Probability - Session 4

Continuous probability distributions

Elizabeth Williamson
with thanks to Jennifer Rogers

Foundations of Medical Statistics

Session objectives

By the end of this session you should be able to:

- explain the concept of a continuous random variable
- define a probability density function
- calculate the expectation and variance for continuous random variables
- describe key continuous distributions which arise in medical statistics, including the normal

Outline

Continuous probability distributions

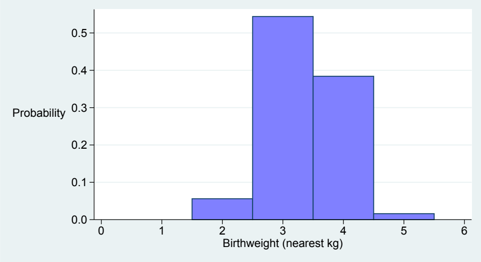
The normal distribution

Some other continuous distributions

Discrete probability distributions

Recall that a discrete random variable X

- takes values in a finite or countably infinite set, and
- is characterised by the probabilities $P(X = x)$ for the possible values x it might take.
  - e.g. $X =$ number of boys in a 4-child family
  - e.g. $X =$ birthweight measured to the nearest kg



Continuous probability distributions

Continuous random variables take values in an uncountably infinite set.

- Measured birthweight (to nearest kg) is discrete
- True birthweight is continuous.

What is the distribution of true birthweight?

- e.g. what is $P(X = 3.48641521654231)$?
- The probability of any value x is zero.
- Although X will take some value, there are so many possible values that the probability of a particular one is zero.

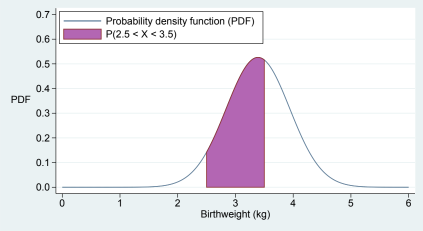
Therefore, we don't assign probabilities to particular values.

Probability density functions

- A continuous random variable X has a **probability density function (PDF)** $f(x)$ associated with it.
- Given two values a and b (with $a < b$):

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

- This is the area under the curve $f(x)$ between $x = a$ and $x = b$:



Probability density functions

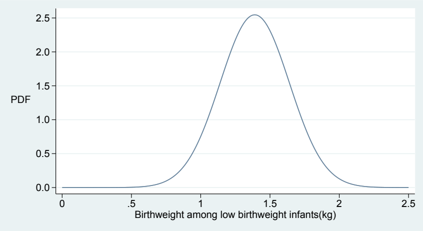
Probability density functions must satisfy a number of conditions.

Continuous X Probability density function	Discrete X Probability distribution function
$f(x) \geq 0$ for all x	$0 \leq P(X = x) \leq 1$ for all x
$\int_{-\infty}^{\infty} f(x)dx = 1$	$\sum_x P(X = x) = 1$

Note that it is possible to have $f(x) > 1$

Probability density functions

The PDF of birthweight among low birth-weight infants (under 2500g)



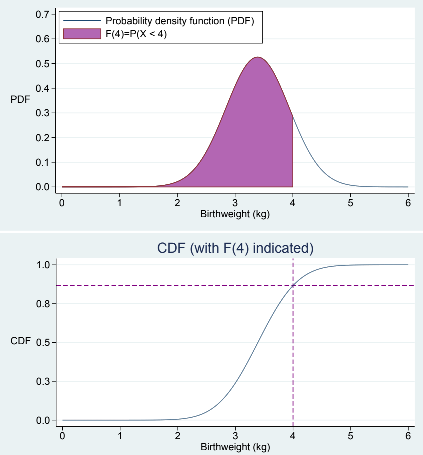
Note $f(x) > 1$ towards the centre of the graph

Cumulative distribution function

- For continuous X, the CDF is defined as:

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

i.e. the area under the PDF curve to the left of x.



Cumulative distribution function

- Probabilities can be expressed using the CDF too:

$$
\begin{aligned}
P(a \leq X \leq b) &= \int_a^b f(x)dx \\
&= \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx \\
&= F(b) - F(a).
\end{aligned}
$$

- The CDF is the integral of the PDF, so we also have

$$f(x) = \frac{d}{dx}F(x).$$

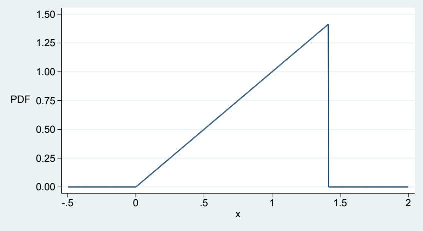
Expectation and variance

Continuous X	Discrete X
Expectation	
$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx.$	$E(X) = \sum_x x\, P(X = x)$
Variance $= E((X-\mu)^2) = E(X^2) - E(X)^2$	
$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx.$	$Var(X) = \sum_x (x-\mu)^2\, P(X = x)$

Exercise

- Let X be the random variable with density function

$$f(x) = \begin{cases} x & \text{for } 0 \leq x \leq \sqrt{2} \\ 0 & \text{elsewhere} \end{cases}$$



- What are $E(X)$ and $Var(X)$?

