The multivariate normal

Anders Nielsen, Ethan Lawler, & Sean Anderson

an@aqua.dtu.dk

The Normal Distribution

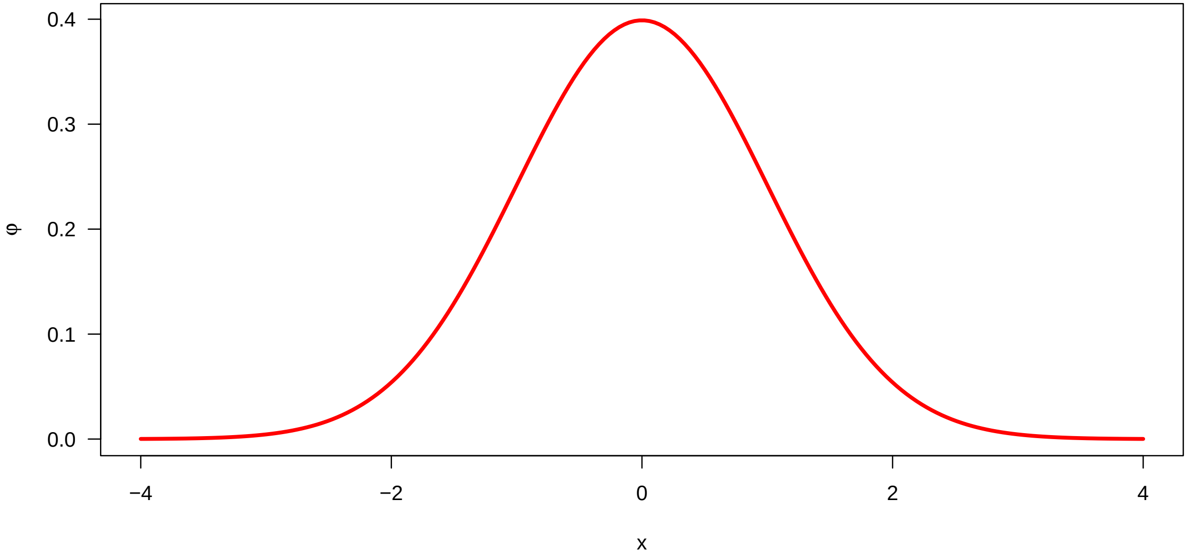
- Easily the most important probability distribution

The Normal Distribution - a few facts

- A continuous probability distribution on $(-\infty, \infty)$
- The probability density function is:

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

- The mean is μ and standard deviation is σ
- The interval $(\mu - 2\sigma, \mu + 2\sigma)$ contains 95%

