MATH5835M Statistical Computing
Exercise Sheet 5 (answers)

http://www1.maths.leeds.ac.uk/~voss/2018/MATH5835M/
Jochen Voss, J.Voss@leeds.ac.uk

2018/19, semester 1

Answer 17.

a) We have

$$\begin{aligned} Z_N &= \frac{1}{N}\sum_{j=1}^{N} f(X_j) \\ &= \frac{1}{N}\sum_{j=1}^{N}\bigl(f(X_j)-\mu\bigr)+\mu \\ &= \frac{\sigma}{\sqrt{N}}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\frac{f(X_j)-\mu}{\sigma}+\mu \\ &= \frac{\sigma}{\sqrt{N}}Y_N+\mu. \end{aligned}$$

b) Since, for large N, we have approximately $Y_N \sim \mathcal{N}(0,1)$ by the central limit theorem, we can conclude that

$$Z_N \sim \mathcal{N}\bigl(\mu, \sigma^2/N\bigr)$$

approximately, for large N.

c) Using the result from part (a), we find

$$\begin{aligned} P\Bigl(\mu \in \bigl[Z_N - \frac{1.96\,\sigma}{\sqrt{N}}, Z_N + \frac{1.96\,\sigma}{\sqrt{N}}\bigr]\Bigr) &= P\Bigl(\bigl|Z_N-\mu\bigr| \leq \frac{1.96\,\sigma}{\sqrt{N}}\Bigr) \\ &= P\Bigl(\bigl|\frac{\sigma}{\sqrt{N}}Y_N\bigr| \leq \frac{1.96\,\sigma}{\sqrt{N}}\Bigr) \\ &= P\bigl(|Y_N| \leq 1.96\bigr) \\ &\to 0.95 \end{aligned}$$

as $N \to \infty$.

Answer 18. a) We have $Y_j = X_{j-1} + \varepsilon_j$ where $\varepsilon_j \sim \mathcal{N}(0,\sigma^2)$. Thus the transition densities for the proposals are given by

$$p(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Bigl(-\frac{(y-x)^2}{2\sigma^2}\Bigr)$$

and, since $p(x,y) = p(y,x)$,

$$\begin{aligned} \alpha(x,y) &= \min\Bigl(\frac{\pi(y)}{\pi(x)}, 1\Bigr) \\ &= \min\Bigl(\frac{\frac{1}{Z}\cdot\frac{\sin(y)^2}{y^2}1_{[-3\pi,3\pi]}(y)}{\frac{1}{Z}\cdot\frac{\sin(x)^2}{x^2}1_{[-3\pi,3\pi]}(x)}, 1\Bigr) \\ &= \min\Bigl(\Bigl(\frac{x\sin(y)}{y\sin(x)}\Bigr)^2 1_{[-3\pi,3\pi]}(y), 1\Bigr). \end{aligned}$$

for all $x \in \mathbb{R}$ with $\pi(x) > 0$.

b) To implement the Random Walk Metropolis algorithm, we can use $\alpha(x,y)$ from part (a). We also have to choose $X_0 \in \mathbb{R}$ with $\pi(X_0) > 0$. The choice $X_0 = 0$ is not possible, since the density

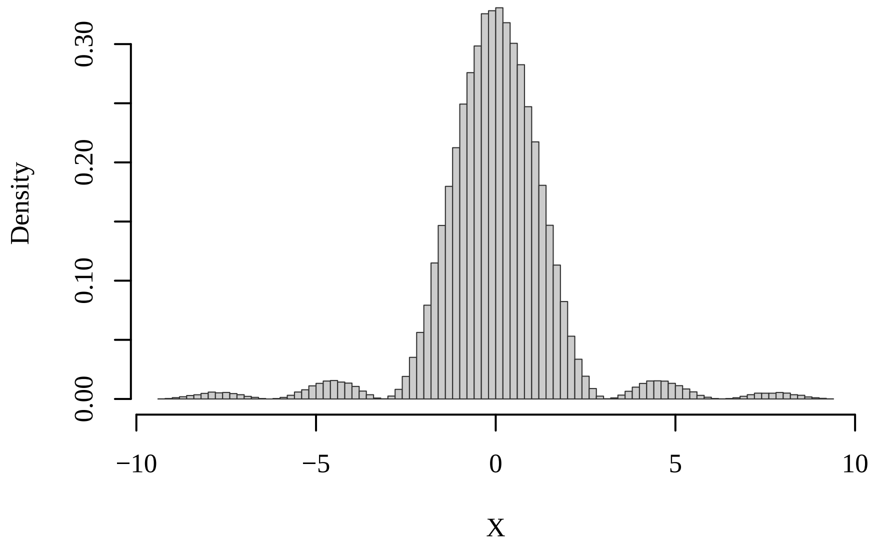

Figure 1. A histogram of the samples from question 18.