Solution

$$E(X) =$$
$$=$$
$$=$$
$$=$$

$$
\begin{aligned}
E(X^2) &= \int_0^{\sqrt{2}} x^2 f(x) dx \\
&= \int_0^{\sqrt{2}} x^3 dx \\
&= \left[\frac{x^4}{4} \right]_0^{\sqrt{2}} \\
&= \frac{2^{4/2}}{4} \\
&= 1.
\end{aligned}
$$

$$Var(X) =$$

Solution

$$\begin{aligned} E(X) &= \int_0^{\sqrt{2}} x f(x) dx \\ &= \int_0^{\sqrt{2}} x^2 dx \\ &= \left[\frac{x^3}{3}\right]_0^{\sqrt{2}} \\ &= \frac{2^{3/2}}{3}. \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_0^{\sqrt{2}} x^2 f(x) dx \\ &= \int_0^{\sqrt{2}} x^3 dx \\ &= \left[\frac{x^4}{4}\right]_0^{\sqrt{2}} \\ &= \frac{2^{4/2}}{4} \\ &= 1. \end{aligned}$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= 1 - \left(\frac{2^{3/2}}{3}\right)^2 = 1 - \frac{2^3}{9} = 1 - \frac{8}{9} = \frac{1}{9} \end{aligned}$$

Properties of expectation and variance

The properties we met in the discrete case hold for continuous random variables.

If a and b are constants, and X is a continuous random variable, then

$$E(aX + b) = aE(X) + b$$
$$Var(aX + b) = a^2 Var(X)$$

For two continuous random variables, X and Y,

$$E(X + Y) = E(X) + E(Y)$$

If X and Y are also *independent* then

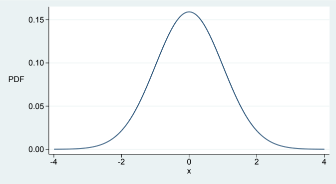
$$E(XY) = E(X)E(Y)$$
$$Var(X + Y) = Var(X) + Var(Y)$$

Outline

The normal distribution

The normal distribution is a very (the most?) important distribution in statistics.

- It is often a useful distribution when modelling data.
- It also features throughout statistical theory because of the central limit theorem (see Session 5).
- Its PDF looks like:

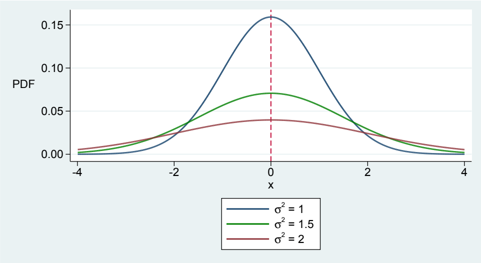


- Referred to as bell shaped curve.

The normal distribution

Unlike the distributions we have met so far, the normal distribution has two parameters, usually denoted μ and σ^2.

- The mean μ
  - determines where the centre of the distribution is.
  - below, the three distributions all have $\mu = 0$.
- The variance σ^2
  - determines how narrow or spread the distribution is.
  - The larger σ^2 is, the more spread out the distribution



Properties of the normal distribution

- If X follows a normal distribution with parameters μ and σ^2, we write
$$X \sim N(\mu, \sigma^2)$$
- The probability density function is given by:
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
- The expectation and variance of a normally-distributed variable X is given by
$$\begin{aligned} E(X) &= \mu \\ Var(X) &= \sigma^2 \end{aligned}$$

The standard normal distribution

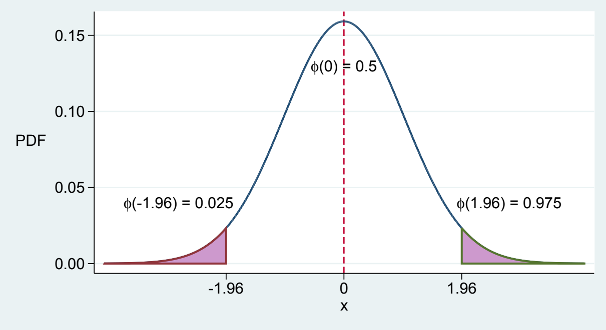
The standard normal distribution (is normal, and) has mean 0 and variance 1.

