Intermediate Econometrics

31st October 2025 - Vincenzo Gioia

Statistical Model in Brief

Statistical Model: The Simplest One

- Independent observations (on n statistical units) about a random variable of interest
- The sample is composed by the random variables $Y_1, \ldots, Y_n$ that are independent and identically distributed according to a certain probability distribution (which is known up to an unknown parameter)

$$Y_i \sim f(\cdot, \theta), \quad \theta \in \Theta \subset \mathrm{R}^d, \quad d \geq 1, \quad i = 1, \ldots, n$$

Assumption

- The random variable is described by a probability mass function (discrete r.v.) or density function (continuous r.v.)
- θ: d-dimensional parameter assuming value in the paramteric space Θ

Statistical Model: The Normal Model

Normal/Gaussian Random Variable

A random variable X is said to be Normal (or Gaussian) with mean $E[X] = \mu$ and variance $V(X) = \sigma^2$ if the density function takes the functional form

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- We denote it with $X \sim \mathcal{N}(\mu, \sigma^2)$

Standard Normal

A particular normal random variable is the standard normal, that is a normal with mean 0 and variance 1

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- We (usually) denote it using $Z \sim \mathcal{N}(0, 1)$ and it represents the core building for all the normal distributions

The Normal Distribution

Relationship between Standard Normal and a Generic Normal

- If $Z \sim \mathcal{N}(0, 1)$ then $X = \mu + \sigma Z$ is distributed as $X \sim \mathcal{N}(\mu, \sigma^2)$
- Viceversa, if $X \sim \mathcal{N}(\mu, \sigma^2)$ then the r.v. $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

Cumulative Distribution Function

- $F_X(x) = P(X \leq x) = P(\mu + \sigma Z \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi(\frac{x-\mu}{\sigma})$ where $\Phi(\cdot)$ is the cumulative distribution function of Z (whose values are obtained numerically)

Quantiles of Order $p \in (0, 1)$ of $X \sim \mathcal{N}(\mu, \sigma^2)$

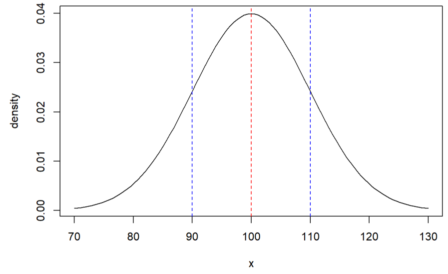
- $x_p = \mu + \sigma z_p$ where $z_p = \Phi^{-1}(p)$ is the quantile of order p of Z

The Normal distribution

Characteristics of $f_X(x)$

- It is symmetric, with the line $x = \mu$ as its axis of symmetry
- It is increasing on the interval $(-\infty, \mu)$ and decreasing on the interval $(\mu, +\infty)$
- It has two inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$
- It is concave on the interval $(\mu - \sigma, \mu + \sigma)$ and convex elsewhere
- It has the x-axis as an asymptote

```r
curve(dnorm(x, mean = 100, sd = 10), xlim = c(70, 130), ylab = "density")
abline(v = 100, lty = "dashed", col = "red")
abline(v = c(90, 110), lty = "dashed", col = "blue")
```

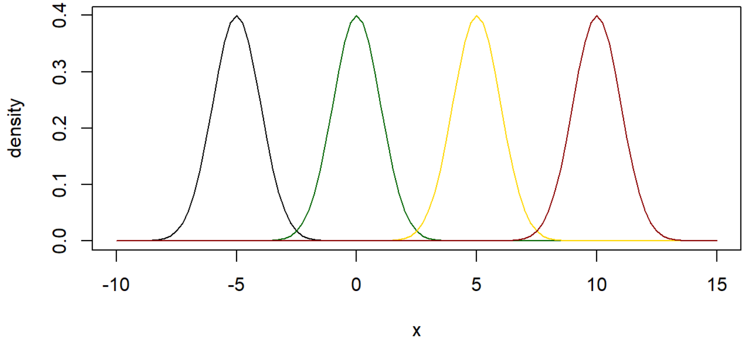


The Normal distribution

Location Shift

- $Z \sim \mathcal{N}(0, 1)$
- $X = \mu + Z \sim \mathcal{N}(\mu, 1)$
- Example considering $\mu \in \{-5, 0, 5, 10\}$

```
curve(dnorm(x, mean = -5, sd = 1), xlim = c(-10, 15), ylab = "density")
curve(dnorm(x, mean = 0, sd = 1), add = TRUE, col = "darkgreen")
curve(dnorm(x, mean = 5, sd = 1), add = TRUE, col = "gold")
curve(dnorm(x, mean = 10, sd = 1), add = TRUE, col = "darkred")
```

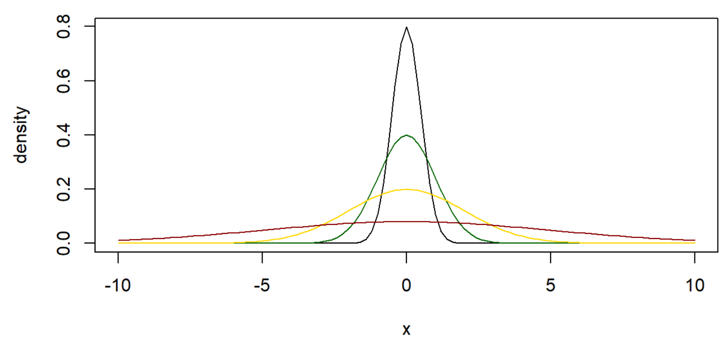


The Normal distribution

Scaling

- $Z \sim \mathcal{N}(0, 1)$
- $X = \sigma Z \sim \mathcal{N}(0, \sigma^2)$
- Example considering $\sigma \in \{0.5, 1, 2, 5\}$

```r
curve(dnorm(x, mean = 0, sd = 0.5), xlim = c(-10, 10), ylab = "density")
curve(dnorm(x, mean = 0, sd = 1), add = TRUE, col = "darkgreen")
curve(dnorm(x, mean = 0, sd = 2), add = TRUE, col = "gold")
curve(dnorm(x, mean = 0, sd = 5), add = TRUE, col = "darkred")
```

