Likelihood and maximum likelihood estimation

The ASTA team

Contents

1 The probability density/mass function 1

2 The likelihood function for a single observation 2

3 The likelihood function for n observations 3

4 Example 5

1 The probability density/mass function

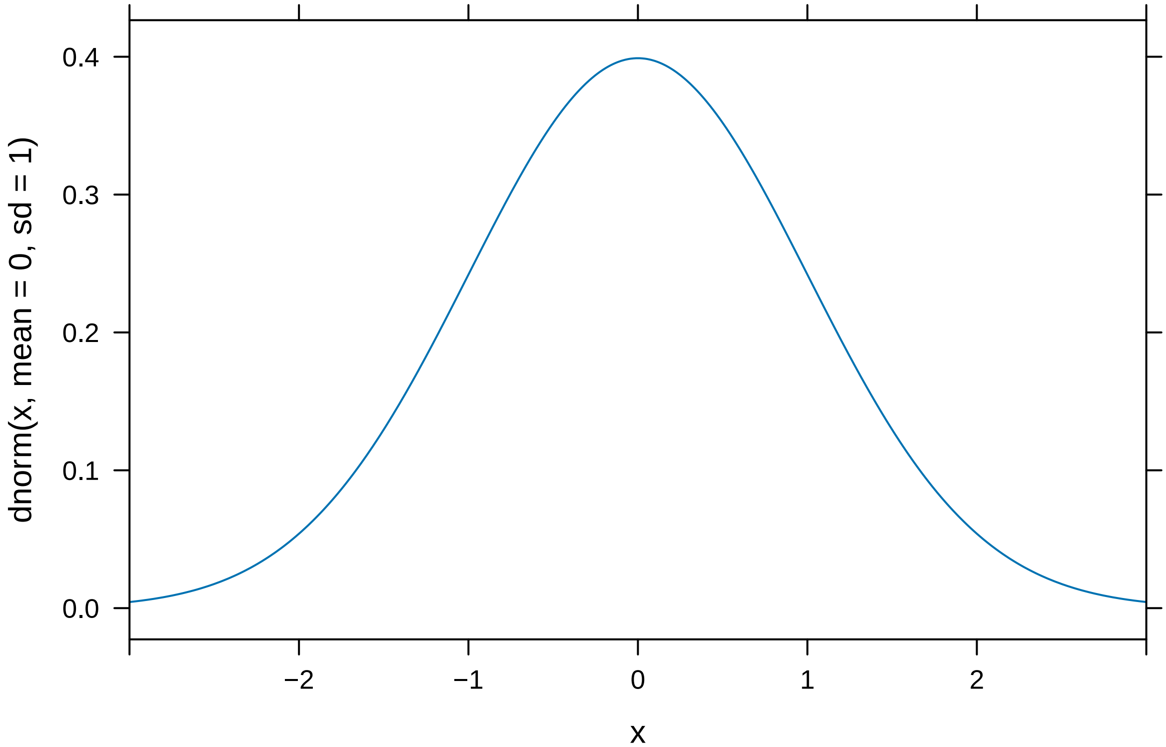
$$X \sim N(\mu, \sigma)$$

Assumes μ and σ are some fixed values:

$$f(x) = f(X = x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

E.g. for $\mu = 0$ and $\sigma = 1$:

```
plotFun(dnorm(x, mean = 0, sd = 1) ~ x, xlim = c(-3, 3))
```