- We write $Z \sim N(0, 1)$.
- The probability density function of the standard normal distribution is:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{z^2}{2} \right)$$

- Its CDF is often denoted $\Phi(z)$.

The standard normal distribution



- 95% of the area is contained within 1.96 units of the mean.
- Note that $\phi(-x) = 1 - \phi(x)$

Transforming to the standard normal

Any normally distributed variable X can be transformed to a standard normal variable using the transformation:

$$Z = \frac{X - \mu}{\sigma}$$

- Then $Z \sim N(0, 1)$

For example, suppose heart rate (X) is normally distributed, $X \sim N(74, 7.5^2)$, then $Z \sim N(0, 1)$, where

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 74}{7.5}$$

Calculating probabilities for the normal

- Probabilities for $X \sim N(\mu, \sigma^2)$ could be calculated from the CDF:

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt.$$

- Unfortunately the integral cannot be solved in closed form
- Can approximate via numerical integration

- Statistical tables usually contain $F(z) = \Phi(z)$ for useful values of z (i.e. -4 to 4).
- This allows us to find $P(Z \leq z)$ where $Z \sim N(0, 1)$.
- By transforming normal variables to a standard normal, this also allows us to find $P(X \leq z)$

Example: Calculating normal probabilities

What is the probability of having a heart rate less than or equal to 60 bpm, i.e. $P(X \leq 60)$?

We know that

$$Z = \frac{X - 74}{7.5} \sim N(0, 1)$$

We can write

$$\begin{aligned} P(X \leq 60) &= P\left(\frac{X - 74}{7.5} \leq \frac{60 - 74}{7.5}\right) \\ &= P\left(Z \leq \frac{60 - 74}{7.5}\right) \\ &= P(Z \leq -1.867) \\ &= \Phi(-1.867) \\ &= 1 - \Phi(1.867) \\ &= 0.031 \text{ from Neave tables.} \end{aligned}$$

Example: Calculating normal probabilities

What is the probability of having a heart rate between 50 and 60 bpm?

This is

$$Pr(50 \leq X \leq 60) = Pr(X \leq 60) - Pr(X \leq 50)$$

Then we apply the method of the previous slide separately to each term.

In general: Calculating normal probabilities

Suppose $X \sim N(\mu, \sigma^2)$, and we want to find $P(a \leq X \leq b)$.

- Transform X to Z,

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

- Then:

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

Normal expectation - I

The expectation of a normally distributed variable, X, is

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

Normal expectation - I

The expectation of a normally distributed variable, X, is

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

- To calculate this quantity, it is useful to make a transformation. Let

$$z = \frac{(x-\mu)}{\sigma}$$

Normal expectation - I

The expectation of a normally distributed variable, X, is

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

- To calculate this quantity, it is useful to make a transformation. Let
$$z = \frac{(x-\mu)}{\sigma}$$
- Then $x = \sigma z + \mu$ and $dx = \sigma dz$.

Normal expectation - I

The expectation of a normally distributed variable, X, is

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

- To calculate this quantity, it is useful to make a transformation. Let
$$z = \frac{(x-\mu)}{\sigma}$$
- Then $x = \sigma z + \mu$ and $dx = \sigma dz$.
- Then,

$$E(X) = \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma dz$$

Normal expectation - II

$$E(X) = \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz$$

Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz
\end{aligned}
$$

Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz
\end{aligned}
$$

Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma E(Z) + \mu
\end{aligned}
$$

Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma E(Z) + \mu \\
&= \sigma E\left(\frac{X-\mu}{\sigma}\right) + \mu
\end{aligned}
$$

Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma E(Z) + \mu \\
&= \sigma E\left(\frac{X-\mu}{\sigma}\right) + \mu \\
&= \sigma \frac{(E(X)-\mu)}{\sigma} + \mu
\end{aligned}
$$

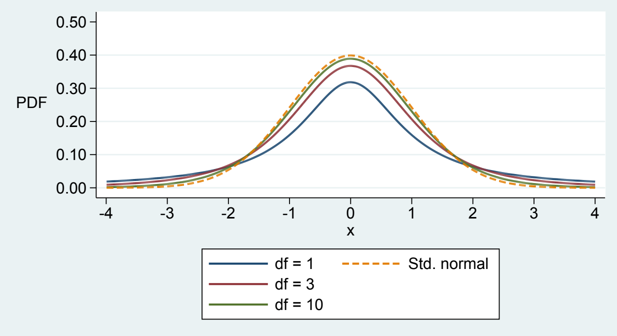
Normal expectation - II

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} \frac{\sigma z + \mu}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \sigma\, dz \\
&= \int_{-\infty}^{\infty} \frac{\sigma z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \sigma E(Z) + \mu \\
&= \sigma E\left(\frac{X-\mu}{\sigma}\right) + \mu \\
&= \sigma \frac{(E(X)-\mu)}{\sigma} + \mu \\
&= \mu
\end{aligned}
$$

Outline

Student's t-distribution

- Student's t-distribution is like a normal with heavy tails
- Additional parameter known as degrees of freedom (df).
- As the df increase, the distribution gets closer to the normal.



