Exercises for the 5th Morning

Exercise 0: Minimise auto-correlation

Write a function in `R` that uses a Metropolis-Hastings algorithm to simulate from a standard normal distribution using normal distributed proposals centred at the current value. Make sure your code allows you to adjust the standard deviation of the proposal distribution and monitor the acceptance probability.

To plot the auto-correlation function use the function `acf` in R. To calculate $\tau = 1 + 2\sum_{m=1}^{\infty} \rho_m$ use the following approach in R: `tau = 2 * sum(acf(x)$acf) - 1`.

Find the standard deviation for the proposal distribution that gives the smallest value of `tau`. What is the acceptance probability in this case? Make sure that you have run the MH-algorithm long enough, so that the auto-correlation function is well estimated.

Exercise 1: Mining Disasters

This examples is concerned with mining disasters in British mines from 1851 and 26,500 days forward. The time for 109 accidents measured in days since 1851. Download the data set `mining.dat` here

http://people.math.aau.dk/~kkb/Undervisning/Bayes/data/

Load the data into `R` using `data = scan(file="mining.dat")` (you might have to change the path depending on where you saved the data).

We model the data as a Poisson process with intensity $\alpha(t)$, $t \in (0, 26500)$. This means that the number of disasters in an interval (t_1, t_2) is Poisson distributed with parameter $\int_{t_1}^{t_2} \alpha(t)dt$.

Let t_i denote the time of the ith disaster. The log likelihood is in this case given by

$$\sum_i \log(\alpha(t_i)) - \int_0^{26500} \alpha(t)dt.$$

During the period which the data covers, a law was introduced which greatly improved the safety in mines. This leads us to assume that α takes one value before this law was introduced and another after. Let j denote the time (measured in days since 1851) at which the change takes place. Assume $\alpha(t) = a$ if $t < j$ and $\alpha(t) = a + d$ if $t \geq j$. If the the times of the disasters are in a vector `data`, you can define a log likelihood function in `R` as

```
log.likelihood = function(j,a,d,data){
    sum(log(a+(data>j)*d))- (j*a + (26500-j)*(a+d))
}
```

Assume a uniform prior on j on the interval $(0, 26500)$. Furthermore, assume that $a = 0.007$ and $d = -0.004$, now construct a Metropolis-Hastings algorithm which generates a sample from the posterior distribution of j.

You may consider how to extend your code to also include an analysis of a and d.

A further extension is to add a further change point.

Exercise 2: Beetles

This example is concerned with the survival beetles after being exposed to 5 hour carbon disulphide. Different beetles are exposed to different concentrations. Download the data set `beetles.dat` here

http://people.math.aau.dk/~kkb/Undervisning/Bayes/data/

Load the data into `R` using `data = read.table(file="beetles.dat")` (you might have to change the path depending on where you saved the data).

The data contains data for 481 beetles. First column of the data contains the concentration. The second column contains the status of the beetles after the exposure coded as zeros and ones. A zero corresponds to death and one to survival.

Let x_i and y_i denote the concentration and status for the ith beetles, respectively. We use a logistic regression, hence we assume that the probability that the ith beetle survives the dose x_i is

$$\frac{\exp(a + bx_i)}{1 + \exp(a + bx_i)}.$$

In `R` you can define `x = data[,1]` and `y = data[,2]`. Further, construct a log-likelihood function

```
log.likelihood = function(a,b,x,y){
    return(sum(y*(a+b*x)) - sum(log(1+exp(a+b*x))))
}
```

Now construct a Metropolis-Hastings algorithm which generates a sample from the posterior distribution of a and b.

Exercise 3: The Banana

Construct a Metropolis-Hastings algorithm which has the following (unnormalised) invariant two-dimensional density

$$\pi(x, y) \propto \exp(-x^2/200 - (y + 0.1 * x^2 - 10)^2).$$

Notice that $p(x, y)$ is symmetric in x. Does your simulations reproduce this symmetry?

Exercise 4: Speed of light

This example you are asked to reproduce the model checking results for the speed of light example. The 66 measurements of time can be found in the data fie `newcomb.csv` which can be downloaded here

http://people.math.aau.dk/~kkb/Undervisning/Bayes/data/

Load the data into `R` using `newcomb = read.table("newcomb.csv",header=TRUE,sep=",")` (you might have to change the path depending on where you saved the data).