Exercises, STK4040, week 38

September 14, 2007

1 Drawing from $N_p(\mu, \Sigma)$

Not all statistical software has the possibility to draw random observations directly from the multinormal distribution. Then one has to draw from the univariate normal distribution, and transform the data so they get the correct distribution: if X is a data matrix from $N_p(\mathbf{0}, \mathrm{I})$, then $\mathrm{X}\Sigma^{1/2} + \mathbf{1}\mu^T$ is a data matrix from $N_p(\mu, \Sigma)$.

a)

Draw 1000 observations from $N_2(\mu, \Sigma)$, where $\mu^T = (1, 2)$ and $\Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, by drawing from $N(0, 1)$ and using the transformation above. (Hint: use the Eigenvalue decomposition (`eigen()` in `R`) to calculate $\Sigma^{1/2}$.

Calculate $\bar{\boldsymbol{x}}$ and S, and compare with μ and Σ as a check.

b)

Draw 1000 observations from the above distribution by using a ready made function to draw from multi normal distributions. (In `R` one can use `library(MASS)` to load the package `MASS`, which contains the function `mvrnorm()`.)

Calculate $\bar{\boldsymbol{x}}$ and S, and compare with the results in 1a).

Extra question: Is the built-in covariance function unbiased (i.e., does it calculate S or $\mathrm{S_u}$)?

2 Plotting multinormally distributed data

Draw $n = 100$ observations (x_1, x_2) from $N_2(\mu, \Sigma)$, with $\mu^T = (1, 2)$ and different Σ: I_2, $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$, and $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

In all cases:

- Plot x_2 against x_1
- Find the Eigenvalue decomposition of Σ

- Compare the Eigenvectors with the directions of the half axes of the plot
- Compare the Eigenvalues with the lengths of the half axes
- Extra: estimate Σ from the data, and compare its Eigenvalue decomposition with the 'truth'

Draw several times and watch the variations in plot and Eigenvalue structure. You can also try with other Σ and n.

3 Hotelling's T^2

Draw $n = 100$ observations from

$$N_3\left(\begin{pmatrix}0.1\\0.1\\0.2\end{pmatrix}, \begin{pmatrix}1 & 0.2 & 0.7\\0.2 & 1.5 & -0.4\\0.7 & -0.4 & 2\end{pmatrix}\right).$$

Test the null hypothesis H_0: $\mu = \mathbf{0}$ against the alternative hypothesis H_1: $\mu \neq \mathbf{0}$. (Tip: `R` does not have the T^2 distribution, but it has the F distribution; see `?FDist`.) Draw several times. Why does the p value change?

Some useful `R` commands

- `?func` or `help(func)` gives help about a function `func`. `help.start()` opens the whole `R` documentation in a browser. `help.search("whatever")` searches the (installed) documentation for `whatever`. `RSiteSearch("something")` searches the `R` website and e-mail lists for `something` (you need to be connected to the Internet).
- Assignment in `R` is done with `<-`; e.g., `mu <- c(1, 2)` stores the vector $(1, 2)$ in the variable `mu`.
- To create an $m \times n$ matrix from an nm vector `v`: `matrix(v, nrow = n)`
- Use `solve(M)` to invert a matrix `M`
- Matrix product in `R` is `%*%`; e.g., `X %*% S`.
- `rep(1, n)` gives you $\mathbf{1}_n$, and `diag(n)` is I_n