MATH612.2 2 Elements of calculus

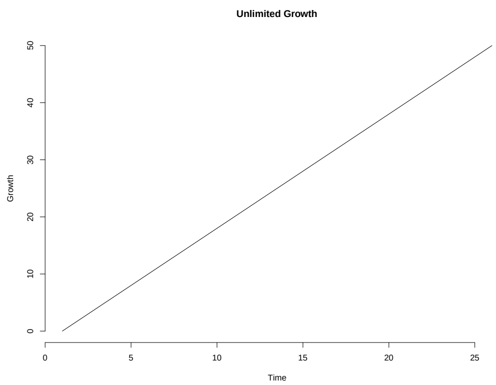
Anonymous

May 10, 2013

Contents

1 Continuity and limits

1.1 The concept of continuity



A function is continuous if it has no jumps. Thus, small changes in each x_0, the input, correspond to small changes in the output, $f(x_0)$.

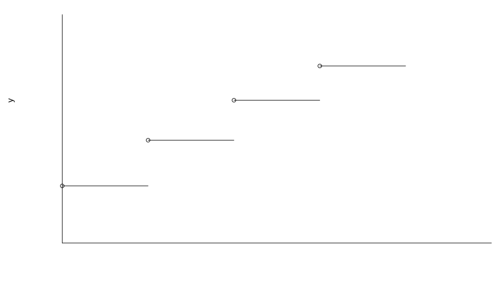
Figure 1: The figure above is an example of linear growth. Thomas Robert Malthus (1766-1834) warned about the dangers of inhibited population growth.

A function is said to be discontinuous if it has jumps. The function is continuous if it has no jumps.

It follows that for a continuous function, small changes in each x_0, the input, thus correspond to small changes in the output, $f(x_0)$.

Polynomials are continuous as are logarithms (for positive numbers).

1.2 Discrete probabilities and cumulative distribution functions



The cumulative distribution function for a discrete random variable is discontinuous.

If X is a random variable with a discrete probability distribution, with probability mass function

$$p(x) = P[X = x]$$

then the *cumulative distribution function*, defined by

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

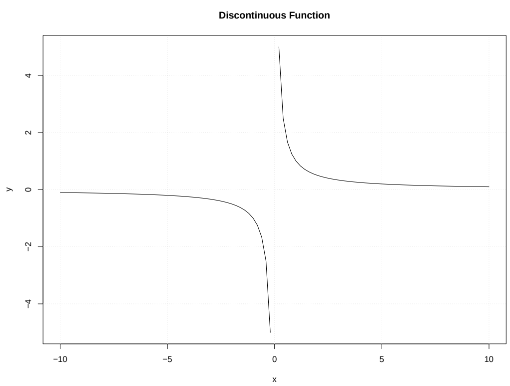
is discontinuous. It's jumps occur at the points which have positive probability.

Example: If a coin is tossed 3 independent times and X denotes the number of heads, then X can only take on the values 0, 1, 2 and 3. The probability of landing exactly x heads, $P(X = x)$, is $p(x) = \binom{n}{x} p^n (1-p)^{n-x}$. The probabilities are

x	p(x)	F(x)
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

The cumulative distribution function, $F(x) = P[X \leq x] = \sum_{t \leq x} p(t)$ has jumps and is therefore discontinuous.

1.3 Notes on discontinuous function



A function is discontinuous for certain values or between certain values of the variable that does not vary continuously as the variable increases. In other words, "breaks" or "jumps."

A function can be discontinuous in a number of different ways. Most commonly, it may jump at certain points or increase without bound in certain places.

Consider the function f, defined by $f(x) = 1/x$ when $x \neq 0$. Naturally, $1/x$ is not defined for $x = 0$. This function increases towards $+\infty$ as x goes to zero from the right but decreases to $-\infty$ as x goes to zero from the left. Since the function does not have the same limit from the right and the left, it follows that it can not be made continuous at $x = 0$ even if one tries to define $f(0)$ as some number.

1.4 Continuity of polynomials

It is easy to show that simple polynomials such as $p(x) = x$, $p(x) = a + bx$, $p(x) = ax^2 + bx + c$ are continuous functions.

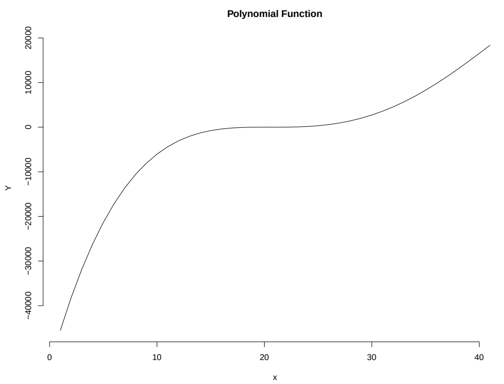
It is generally true that a polynomial of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

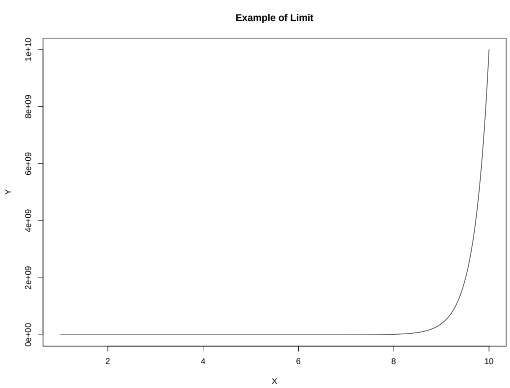
is a continuous function.

1.5 Simple Limits

In mathematics, the concept of a "limit" is used to describe the value that a function or sequence "approaches" as the input or index approaches some value. Limits are essential to calculus (and mathemat-



All polynomials, $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, are continuous.



A "limit" is used to describe the value that a function or sequence "approaches" as the input or index approaches some value. Limits are used to define continuity, derivatives and integrals.

ical analysis in general) and are used to define continuity, derivatives and integrals.

Consider a function and a point x_0 .

If $f(x)$ gets steadily close to some number c as x gets close to a number x_0, then c is called the limit of $f(x)$ as x goes to x_0, written

$$c = \lim_{x \to 0} f(x)$$

If $c = f(x_0)$ then f is called continuous at x_0.

Example:

Consider the function

$$g(x) = \frac{1}{x}$$

where x is a positive real number. As x increases, $g(x)$ decreases, approaching 0 but never getting there since $\frac{1}{x} = 0$ has no solution. One can therefore say, "The limit of $g(x)$, as x approaches infinity, is 0," and write

$$\lim_{x \to \infty} g(x) = 0.$$

1.6 More on limits

Example 1:

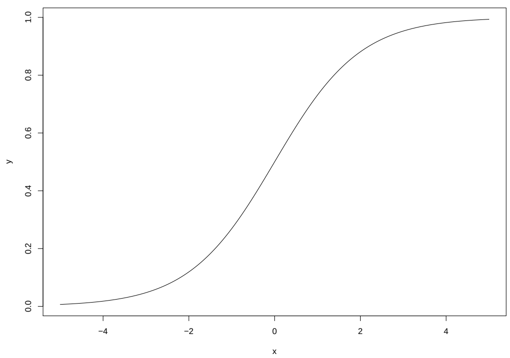


Figure 4: The function $f(x) = \frac{1}{1+e^{-x}}$.

Limits impose a certain range of values that may be applied to the function.

