Statistical Computation - Assignment 3

Philip Dahlqvist-Sjöberg

23th of April 2020

1 Integration with trapezoidal rule

a) See R-script

b) See Figure 1, where the red verticle lines indicate where the function seems to not change anymore if you go further out on the tails. This indicates that if we integrate over the interval [-4,5], we would get approximately the same result as for the infinite interval.

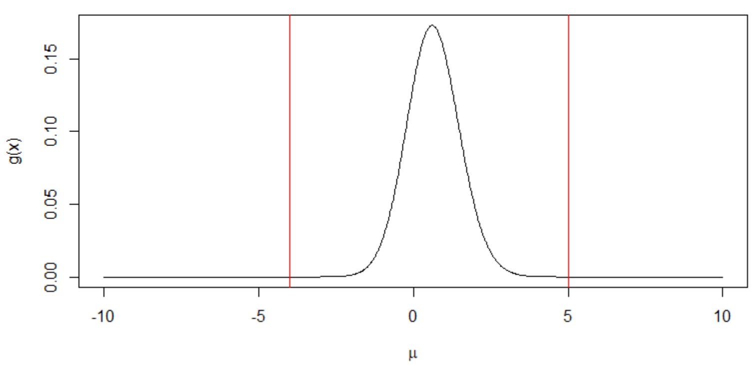

Figure 1: g($\mu|\bar{x}$) function plotted, with red lines where the function is flat.

c) Since,

$$f_{post}(\mu|\bar{x}) = \frac{1}{c}\int_{-\infty}^{\infty} g(\mu|\bar{x}) = 1$$
$$c \approx \int_{-4}^{5} g(\mu|\bar{x})d\mu = 0.3656$$

d) From the first and second order moment, we can find the expected value and variance of our parameter μ.

$$\int_{-4}^{5} \mu f_{post}(\mu|\bar{x})d\mu = 0.6658 \tag{1}$$
$$\int_{-4}^{5} \mu^2 f_{post}(\mu|\bar{x})d\mu = 1.177 \tag{2}$$
$$1.177 - 0.6658^2 = 0.7337 \tag{3}$$

In equation (1) the expected value is integrated to 0.6658, and in equation (3) the variance is integrated to 0.7337.

2 Sampling algorithms

a) We have the distribution function in equation (4),

$$f(x) = \begin{cases} 0 & \text{, if } 1 < x < -1 \\ x+1 & \text{, if } -1 \leq x \leq 0 \\ 1-x & \text{, if } 0 < x <\leq 1 \end{cases} \tag{4}$$

$$F(x) = \begin{cases} 0 & \text{, if } x < -1 \\ 0.5 + x^2 + x & \text{, if } -1 \leq x \leq 0 \\ 0.5 + x - x^2 & \text{, if } 0 < x <\leq 1 \\ 1 & \text{, if } 1 < x \end{cases} \tag{5}$$

$$F^{-1}(y) = \begin{cases} -1 & \text{, if } u \leq 0 \\ -1 + \sqrt{2}\sqrt{u} & \text{, if } 0 < u \leq 0.5 \\ -\frac{-10+\sqrt{-200u+200}}{10} & \text{, if } 0.5 < u < 1 \\ 1 & \text{, if } 1 \leq u \end{cases} \tag{6}$$

with the cumulative distribution function in (5), which is the integral of (4). Then the inverse in (6) is calculated using quadratic formula,

$$ax^2 + bx + c = 0$$

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

For the inverse transformation program, see R-code.

b) In Figure 2, the envelope, e(x), chosen is shown. The envelope is,

$$g(x) \sim N(0, 0.45^2)$$

$$e(x) = \frac{g(x)}{\alpha}, \text{ where } \alpha = 0.83 \text{ is a scaling factor, such that}$$

$$e(x) > f(x), \forall x$$

For program, see R-code.

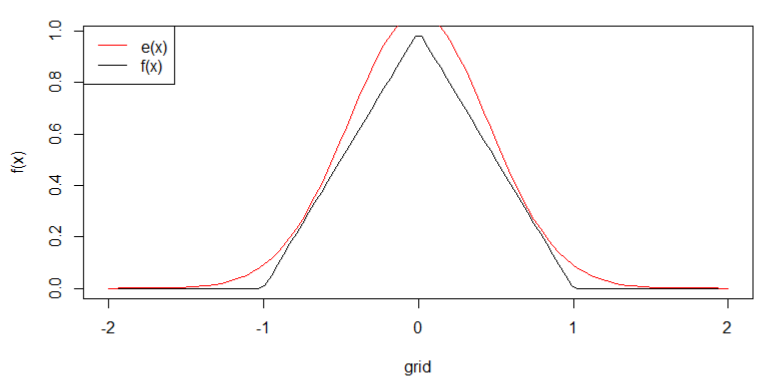

Figure 2: Plot of f(x) and envelope, e(x).

c) In Figure 3, the two triangle functions which together creates the structure for the composition sampling is shown.

