## Contents

# 1 Some properties of the normal distribution

The normal distribution is parameterised by two things: the distribution's **mean** and **variance**. Commonly we denote them with $\mu$ and $\sigma^2$.

Say we have a random variable that follows the normal distribution, say $X \sim N(\mu, \sigma^2)$. This variable has mean $\mu$ and variance $\sigma^2$, or equivalently, standard deviation $\sigma$.

We can shift the mean left or right (negatively or positively) by adding or subtracting a fixed constant to $X$. Consider $(X + c)$, which will have mean $\mu + c$. The variance remains unchanged.

The variance can be changed by multiplying a fixed constant to $X$. $(cX)$ will have variance $c^2\sigma^2$. The mean is also scaled accordingly to $c\mu$.

So, to transform this variable $X \sim N(\mu, \sigma^2)$ into the standard normal distribution $Z \sim N(0, 1)$, we have to subtract the mean $\mu$ and divide by $\sigma$.

$$X \sim N(\mu, \sigma^2) \iff (X - \mu) \sim N(0, \sigma^2) \iff \frac{X - \mu}{\sigma} \sim N(0, 1)$$

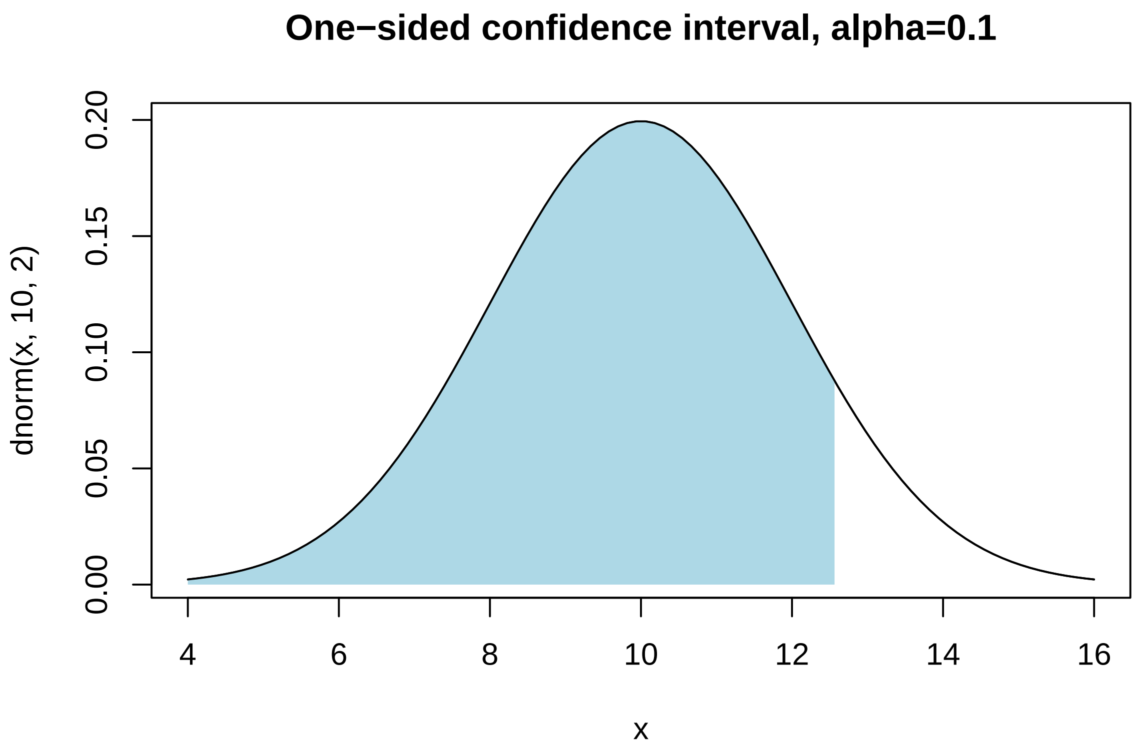
This actually corresponds with the properties of expectation and variance, i.e. $E(aX+b) = aE(X)+b$ and $Var(aX + b) = a^2Var(X)$.