The Beverton-Holt stock recruitment curve is given by:

$$R = \frac{\alpha S}{1+\frac{S}{K}}$$

where $\alpha, K > 0$ are constants.

and S = biomass, R= recruitment

The behavior of this curve as S increases $S \to \infty$ is

$$\lim_{S\to\infty} \frac{\alpha S}{1+\frac{S}{K}} = \alpha K.$$

This is seen by rewriting the formula as follows:

$$\lim_{S\to\infty} \frac{\alpha S}{1+\frac{S}{K}} = \lim_{S\to\infty} \frac{\alpha}{\frac{1}{S}+\frac{1}{K}} = \alpha K.$$

Example 2:

A popular model for proportions is to use

$$f(x) = \frac{1}{1+e^{-x}}$$

Note: As x increases, e^{-x} decreases which implies that the term $1+e^{-x}$ decreases and hence $\frac{1}{1+e^{-x}}$ increases, from which it follows that f is an increasing function.

Notice that $f(0) = \frac{1}{2}$ and further,

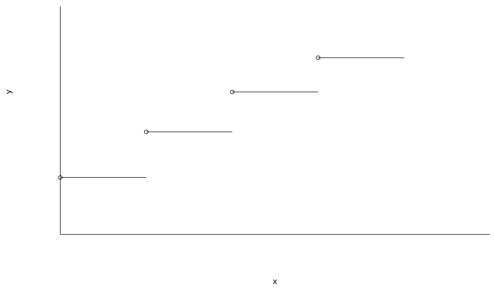
$$\lim_{x\to\infty} f(x) = 1.$$

This is seen from considering the components: Since $e^{-x} = \frac{1}{e^x}$ and the exponential function goes to infinity as $x \to \infty$, e^{-x} goes to 0 and hence $f(x)$ goes to 1.

Through a similar analysis one finds that

$$\lim_{x\to-\infty} f(x) = 0,$$

since, as $x \to \infty$, first $-x \to \infty$ and second $e^{-x} \to \infty$.



$f(x)$ may tend towards different numbers depending on whether $x \to x_0$: from the right ($x \to x_{0+}$) or from the left ($x \to x_{0-}$).

1.7 One-sided limits

Sometimes a function is such that $f(x)$ tends to different numbers depending on whether $x \to x_0$ from the right ($x \to x_{0+}$) or from the left ($x \to x_{0-}$).

If

$$\lim_{x \to x_{0+}} f(x) = f(x_0)$$

then we say that f is continuous from the right at x_0.

2 Sequences and series

2.1 Sequences

A *sequence* is a string of indexed numbers $a_1, a_2, a_3, \ldots$. We denote this sequence with $(a_n)_{n \geq 1}$.

In a sequence the same number can be appeared in several times in different places.

Example 1

$(\frac{1}{n})_{n \geq 1}$ is the sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$.

Example 2

$(n)_{n \geq 1}$ is the sequence $1, 2, 3, 4, 5, \ldots$.

Example 3

$(2^n n)_{n \geq 1}$ is the sequence $2, 8, 24, 64, \ldots$.

2.2 Convergent sequences

A sequence a_n is said to *converge* to the number b if for every $\varepsilon > 0$ we can find an $N \in \mathbb{N}$ such that $|a_n - b| < \varepsilon$ for all $n \geq N$. We denote this with $\lim_{n \to \infty} a_n = b$ or $a_n \to b$, as $n \to \infty$.

A sequence a_n is said to *converge* to the number b if for every $\varepsilon > 0$ we can find an $N \in \mathbb{N}$ such that $|a_n - b| < \varepsilon$ for all $n \geq N$. We denote this with $\lim_{n\to\infty} a_n = b$ or $a_n \to b$, as $n \to \infty$.

If x is a number then,

$(1 + \frac{x}{n})^n \to e^x$ as $n \to \infty$

Example 1

The sequence $(\frac{1}{n})_{n\geq\infty}$ converges to 0 as $n \to \infty$

Example 2

If x is a number then,

$(1 + \frac{x}{n})^n \to e^x$ as $n \to \infty$

2.3 Infinite sums (series)

We are interested in, whether infinte sums of sequences can be defined. Let $(a_n)_{n\to\infty}$ be a sequence of numbers. We define another sequence $(s_n)_{n\to\infty}$, where

$$s_n = \sum_{k=1}^{n} a_k$$

. If $(s_n)_{n\to\infty}$ is convergent with $\lim_{n\to\infty} s_n = S$ we write

$$s_n = \sum_{n=1}^{\infty} a_n = S$$

.

If

$$a_k = x^k, k = 0, 1, ...$$

then

$$s_n = \sum_{k=0}^{n} x^k = x^0 + x^1 + ... + x^n$$

Note also that

$$xs_n = x(x^0 + x^1 + ... + x^n) = x + x^2 + ... + x^{n+1}$$

We have

$$s_n = 1 + x + x^2 + ... + x^n$$

$$xs_n = x + x^2 + ... + x^n + x^{n+1}$$

$$s_n\text{˘}xs_n = 1 - x^{n+1}$$

i.e.

$$s_n(1 - x) = 1 - x^{n+1}$$

and we have

$$s_n = \frac{1 - x^{n+1}}{1 - x}$$

if $x \neq 1$. If $0 < x < 1$ then $x^{n+1} \to 0$ as $n \to \infty$ and we obtain $s_n \to \frac{1}{1-x}$ so $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$.

The exponential function can be written as a series (infinite sum):

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

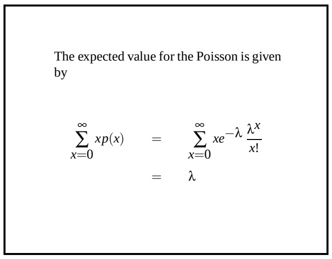
Knowing this we can see why the Poisson probabilities

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

add to one:

$$\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

2.4 Relation to expected values



The expected value for the Poisson is given by

$$\begin{aligned}
\sum_{x=0}^{\infty} xp(x) &= \sum_{x=0}^{\infty} xe^{-\lambda}\frac{\lambda^x}{x!} \\
&= e^{-\lambda}\sum_{x=1}^{\infty}\frac{x\lambda^x}{x!} \\
&= e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} \\
&= e^{-\lambda}\lambda\sum_{x=1}^{\infty}\frac{\lambda^{(x-1)}}{(x-1)!} \\
&= e^{-\lambda}\lambda\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} \\
&= e^{-\lambda}\lambda e^{\lambda} \\
&= \lambda
\end{aligned}$$

3 Slopes of lines and curves

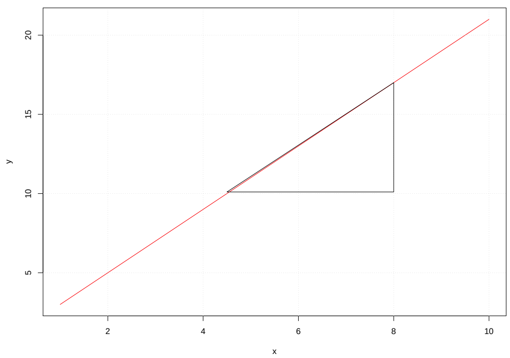
3.1 The slope of a line

The slope of a straight line represents the change in the y coordinate corresponding to a unit change in the x coordinate.