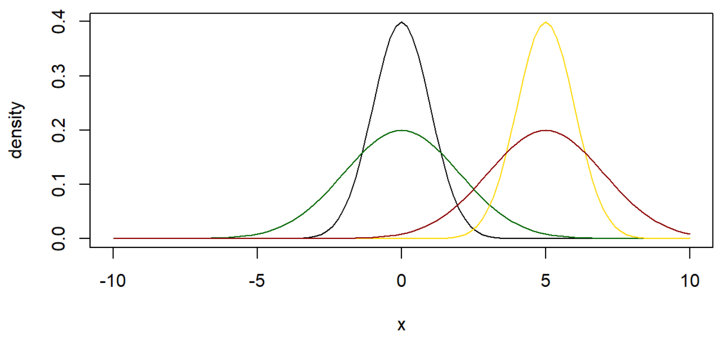


The Normal distribution

Location Shift and Scaling

- $Z \sim \mathcal{N}(0, 1)$
- $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$
- Example considering all the combination of $\mu \in \{0, 5\}$ and $\sigma \in \{1, 2\}$

```
curve(dnorm(x, mean = 0, sd = 1), xlim = c(-10, 10), ylab = "density")
curve(dnorm(x, mean = 0, sd = 2), add = TRUE, col = "darkgreen")
curve(dnorm(x, mean = 5, sd = 1), add = TRUE, col = "gold")
curve(dnorm(x, mean = 5, sd = 2), add = TRUE, col = "darkred")
```

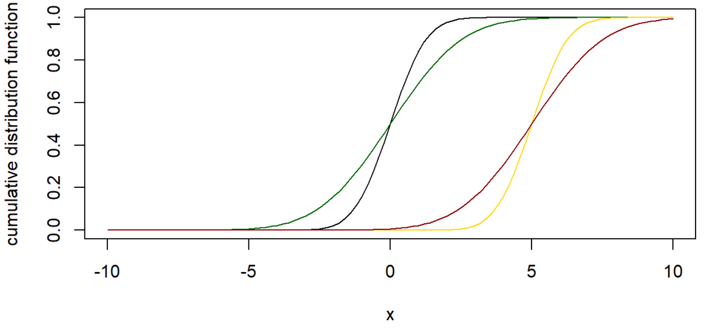


The Normal distribution

The Cumulative Distribution Function $F_X(x)$

- $Z \sim \mathcal{N}(0, 1)$
- $X = \mu + \sigma Z \sim \mathcal{N}(0, \sigma^2)$
- Example considering all the combinations of $\mu \in \{0, 5\}$ and $\sigma \in \{1, 2\}$

```
curve(pnorm(x, mean = 0, sd = 1), xlim = c(-10, 10), ylab = "cumulative distributior
curve(pnorm(x, mean = 0, sd = 2), add = TRUE, col = "darkgreen")
curve(pnorm(x, mean = 5, sd = 1), add = TRUE, col = "gold")
curve(pnorm(x, mean = 5, sd = 2), add = TRUE, col = "darkred")
```



The Normal distribution

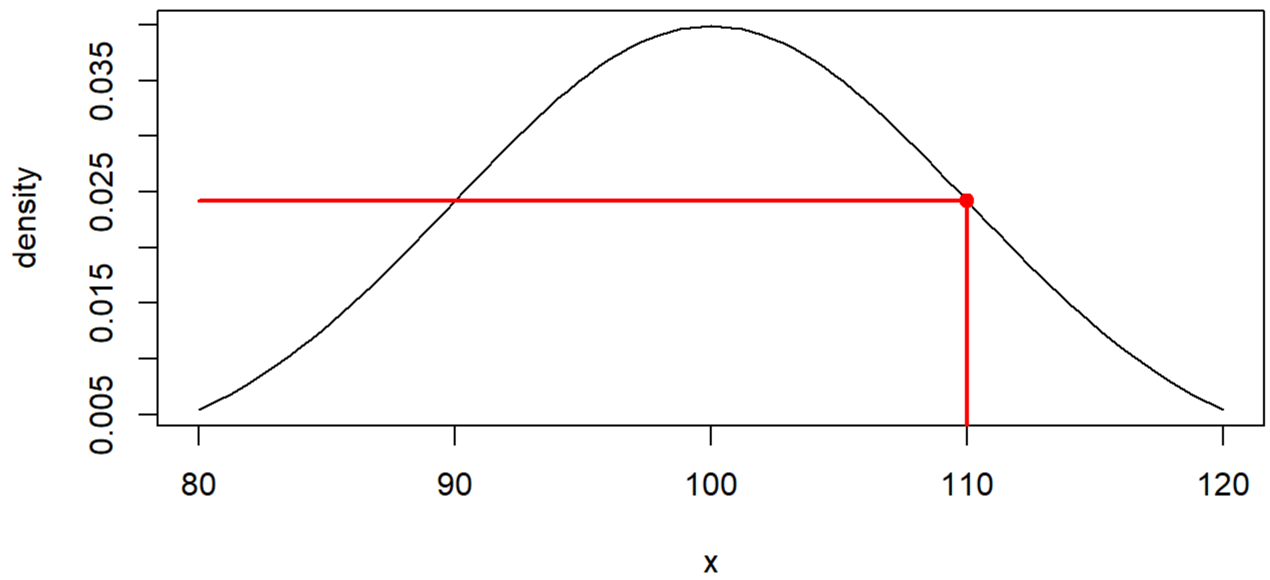
The Density Function $f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

- In R, the density function is obtained by means of the dnorm() function

```r
curve(dnorm(x, mean = 100, sd = 10), xlim = c(80,120), ylab = "density")
dnorm(110, mean = 100, sd = 10)
```

```
[1] 0.02419707
```

```r
y <- dnorm(110, mean = 100, sd = 10)
segments(x0 = 110, y0 = 0, x1 = 110, y1 = y, col = "red", lwd = 2)
segments(x0 = 80, y0 = y, x1 = 110, y1 = y, col = "red", lwd = 2)
points(110, y, col = "red", pch = 19)
```