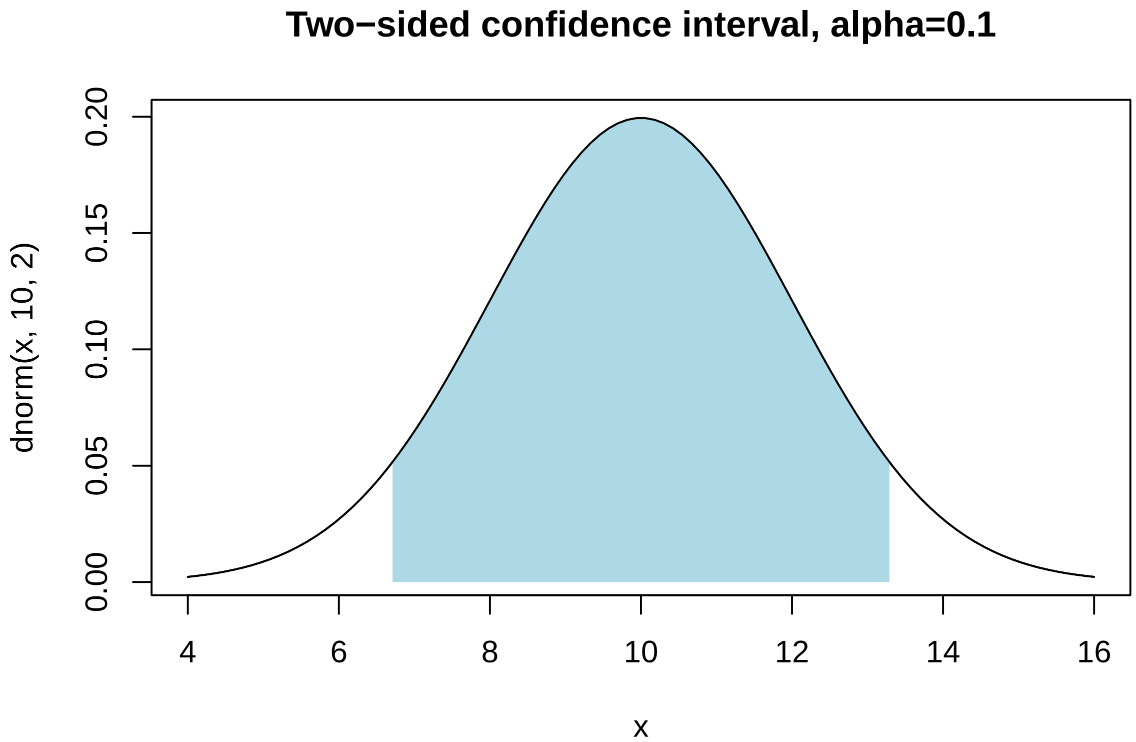
# 2 Confidence intervals

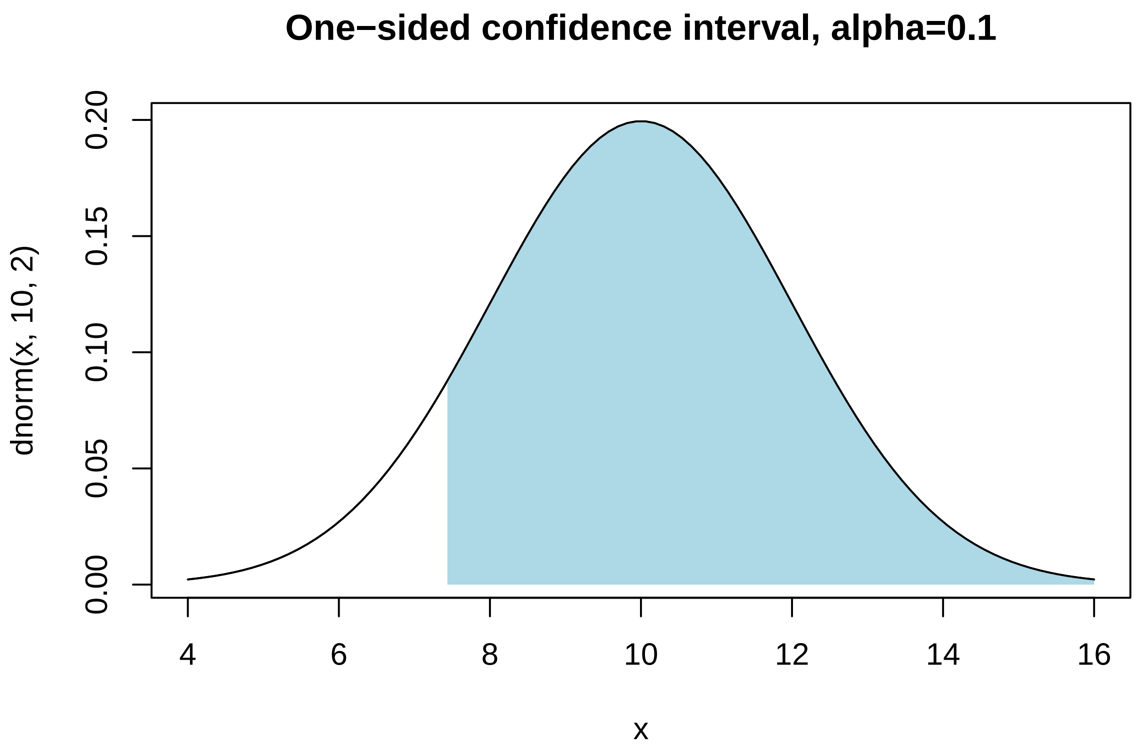
Confidence intervals are expressed with some value between 0 and 1. In the context of estimating the sample mean, for example, a 0.9-confidence interval represents an interval that has a 90% probability of containing the true population mean.

Here, what we really want is the population mean - but we since that is typically infeasible to calculate, we use the sample mean as an estimate. Typically, the sample mean has a nice distribution (normally distributed or t-distributed), thanks to the central limit theorem.

Say we find that our sample mean $\bar{X} \sim N(10, 4)$. To find a 90% confidence interval, we want to find an interval such that the area under the p.d.f. is 0.9. There are infinitely many - typically it is chosen to be one-sided or two-sided. All of the graphs below represent a valid 0.9-confidence interval - all of them have an area under curve of 0.9.


One−sided confidence interval, alpha=0.1
dnorm(x, 10, 2)
0.00
0.05
0.10
0.15
0.20
4
6
8
10
12
14
16
x


Two−sided confidence interval, alpha=0.1
dnorm(x, 10, 2)
0.00
0.05
0.10
0.15
0.20
4
6
8
10
12
14
16
x




These intervals are constructed with the quantile functions. Recall that the quantile function $Q$ is the inverse c.d.f.

If I take the interval $(Q(0), Q(0.9))$, this gives exactly a (one-sided) 0.9-confidence interval. Likewise, the interval $(Q(0.05), Q(0.95))$ also gives a (two-sided) 0.9-confidence interval.

It is common to parameterise this confidence interval, usually with $\alpha$ as the significance level in hypothesis testing, thus forming $(1-\alpha)$-confidence intervals, which are $(Q(0), Q(1-\alpha))$, $(Q(\alpha), Q(1))$, and $(Q(\alpha/2), Q(1-\alpha/2))$.

For example, if $\alpha = 0.05$, we get 0.95-confidence intervals.