3.2 Segment slopes

Consider two points, (x_0, y_0) and (x_1, y_1). The slope of the straight line that goes through these points is

$$\frac{y_1 - y_0}{x_1 - x_0}.$$

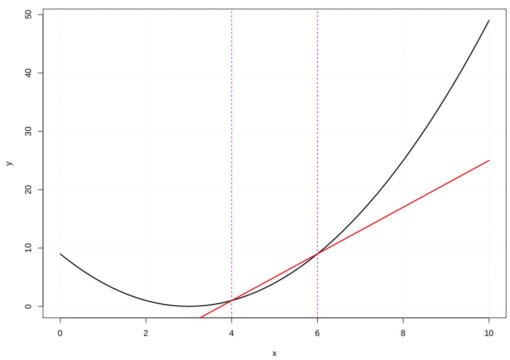


Linear functions produce straight-line graphs. In general, a straight line consist of points in the plane which satisfy an equation of the form

$$y = a + bx,$$

where a and b are fixed numbers.
The graph of the line is the set of points:

$$\{(x, y) : x, y \in \mathbb{R}, y = a + bx\}.$$



Let's assume we have a more general function
$y = f(x)$
To find the slope of line segment, consider 2 x-coordinates x_0 and x_1 and look at the slope between $(x_0, f(x_0))$ and $(x_1, f(x_1))$

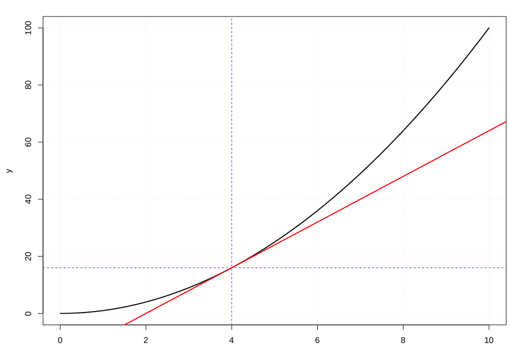
Thus, the slope of a line segment passing throught the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$, for some function f is

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

If we let $x_1 = x_0 + h$ the slope of the segment is

$$\frac{f(x_0 + h) - f(x_0)}{h}.$$

3.3 The slope of $y = x^2$



Consider the task of computing the slope of the function $y = x^2$ **at** a given point.

Consider the following function;

$y = f(x) = x^2$

In order to find the slope at a given point (x_0), we look at

$$y = \frac{f(x_0+h) - f(x_0)}{h}$$

for small values of h.

For this particular function, $f(x_0) = x^2$, and hence

$$f(x_0+h) = (x_0+h)^2 = x^2 + 2hx_0 + h^2.$$

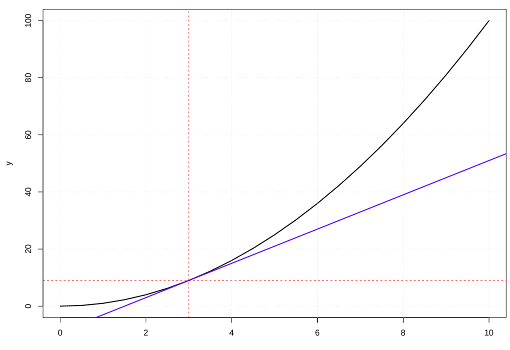
The slope of a line segment is therefore given by

$$\frac{f(x_0+h) - f(x_0)}{h} = \frac{2hx_0 + h^2}{h} = 2x_0 + h.$$

As we make h steadily smaller, the segment slope, $2x_0 + h$, tends towards $2x_0$.

It follows that the slope of the curve **at a general point** x is given by $y = 2x$.

3.4 The tangent to a curve



A *tangent* to a curve is a line that intersects the curve at exactly one point. The slope of a tangent to the graph of the function $y = f(x)$ at the point $(x_0, f(x_0))$ is

$$\lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}.$$

To find the slope of the tangent to a curve at a point, we look at the slope of a line segment between the points $(x_0, f(x_0))$ and $(x_0+h, f(x_0+h))$, which is

$$\frac{f(x_0+h) - f(x_0)}{h}$$

and then we take h to be closer and closer to 0. Thus the slope is

$$\lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$$

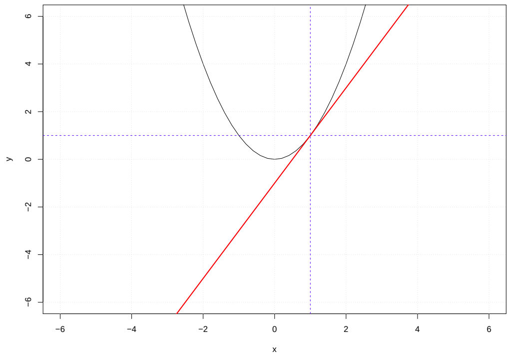
when this limit exists.

Example 1

We wish to find the line that is tangent to the graph of the function $f(x) = x^2$ at the point $(1,1)$. First we need to fin the slope of this tangent, it is given as

$$\lim_{h \to 0} \frac{(1+h)^2 - 1^2}{h} = \lim_{h \to 0} \frac{2h + h^2}{h} = \lim_{h \to 0} (2+h) = 2.$$

Then since we know the tangent goes through the point $(1,1)$ the line is $y = 2x - 1$.



Consider a nonlinear function $y = f(x)$
The slope of the line segment: $\frac{f(x_0+h)-f(x_0)}{h}$
Now find the limit as h goes towards zero, if it exists.

3.5 The slope of a general curve

Imagine a nonlinear function whose graph is a curve describe by the equation,

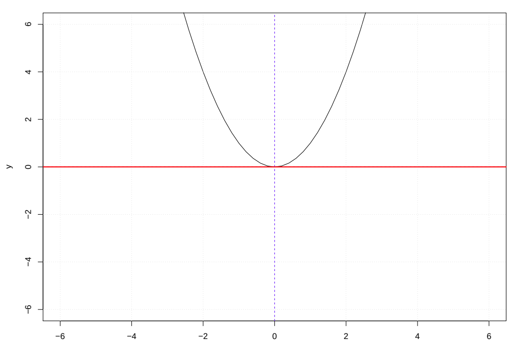
$y = f(x)$

Here we want to find the slope of a line tangent to the curve at a specific point (x_0).

The slope of the line segment is given by following equation as explained earlier

$\frac{f(x_0+h)-f(x_0)}{h}$

Reducing h towards zero, gives the slope of this curve if it exists.



4 Derivatives

4.1 The derivative as a limit

The derivative of the function f at the point x is defined as

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

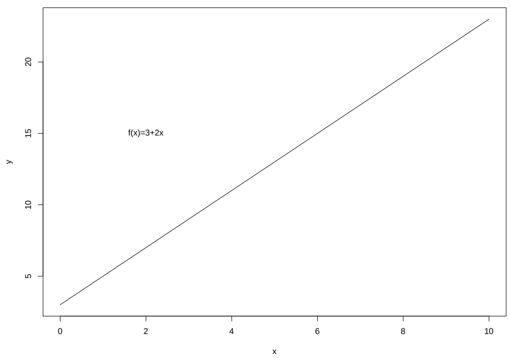
if this limit exists.

The derivative of the function f at the point x is defined as

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$$

if this limit exists. When we write $y=f(x)$, we commonly use the notation $\frac{dy}{dx}$ or $f'(x)$ for this limit.