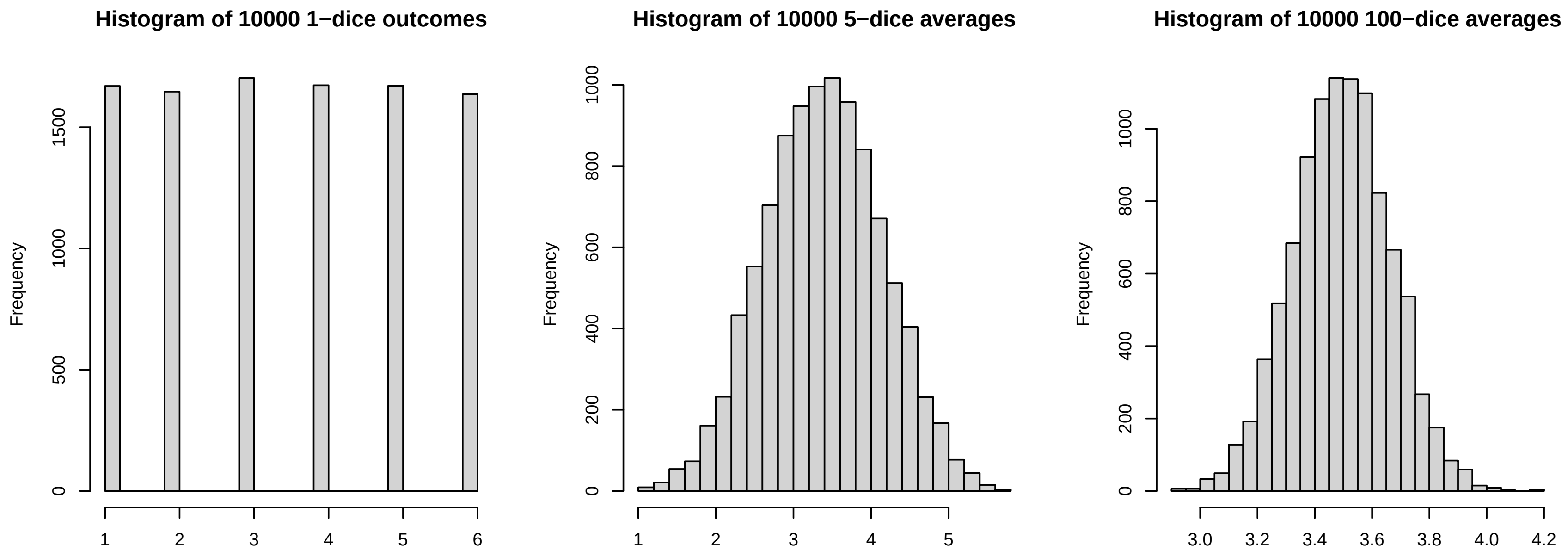


Central Limit Theorem (CLT)

- If $X_1, X_2, \ldots$ are independent identically distributed variables with finite mean μ and variance σ^2, then

$$\sqrt{n}\frac{\frac{1}{n}\sum X - \mu}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \ , \ \text{as } n \to \infty$$

- Notice that nothing is said about the distribution of X (except about mean and variance)



- Why is this important?

Ex: Complete program using normal likelihood

```
library(RTMB)
dat <- list(Y=rnorm(1000,2,.3))
par <- list(mu=0, logsigma=0)
f <- function(par){
  getAll(par,dat)
  sigma <- exp(logsigma)
  nll <- -sum(dnorm(Y, mu, sigma, log=TRUE))
  ADREPORT(sigma)
  nll
}
obj <- MakeADFun(f, par)
opt <- nlminb(obj$par, obj$fn, obj$gr)
summary(sdreport(obj))
```

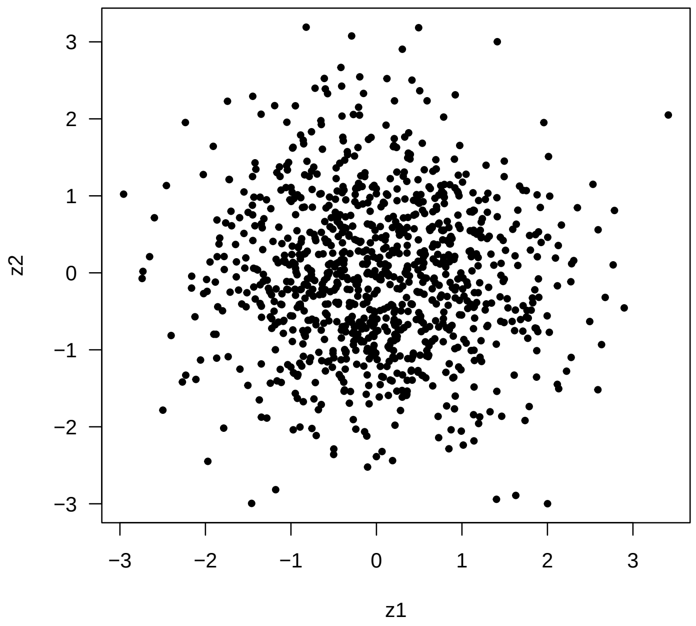
files/norm.R

Two normal random variables

- Imagine we have two univariate normally distributed random variables

$$Z_1 \sim N(0,1) \text{ and } Z_2 \sim N(0,1)$$

- If we plot a lot of simulations (z_1, z_2) we get:



