Mandatory assignment 2

Deadline: Friday October 7th at 12:00 (noon, Norwegian time).

Hand in on Canvas. Submissions should be of **either** of the following types

- Preferred: Submit two files: One R markdown (Rmd) file containing both theory answers and R-code, and a pdf-file with the output you obtain when running (knitting) you R markdown file. See tutorial to get started.
- Submit two files: one pdf-file with a report containing the answers to the theory questions, and one file including the R-code.

The first line of R-code should be: `rm(list=ls())`. Check that the Rmd/R-code file runs before you submit it. Use comments in the R-code to clearly identify which question each part of the R-code belong to. Also try to add some comments to explain important parts of the code. The file ending of the R-code file should be .Rmd, .R or .r. The report can be handwritten and scanned to pdf-file, or written in your choice of text editor and converted to pdf. Cite the sources you use.

Problems marked with an $^{\mathrm{R}}$ should be solved in R, the others are theory questions.

It is OK to use any code from lectures, examples etc. AS LONG AS THESE ARE PROPERLY REFERENCED TO.

Problem 1:

Consider the Laplace distribution, which has probability density function

$$f(x) = \frac{1}{2}\exp(-|x|), \ -\infty < x < \infty$$

(where $|x|$ denotes the absolute value of x).[1]

a) Find the cumulative distribution function associated with $f(x)$. Also show that the inverse cumulative distribution function may be expressed as

$$F^{-1}(u) = \begin{cases} \log(2u) & \text{if } 0 < u \leq 1/2 \\ -\log(2-2u) & \text{if } 1/2 < u < 1 \end{cases}$$

b)$^{\mathrm{R}}$ Implement and test the inverse transform algorithm for generating n random numbers with density $f(x)$. The code should ideally avoid any for-loops.

Now, suppose R is a random variable taking the values -1 and 1 with equal probability (i.e. $P(R=-1) = P(R=1) = 1/2$) and $Y \sim Exp(1)$. Then it can be shown that the random variable

$$X = RY \tag{1}$$

will have the Laplace distribution with density $f(x)$, provided R and Y are independent. Note that this construction may be interpreted as X has a mixture distribution, with equal proportions of $Exp(1)$ and the negative of $Exp(1)$ random variables.

c)$^{\mathrm{R}}$ Use the above information to write an alternative function for generating n random numbers with density $f(x)$. Check your code, and compare with the inverse transform-method - which one would you prefer?

d) Explain why you cannot use an Accept-Reject algorithm with a $N(0, \sigma^2)$ proposal distribution $g(x)$ to sample from $f(x)$.

---

[1] When calculating integrals with respect to $f(x)$, it is a good idea to split the integration range into $(\infty, 0)$ and $(0, \infty)$ in order handle/avoid the absolute value. E.g.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx =$$

$$\int_{-\infty}^{0} \exp(x)/2dx + \int_{0}^{\infty} \exp(-x)/2dx = 1/2 + 1/2 = 1.$$

(first integral in the second line uses that $-|x| = x$ when $x < 0$).

Problem 2:

Consider two independent non-homogenous Poisson processes (NHPP) $N_1(t)$ and $N_2(t)$ with intensities $\lambda_1(t) = 1 + \frac{1}{2}\cos(t\pi/5)$ and $\lambda_2(t) = (1 + \min(10, t))/6$. In both cases $t \geq 0$.

a)$^{\text{R}}$ Implement functions for simulating from the two NHPPs (e.g. using the method of thinning, if so be explicit about which value of λ_{max} you are using). Check your routines agains the fact that both $N_1(10)$ and $N_2(10)$ have Poisson distributions with expectation 10.

b)$^{\text{R}}$ Using the routines from a), estimate $E(N_1(5))$, $E(N_2(5))$.
Let $T_{3,i}$ be the time until the third event with under rate λ_i, $i = 1, 2$. Make graphical representations of the distributions of $T_{3,i}$, $i = 1, 2$, and estimate $E(T_{3,i})$, $i = 1, 2$.
Suppose you wish to choose the process that attains the third event the fastest; which one would you choose?