density
0.005
0.015
0.025
0.035
80
90
100
110
120
x

The Normal distribution

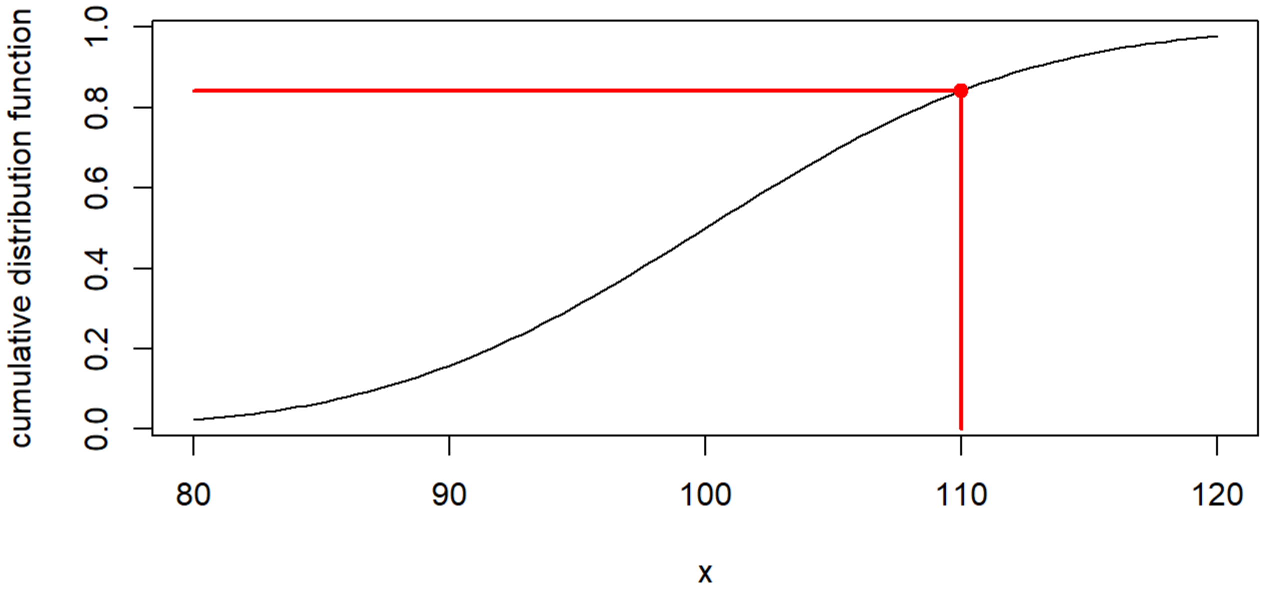
The Cumulative Distribution Function $F_X(x; \mu, \sigma^2) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

- In R, the cumulative distribution function is obtained by means of the pnorm() function

```
curve(pnorm(x, mean = 100, sd = 10), xlim = c(80,120), ylab = "cumulative distributi
pnorm(110, mean = 100, sd = 10)
```

```
[1] 0.8413447
```

```
y <- pnorm(110, mean = 100, sd = 10)
segments(x0 = 110, y0 = 0, x1 = 110, y1 = y, col = "red", lwd = 2)
segments(x0 = 80, y0 = y, x1 = 110, y1 = y, col = "red", lwd = 2)
points(110, y, col = "red", pch = 19)
```



cumulative distribution function
0.0
0.2
0.4
0.6
0.8
1.0
80
90
100
110
120
x

The Normal distribution

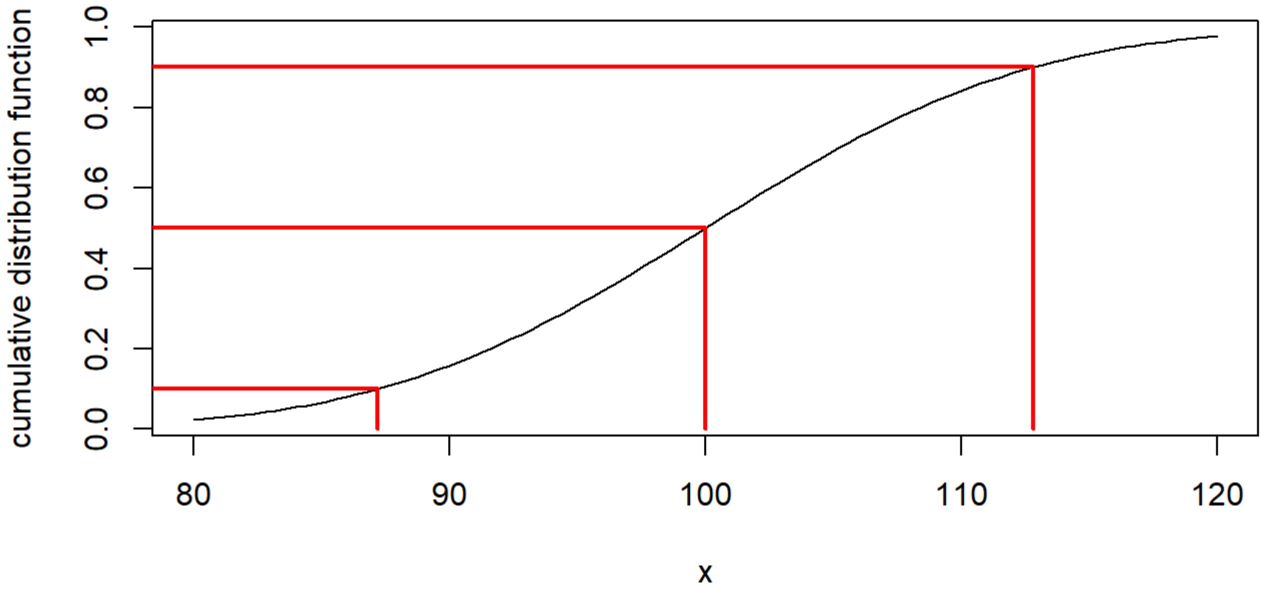
Quantiles of Order $p \in (0,1)$ of $X \sim \mathcal{N}(\mu, \sigma^2)$

- $x_p = \mu + \sigma z_p$ where $z_p = \Phi^{-1}(p)$ is the quantile of order p of Z
- In R, the quantiles are obtained by means of the qnorm() function

```
curve(pnorm(x, mean = 100, sd = 10), xlim = c(80,120), ylab = "cumulative distributi
qnorm(c(0.1,0.5,0.9), mean = 100, sd = 10)
```

```
[1]  87.18448 100.00000 112.81552
```

```
x <- qnorm(c(0.1,0.5,0.9), mean = 100, sd = 10)
segments(x0 = c(0,0,0), y0 = c(0.1, 0.5, 0.9), x1 = x, y1 = c(0.1, 0.5, 0.9), col = 
segments(x0 = x, y0 = 0, x1 = x, y1 = c(0.1, 0.5, 0.9), col = "red", lwd = 2)
```