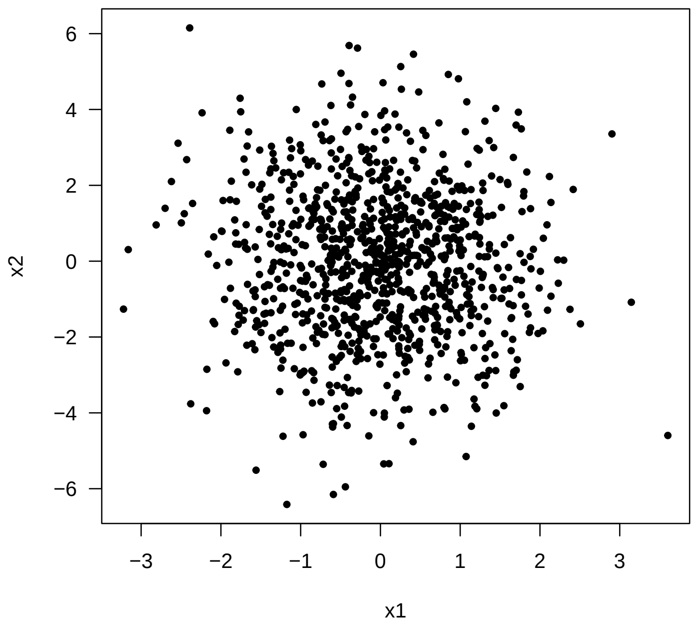
- The marginal distribution on each axis is a $N(0,1)$

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$X = \begin{pmatrix} Z_1 \\ 2Z_2 \end{pmatrix}$$

- If we plot a lot of simulations (x_1, x_2) we get:



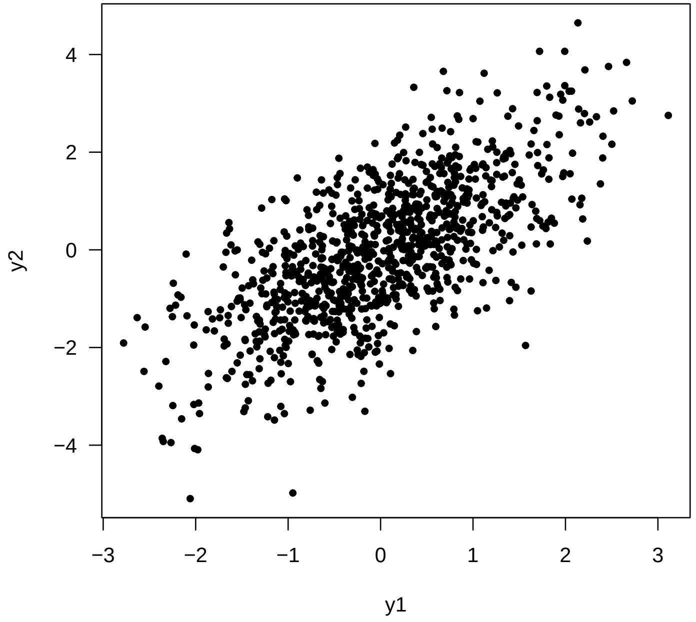
- The marginal distribution is a $N(0,1)$ on first axis and $N(0,4)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$Y = \begin{pmatrix} Z_1 \\ Z_1 + Z_2 \end{pmatrix}$$

- If we plot a lot of simulations (y_1, y_2) we get:



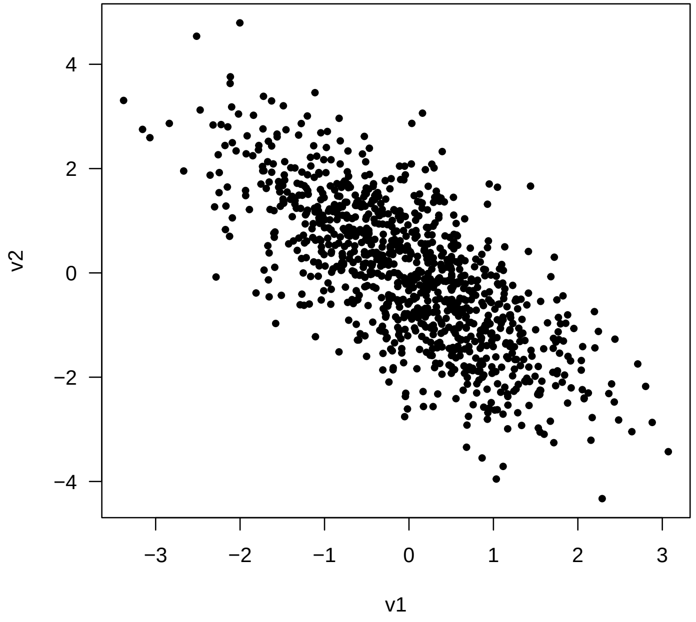
- The marginal distribution is a $N(0,1)$ on first axis and $N(0,2)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$V = \begin{pmatrix} Z_1 \\ Z_2 - Z_1 \end{pmatrix}$$