4.2 The derivative of $f(x)=a+bx$



If $f(x)=a+bx$ then $f(x+h)=a+b(x+h)=a+bx+bh$ and thus

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{bh}{h}=b$$

If $f(x)=a+bx$ then $f(x+h)=a+b(x+h)=a+bx+bh$ and thus

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=\lim_{h\to 0}\frac{bh}{h}=b.$$

Thus $f'(x)=b$.

4.3 The derivative of $f(x)=x^n$

Let $f(x)=x^n$, where n is a positive integer. To find f' we calculate, using the binomial theorem in the third step:

$$\begin{aligned}\frac{f(x+h)-f(x)}{h} &= \frac{(x+h)^n-x^n}{h}\\ &= \frac{\sum_{q=0}^{n-1}x^q h^{n-q}}{h}\\ &= \sum_{q=0}^{n-1}x^q h^{n-q-1}\to\binom{n}{n-1}x^{n-1}=nx^{n-1}\end{aligned}$$

Thus, we obtain $f'(x)=nx^{n-1}$.

4.4 The derivative of ln and exp<Ólöf BM>

The derivatives of the exponential function is the exponential function itself i.e. if

$$f(x) = e^x$$

then

$$f'(x) = e^x$$

The derivatives of the natural logarithm, $\ln(x)$, is $\frac{1}{x}$, i.e. if

$$g(x) = \ln(x)$$

then

$$g'(x) = \frac{1}{x}$$

4.5 The derivative of a sum and linear combination

If f and g are functions then the derivative of $f+g$ is given by $f'+g'$

Similarly, the derivative of a linear combination is the linear combination of the derivatives.

If f and g are functions and $k(x) = af(x) + bg(x)$ then $k'(x) = af'(x) + bg'(x)$

Example:
If $f(x) = 2+3x$ and $g(x) + x^3$
then we know that
$f'(x) = 3$, $g(x) = 3x^2$
and if we write
$h(x) = f(x) + g(x) = 2+3x+x^3$
then
$h'(x) = 3+3x^2$

4.6 The derivative of a polynomial

The derivative of a polynomial is the sum of the derivatives of the terms of the polynomial.

If

$$p(x) = a_0 + a_1x + \ldots + a_nx^n$$

then

$$p'(x) = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \ldots + na_nx^{(n-1)}$$

If

$$p(x) = 2x^4 + x^3$$

then

$p'(x) = 2\frac{dx^4}{dx} + \frac{dx^3}{dx} = 2 \cdot 4x^3 + 3x^2 = 8x^3 + 3x^2$

4.7 The derivative of a product

If

$$h(x) = f(x) \cdot g(x)$$

then

$$h'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Consider two functions, f and g and their product, h:

$$h(x) = f(x) \cdot g(x).$$

The derivative of the product is given by

$$h'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Example: Suppose the function f is given by

$$f(x) = xe^x + x^2 \ln x.$$

Then the derivative can be computed step by step as

$$\begin{aligned} f(x) &= \frac{dx}{dx}e^x + x\frac{de^x}{dx} + \frac{dx^2}{dx}\ln x + x^2\frac{d\ln x}{dx} \\ &= 1 \cdot e^x + x \cdot e^x + 2x \cdot \ln x + x^2 \cdot \frac{1}{x} \\ &= e^x(1+x) + 2x\ln x + x \end{aligned}$$

4.8 Derivatives of composite functions

If f and g are functions and $h = f \circ g$ so that

$h(x) = f(g(x))$ then

$h'(x) = \frac{dh(x)}{dx} = f'(g(x))g'(x)$

1. For fixed x consider ;

$$\begin{aligned} f(p) &= \ln(p^x(1-p)^{n-x}) \\ &= \ln p^x + \ln(1-p)^{n-x} \\ &= x\ln p + (n-x)\ln(1-p) \end{aligned}$$

$$\begin{aligned} f'(p) &= x\frac{1}{p} + \frac{n-x}{1-p}(-1) \\ &= \frac{x}{p} - \frac{n-x}{1-p} \end{aligned}$$

2. $f(b) = (y - bx)^2$ (y, x fixed)

$$\begin{aligned} f'(b) &= 2(y - bx)(-x) \\ &= -2x(y - bx) \\ &= (-2xy) + (2x^2)b \end{aligned}$$

5 Applications of differentiation

5.1 Tracking the sign of the derivative

If f is a function, then the sign of it's derivative, f', indicates whether f increasing ($f' > 0$), decreasing ($f' < 0$) or f' can be zero at points where f has a maximum, minimum or a saddle point.
If $f'(x) > 0$ for $x < x_0$, $f'(x_0) = 0$ and $f'(x) < 0$ for $x > x_0$ then f has a maximum at x_0
If $f'(x) < 0$ for $x < x_0$, $f'(x_0) = 0$ and $f'(x) > 0$ for $x > x_0$ then f has a minimum at x_0
If $f'(x) > 0$ for $x < x_0$, $f'(x_0) = 0$ and $f'(x) > 0$ for $x < x_0$ then f has a saddle point at x_0
If $f'(x) < 0$ for $x < x_0$, $f'(x_0) = 0$ and $f'(x) < 0$ for $x < x_0$ then f has a saddle point at x_0

Example 1:

If f is a function such that it's derivative

$$f'(x) = (x-1)(x-2)(x-3)(x-4),$$

then applying the above criteria for maxima and minima, we see that f has maxima at 1 and 3 and f has minima at 2 and 4.

5.2 Describing extrema

x_0 with $f'(x_0) = 0$ corresponds to a maximum, if $f''(x_0) < 0$
x_0 with $f'(x_0) = 0$ corresponds to a minimum, if $f''(x_0) > 0$

If $f'(x_0) = 0$ corresponds to a maximum, then the derivative is decreasing and the second derivative can not be positive, (i.e. $f''(x_0) \leq 0$). In particular, if the second derivative is strictly negative, ($f''(x_0) < 0$), then we are assured that the point is indeed a maximum, and not a saddle point.

If $f'(x_0) = 0$ corresponds to a minimum, then the derivative is increasing and the second derivative can not be negative, (i.e. $f''(x_0) \geq 0$).

If the second derivative is zero, then the point may be a saddle point, as happens with $f(x) = x^3$ at $x = 0$.

5.3 The likelihood function

Recall that the probability mass function (p.m.f) is a function, typically denoted p so $p(x)$ gives the probability of a given outcome, x, of an experiment, based on some parameter. We often write,
$p(x) = P[X = x]$
when we are going to take a sample of independent measurements, all from p, then the joint probability of a given set of numbers is,
$p(x_1) \cdot p(x_2) \cdot p(x_3) \ldots p(x_n)$
Suppose each probability includes same parameter θ, then this is typically written,
$p_\theta(x_1), \ldots p_\theta(x_n)$
Now consider the set of outcomes $x_1, x_2 \ldots, x_n$ from the experiment. We can now take the probability of this outcome as a function of the parameters.
$L_{\mathbf{x}}(\theta) = p_\theta(x_1), \ldots p_\theta(x_n)$
This is the **likelihood function** and we often seek to maximize it given outcomes from an experiment.

Recall that the probability mass function (p.m.f) is the function of p and p(x) gives the probability of a given outcome of an experiment, based on same parameter. We often write,

$P(x) = P[X = x]$

when we are going to take a sample of independent measurements, all from p, then the joint probability of a given set of number is,

$p(x_1) \cdot p(x_2) \cdot p(x_3) \ldots p(x_n)$

Suppose each probability includes same parameter θ, then this is typically written,

$p_\theta(x_1), \ldots p_\theta(x_n)$

Now consider the set of outcomes $x_1, x_2 \ldots, x_n$ from the experiment. We can now take the probability of this outcome as a function of the parameters.

