

STAT 513/413, ASSIGNMENT 4

1. (a) For any random variables X, Y , and any a, b, c, d , calculate $\text{Cor}(aX + b, cY + d)$, in terms of $\text{Cor}(X, Y)$.

(b) Refer to Lecture 14, pages 17–20, about control variates: that is, consider the integral

$$\int_0^1 g_2(x)dx = \int_0^1 \frac{e^{-x}}{1+x^2}dx,$$

and implement the control variate variance reduction approach: first with the control variate $f_2(x) = 1 - x$, as in the lecture, and then with $f_3(x) = 1 - 0.9x$. For the sake of comparison, start with `set.seed(007)`, use 10000 uniform random numbers and then 100000, *exactly in the same way* as on page 21 in Lecture 14; the case $f_2(x)$ is to serve you as a control whether your code is correct – you are not supposed to proceed any further unless you obtain the very same results as in the lecture.

Once you are through this, set the seed up again, but now to be your personal number; then recompute the result for $f_2(x)$; then set up seed again (again to your personal number); and then compute, in the very same way, the result for $f_3(x)$. If f_3 is doing better or worse than f_2 ?

(c) After answering the last question, give an explanation why.

2. *The importance of importance sampling.* We know that when a random variable Z has standard normal distribution, the probability $P[Z > 4.5]$ is quite small. Suppose that instead of a numerical calculation of the cumulative distribution function of Z , we would like to estimate it via Monte Carlo. We will then compare the methods to the numeric value obtained via `pnorm()`.

(a) Try to do it first using the “straightforward” approach: estimate the probability via generating, say, $N = 10000$ numbers. Repeat several times: what kind of results do you get? Experiment with N : when will you start getting positive values?

(b) Implement now a method using importance sampling and compare the results (in particular with regard to the number N of required random numbers). Hint: exponential distribution (as readily available in R), or its modification. (But that does not say that you may not propose something other or better.)

3. *An easy proof omitted in the lectures.* Prove that the transition probabilities of the (discrete version of) the Metropolis method

$$p_x(y) = \begin{cases} q_x(y) \min \left\{ 1, \frac{\ell(y)}{\ell(x)} \right\} & \text{for } y \neq x, \\ 1 - \sum_{y \neq x} p_x(y) & \text{for } y = x, \end{cases}$$

yield a Markov chain with an invariant probability proportional to $\ell(x)$. State required assumptions.

4. *Can we together challenge the book?* Refer to Example 11.1 of the textbook (2019 edition; in the old edition it is 9.1) in the textbook that uses “Metropolis-Hastings sampler” to generate a sample from Rayleigh density. The latter is defined there to be

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad \text{for } x \geq 0,$$

and equal to 0 otherwise.

(a) Reuse (or rewrite, if you wish) the code given for this example and run a test run of it, comparing it via a `qqplot()` to the theoretical quantiles of the Rayleigh distribution – with one or two selected values of σ^2 , and also with one or two variants of the initial “burn-in”. Comment (briefly) on your findings. For the alternative you find most successful, set the random seed again, run the Markov chain several times (three or four times, say), aggregate the results and observe whether there is any change to the previous single run of the Markov chain. *(You may need a quantile function of the Rayleigh distribution for this; it is not absolutely necessary to derive it yourself, but do not use any formula without proper reference.)*

(b) Create your own implementation of the Metropolis method using the normal distribution centered at x (“random walk Metropolis sampler”) for the transition probabilities of the “proposal distribution”: that is, take

$$q_x(y) = e^{-\frac{(y-x)^2}{2s^2}}.$$

Run it for about two or three values s , and also experiment with the length of the burn-in (two or three values). Perform the same type of comparisons with the target Rayleigh distribution as you did in (a). Comment briefly on the results, also possibly comparing them with (a).

Due Friday, March 26, 2021