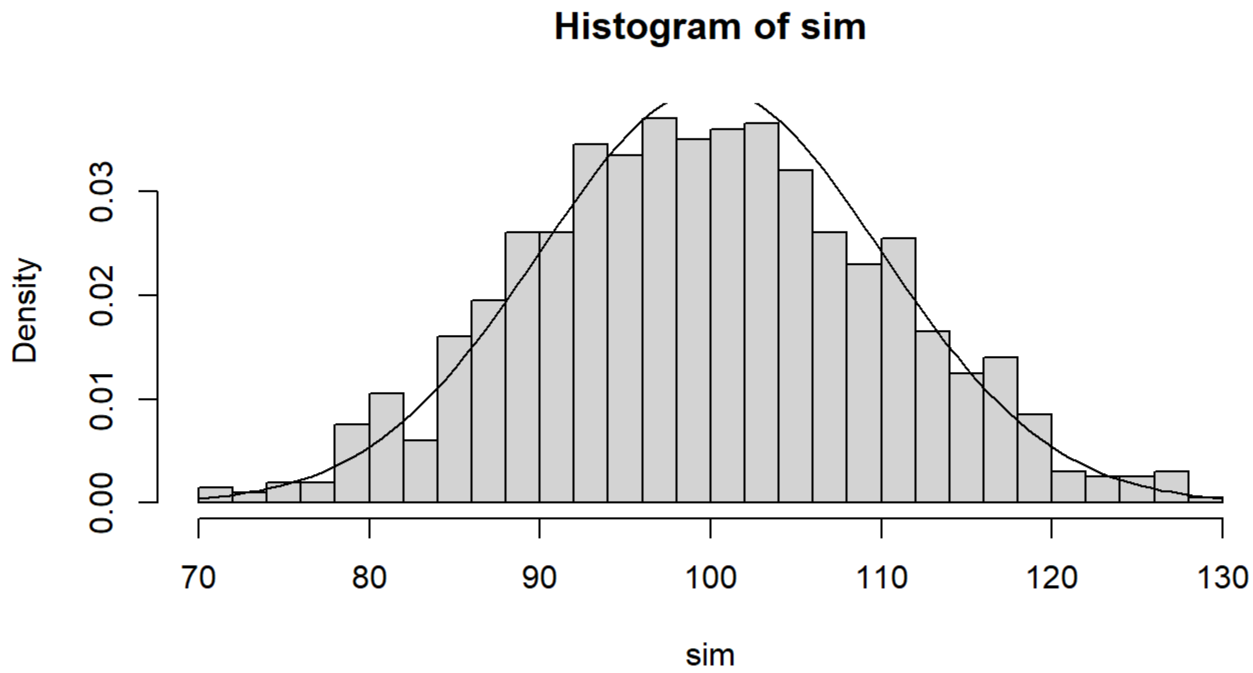

cumulative distribution function
0.0
0.2
0.4
0.6
0.8
1.0
80
90
100
110
120
x

The Normal distribution

Random Generation

- We can generate a number of random values from the $\mathcal{N}(\mu, \sigma^2)$

```r
set.seed(14)
sim <- rnorm(1000, mean = 100, sd = 10)
hist(sim, breaks = 40, freq = FALSE)
curve(dnorm(x, mean = 100, sd = 10), add = TRUE)
```


Histogram of sim
Density
0.00
0.01
0.02
0.03
70
80
90
100
110
120
130
sim

Statistical model: The Normal Model

The Normal (or Gaussian) Model

- We assume that $Y_1, \ldots, Y_n$ are random variables independent and identically distributed according to the $\mathcal{N}(\mu, \sigma^2)$ distribution
- Parameter: $\theta = (\mu, \sigma^2) \in \Theta = \mathrm{R} \times \mathrm{R}_+$
- The most popular estimators of μ and σ^2 are respectively

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

  that is the sample mean and the sample variance, respectively
- The estimators above are unbiased and consistent (meaning that the variance tends to zero when $n \to \infty$)

The Normal Model: Inferential Results

Inference on $\theta = (\mu, \sigma^2)$

- $\bar{Y}$ and S^2 are independent
- $\bar{Y} \sim \mathcal{N}(\mu, \sigma^2/n)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, where χ^2_v is the Chi-square distribution with v degrees of freedom

Pivotal Quantity (Focus on μ)

- $T = \frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim \mathrm{t}_{n-1}$ where t_v denotes the Student's t distribution with v degrees of freedom

Confidence Interval for μ at level $1-\alpha$

- For $0 < \alpha < 1$

$$\Pr(\mathrm{t}_{\mathrm{n}-1\,;\,\alpha/2} \leq \mathrm{T} \leq \mathrm{t}_{\mathrm{n}-1\,;\,1-\alpha/2}) = 1 - \alpha$$

- The realization of the confidence interval is

$$IC_{\mu}^{1-\alpha} = (\bar{y} - t_{n-1\,;\,1-\alpha/2} s/\sqrt{n}, \bar{y} - t_{n-1\,;\,1-\alpha/2} s/\sqrt{n})$$

- where $-t_{n-1\,;\,1-\alpha/2} = t_{n-1\,;\,\alpha/2}$ and $t_{n-1\,;\,1-\alpha/2}$ represent the quantiles of order $\alpha/2$ and $1 - \alpha/2$ of a Student's t distribution with $n - 1$ degrees of freedom, respectively.

Statistical Model: Inferential Results

Confidence Intervals: Example

- Let us to generate $B = 1000$ samples of size $n = 10$ from a $N(\mu = 5, \sigma^2 = 4)$ and fix $\alpha = 0.05$