$L_\theta = p_\theta(x_1), \ldots p_\theta(x_n)$

This is the **likelihood function**.

Suppose we toss a biased coin n independent times and obtain x heads, we know the probability of obtaining x heads is,

$\binom{n}{x} p^x (1-p)^{n-x}$.

The parameter of interest is p and the likelihood function is,

$L(p) = \binom{n}{x} p^x (1-p)^{n-x}$. If p is unknown we sometimes wish to maximize this function with respect to p in order to estimate the *real* probability p.

5.4 Plotting the likelihood <Chiara>

5.5 Maximum likelihood estimation

If L is a likelihood function for a p.m.f p_θ, then the value $\hat{\theta}$ which gives the maximum of L:

$$L(\hat{\theta}) = \max_\theta (L_\theta)$$

is the maximum likelihood estimator (MLE) of θ

If L is a likelihood function for a p.m.f p_θ, then the value $\hat{\theta}$ which gives the maximum of L:

$$L(\hat{\theta}) = \max_{\theta}(L_\theta)$$

is the maximum likelihood estimator of θ

If x is the number of heads from n independent tosses of a coin, the likelihood function is;

$$L_x(p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Maximizing of this is equivalent to maximizing the logarithm of the likelihood, since logarithmic functions are increasing. The log-likelihood can be written as;

$$\ln(L(p)) = \ln\binom{n}{x} + x\ln(p) + (n-x)\ln(1-p).$$

To find possible maxima , we need to differentiate this formula and set the derivative to zero

$0 = \frac{dl(p)}{dp} = 0 + \frac{x}{p} + \frac{n-x}{1-p}(-1)$

$0 = p(1-p)\frac{(x)}{p} - p(1-p)\frac{n-x}{1-p}$

$0 = (1-p)x - p(n-x)$

$0 = x - px - pn + px = x - pn$

So

$0 = x - pn$

$p = \frac{x}{n}$ is the extremum and we can write

$\hat{p} = \frac{x}{n}$ for the MLE

5.6 Least squares estimation

Least squares: Estimate the parameters θ by minimizing

$$\sum_{i=1}^{n}(y_i - g_i(\theta))^2$$

Suppose we have a model linking data to parameters. In general we are predicting y_i as g_i (θ).

In this case it makes sense to estimate parameters θ by minimizing

$$\sum_{i=1}^{n}(y_i - g_i(\theta))^2.$$

Example 1: One may predict numbers, x_i, as a mean, μ, plus error. Consider the simple model $x_i = \mu + \varepsilon_i$, where μ is an unknown parameter (constant) and ε_i is the error in measurement when obtaining the i'th observations, x_i, $i = 1, \ldots, n$.

A natural method to estimate the parameter is to minimize the squared deviations

$$\min_{\mu} \sum_{i=1}^{n} (x - \mu)^2 .$$

It is not hard to see that the $\hat{\mu}$ that minimizes this is the mean:

$$\hat{\mu} = \overline{x}.$$

Example 2: One also commonly predicts data $y_1, \cdots, y_n$ with values on straight line, i.e. with $\alpha + \beta x_i$, where $x_1, \ldots, x_n$ are fixed numbers.

This leads to the regression problem of finding those parameter values, $\hat{\alpha}$ and $\hat{\beta}$, which give the best fitting straight line in the sense that of ordinary least squares:

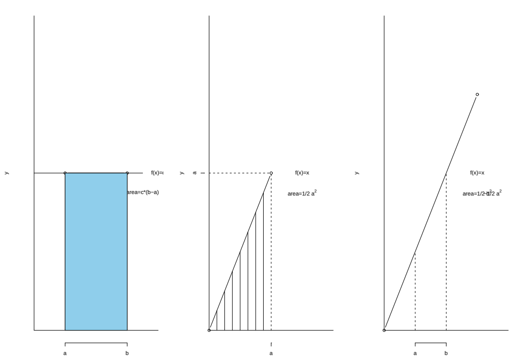
$$\min_{\alpha,\beta} \sum (y_i - (\alpha + \beta x_i))^2$$

Example 3: As a general exercise in finding the extrema of a function, let's look at the function $f(\theta) = \sum_{i=1}^{n} (x_i\theta - 3)^2$ where x_i are some constants. We wish to find the θ that minimizes this sum. We simply differentiate w.r.t. θ to obtain $f'(\theta) = \sum_{i=1}^{n} 2(x_i\theta - 3)x_1 = 2\sum_{i=1}^{n} x_i^2\theta - 2\sum_{i=1}^{n} 3x_i$. Thus

$$f'(\theta) = 2\theta \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} 3x_i = 0$$
$$\Leftrightarrow \theta = \frac{\sum_{i=1}^{n} 3x_i}{\sum_{i=1}^{n} x_i^2}.$$

6 Integrals and probability density functions

6.1 Area under a curve



The area under a curve between x=a and x=b (for a positive function) is called the integral of the function.

Figure 5: Example 1, 2 and 3

The area under a curve between x=a and x=b (for a positive function) is called the integral of the function denoted: $\int_a^b f(x)dx$ when this exists.

6.2 The antiderivative

Given a function f, if there is another function F such that $F' = f$, we say that F is the *antiderivative* of f. For a function f the antiderivative is denoted by $\int f dx$.

Example 1

$\int x^n dx = \frac{1}{n+1}x^{n+1}$.

Example 2

$\int e^x dx = e^x$.

Example 3

$\int \frac{1}{x} dx = \ln(x)$.

Example 4

$\int 2xe^{x^2} dx = e^{x^2}$.

6.3 The fundamental theorem of calculus

The fundamental theorem of calculus states: The area under the graph of the function f on the interval $[a,b]$ is equal to the difference of the values of its antiderivative at a and b. That is, if F is the antiderivative of f, then the area under the graph of f on the interval $[a,b]$ is $F(b) - F(a)$. This difference is often written as $\int_a^b f dx$ or $[F(x)]_a^b$.

Example 1

The area under the graph of x^n between 0 and 3 is $\int_0^3 x^n dx = [\frac{1}{n+1}x^{n+1}]_0^3 = \frac{1}{n+1}3^{n+1} - \frac{1}{n+1}0^{n+1} = \frac{3^{n+1}}{n+1}$

Example 2

The area under the graph of e^x between 3 and 4 is $\int_3^4 e^x dx = [e^x]_3^4 = e^4 - e^3$

Example 3

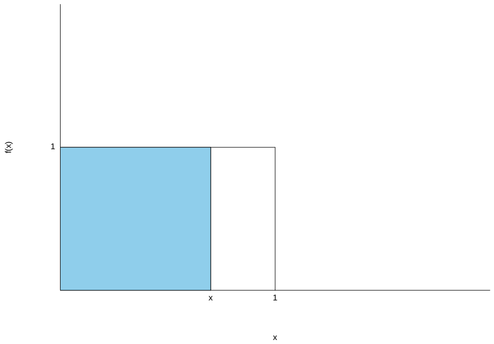
The area under the graph of $\frac{1}{x}$ between 1 and a is $\int_1^a \frac{1}{x} dx = [\ln(x)]_1^a = \ln(a) - \ln(1) = \ln(a)$.