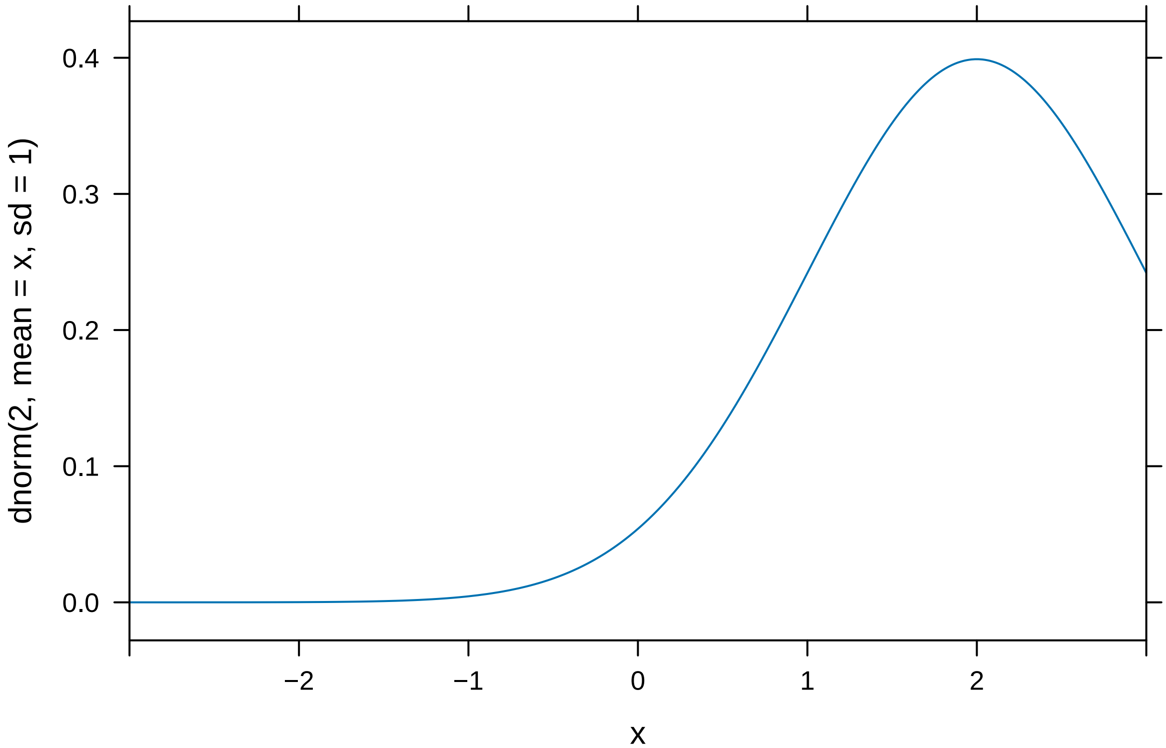
2 The likelihood function for a single observation

What if instead the data is fixed and we let μ and σ vary?

$$L(\mu, \sigma) = L(\mu, \sigma; X = x) = f(X = x; \mu, \sigma)$$

E.g. assume that $X = 2$ was observed:

```
plotFun(dnorm(2, mean = x, sd = 1) ~ x, xlim = c(-3, 3))
```



The value of μ that gives the highest probability (density) of observing $X = 2$ is $\mu = 2$.

3 The likelihood function for n observations

Assume independence for n observations, then

$$L(\mu, \sigma) = L(\mu, \sigma; X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} L(\mu, \sigma; X_i) = \prod_{i=1}^{n} L(\mu, \sigma; X_i). \quad (1)$$

- Maximum likelihood estimation (MLE):
  - Find the values of μ and σ that gives the largest value of L for fixed X's.
- Would you rather differentiate products or sums?

Do you know a function that turns products into sums?