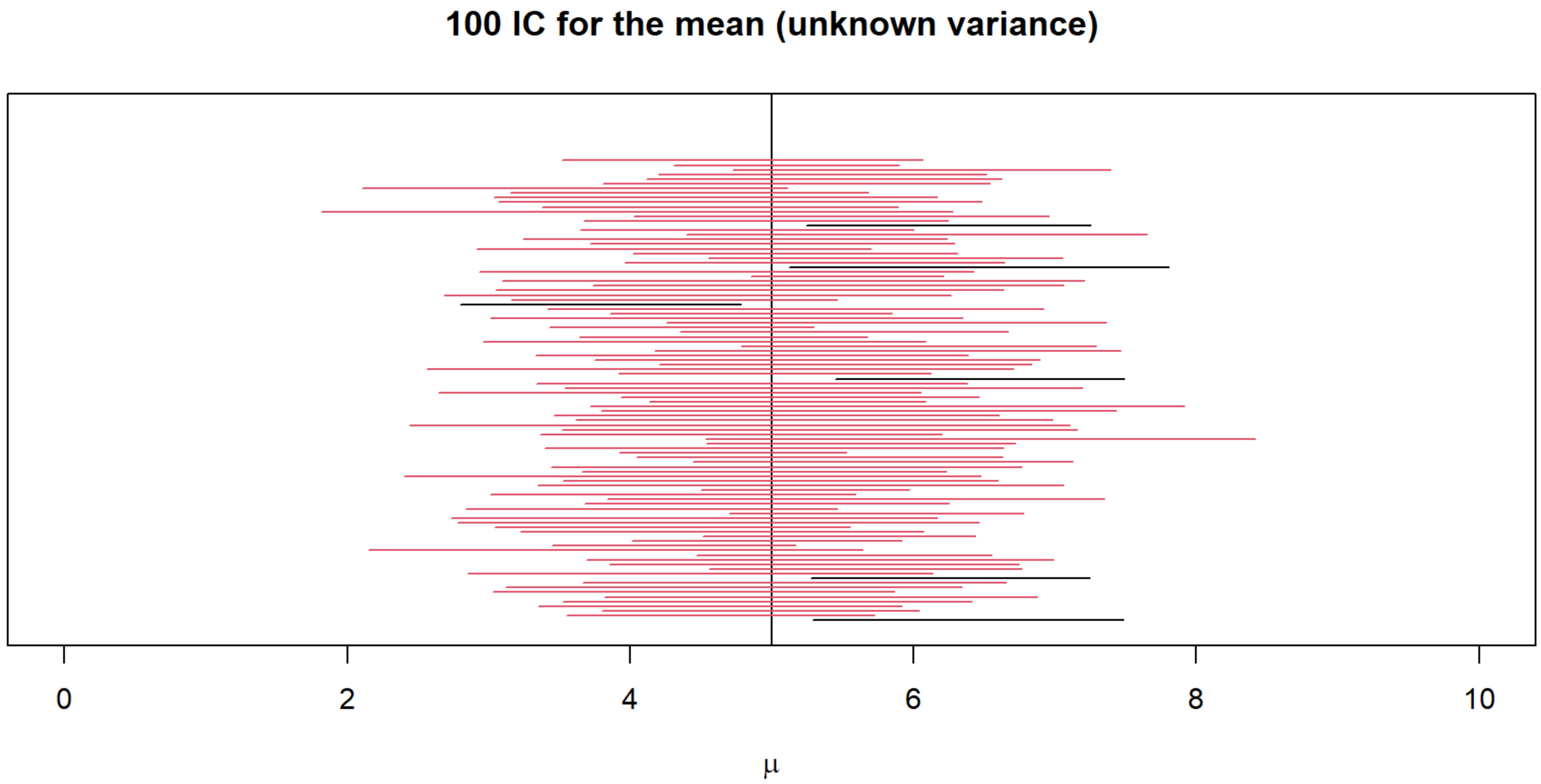
```r
mu <- 5; sigma <- 2; B <- 1000; n <- 10; alpha <- 0.05
CI <- matrix(0, B, 2)
l <- rep(0, B)

set.seed(23)
q0975 <- qt(1 - alpha/2, n - 1)

for (i in 1 : B){
 x <- rnorm(n, mu, sigma)
 CI[i, ] <- mean(x) + c(-q0975, q0975) * sd(x)/sqrt(n)
 l[i] <- (mu > CI[i,1] & mu < CI[i,2])
}
```

Statistical Model: Inferential Results

```
#Empirical coverage probability
mean(l)
```

```
[1] 0.956
```

```
# Plot of the (first 100) CIs
plot(1, xlim = c(0,10), ylim = c(0,11), type = "n",
     xlab = expression(mu), ylab = "", yaxt  = "n",
     main = paste("100 IC for the mean (unknown variance)"), cex.main = 1.2)
abline(v = mu)

d <- 0
for (i in 1 : 100){
  d <- d + 0.1
  lines(seq(CI[i, 1], CI[i, 2], length = 100), rep(d, 100), col = (l[i] + 1))
}
```




```
# number of intervals (out the 100) including the true parameter value
sum(l[1 : 100])
```

```
[1] 94
```

The Normal Model: Inferential Results

Hypothesis test

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

- The reject region of level α associated to the test is

$$R_\alpha = \left\{ (y_1, \ldots, y_n) : \left| \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \right| > t_{n-1\,;\,1-\alpha/2} \right\}$$

- The p-value (the probability under a specified statistical model that a statistical summary of the data would be equal to or more extreme than its observed value) is

$$p - value = \Pr\left(|T| > \left| \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \right| \right)$$

The Normal Model: Inferential Results

Example

To check whether the declared weight on a package corresponds to the truth, a pasta manufacturer analyzes the weight of a sample of $n = 10$ packages of pasta. The packages have a declared weight of $500g$, and it is reasonable to assume that the weight of a package is a normally distributed random variable. To verify that the machine does not produce packages with a different weight, it is checked whether the sample mean weight deviates too much from 500; if it does, production is suspended.

To decide when the sample mean weight is too far from 500, a 5% rejection region is used. The hypothesis system is:

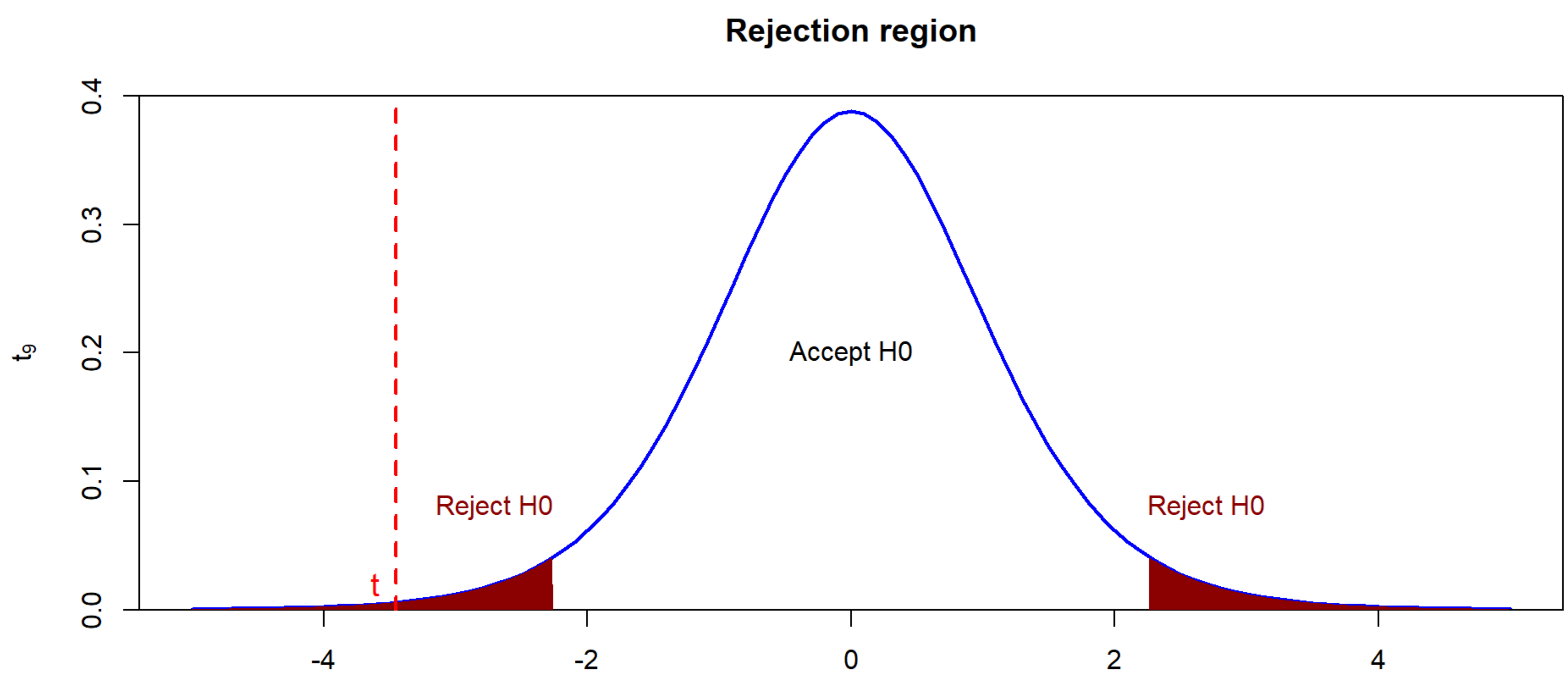
$$\begin{cases} H_0 : \mu = 500 \\ H_1 : \mu \neq 500 \end{cases}$$

