1$), $X \sim \mathrm{Bin}(n, p)$, if

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x \in \{0, 1, \ldots, n\}$$

(zero otherwise).

We have $\mathbb{E}(X) = np$, $\mathrm{var}(X) = np(1-p)$.

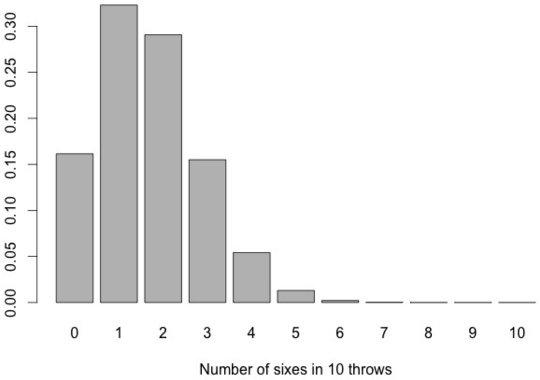
This is the distribution of the number of successes out of n independent Bernoulli trials, each of which has success probability p.

Example: throwing dice

let X = number of sixes when throw 10 fair dice, so $X \sim \text{Bin}(10, \frac{1}{6})$

R code:

```
barplot(  dbinom(0:10, 10, 1/6), names.arg=0:10,
               xlab="Number of sixes in 10 throws" )
```



Some important discrete distributions: Poisson

X has a **Poisson** distribution with parameter μ ($\mu > 0$), $X \sim \text{Poisson}(\mu)$, if

$$\mathbb{P}(X = x) = e^{-\mu}\mu^x/x!, \text{ for } x \in \{0, 1, 2, \ldots\},$$

(zero otherwise).

Then $\mathbb{E}(X) = \mu$ and $\text{var}(X) = \mu$.

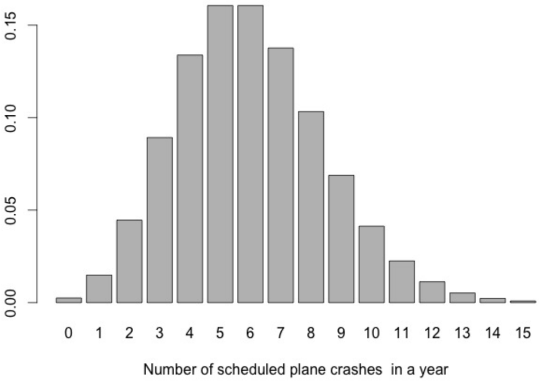
In a *Poisson process* the number of events $X(t)$ in an interval of length t is Poisson(μt), where μ is the rate per unit time.

The Poisson(μ) is the limit of the Bin(n,p) distribution as $n \to \infty$, $p \to 0$, $\mu = np$.

Example: plane crashes. Assume scheduled plane crashes occur as a Poisson process with a rate of 1 every 2 months. How many (X) will occur in a year (12 months)?

Number in two months is Poisson(1), and so $X \sim \text{Poisson}(6)$.

```
barplot(  dpois(0:15, 6), names.arg=0:15,
               xlab="Number of scheduled plane crashes  in a year" )
```



Some important discrete distributions: Negative Binomial

X has a **negative binomial** distribution with parameters k and p ($k \in \mathbb{N}$, $0 \leq p \leq 1$), if

$$\mathbb{P}(X = x) = \binom{x-1}{k-1}(1-p)^{x-k}p^k, \text{ for } x = k, k+1, \ldots,$$

(zero otherwise). Then $\mathbb{E}(X) = k/p$, $\text{var}(X) = k(1-p)/p^2$. This is the distribution of the number of trials up to and including the kth success, in a sequence of independent Bernoulli trials each with success probability p.

The negative binomial distribution with $k = 1$ is called a **geometric** distribution with parameter p.

The r.v $Y = X - k$ has

$$\mathbb{P}(Y = y) = \binom{y+k-1}{k-1}(1-p)^y p^k, \text{ for } y = 0, 1, \ldots.$$

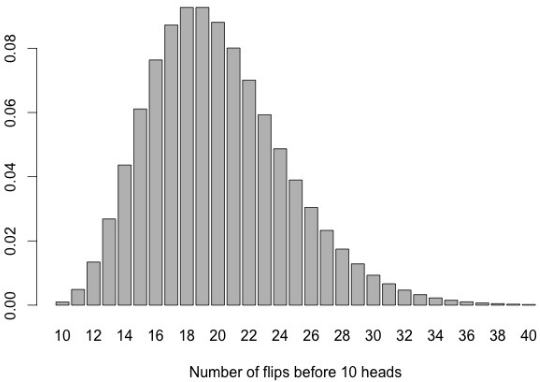
This is the distribution of the number of failures before the kth success in a sequence of independent Bernoulli trials each with success probability p. It is *also* sometimes called the negative binomial distribution: be careful!