$$\log(a \cdot b) = \log(a) + \log(b)$$

Assume independence for n observations, then

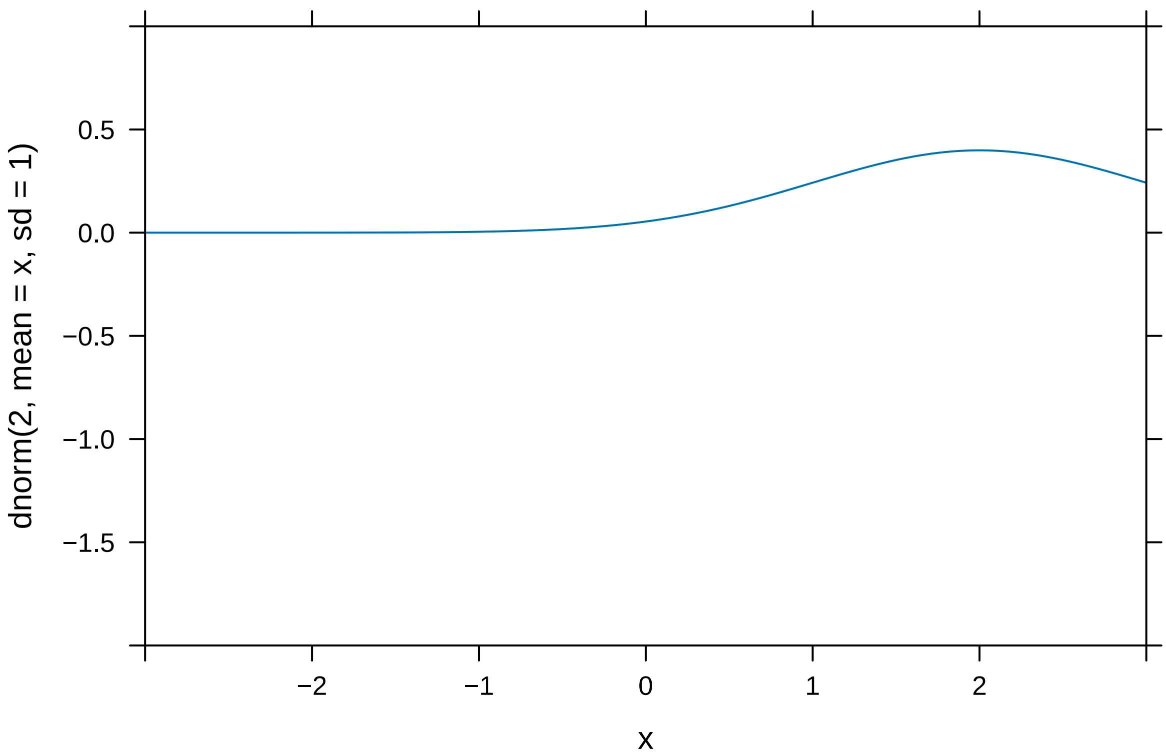
$$L(\mu,\sigma) = L(\mu,\sigma;X_1,X_2,\ldots,X_n) = \prod_{i=1}^{n} L(\mu,\sigma;X_i) = \prod_{i=1}^{n} L(\mu,\sigma;X_i) \tag{2}$$

$$\begin{aligned} l(\mu,\sigma) = \log L(\mu,\sigma) &= \log L(\mu,\sigma;X_1,X_2,\ldots,X_n) && (3)\\ &= \log\left(\prod_{i=1}^{n} L(\mu,\sigma;X_i)\right) && (4)\\ &= \sum_{i=1}^{n} \log L(\mu,\sigma;X_i) && (5)\\ &= \sum_{i=1}^{n} l(\mu,\sigma;X_i). && (6)\end{aligned}$$