6.4 Density functions

If X is a random variable such that

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

for some function f which satisfies $f(x) \geq 0$ for all x and



The probability density function (p.d.f.) and the cumulative distribution function (c.d.f.).

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

then f is said to be a probability density function (p.d.f.) for X.

The function

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

is the cumulative distribution function (c.d.f.).

Example 1: Consider a random variable X from the uniform distribution, denoted by $X \sim U(0,1)$. This distribution has density

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{e.w.} \end{cases}$$

The cumulative distribution function is given by

$$P[X \le x] = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \end{cases}$$

Example 2: Suppose $X \sim P(\lambda)$, where X may denote the number of events per unit time. The p.m.f. of X is described by $p(x) = P[X = x] = e^{-\lambda}\frac{\lambda^x}{x!}$ for $x = 0,1,2,\ldots$. Consider now the waiting time, T, between events, or simply until the first event. Consider the event $T > t$ for some number t>o. If $X \sim p(\lambda)$ denotes the number of events per unit time, then let X_t denote the number of events during the time period for 0 through t. The it is natural to assume

$X_t \sim P(\lambda t)$ and it follows that $T > t$ if and only if $X_t = 0$ and we obtain $P[T > t] = P[X_t = 0] = e^{-\lambda t}$. It follows that the c.d.f. of T is $F_T(t) = P[T \le t] = 1 - P[T > t] = 1 - e^{-\lambda t}$ for $t > 0$.

The p.d.f. of T is therefore $f_T(t) = F_T'(t) = \frac{d}{dt}F_T(t) = \frac{d}{dt}(1 - e^{-\lambda t} = 0 - e^{-\lambda t} * (-\lambda) = \lambda e^{-\lambda t}$ for $t \ge 0$ and $f_T(t) = 0$ for $t = 0$.

This resulting density $f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for} \quad t \ge 0 \\ 0 & \text{for} \quad t < 0 \end{cases}$
describes the exponential distribution.

This distribution has expected value $E[T] = \int_{-\infty}^{\infty} t f(t)dt = \int_{0}^{\infty} t\lambda e^{-\lambda t}dt$

We set $u = \lambda t$ and $du = \lambda dt$

$\int ue^{-u}du = \frac{1}{\lambda}\int\limits_0^\infty ue^{-u}du = \frac{1}{\lambda}\begin{cases} \int\limits_0^\infty 1 \cdot e^{-u}du \\ [-ue^{-u}]_0^\infty \end{cases}$

$$= \left[\frac{1}{\lambda}(-e^{-u})\right]_0^\infty - 0 = \frac{1}{\lambda}$$

6.5 Probabilities in R: The normal distribution

> R has functions to compute values of probability density functions (p.d.f.) and cumulative distribution functions (c.m.d.) for most common distributions.

The p.d.f. for the normal distribution is

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

The c.d.f. for the normal distribution is

$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

dnorm() gives the value of the normal p.d.f.

pnorm() gives the value of the normal c.d.f.

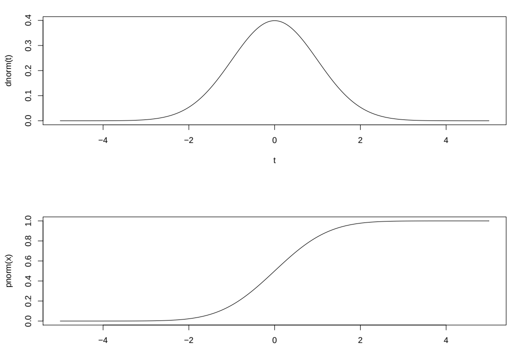


Figure 6: Top: The probability density function for the normal distribution. Bottom: The cumulative distribution function for the normal distribution.

6.6 Some rules of integration

Example 1

By integration by parts we obtain $\int \ln(x)xdx = \frac{1}{2}x^2\ln(x) - \int \frac{1}{2}x^2 \cdot \frac{1}{x}dx = \frac{1}{2}x^2\ln(x) - \int \frac{1}{2}xdx = \frac{1}{2}x^2\ln(x) - \frac{1}{4}x^2$.

Example 2

Consider $\int_1^2 2xe^{x^2}dx$. By setting $x = g(t) = \sqrt{t}$ we obtain

$$\int_1^2 2xe^{x^2}dx = \int_1^4 2\sqrt{t}e^t \frac{1}{2\sqrt{t}}dt = \int_1^4 e^t dt = e^4 - e.$$

The two most common "tricks" applied in integration are a) integration by parts and b) integration by substitution.

a) **Integration by parts**

Since $(fg)' = f'g + fg'$, by integrating both sides of the equation we obtain

$$fg = \int f'gdx + \int fg'dx \Leftrightarrow \int fg'dx = fg - \int f'gdx.$$

b) **Integration by substitution**

Consider the definite integral $\int_a^b f(x)dx$ and let g be a one-to-one, differential function from the interval (c,d) to (a,b). Then

$$\int_a^b f(x)dx = \int_c^d f(g(y))g'(y)dy.$$

7 Principles of programming

7.1 Modularity

Modularity is designing a system that is divided into a set of functional units (named modules) that can be composed into a larger application. So any programming project could be split into logical modules piece of codes which become together.

Typically input, initialization, analysis and output commands are grouped into separate parts.

Example
Input

```
dat<-read.table("http://notendur.hi.is/~gunnar/kennsla/alsm/data/set115.dat", header=T)
cols<- c("le", "osl")
```

Analysis

```
Mn<-mean(dat[, cols[1]])
```

Output

```
print (Mn)
```

7.2 Modularity and functions

> In many cases groups of commands can be collected together into a function.

Typically a project has several such functions.

Example:
Suppose you want to plot the weight vs. length for several datasets in
http://hi.is/ gunnar/kennsla/alsm/data
A function can then be set up with the number as an argument:

```
plotwtle<-function (fnum){
fname<-paste(
"http://hi.is/~gunnar/kennsla/alsm/data/set",fnum,".dat",sep="")
 cat("The URL B", fname,"\n")
 dat<-read.table(fname,header=T)
 ttl<-paste("Data from file number", fnum)
 plot(dat$le,dat$osl,main=ttl)
 }
```

Now call this with

```
plotwtle(105)
```

7.3 Modularity and files

> It is advisable to split larger projects into several manageable files.

Once a project reaches more than five lines of codes, these should be stored in one or more separate files. In order to combine these files which are having different commands, a single “source” command file can be created.

Typically function definitions are stored in separate files, so one may have several separate files like;

"input.r" "function.r" "analysis.r" "output.r"

While developing the analysis, the data would only be read once with

source(“input.r”)

The goal of this practice is to end up with a set of files which are completely self-contained, so one can start with an empty R session and give only the commands like

source (“input.r”) source (“functions.r”) source (“analysis.r”)

Further, this ensures the repeatability.