- If we plot a lot of simulations (v_1, v_2) we get:



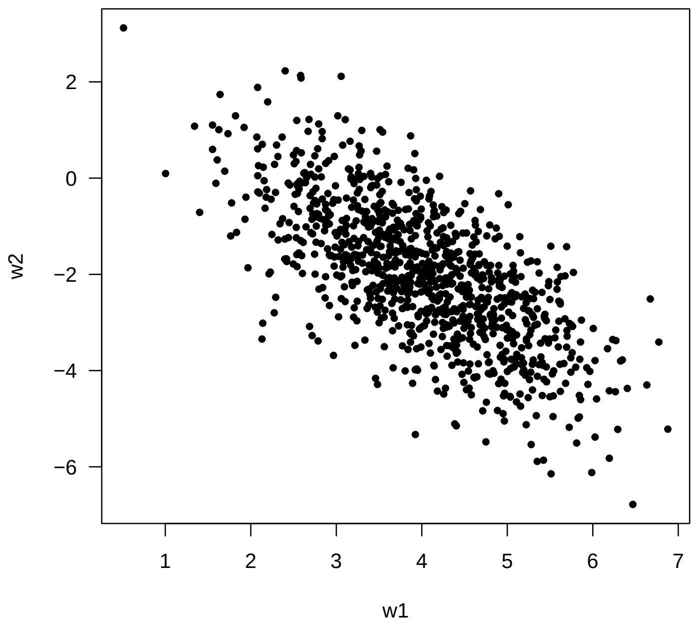
- The marginal distribution is a $N(0,1)$ on first axis and $N(0,2)$ in the second axis.

Two normal random variables

- Imagine again we have the same two univariate normally distributed random variables, but now we look at:

$$W = \begin{pmatrix} Z_1 + 4 \\ Z_2 - Z_1 - 2 \end{pmatrix}$$

- If we plot a lot of simulations (w_1, w_2) we get:



- The marginal distribution is a $N(4,1)$ on first axis and $N(-2,2)$ in the second axis.

Two normal random variables

- If we define Z as:

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

- Then we can write all the cases as:

$$AZ + b$$

- where A is a matrix and b is a vector.
- E.g. the last example:

$$W = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} Z + \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

Multivariate normal distribution

- We say that the k-dim random variable X follows a multivariate normal distribution $X \sim N_k(\mu, \Sigma)$ if there exsists random l-dim random varible Z where each component follows a $N(0,1)$ distribution, such that $X = AZ + b$.
- In that case $\Sigma = AA^t$ and $\mu = b$
- The density for a k-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ is:

$$L(x) = \frac{1}{(2\pi)^{k/2}\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

- We write $X \sim N_k(\mu, \Sigma)$.

Covariance and correlation

- The covariance between two random variables is defined as:

$$\text{cov}(X, Y) = E\left((X - \mu_x)(Y - \mu_y)\right)$$

- For a multivariate normal $X \sim N_k(\mu, \Sigma)$ we have arranged all the covariances in the matrix Σ, such that:

$$\Sigma_{ij} = \text{cov}(X_i, X_j)$$

- The covariance between a variable and itself is the variance of that variable, so

$$\Sigma_{ii} = \text{cov}(X_i, X_i) = \text{var}(X_i)$$

- The correlation coefficient is defined as:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

- Mini exercise: Find the correlation coefficient of W on previous page.
- Mini exercise: Can you construct an A, such that $\rho = 0.9$?

Ex: Using multivariate normal likelihood

```
library(RTMB)
z1 <- rnorm(1000)
z2 <- rnorm(1000)
Z <- rbind(z1,z2)
A <- rbind(c(1,0),c(-1,1))
b <- c(4,-2)
X <- A%*%Z+b

dat <- list(X=t(X))
par <- list()
par$mu <- c(0,0)
par$logSigma <- c(0,0)
par$tRho <- 1

f <- function(par){
  getAll(par,dat)
  sigma <- exp(logSigma)
  rho <- 2*plogis(tRho)-1
  Sigma <- rbind( c( sigma[1]^2,             sigma[1]*sigma[2]*rho),
                  c( sigma[1]*sigma[2]*rho,  sigma[2]^2)            )
  REPORT(Sigma)
  -sum(dmvnorm(X, mu, Sigma=Sigma, log=TRUE))
}

obj <- MakeADFun(f, par)
opt <- nlminb(obj$par, obj$fn, obj$gr)
summary(sdreport(obj))
```