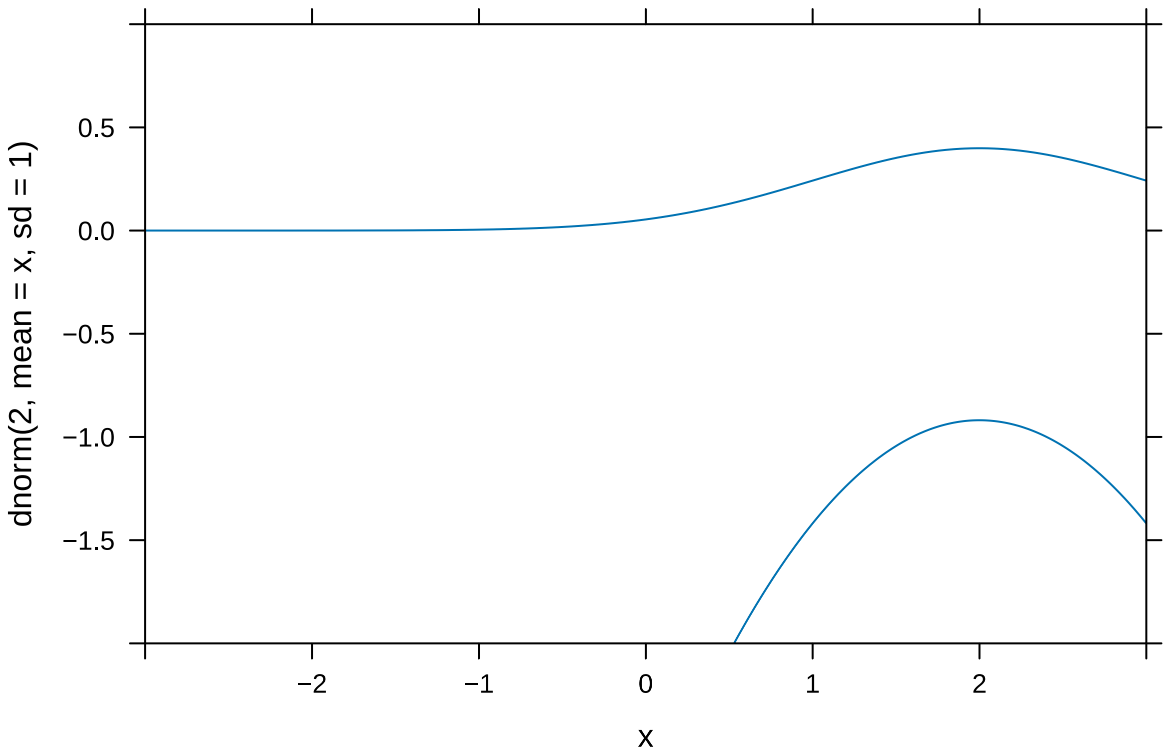
Note that $\log$ is a monotonic (increasing) function, so maximum of $l = \log L$ is the same as that of L.

$$\begin{aligned} L(\mu,\sigma) = L(\mu,\sigma;X=x) &= f(X=x;\mu,\sigma) && (7)\\ l(\mu,\sigma;X) = \log L(\mu,\sigma;X) &= \log f(X;\mu,\sigma) && (8)\end{aligned}$$

```
plotFun(dnorm(2, mean = x, sd = 1) ~ x, xlim = c(-3, 3), ylim = c(-2, 1));
```




```
plotFun(log(dnorm(2, mean = x, sd = 1)) ~ x, xlim = c(-3, 3), add = TRUE)
```