Example:
For a given project “input”, “functions” “analysis” and “output” files can be created as below.

input.r

```
dat<-read.table("http://notendur.hi.is/~gunnar/kennsla/alsm/data/set115.dat", header=T)
```

functions.r

```
plotwtle<-function(fnum){
 fname<-paste("http://notendur.hi.is/~gunnar/kennsla/alsm/data/set",fnum,".dat",sep="")
  cat("The URL is",fname,"\n")
  dat<-read.table(fname,header=T)
  ttl<-paste("My data set was",fnum)
  plot(dat$le,dat$osl,main=ttl,xlab="Length(cm)",ylab="Live weight (g)")
}
```

output.r

```
source("functions.r")
for(i in 101:150){
  fnam<-paste("plot",i,".pdf",sep="")
  pdf(fnam)
  plotwtle(i)
  dev.off()
}
```

These files can be executed with source commands as below;

source (“input.r”)

source (“functions.r”)

source (“output.r”)

7.4 Structuring an R program

In order to control for organization, when larger projects are undertaken, they need to be split into manageable pieces described as modules, functions, and files. In order to link these files that hold different commands, one should use the "source" commands together in one file.

Example:

The file "run.r" could contain the sequence of commands:

source("setup.r")

source("analysis.r")

source("plot.r")

The benefits of this type of organization is that within the R interface there will not be 1000 lines of code increasing the likelihood of misplacement or loss of methodology. In addition, this permits one to edit in the file and to add comments.

7.5 Loops, for

> If a piece of code is to be run repeatedly, the for-loop is normally used. This is of the form: `for`(index in sequence) commands

Example 1:
To add numbers we can do e.g.

```
tot <- 100
for(i in 1:100){
tot <- tot + i
}
cat ("the sum is ", tot, "\n")
```

Example 2:
Define the plot function

```
plotwtle <- AS BEFORE
```

To plot several of these we can use a sequence:

```
plot wtle(101)
plot wtle(102)
.
.
.
```

or a loop

```
for (i in 101:150){
fname<- paste("plot", i, ".pdf", sep="")
pdf(fname)
plotwtle(i)
dev.off()
}
```

7.6 The if and ifelse commands

> The "if" statement is used to conditionally execute statements.
> The "ifelse" statement conditionally replaces elements of a structure.

Example 1:
If we want to compute x^x for x-values in the range 0 through 5, we can use

```
xlist<-seq(0,5,0.01)
```

```
y<-NULL
for(x in xlist){
  if(x==0){
    y<-c(y,1)
  }else{
    y<-c(y,x**x)
  }
}
```

Example 2:

```
x<-seq(0,5,0.01)
y<-ifelse(x==0,1,x^x)
```

Example 3:

```
dat<-read.table ("file")
dat<-ifelse (dat==0,0.01,dat)
```

Example 4:

```
x<-ifelse (is.na(x),0,x)
```

7.7 Indenting

Code should be properly indented!

That is, functions, for-loops, if-statements should always be indented.

Example:
.....

7.8 Comments

All code should contain informative comments.

```
###################
####SETUP DATA####
###################

dat<-read.table(filename)
x<-log(dat$le) #log-transformation of length
```

```
y<-log(dat$wt) #log-transformation of weight


######################
####THE ANALYSIS####
######################
```

8 The Central Limit Theorem and related topics

8.1 The Central Limit Theorem

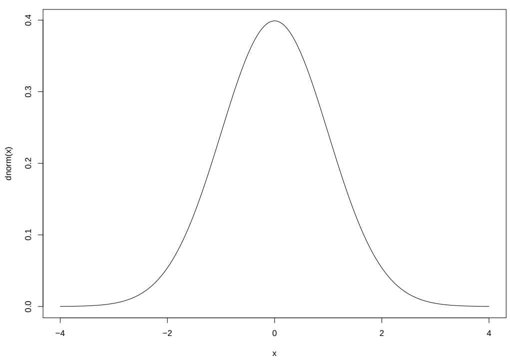


Figure 7: The standard normal density

> If measurements are obtained independently and come from a process with finite variance, then the distribution of the mean of these data tends towards a Gaussian (normal) distribution as the sample size increases.

The Central Limit Theorem states that if $X_1, X_2, \ldots$ are i.i.d. random variables with mean μ and variance σ^2, then the distribution of $\bar{X}_n := \frac{X_1+\cdots+X_n}{n}$ tends towards a normal distribution. The random variable $\bar{X}_n$ can be approximated by $N(\mu, \sigma^2/n)$.

The Gaussian distribution is given by the p.d.f.

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}$$

for $z \in \mathbb{R}$

The distribution has an expected value of zero;

$$\mu = \int z\varphi(z)dz = 0$$

and a variance of

$$\sigma^2 = \int (z-\mu)^2 \varphi(z)dz = 1$$

The general normal distribution, with arbitrary mean μ and variance σ^2 has the p.d.f.

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

Example 1:
If we collect measurements on waiting times, typically from an exponential distribution with density

$$\lambda e^{-\lambda t}, t > 0$$

then the mean of several such waiting times will tend to have a normal distribution.

Example 2:
We are often interested in computing

$$w = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

which has a t-distribution if the x_i are independent outcomes from a normal distribution. If n is large and σ^2 is finite then w will look as it came from a normal distribution.

8.2 Properties of the binomial and Poisson distributions

The binomial distribution is really a sum of 0 and 1 values (counts of failures = 0 and successes =1) so a simple, single binomial outcome will correspond to coming from a normal distribution if the count is large enough.

Consider the binomial probabilities:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, 2, 3, \cdots, n$

Where n is a non-negative integer. Suppose p is a small positive number, specifically consider a sequence of decreasing p-values, specified with $p_n = \frac{\lambda}{n}$ and consider the behavior of the probability as $n \to \infty$

we obtain:

$$\binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{n!}{x!(n-x!)} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \quad (1)$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \frac{\frac{\lambda}{n}^x}{\left(1-\frac{\lambda}{n}\right)^x} \left(1 - \frac{\lambda}{n}\right)^n \quad (2)$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x! n^x} \frac{\lambda^x}{\left(1-\frac{\lambda}{n}\right)^x} \left(1 - \frac{\lambda}{n}\right)^n \quad (3)$$

$$(4)$$

Notice that $\frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \to 1$ as $n \to \infty$. Also notice that $(1-\frac{\lambda}{n})^x \to 1$ as $n \to \infty$. Also

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

and it follows that

$$\lim_{n\to\infty} \binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \cdots, n$$

and hence the binomial probabilities may be approximated with the corresponding Poisson.

Example 1:

The mean of a binomial (n,p) variable is $\mu = n \cdot p$ and the variance is $\sigma^2 = np(1-p)$

The R command pbinom(q,n,p) calculates the probability of q successes in n trials assuming that the probability of a success is p in each trial (binomial distribution). The normal approximation of this distribution can be calculated with pnorm(q,mu,sigma) which becomes pnorm(q,n*p,sqrt(n*p(1-p)). Three numerical examples (note that pbinom and pnorm give similar values for large n):

```
> pbinom(3,10,0.2)
[1] 0.8791261
> pnorm(3,10*0.2,sqrt(10*0.2*(1-0.2)))
[1] 0.7854023

> pbinom(3,20,0.2)
[1] 0.4114489
> pnorm(3,20*0.2,sqrt(20*0.2*(1-0.2)))
[1] 0.2880751

> pbinom(30,200,0.2)
[1] 0.04302156
> pnorm(30,200*0.2,sqrt(200*0.2*(1-0.2)))
[1] 0.03854994
```

Example 2:

We are often interested in computing $w = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

which has a t distribution if the x_i are independent outcomes from a normal distribution. If n is large and σ^2 is finite, this will look as if it comes from a normal distribution.