- Student's t-distribution arises commonly in statistics.
- Ratio of sample mean to standard error (Inference).

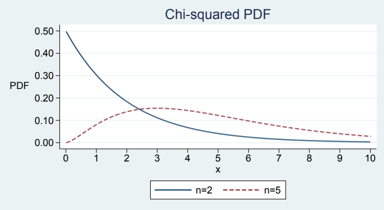
Chi-squared distribution

The sum of squared independent standard normal distributions

- If $X_1, .., X_n \sim N(0, 1)$ and are independent, then:

$$Q = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

  follows a chi-squared distribution on n degrees of freedom.
- $E(Q) = n,\ Var(Q) = 2n$



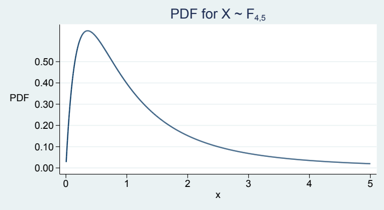
The F-distribution

Used in hypothesis tests in ANOVA and linear regression.

- $U_1 \sim \chi^2_n, \quad U_2 \sim \chi^2_m$
- U_1, U_2 are independent (see Session 5)
- Then:

$$F = \frac{U_1/n}{U_2/m}$$

- We write $F \sim F_{n,m}$.



The exponential distribution

- Suppose events follow a Poisson process and occur at rate λ
- Suppose X is the time between two consecutive events.
- Then X follows the exponential distribution, $X \sim exp(\lambda)$

- It has PDF:

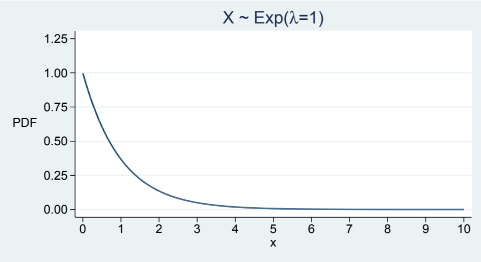
$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0$$

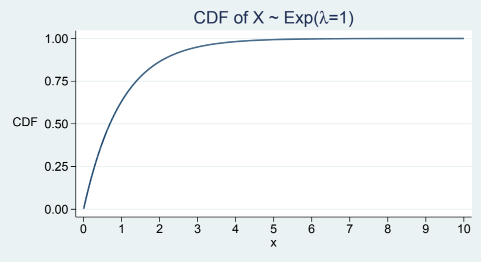
and CDF:

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

- The mean time is $E(X) = 1/\lambda$
- The variance is $Var(X) = 1/\lambda^2$

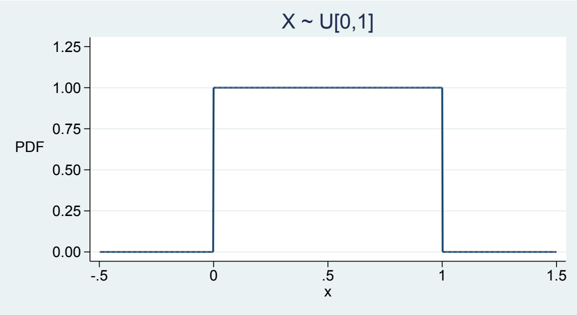
The exponential distribution





The (continuous) uniform distribution

The PDF of the standard continuous uniform distribution is shown below



The (continuous) uniform distribution

- $X \sim U(0, 1)$ if

$$f(x) = \begin{cases} 1 & \text{for} \quad 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

- More generally, $X \sim U(a, b)$ if

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for} \quad a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

- The mean is $E(X) = \frac{1}{2}(a + b)$
- The variance is $Var(X) = \frac{1}{12}(b - a)^2$

Summary

- Continuous random variables are characterised by a probability density function.
- Expectations and variances for continuous random variables are calculated via integration.
- Introduced the normal distribution, and how we can calculate probabilities based on the CDF of the standard normal.
- Other continuous distributions which arise in medical statistics: t-distribution, chi-squared, F, exponential, uniform.