Example: How many times do I have to flip a coin before I get 10 heads?

This is first (X) definition of the Negative Binomial since it includes all the flips.

R uses second definition (Y) of Negative Binomial, so need to add in the 10 heads:

```
barplot(  dnbinom(0:30, 10, 1/2), names.arg=0:30 + 10,
          xlab="Number of flips before 10 heads" )
```



Some important discrete distributions: Multinomial

Suppose we have a sequence of n independent trials where at each trial there are k possible outcomes, and that at each trial the probability of outcome i is p_i.

Let N_i be the number of times outcome i occurs in the n trials and consider $N_1, \dots, N_k$. They are discrete random variables, taking values in $\{0, 1, \dots, n\}$.

This **multinomial** distribution with parameters n and $p_1, \dots, p_k$, $n \in \mathbb{N}$, $p_i \geq 0$ for all i and $\sum_i p_i = 1$ has joint pmf

$$\mathbb{P}(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}, \qquad \text{if } \sum_i n_i = n,$$

and is zero otherwise.

The rv's $N_1, \dots, N_k$ are not independent, since $\sum_i N_i = n$.

The marginal distribution of N_i is Binomial(n,p_i).

Example: I throw 6 dice: what is the probability that I get one of each face 1,2,3,4,5,6? Can calculate to be $\frac{6!}{1! \dots 1!} \left(\frac{1}{6}\right)^6 = 0.015$

```
dmultinom( x=c(1,1,1,1,1,1), size=6, prob=rep(1/6,6))
```

Some important continuous distributions: Normal

X has a **normal** (Gaussian) distribution with mean μ and variance σ^2 ($\mu \in \mathbb{R}$, $\sigma^2 > 0$), $X \sim N(\mu, \sigma^2)$, if it has pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

We have $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$.

If $\mu = 0$ and $\sigma^2 = 1$, then X has a **standard normal** distribution, $X \sim N(0,1)$. We write ϕ for the standard normal pdf, and Φ for the standard normal distribution function, so that

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right), \qquad \Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

The upper $100\alpha\%$ point of the standard normal distribution is z_α where

$$\mathbb{P}(Z > z_\alpha) = \alpha, \text{ where } Z \sim N(0,1).$$

Values of Φ are tabulated in normal tables, as are percentage points z_α.

Some important continuous distributions: Uniform

X has a **uniform** distribution on $[a, b]$, $X \sim U[a, b]$ $(-\infty < a < b < \infty)$, if it has pdf

$$f_X(x) = \frac{1}{b-a}, \qquad x \in [a, b].$$

Then $\mathbb{E}(X) = \frac{a+b}{2}$ and $\text{var}(X) = \frac{(b-a)^2}{12}$.

Some important continuous distributions: Gamma

X has a **Gamma** (α, λ) distribution ($\alpha > 0$, $\lambda > 0$) if it has pdf

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \qquad x > 0,$$

where $\Gamma(\alpha)$ is the gamma function defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ for $\alpha > 0$.
We have $\mathbb{E}(X) = \frac{\alpha}{\lambda}$ and $\text{var}(X) = \frac{\alpha}{\lambda^2}$.
The moment generating function $M_X(t)$ is

$$M_X(t) = \mathbb{E}\left(e^{Xt}\right) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \qquad \text{for } t < \lambda.$$

Note the following two results for the gamma function:
(i) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$,
(ii) if $n \in \mathbb{N}$ then $\Gamma(n) = (n - 1)!$.

Some important continuous distributions: Exponential

X has an **exponential** distribution with parameter λ ($\lambda > 0$) if $X \sim \text{Gamma}(1, \lambda)$, so that X has pdf

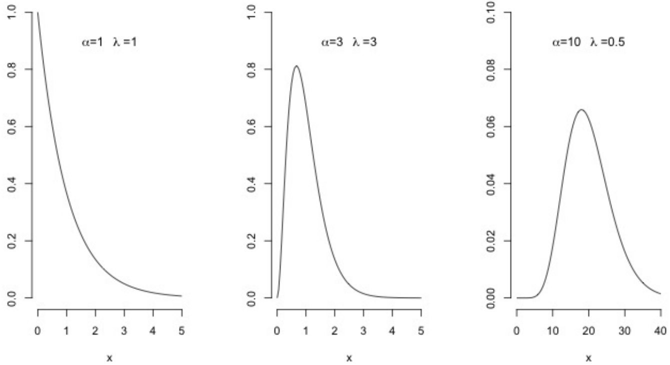
$$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

Then $\mathbb{E}(X) = \frac{1}{\lambda}$ and $\text{var}(X) = \frac{1}{\lambda^2}$.
Note that if $X_1, \ldots, X_n$ are iid Exponential(λ) r.v's then $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$.