The numerical examples below demonstrate how the t distribution approaches the normal distribution.

```
> qnorm(0.7)
[1] 0.5244005
#This is the value which gives the cumulative probability of p=0.7 for a n~(0,1)
> qt(0.7,2)
[1] 0.6172134
#The value, which gives the cumulative probability of p=0.7 with n=2 for the t-distribution.
> qt(0.7,5)
[1] 0.5594296
> qt(0.7,10)
[1] 0.541528
> qt(0.7,20)
[1] 0.5328628


> qt(0.7,100)
[1] 0.5260763
```

8.3 Monte Carlo simulation <Warsha>

Example:

Suppose our measurements come from an exponential distribution and we want to compute

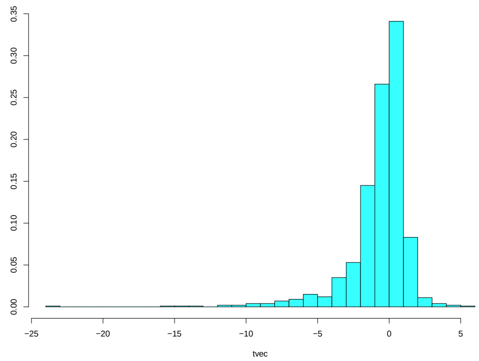


Figure 8: A simulated set of t-values based on data from an exponential distribution.

If we know an underlying process we can simulate data from the process and evaluate the distribution of any quantity based on such data.

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

but we want to know the distribution of those when μ is the true mean.

For instance, $n = 5$ and $\mu = 1$, we can simulate (repeatedly) $x_1, \ldots, x_5$ and compute a t-value for each. The following R commands can be used for this:

```
> library(MASS)
> n<-5
> mu<-1
> lambda<-1
> tvec<-NULL
> for(sim in 1:10000){
>   x<-rexp(n,lambda)
>   xbar<-mean(x)
>   s<-sd(x)
>   t<-(xbar-mu)/(s/sqrt(n))
>   tvec<-c(tvec,t)
> }

#then do...

> truehist(tvec)  #truehist gives a better histogram
> sort(tvec)[9750]
> sort(tvec)[250]
```

newpage inserted

newpage inserted

9 Miscellanea

9.1 Simple probabilities in R

R has functions to compute probabilities based on most common distributions.
If X is a random variable with a known distribution, then R can typically compute values of the c.d.f. or:

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

newpage inserted in example

newpage inserted in example

Example 1

If $X \sim b(n, p)$ has binomial distribution, i.e.

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

then cumulative probabilities can be computed with pbinom, e.g.

```
>pbinom(5,10,0.5)
```

gives

$$P[X \leq 5] = 0.623$$

where

$$X \sim b(n = 10, p = \frac{1}{2})$$

Further,

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

and

```
>pbinom(0,10,0.5)
[1] 0.0009765625
```

or $\frac{1}{2}^{10} = \frac{1}{1024}$ $(\frac{1}{2} * \frac{1}{2} * \ldots\ldots * \frac{1}{2} = \frac{1}{2^{10}} = \frac{1}{2024})$

It is sometime of interest to compute $P[X = x]$ in this case, and this is given by the dbinom function, e.g.

```
>dbinom(1,10,0.5)
[1] 0.009765625
```

or $\frac{10}{1024}$

newpage inserted in example

newpage inserted in example

Example 2

Suppose X has a uniform distribution between 0 and 1, i.e. $X \sim U(0,1)$.
Then the punif function will return probabilities of the form

$$P[X \leq x] = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} f(t)dt$$

where $f(t) = 1$ if $0 \leq t \leq 1$ and $f(t) = 0$ otherwise. For example:

```
>punif(0.75)
[1]  0.75
```

To obtain $P[a \leq X \leq b]$, we use punif twice, e.g.

```
>punif(0.75)-punif(0.25)
[1]  0.5
```

newpage inserted in example

newpage inserted in example

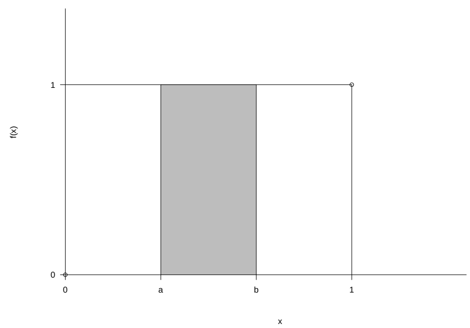


Figure 9: Example 2

newpage inserted

newpage inserted

9.2 Computing normal probabilities in R

> To compute probabilities $X \sim n(\mu, \sigma^2)$ is usually transformed, since we know that
>
> $$Z := \frac{X-\mu}{\sigma} \sim (0,1)$$
>
> The probabilities can then be computed for either X or Z with the pnorm function in R.

Suppose X has a normal distribution with mean μ and variance

$$X \sim n(\mu, \sigma^2)$$

then to compute probabilities, X is usually transformed, since we know that

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

and the probabilities can be computed for either X or Z with the pnorm function.

Example 1:
If $Z \sim n(0,1)$ then we can e.g. obtain $P[Z \leq 1.96]$ with

```
> pnorm(1.96)
[1] 0.9750021

> pnorm(0)
[1] 0.5

> pnorm(1.96)-pnorm(1.96)
[1] 0

> pnorm(1.96)-pnorm(-1.96)
[1] 0.9500042
```

The last one gives the area between -1.96 and 1.96.

Example 2:
If $X \sim n(42, 3^2)$ then we can compute probabilites either by transforming

$$\begin{aligned} P[X \leq x] &= P[\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}] \\ &= P[Z \leq \frac{x-\mu}{\sigma}] \end{aligned}$$

and calling pnorm with the computed value $z = \frac{x-\mu}{\sigma}$, or call pnorm with x and specify μ and σ.

To compute $P[X \leq 48]$, either set $z = (48-42)/3 = 2$ and obtain

```
> pnorm(2)
[1] 0.9772499
```

or specify μ and σ

```
> pnorm(42,42,3)
[1] 0.5
```

newpage inserted

newpage inserted

9.3 Introduction to hypothesis testing

If we have a random sample $x_1,...,x_n$ from a normal distribution, then we consider them to be outcomes of independent random variables $X_1,...,X_n$ where

$X_i \sim n(\mu,\sigma^2)$ Typically, μ and σ^2 are unknown but assume for now that σ^2 is known.

Consider the hypothesis

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$

where μ_0 is a specified number.

Under the assumption of independence, the sample mean

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

is also an observation from a normal distribution, with mean μ but a smaller variance.

Specifically, $\overline{x}$ is the outcome of

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

and

$X \sim n(\mu, \frac{\sigma^2}{n})$

so the standard deviation of X is $\frac{\sigma}{\sqrt{n}}$,

so the appropriate error measure for $\overline{x}$ is $frac\sigma\sqrt{n}$, when σ is unknown.

If H_0 is true, then

$z := \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$

is an observation from an $n \sim n(0,1)$ distribution, i.e. an outcome of

$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$

where $Z \sim n(0,1)$ when H_0 is correct. It follows that e.g. $P[|Z| > 1.96] = 0.05$ and if we observe $|Z| > 1.96$ then we reject the null hypothesis.

Note that the value z* = 1.96 is quantile of the normal distribution and we can obtain other quantiles with the pnorm function, e.g. qnorm(0.975) gives 1.96.

newpage inserted

newpage inserted