---

is not defined in this point; here we choose $X_0 \sim \mathcal{U}[-3\pi, 3\pi]$ instead. The resulting method can be implemented as follows.

```
GenerateMCMCSample <- function(N, sigma) {
  res <- numeric(length=N)
  X <- runif(1, -3*pi, +3*pi)
  for (j in 1:N) {
    Y <- rnorm(1, X, sigma)
    if (abs(Y) <= 3 * pi) {
      alpha <- (X * sin(Y) / (Y * sin(X))) ^ 2
    } else {
      alpha <- 0
    }
    U <- runif(1)
    if (U <= alpha) {
      X <- Y
    }
    res[j] <- X
  }
  return(res)
}
```

A sample of (dependent) values with (approximately) density π can now be generated as follows:

```
X <- GenerateMCMCSample(1000000, 6)
hist(X, breaks=100, prob=T, main=NULL, col="gray80", border="gray20")
```

The second line in the above listing plots a histogram of the sample, the result is shown in figure 1. The required moments $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$ and $\mathbb{E}(X^2)$ of X can be estimated using the sample generated above:

```
> mean(X)
[1] 0.0003257483
> mean(X^2)
[1] 3.117046
> mean(X^3)
[1] -0.01975976
> mean(X^4)
[1] 90.84923
```

Since π is symmetric, *i.e.* $\pi(x) = \pi(-x)$, the exact value for $\mathbb{E}(X)$ and $\mathbb{E}(X^3)$ is 0. Our estimates for the remaining two moments are $\mathbb{E}(X^2) \approx 3.117046$ and $\mathbb{E}(X^4) \approx 90.84923$. The probability $P\big(|X| \leq 2\big)$ can be estimated similarly:

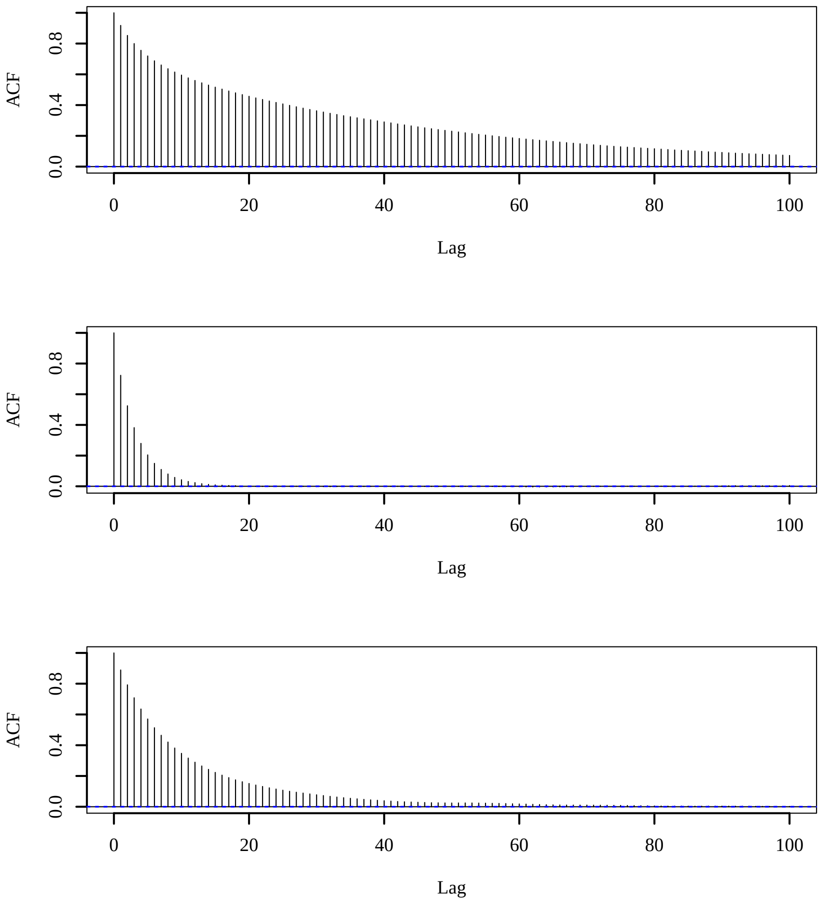

Figure 2. The auto-correlation functions for the Markov chain from question 18(c), for $\sigma = 1$, $\sigma = 6$ and $\sigma = 36$ (top to bottom).

```
> mean(abs(X) <= 2)
[1] 0.885983
```