Now consider the NHPP $N_3(t)$ with intensity $\lambda_3(t) = 1 + t/10$, $t \geq 0$. It can be shown that the event times $\{t_i\}$ of this process may be simulated as

$$t_i = \sqrt{t_{i-1}^2 + 20\varepsilon_i + 20t_{i-1} + 100} - 10, \tag{2}$$

where $t_0 = 0$ and $\varepsilon_i \sim$ iid $Exp(1)$.

c) Show that $N_3(20) \sim \text{Poisson}(40)$.

d)$^{\text{R}}$ Implement a function that simulates event times from the NHPP with intensity $\lambda_3(t)$ in the time interval $(0, b)$ based on (2).
Test your function against the information in point c).

Problem 3:

Consider the Gaussian AR(1) process (notice slightly different parameterization than in the Markov process notes)

$$X_{n+1} = \mu + \phi(X_n - \mu) + \varepsilon_{n+1},\ \varepsilon_n \sim \text{ iid } N(0, (1-\phi^2)\sigma^2)$$

for parameters μ, $-1 < \phi < 1$ and $\sigma > 0$. It can be shown that this process has stationary distribution $\pi = N(\mu, \sigma^2)$. This problem is intended to illustrate possibilities (and limitations) of using Markov processes to estimate moments of stationary distribution, which will be discussed more when we consider MCMC methods later in the course.[2]

a)$^{\text{R}}$ Write an R-function that simulates realizations of this process $X_1, \dots, X_T$ when $X_1 \sim N(\mu, \sigma^2)$ for general values of μ, ϕ, σ, and $T = 1, 2, \dots$.

b)$^{\text{R}}$ Set $T = 1000$, $\mu = 0.0$, $\sigma = 1$. Use the function from point a) to determine (via simulation) the mean and variance of M_1 and M_2, given by

$$M_1 = \frac{1}{T}\sum_{n=1}^{T} X_n, \text{ and } M_2 = \frac{1}{T}\sum_{n=1}^{T} X_n^2.$$

for the values of $\phi = -0.8, 0.0, 0.8$. Comment on what you find.

For any stationary Markov process X_n, $n = 1, \dots, T$, we may calculate the Effective Sample Size (ESS), T_{eff}. The ESS is defined implicitly so that

$$Var\left(\frac{1}{T}\sum_{n=1}^{T} X_n\right) = \frac{1}{T_{\text{eff}}} Var(X_n).$$

Roughly speaking, T_{eff} measures how correlated/dependent the samples of X_n are, and if the samples are iid (e.g. for $\phi = 0$ in the AR(1)-process), then $T_{\text{eff}} = T$. For the AR(1) process above it has an explicit form

$$T_{\text{eff}} = T\frac{1-\phi}{\phi+1}$$

c)$^{\text{R}}$ For $\sigma = 1$, how many samples from the AR(1)-process with either $\phi = 0.98$ or $\phi = -0.2$ would you need in order to get $Var(M_1) < (0.25)^2$. Check your answer using simulations.

In general, the ESS is calculated as

$$T_{\text{eff}} = \frac{T}{1 + 2\sum_{t=1}^{T} \rho(t)},\ \text{where } \rho(t) = corr(X_n, X_{n+t}),\ t = 1, 2, \dots$$

Here $\rho(t)$ is the autocorrelation function[3] associated with X_n. In practice, we may estimate the ESS (the details of estimation of the ESS is beyond the scope of this course) from a realization X_n, $n = 1, \dots, T$ using e.g. the R-function ESS given as

[2] In this and the next problem, we consider Markov processes with a known (Gaussian) stationary distribution, and therefore estimating e.g. the mean using this rather heavy machinery is not really necessary, i.e. we know already that the mean is μ. However, as will be clear later, we can construct Markov processes for stationary distributions that are arbitrary and complicated, and thus knowing how to obtain useful information from such Markov processes will turn out to be really useful.

[3] See e.g. https://en.wikipedia.org/wiki/Autocorrelation

```r
#install.packages("coda") # install the coda package once
#input x is a vector containing a realization of the process
ESS <- function(x){ return(as.numeric(coda::effectiveSize(x))) }
```

Notice however that this function produces *estimates* of the ESS, which are often subject to substantial variance.

d)[R] Simulate realizations of the AR(1)-process with $T = 1000$, $\mu = 0.0$, $\sigma = 1$, different values $\phi = -0.5, 0.0, 0.5, 0.9$ and check that on average, the ESS estimator for M_1 above gives results consistent with the explicit expression for the ESS above. E.g. check the mean, standard deviation and make a histogram.