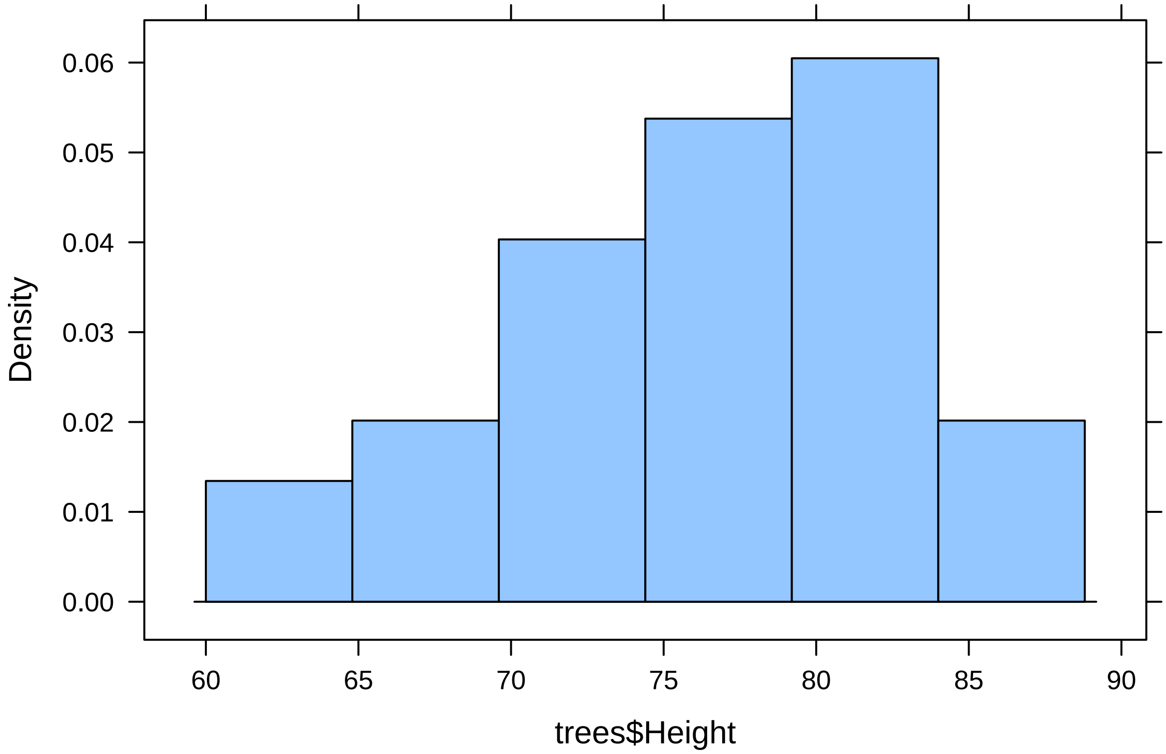


4 Example

```
trees <- read.delim("https://asta.math.aau.dk/datasets?file=trees.txt")
head(trees)
```

```
##   Girth Height Volume
## 1   8.3     70     10
## 2   8.6     65     10
## 3   8.8     63     10
## 4  10.5     72     16
## 5  10.7     81     19
## 6  10.8     83     20
```

```
histogram(trees$Height)
```



If Height is normally distributed, what are the parameters (mean and standard deviation)?

```
sd(trees$Height)
```

```
## [1] 6.4
```

```
single_loglik <- function(mu) {
  sum(log(dnorm(trees$Height, mean = mu, sd = 6)))
}
single_loglik(65)
```

```
## [1] -153
```