The sample has given the following measurements:

$$497.0621,\ 497.7690,\ 497.2631,\ 498.6688,\ 494.6227,\ 496.2185,\ 498.2635,\ 502.7199,\ 497.4339,\ 494.2883$$

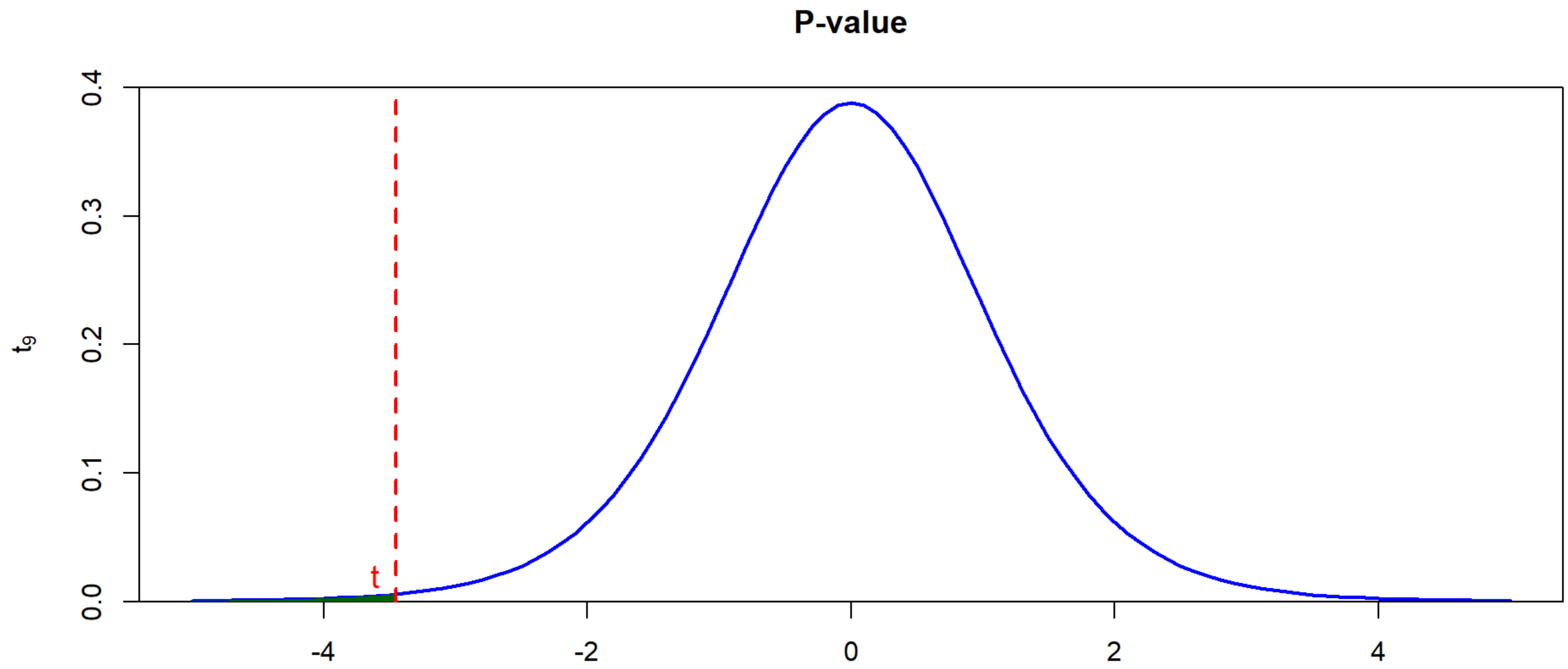
The Normal Model: Inferential Results

```r
y <- c(497.0621, 497.7690, 497.2631, 498.6688, 494.6227, 496.2185, 498.2635, 502.719
t_obs <- (mean(y) - 500)/(sd(y)/sqrt(length(y)))
curve(dt(x, 9), xlim = c(-5, 5), ylim = c(0, 0.4),
      main = "Rejection region", col = "blue",
      lwd = 2, xlab = "",  ylab = expression(t[9]),  yaxs="i")
cord.x0 <- c(-5, seq(-5, qt(0.025, 9), 0.01),  qt(0.025, 9))
cord.y0 <- c(0, dt(seq(-5,qt(0.025, 9), 0.01), 9), 0)
cord.x1 <- c(qt(0.975, 9),seq(qt(0.975, 9), 5, 0.01), 5)
cord.y1 <- c(0, dt(seq(qt(0.975, 9), 5, 0.01), 9), 0)
polygon(cord.x0, cord.y0, col = "darkred", border = NA )
polygon(cord.x1, cord.y1, col = "darkred", border = NA )

abline(v = t_obs, lty = 2, lwd = 2, col = "red")
text(0, 0.2, paste("Accept", expression(H0)))
text(c(-2.7,2.7), 0.08, paste("Reject", expression(H0)), col = "darkred")
text(as.double(t_obs) - 0.15, 0.02, "t", col = "red", cex = 1.2)
```


Rejection region
0.0
0.1
0.2
0.3
0.4
t_9
-4
-2
0
2
4
Accept H0
Reject H0
Reject H0
t

The Normal Model: Inferential Results

```
curve(dt(x, 9), xlim = c(-5, 5), ylim = c(0, 0.4),
      main = "P-value", col = "blue",
      lwd = 2, xlab = "",  ylab = expression(t[9]),  yaxs="i")
cord.x0 <- c(-5, seq(-5, t_obs, 0.01), t_obs)
cord.y0 <- c(0, dt(seq(-5,t_obs, 0.01), 9), 0)
polygon(cord.x0, cord.y0, col = "darkgreen", border = NA )

abline(v = t_obs, lty = 2, lwd = 2, col = "red")
text(as.double(t_obs) - 0.15, 0.02, "t", col = "red", cex = 1.2)
```



