# Examination Computational Statistics

Linköpings Universitet, IDA, Statistik

| | |
|---|---|
| Course code and name: | 732A90 Computational Statistics |
| Date: | 2017/05/09 |
| Assisting teacher: | Krzysztof Bartoszek |
| Allowed aids: | Printed books and 100 page computer document |
| Grades: | A= $[18-20]$ points<br>B= $[15.5-18)$ points<br>C= $[10.5-15.5)$ points<br>D= $[8.5-10.5)$ points<br>E= $[7-8.5)$ points<br>F= $[0-7)$ points **(FAIL)** |
| Instructions: | Provide a detailed report that includes plots, conclusions and interpretations.<br>Give motivated answers to the questions. If an answer is not motivated,<br>the points are reduced. Provide all necessary codes in an appendix.<br>In a number of questions you are asked to do plots. Make sure that<br>they are informative, have correctly labelled axes, informative<br>axes limits and are correctly described.<br>Points may be deducted for poorly done graphs. |

## Assignment 1 (10p)

Recall that the binomial coefficient "$n$ choose $k$" is defined as

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{(k+1)(k+2)\cdots(n-1)n}{(n-k)!},$$

where $n$ and $k$ are an arbitrary pair of integers satisfying $0 \leq k \leq n$. Consider the three below `R` expressions for computing the binomial coefficient. They all use the `prod()` function, which computes the product of all the elements of the vector passed to it.

```
A) prod(1:n) / (prod(1:k) * prod(1:(n-k)))
B) prod((k+1):n) / prod(1:(n-k))
C) prod(((k+1):n) / (1:(n-k)))
```

### Question 1.1 (3p)

Even if overflow and underflow would not occur these expressions will not work correctly for all values of $n$ and $k$. Explain what is the problem in A, B and C respectively.

### Question 1.2 (4p)

In mathematical formulae one should suspect overflow to occur when parameters, here $n$ and $k$, are large. Experiment numerically with the code of A, B and C, for different values of $n$ and $k$ to see whether overflow occurs. Graphically present the results of your experiments.

### Question 1.3 (3p)

Which of the three expressions have the overflow problem? Explain why.

**TIP:** Recall special values of the binomial coefficient

$$\binom{n}{0} = 1, \quad \binom{n}{1} = n, \quad \binom{n}{2} = \frac{n(n-1)}{2}, \ldots, \binom{n}{n} = 1,$$

experiment with other values of $k$.

## Assignment 2 (10p)

The goal is to draw an i.i.d sample from the standard normal distribution $\mathcal{N}(0,1)$, whose probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right).$$

The function $\log(f(x))$ is dominated by e.g. the function

$$\log(g(x)) = \begin{cases} x + \frac{1}{2} - \frac{1}{2}\log(2\pi) & \text{for } x \leq -\frac{1}{2}, \\ -\frac{1}{2}\log(2\pi) & \text{for } -\frac{1}{2} \leq x \leq \frac{1}{2}, \\ -x + \frac{1}{2} - \frac{1}{2}\log(2\pi) & \text{for } x \geq -\frac{1}{2}. \end{cases}$$

The cumulative distribution function associated with $g(x)$ is

$$F_g(x) = \begin{cases} \frac{1}{3}(\sqrt{e})e^{x} & \text{for } x \leq -\frac{1}{2}, \\ \frac{1}{2} + \frac{x}{3} & \text{for } -\frac{1}{2} < x \leq \frac{1}{2}, \\ 1 - \frac{1}{3}(\sqrt{e})e^{-x} & \text{for } x > \frac{1}{2}. \end{cases}$$

**NOTE:** If you fail to do a part on which subsequent question(s) depend on describe (maybe using dummy data, partial code e.t.c.) how you would do them given you had done that part. You *might* be eligible for partial points.

## Question 2.1 (2p)

Write an R function to sample from $F_g$ using the inverse cumulative distribution function method.

**TIP:** $F_g$'s domain (i.e. $x$ values) is split into three regions $(-\infty, -0.5]$, $(-0.5, 0.5]$ and $(0.5, \infty)$. If you look at the boundary values of $F_g$ on these subintervals of $\mathbb{R}$ you will see that it splits the domain of $F_g^{-1}$, i.e. the inverse function to $F_g$, into the subintervals $(0, 1/3]$, $(1/3, 2/3]$ and $(2/3, 1)$.

## Question 2.2 (4p)

Implement a rejection sampling algorithm from $f(x)$, i.e. from the standard normal, using $g(x)$ as the majorizing density.

## Question 2.3 (2p)

In each step of the rejection sampling algorithm a proposed value may be accepted or rejected. Find the rejection rate, i.e. sample many points from $f(x)$ and for each point see how many times a rejection event took place. In particular present the histogram of the amount of rejections per sampled point, the expected value and variance of the amount of rejections per sampled point.

## Question 2.4 (2p)

Evaluate how good your rejection sampler is by looking at the expectation, variance and histogram of your generated sample. Compare to the mean, variance and histogram of the standard normal. You may use `rnorm()` to simulate from the standard normal. Use `t.test()` to compare the mean and `var.test()` for the variance.

**TIP:** Your generated sample should be of size about 10000 to get decently looking histograms. If your implementation is OK, a sample of this size is generated very quickly.