Thus we have $P\big(|X| \leq 2\big) \approx 0.885983$.

c) Now we need to run the algorithm with different values of σ. To compute and plot the auto-correlations $\rho_k = \mathrm{Corr}(X_j, X_{j+k})$ we can use the R command `acf`:

```
par(mfrow=c(3,1))

X <- GenerateMCMCSample(1000000, 1)
acf(X, lag.max=100)

X <- GenerateMCMCSample(1000000, 6)
acf(X, lag.max=100)

X <- GenerateMCMCSample(1000000, 36)
acf(X, lag.max=100)
```

The result is shown in figure 2.

d) From lectures we know that

$$\mathrm{MSE}(Z_N) \approx \frac{\mathrm{Var}\big(X^4\big)}{N}\Big(1 + \sum_{k=1}^{\infty} \rho_k\Big).$$

This can be used to get the value of N required to get the MSE down to 0.01:

```
GetRequiredN <- function(sigma) {
  N <- 1000000
  X <- GenerateMCMCSample(N, sigma)
  a <- acf(X^4, lag.max=100, plot=F)
  sum.rho <- sum(a$acf[-1])
  MSE <- var(X^4) / N * (1 + sum.rho)
  cat("MSE =", MSE, "for N =", N, "\n")
  cat("required N = ", N * MSE / 0.01, "\n")
}
```

This function can be called as follows:

```
> GetRequiredN(1)
MSE = 9.835187 for N = 1e+06
required N =  983518666
> GetRequiredN(6)
MSE = 1.071058 for N = 1e+06
required N =  107105817
> GetRequiredN(36)
MSE = 2.811268 for N = 1e+06
required N =  281126766
```

Answer 19.

a) Let $X^* = X_K$ where $K \sim \mathcal{U}\{1, \ldots, N\}$, independently of the X_i. Then we have

$$\mathbb{E}^*_X(X^*) = \mathbb{E}^*_X(X_K) = \sum_{i=1}^n X_i P(K = i) = \frac{1}{n} \sum_{i=1}^n X_i.$$

Similarly, we can find

$$\mathbb{E}^*_X\big((X^*)^2\big) = \frac{1}{n} \sum_{i=1}^n X_i^2$$

and thus

$$\mathrm{Var}^*_X(X^*) = \mathbb{E}^*_X\big((X^*)^2\big) - \mathbb{E}^*_X(X_K)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \Big(\frac{1}{n} \sum_{i=1}^n X_i\Big)^2.$$

b) Using the results from part (a), we get

$$\mathrm{Var}\Big(\mathbb{E}^*_X(X^*)\Big) = \mathrm{Var}\Big(\frac{1}{n} \sum_{i=1}^n X_i\Big) = \frac{\sigma^2}{n}$$

and

$$\mathbb{E}\Big(\mathrm{Var}^*_X(X^*)\Big) = \mathbb{E}\Big(\frac{1}{n} \sum_{i=1}^n X_i^2 - \Big(\frac{1}{n} \sum_{i=1}^n X_i\Big)^2\Big) = \frac{n-1}{n}\sigma^2,$$

where we know the equation in the last equality sign, since we recognise the value of $\mathrm{Var}^*_X(X^*)$ as an estimator for the variance σ^2, and can remember that the bias of this estimator is $-\sigma^2/n$.

c) Using the results from part (b), the relation

$$\mathrm{Var}(X^*) = \mathbb{E}\big(\mathrm{Var}^*_X(X^*)\big) + \mathrm{Var}\big(\mathbb{E}^*_X(X^*)\big).$$

becomes

$$\sigma^2 = \frac{n-1}{n}\sigma^2 + \frac{\sigma^2}{n}.$$

This formula shows how the total variance of a bootstrap sample X^* splits into variance caused by picking X^* randomly from the list $(X_1, \ldots, X_n)$ of samples, and variance caused by the random choice of $(X_1, \ldots, X_n)$. For small n, the first term dominates. In particular, for $n = 1$, only one choice for X^* is possible, and thus $\mathrm{Var}^*_X(X^*) = 0$. At the same time, choice of the sample X_1 has a large influence on the bootstrap estimate, so $\mathbb{E}^*_X(X^*)$ has large variance. For $n = 1$, all of the variance falls into this part.

For large n, choosing a random X_i is nearly the same as choosing a random value from the underlying distribution, and thus $\mathrm{Var}^*_X(X^*) \approx \mathrm{Var}(X) = \sigma^2$. For the same reason we have $\mathbb{E}^*_X(X^*) \approx \mathbb{E}(X) = \mu$, and the choice of the X_i has only little effect, leading to a negligible variance.