P-value
t_9
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
t

Statistical Model: Inferential Results

Likelihood Inference

- Let's assume to be interested on the inference on a parameter θ (unknown)
- Given a random sample $y = (y_1, \ldots, y_n)$ of size n from Y which is distributed according to the model $f(y; \theta)$, with $\theta \in \Theta \subset R^d, d \geq 1$, we define the likelihood function for θ the function as $\mathcal{L}(\theta) = f(y_1, \ldots, y_n; \theta)$
- It's not a probability function: It's a function of θ which is defined on Θ getting values in R^+
- Under a random sampling, due to the independence condition of the random variables (and for the condition of identical distribution), we have

$$\mathcal{L}(\theta) = f(y_1, \ldots, y_n; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$$

- For convenience, it is better to work with the log-likelihood function

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log f(y_i; \theta)$$

Statistical Model: Inferential Results

Maximum Likelihood

- A value $\hat{\theta} \in \Theta$ such that $\ell(\hat{\theta}) \geq \ell(\theta)$ for all $\theta \in \Theta$ is said maximum likelihood estimate (the random variable $\hat{\theta}(Y)$ is said maximum likelihood estimator)
- Under regularity conditions (what we consider) the maximum likelihood estimate is obtained by solving

$$\ell^{'}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = 0$$

where the vector $\ell^{'}(\theta)$ is the score vector

- Finally, $\hat{\theta}$ is consistent and has asymptotic distribution $\hat{\theta} \stackrel{\cdot}{\sim} \mathcal{N}(\theta, j(\hat{\theta})^{-1})$ where

$$j(\theta) = -\ell^{''}(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^{\mathsf{T}}}$$

Statistical Model: Inferential Results

Hypothesis Test

- To verify the hypothesis $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta$ the test statistic is based on the test of the log-likelihood ratio, that is

$$W(\theta_0) = 2(\ell(\hat{\theta}) - \ell(\theta_0))$$

  whose asymptotic distribution is χ^2_d

- Confidence Region: $\{\theta: W(\theta) \leq \chi^2_{d;1-\alpha}\}$
- In the normal linear model, by assuming the normality of errors the function $W(\theta)$ is a monotone increasing function of a statistic F (Fisher) or t (Students' t) and the results are reported in that terms

The Normal model: Likelihood Inference

Likelihood Function

$$\mathcal{L}(\theta) = \mathcal{L}(\mu, \sigma^2) = \prod_{i=1}^{n} f(y_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$\ell(\mu, \sigma^2) = \log\mathcal{L}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

Score Vector: $\ell'(\theta) = (s_1, s_2)$

$$s_1 = \frac{\partial\ell(\theta)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)$$

$$s_2 = \frac{\partial\ell(\theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2$$

The Normal model: Likelihood Inference

Maximum Likelihood Estimate

- From $s_1 = 0$ we obtain $\hat{\mu} = \bar{y}$
- From $s_2 = 0$ (and substituting μ with $\hat{\mu}$), we obtain $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$

Note

- The latter is biased but it can be expressed in terms of the unbiased variance estimator using the relationship $\hat{\sigma}^2 = \frac{S^2(n-1)}{n}$
- In this case, we have an explicit solution for the maximum likelihood estimates. In other cases, we must obtain the estimates numerically (for instance considering the Newton-Raphson algorithm)

The Normal model: Likelihood Inference

Optimization

- In R, there are several routines for optimization purposes (usually minimize a function): optim, nlminb, …
- We analyze one of them in order to obtain the maximum likelihood estimates, comparing it with the analytic solution
- First, we need to write down the likelihood function

```
nllik <- function(par, data) {
  mu <- par[1]
  sigma2 <- par[2]
  n <- length(data)
  nll <-  0.5 * n * log(sigma2) + sum((data - mu)^2) / (2 * sigma2)
  return(nll)
}
```

Statistical Model: The Normal model

Optimization

- Let's generate a sample of size n from a $\mathcal{N}(\mu = 10, \sigma^2 = 4)$
- Optimize the likelihood function using nlminb(), which requires the starting values for the algorithm, the function to minimize, the data, and the limits of the parametric space (if there are some constraints on the prametric space)

```r
set.seed(13)
n <- 50
x <- rnorm(n, mean = 10, sd = 2)
start_vals <- c(mu = 0, sigma = 1)

fit <- nlminb(start = start_vals,
              objective = nllik,
              data = x,
              lower = c(-Inf, 1e-6))
fit$par
```

```
      mu    sigma
9.947162 3.726903
```

```r
mean(x)
```

```
[1] 9.947163
```

```
var(x)*(n-1)/n
```

```
[1] 3.726902
```

The Normal Model: Linear Regression

Linear Regression Model

- The settings above can be straightforwardly extended to linear regression models
- Let's consider a simple linear regression $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ independently
- Note that the difference is that we do not have the identically distribution because $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ where $\mu_i = \beta_1 + \beta_2 x_i$
- Let's adapt the negative log-likelihood function to this case

```r
nllik_lm <- function(par, x, y) {
  beta0 <- par[1]
  beta1 <- par[2]
  sigma2 <- par[3]
  mu <- beta0 + beta1 * x
  n <- length(y)
  nll <-  0.5 * n * log(sigma2) + sum((y - mu)^2) / (2 * sigma2)
  return(nll)
}
```

The Normal Model: Linear Regression

Linear Regression Model

- Let's generate some data and fit the model

```
n <- 100
x <- runif(n, 0, 10)
beta0 <- 2; beta1 <- 0.8; sigma <- 1.5
y <- beta0 + beta1 * x + rnorm(n, 0, sigma)
start_vals <- c(beta0 = 0, beta1 = 0, sigma2 = 1)

fit <- nlminb(
  start = start_vals,
  objective = nllik_lm,
  x = x,
  y = y,
  lower = c(-Inf, -Inf, 1e-6)
)

fit$par
```

```
    beta0     beta1    sigma2 
2.1476594 0.7798585 2.6494339 
```

```
fitLM <- lm(y ~ x)
```

```r
fitLM$coefficients
```

```
(Intercept)           x 
  2.1476602   0.7798585 
```

```r
sum(fitLM$residuals^2)/n
```

```
[1] 2.649434
```

The Normal Model: Linear Regression

Linear Regression Model

- As we noticed the β estimates obtained via maximum likelihood are the same of those obtained via OLS.
- Indeed, from

$$\mathcal{L}(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(y_i - \beta_1 x_{i1} - \ldots - \beta_p x_{ip})^2}$$

$$\propto (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \ldots - \beta_p x_{ip})^2}$$

$$\propto (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)}$$

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)$$

- Note that the term that does not depend on the parameter has been removed and we introduce the proprionality symbol $\propto$