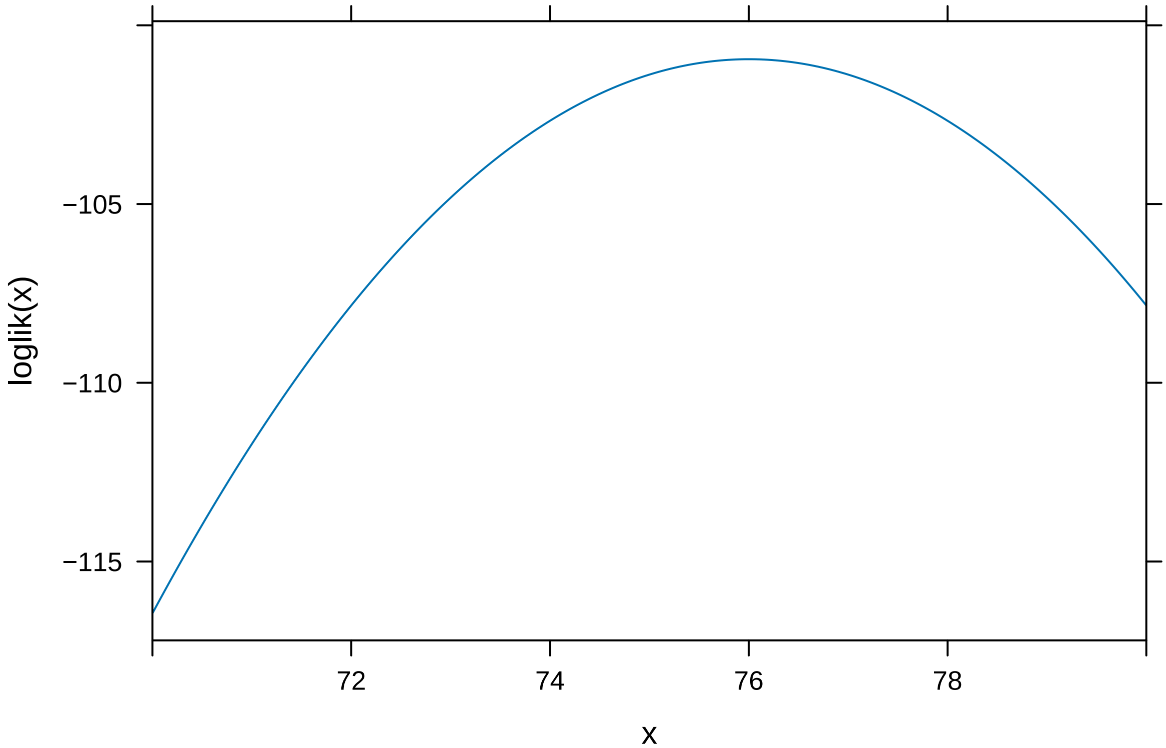
```
single_loglik(c(60, 65, 70))
```

```
## [1] -159
```

```
loglik <- Vectorize(single_loglik)
loglik(c(60, 65, 70))
```

```
## [1] -211 -153 -116
```

```
plotFun(loglik(x) ~ x, xlim = c(70, 80))
```



Maximum around 76.

```r
optimise(single_loglik, interval = c(0, 100))
```

```
## $minimum
## [1] 4.6e-05
##
## $objective
## [1] -2588
```

```r
optimise(single_loglik, interval = c(0, 100), maximum = TRUE)
```

```
## $maximum
## [1] 76
##
## $objective
## [1] -101
```

Maximum of l is minimum of $-l$.

```r
single_loglik <- function(mu) {
  -sum(log(dnorm(trees$Height, mean = mu, sd = 6)))
}
optimise(single_loglik, interval = c(0, 100))
```

```
## $minimum
## [1] 76
##
## $objective
## [1] 101
```

Both parameters:

```
single_loglik <- function(pars) {
  mu <- pars[1]
  sigma <- pars[2]
  -sum(log(dnorm(trees$Height, mean = mu, sd = sigma)))
}
optim(c(1, 50), single_loglik)
```

```
## Warning in dnorm(trees$Height, mean = mu, sd = sigma): NaNs produced

## Warning in dnorm(trees$Height, mean = mu, sd = sigma): NaNs produced

## Warning in dnorm(trees$Height, mean = mu, sd = sigma): NaNs produced
```

```
## $par
## [1] 76.0  6.3
##
## $value
## [1] 101
##
## $counts
## function gradient
##       87       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
single_loglik <- function(pars) {
  mu <- pars[1]
  sigma <- exp(pars[2])
  -sum(log(dnorm(trees$Height, mean = mu, sd = sigma)))
}
mle <- optim(c(1, 50), single_loglik)
mle
```

```
## $par
## [1] 76.0  1.8
##
## $value
## [1] 101
##
## $counts
## function gradient
##      113       NA
##
## $convergence
## [1] 0
##
## $message
```

```
## NULL
```

```r
exp(mle$par[2])
```

```
## [1] 6.3
```

```r
mean(trees$Height)
```

```
## [1] 76
```

```r
sd(trees$Height)
```

```
## [1] 6.4
```

```r
n <- length(trees$Height)
sd(trees$Height)
```

```
## [1] 6.4
```