files/mvnorm.R

Exercise: To investigate the effect of a certain type of exposure in three doses (1,2,and 3) the following experiment was carried out. The experimental unit was a cage with 2 rats. Once per month in 10 months the activity was measured as number of crossing of a light beam. The data can be seen on the next page. It must be expected that measurements from same cage are correlated, and even that measurements close in time have higher correlations. The following model was proposed:

$$\log(\text{count}) \sim \mathcal{N}(\mu, \Sigma), \quad \text{where}$$

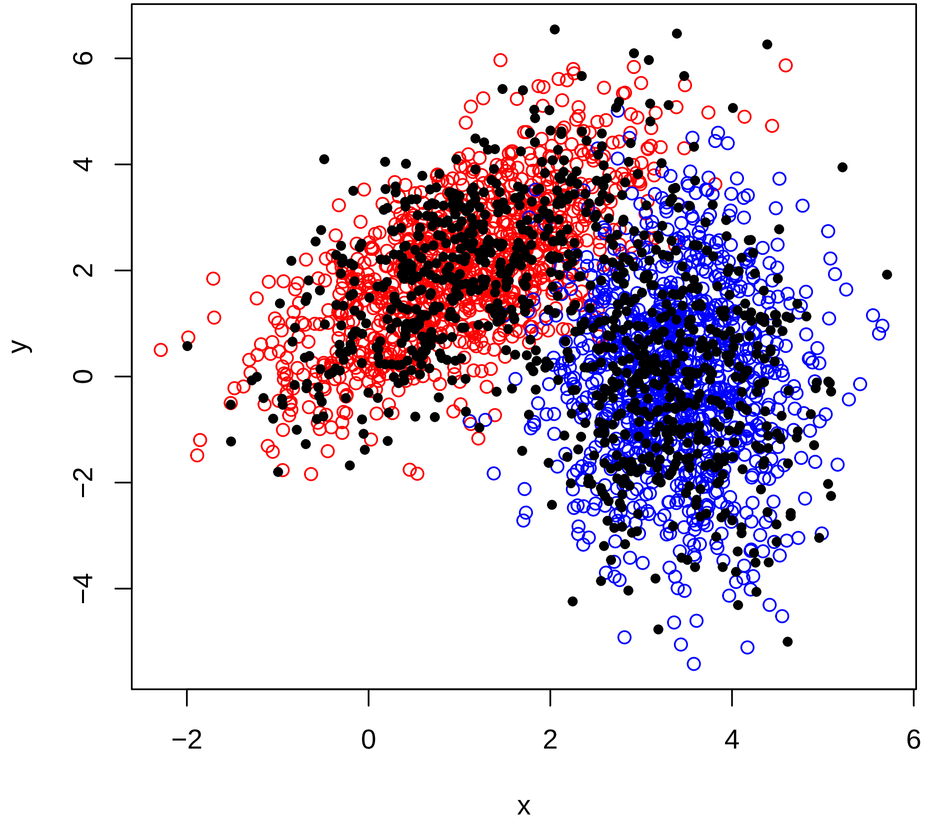
$$\mu_i = \alpha(\text{dose}_i, \text{month}_i), \quad i = 1 \dots 300$$

$$\Sigma_{i,j} = \begin{cases} 0, & \text{if cage}_i \neq \text{cage}_j \\ \nu^2 + \tau^2 \exp\{\frac{-(\text{month}_i - \text{month}_j)^2}{\rho^2}\}, & \text{if cage}_i = \text{cage}_j \text{ and } i \neq j \\ \nu^2 + \tau^2 + \sigma^2, & \text{if } i = j \end{cases}$$

? Implement the model and remember that the variance parameters should be positive.

! The data has been prepared in the file `rats.RData`

Exercise: In a classification setup we have 1000 points from each of two groups ("red" and "blue"). We are then given 1000 additional points with unknown class. Each of the groups are well described by a 2 dimensional normal distribution. Write the code to estimate these two normal distributions and assign the most likely class to each of the "black" points. The data set is in the file `lda.RData`.