Proof: mgf of X_i is $\left(\frac{\lambda}{\lambda - t}\right)$, and so mgf of $\sum_{i=1}^n X_i$ is $\left(\frac{\lambda}{\lambda - t}\right)^n$, which we recognise as the mgf of a Gamma(n, λ).$\square$

Some Gamma distributions:

```
a<-c(1, 3, 10);   b<-c(1, 3, 0.5)
for(i in 1:3){
  y= dgamma(x, a[i],b[i])
  plot(x,y,.......) }
```



Some important continuous distributions: Chi-squared

If $Z_1, \ldots, Z_k$ are iid $N(0,1)$ r.v.'s, then $X = \sum_{i=1}^k Z_i^2$ has a **chi-squared** distribution on k degrees of freedom, $X \sim \chi^2_k$.

Since $\mathbb{E}(Z_i^2) = 1$ and $\mathbb{E}(Z_i^4) = 3$, we find that $\mathbb{E}(X) = k$ and $\mathrm{var}(X) = 2k$.

Further, the moment generating function of Z_i^2 is

$$M_{Z_i^2}(t) = \mathbb{E}\left(e^{Z_i^2 t}\right) = \int_{-\infty}^{\infty} e^{z^2 t} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = (1-2t)^{-1/2} \text{ for } t < 1/2$$

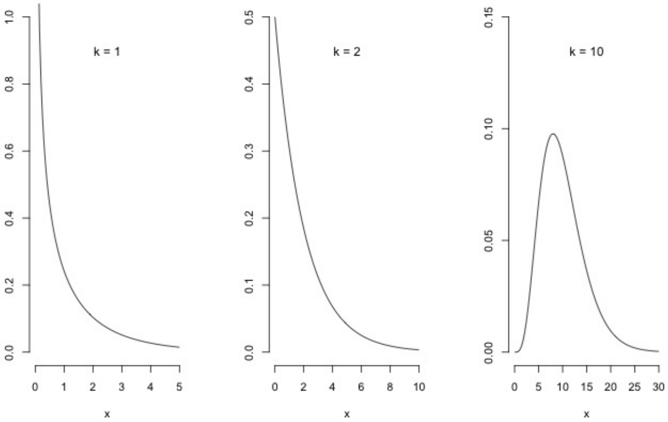
(check), so that the mgf of $X = \sum_{i=1}^k Z_i^2$ is $M_X(t) = (M_{Z^2}(t))^k = (1-2t)^{-k/2}$ for $t < 1/2$.

We recognise this as the mgf of a $\mathrm{Gamma}(k/2, 1/2)$, so that X has pdf

$$f_X(x) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} x^{k/2-1} e^{-x/2}, \qquad x > 0.$$

Some chi-squared distributions: $k = 1,2,10$:

```
k<-c(1,2,10)
for(i in 1:3){
y=dchisq(x, k[i])
  plot(x,y,.......) }
```



Note:

1. We have seen that if $X \sim \chi^2_k$ then $X \sim \mathrm{Gamma}(k/2, 1/2)$.
2. If $Y \sim \mathrm{Gamma}(n, \lambda)$ then $2\lambda Y \sim \chi^2_{2n}$ (prove via mgf's or density transformation formula).
3. If $X \sim \chi^2_m$, $Y \sim \chi^2_n$ and X and Y are independent, then $X + Y \sim \chi^2_{m+n}$ (prove via mgf's). This is called the additive property of χ^2.
4. We denote the upper $100\alpha\%$ point of χ^2_k by $\chi^2_k(\alpha)$, so that, if $X \sim \chi^2_k$ then $\mathbb{P}(X > \chi^2_k(\alpha)) = \alpha$. These are tabulated. The above connections between gamma and χ^2 means that sometimes we can use χ^2-tables to find percentage points for gamma distributions.

Some important continuous distributions: Beta

X has a **Beta(**α, β) distribution ($\alpha > 0$, $\beta > 0$) if it has pdf

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \qquad 0 < x < 1,$$

where $B(\alpha, \beta)$ is the beta function defined by

$$B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta).$$

Then $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ and $\text{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
The mode is $(\alpha-1)/(\alpha+\beta-2)$.
Note that Beta(1,1)$\sim U[0,1]$.

Some beta distributions :

```
k<-c(1,2,10)
for(i in 1:3){
y=dbeta(x, a[i],b[i])
  plot(x,y,.......) }
```