The Normal Model: Linear Regression

β **Estimate**

- For fixed σ^2, $\ell(\beta, \sigma^2)$ is maximum when the residual sum of square $RSS(\beta) = (y - X^\top \beta)^\top (y - X\beta)$ is minimum, that is

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \text{RSS}(\beta) = \underset{\beta}{\text{argmax}} \ell(\beta, \sigma^2)$$

- Then, the maximum likelihood estimate is equal to the OLS estimate

$$\hat{\beta}_{ML} = \hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y$$

The Normal model: Linear Regression

σ^2 Estimate

- By substituting $\hat{\beta}$ in $\ell(\beta, \sigma^2)$ we have

$$\ell(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X^\top\hat{\beta})^\top(y - X\hat{\beta})$$

$$= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}RSS(\hat{\beta}) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} e_i^2$$

- By getting the derivative with respect to σ^2 and equating to zero, we obtain

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}e^\top e$$

The Normal model: Linear Regression

Distribution of $\hat{\beta}$

- Since $Y \sim \mathcal{N}_n(\mu, \sigma^2 I)$, the estimator $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ is a linear transformation of a multivariate normal. So

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2 (X^\top X)^{-1})$$

- Due to the results on the multivariate normal distribution, we know that also the marginals are normally distributed and in particular

$$\hat{\beta}_r \sim \mathcal{N}(\beta_r, \sigma^2 (X^\top X)^{-1}_{rr}), \quad r = 1, \ldots, p$$

having denoted with $(X^\top X)^{-1}_{rr}$ the r-th diagonal element of $(X^\top X)^{-1}$.

The Normal model: Linear Regression

Distribution of $\hat{\sigma}^2$

- The estimator σ^2 is biased, while the unbiased estimator is

$$S^2 = \frac{n}{n-p}\hat{\sigma}^2$$

for which

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{n-p}$$

Joint Distribution of $(\hat{\beta}, \hat{\sigma}^2)$

- $\hat{\beta}$ is independent from $\hat{\sigma}^2$, and so $\hat{\beta}$ is independent from S^2

Normal Linear Regression: Inference

Inference on a Single Coefficient

- From the results just outlined, we can obtain pivotal quantities to make inference on a single coefficient

$$t_r = \frac{\hat{\beta}_r - \beta_r}{\sqrt{S^2 (X^\top X)^{-1}_{rr}}} \sim t_{n-p}$$

- Confidence interval of level $1-\alpha$ for β_r

$$IC^{1-\alpha}_{\beta_r} = (\hat{\beta}_r - t_{n-p\,;\,1-\alpha/2}\sqrt{S^2((X^\top X)^{-1}_{rr})}, \hat{\beta}_r + t_{n-p\,;\,1-\alpha/2}\sqrt{S^2((X^\top X)^{-1}_{rr})})$$

- Hypothesis test (nullity of the single coefficient): *[Math Processing Error]*

$$t_r = \hat{\beta}_r / \sqrt{S^2 (X^\top X)^{-1}_{rr}} \overset{H_0}{\sim} t_{n-p}$$

$$\mathcal{R}_\alpha = \{ |t_r^{obs}| > t_{n-p\,;\,1-\alpha/2} \}$$

$$p - value = P(|t_{n-p}| > |t_r^{obs}|) = 2P(t_{n-p} \leq -|t_r^{obs}|)$$

Normal Linear Regression: Inference

```r
library(ISLR)
data("Credit")
lmFit <- lm(Balance ~ Limit + Student, data = Credit)
summary(lmFit)$coefficients
```

```
                Estimate   Std. Error   t value      Pr(>|t|)
(Intercept) -334.7299372 23.069301674 -14.50976  1.417949e-38
Limit          0.1719538  0.004330826  39.70463 2.558391e-140
StudentYes   404.4036438 33.279679039  12.15167  4.181612e-29
```

```r
X <- model.matrix(Balance ~ Limit + Student, data = Credit)
n <- nrow(X)
p <- ncol(X)
hat_s2 <- sum(residuals(lmFit)^2)/(nrow(Credit) - p)

tval <- lmFit$coefficients/(sqrt(hat_s2 * diag(solve(t(X) %*% X))))
tval
```

```
(Intercept)       Limit  StudentYes
  -14.50976    39.70463    12.15167
```

```r
2*pt(-abs(tval), n - p)
```

```
  (Intercept)         Limit    StudentYes
 1.417949e-38 2.558391e-140  4.181612e-29
```

Normal Linear Regression: Inference

Inference on a Pool of Coefficients (A Special Case)

- Let’s suppose we want assess the model overall, that is no one of the predictors are explaining the variability of Y. The hypothesis system is *[Math Processing Error]*
- In this case we distinguish between

$$\text{Full model}(\mathcal{M}_1): Y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta x_{ip} + \varepsilon_i$$

$$\text{Reduced model}(\mathcal{M}_0): Y_i = \beta_1 + \varepsilon_i$$

- Under $\mathcal{M}_1$ the estimate of the variance of the error term component is $\hat{\sigma}^2 = \frac{1}{n} e^\top e$, while Under $\mathcal{M}_0$ is simply $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$
- The test statistic is

$$F = \frac{\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{p-1}}{\frac{\hat{\sigma}^2}{n-p}} \overset{H_0}{\sim} F_{p-1,n-p}$$

- The p-value is given by $p - value = Pr(F_{p-1,n-p} > f^{obs})$

Normal Linear Regression: Inference

```r
hat_sigma2 <- sum(residuals(lmFit)^2)/n
tilde_sigma2 <- var(Credit$Balance)*(n-1)/n
fval <- ((tilde_sigma2 - hat_sigma2)/(p-1))/((hat_sigma2)/(n-p))
fval
```

```
[1] 859.1893
```

```r
summary(lmFit)$fstatistic
```

```
   value    numdf    dendf 
859.1893   2.0000 397.0000 
```

```r
pf(fval, p-1, n-p, lower = FALSE)
```

```
[1] 5.893567e-145
```

```r
summary(lmFit)
```

```

Call:
lm(formula = Balance ~ Limit + Student, data = Credit)

Residuals:
    Min      1Q  Median      3Q     Max 
-637.77 -116.90    6.04  130.92  434.24 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.347e+02  2.307e+01  -14.51   <2e-16 ***
Limit        1.720e-01  4.331e-03   39.70   <2e-16 ***
```

```
StudentYes   4.044e+02  3.328e+01   12.15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 199.7 on 397 degrees of freedom
```