Multivariate normal distribution of estimator

- Asymptotically (as we gather more data) the distribution of our estimator will become

$$\widehat{\theta} \sim N_k(\theta_{\text{true}}, H(\theta_{\text{true}})^{-1})$$

- So we can (and will later) use this to construct confidence regions of our estimates.

Linear transformation of multivariate normal

- Assume:

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

- The distribution of

$$(AX + b) \sim N(A\mu + b, A\Sigma A^t)$$

The delta method

- Let's be willing to assume:

$$\theta \sim N(\hat{\theta}, \Sigma)$$

- Are interested in a quantity, which is a non-linear function of θ:

$$Q = f(\theta)$$

- The linear approximation is:

$$Q \sim N\left(f(\hat{\theta}), \nabla f(\hat{\theta})^T \Sigma \nabla f(\hat{\theta})\right)$$

- Mini Exercise: Assume we have estimated $(\log F_2, \log F_3, \log F_4)$ to (-1.13, -0.75, -0.94) with a covariance matrix of:

$$\begin{pmatrix} 0.0222 & 0.0135 & 0.0114 \\ 0.0135 & 0.0169 & 0.0137 \\ 0.0114 & 0.0137 & 0.0191 \end{pmatrix}$$

setup a confidence interval for $\log(\overline{F}_{2-4})$

Conditional of multivariate normal

- Assume:

$$X \sim N_k(\mu, \Sigma)$$

- where the first $m < k$ elements define a a block, such that $X = (X_{1:m}, X_{(m+1):k})'$
- Similarly the mean vector can be devided into $\mu = (\mu_{1:m}, \mu_{(m+1):k})'$ and the covariance into four blocks

$$\Sigma = \left(\begin{array}{c|c} \Sigma_{1:m,1:m} & \Sigma_{1:m,(m+1):k} \\ \hline \Sigma_{(m+1):k,1:m} & \Sigma_{(m+1):k,(m+1):k} \end{array}\right)$$

- Then the conditional distribution of $X_{1:m}$ given that $X_{(m+1):k} = x_{(m+1):k}$ is given by:

$$(X_{1:m} | X_{(m+1):k} = x_{(m+1):k}) \sim N_m(\widetilde{\mu}, \widetilde{\Sigma}) \text{ where}$$
$$\widetilde{\mu} = \mu_{1:m} + \Sigma_{1:m,(m+1):k}\Sigma^{-1}_{(m+1):k,(m+1):k}(x_{(m+1):k} - \mu_{(m+1):k})$$
$$\widetilde{\Sigma} = \Sigma_{1:m,1:m} - \Sigma_{1:m,(m+1):k}\Sigma^{-1}_{(m+1):k,(m+1):k}\Sigma_{(m+1):k,1:m}$$

- Mini exercise: Find the conditional distribution of W_1 given that $W_2 = 0$.

Multivariate normal

A couple of functions to help define multivariate normal densities are:

`dmvnorm` Multivariate normal density specified via mean vector and covariance matrix

`dgmrf` Multivariate normal density specified via mean vector and sparse inverse covariance

`dautoreg` Multivariate normal density with AR(k) covariance structure specified via mean vector and ϕ vector. The order (k) is determined by the length of the ϕ vector

`dseparable` Multivariate normal density defined as separable extentions of 2 or more already defined densities

In addition to these there is a function `unstructured(k)` to help setup unstructured covariances to use with `dmvnorm` and further the functions `dmvnorm`, `dgmrf`, and `dautoreg` have an argument `scale` (which can be a single element or a vector) to scale the standard deviation.

Implement an AR(1) process

- Let's say we have $n = 100$ observations $x_1, ..., x_n$ from a mean zero AR(1) process:

$$x_{i+1} = \phi x_i + \varepsilon_i \quad , \quad \text{where } \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- If the process is in equilibrium when we start observing it, then the distribution of the first observation will be $x_1 \sim \mathcal{N}(0, \sigma^2/(1-\phi^2))$
- The observations are in the file `ar1.dat`.
- Implement the model only via the univariate normal distribution function (`dnorm`)
- Implement the model via the multivariate normal functions (e.g. `dautoreg`, `dgmrf`, ...)
- Verify that you get identical results.