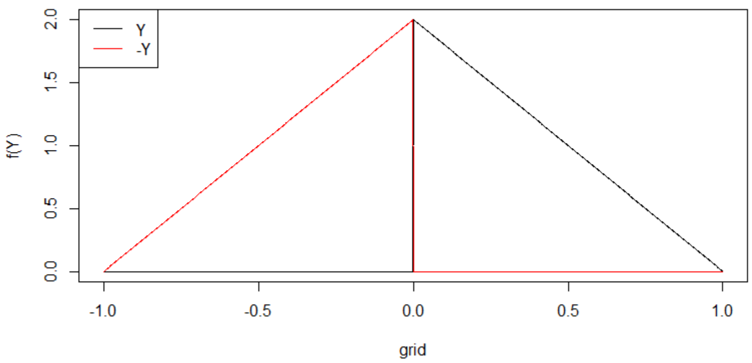

Figure 3: Plot of the two triangles for Y and -Y.

For program, see R-code.

d) For the uniformly difference linear combination method, the function is,

$$X_{sample} = U_1 - U_2$$
$$U_1 \text{ and } U_2 \sim U(0,1)$$

e) In Figure 4 the true distribution of X is shown. It has a clear peak at zero, and is defined for [-1,1]. For a good method of sampling, we are looking to match this shape of x.

In Figure 5, the four different methods of sampling is shown. They all have a relatively close shape as true x, but the method I would choose is c), composition sampling. In this case, it looks to fit the true shape best. This might be because, the triangles that we sample from, looks almost identical, together, to the true shape. However, this method might not be best in every circumstance, since it requires knowledge of the distribution of the similar functions.

If choosing for simplicity, the envelope with rejection sampling is a good and easier method to pull of.

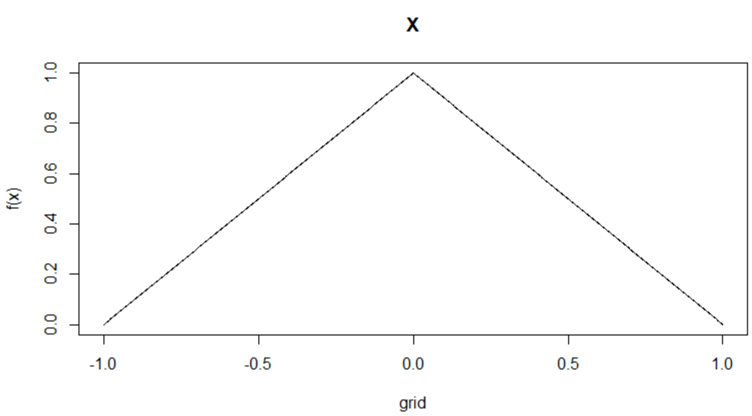

Figure 4: True distribution of X.

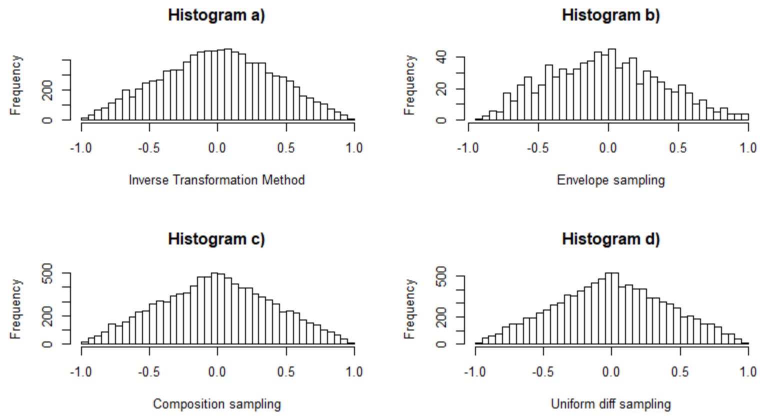

Figure 5: Histogram of the four methods of sampling.

Table 1: Power table for different μ with t-test.

μ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Power	0.1027	0.2088	0.3465	0.5209	0.6949	0.8293	0.92	0.9683	0.9898	0.9965	0.9995

3 Bootstrap confidence interval for simulation results

a) The analysis has been redone, see R-code, and the 95% confidence interval is shown in equation (7).

$$CI_{95\%} = (0.0735 \leq \hat{p} = 0.0737 \leq 0.0738) \tag{7}$$

$$CI^{B}_{95\%} = (0.0733 \leq \hat{p} = 0.0736 \leq 0.0739) \tag{8}$$

b) A bootstrap 95% confidence interval has been done with 500 replicates. See equation (8) for interval, and R-code for program.

c) The confidence interval in a) is shorter than in b). This means that the estimated parameter using the antithetic importance sampling of parametric design is more accurate than the non parametric design, that is bootstrap. However, the bootstrap method requires less assumptions, i.e., can be applied to any function.

That is a huge advantage for the bootstrap method, and as we can see, the interval is relatively close to the parametric design.

4 Simulation of power curves

a) See R-code for simulation and calculations of power of test. In Table 1 and 2, the power is shown for different μ.

b) In Figure 6, the power curves are ploted together. We can see that they follow the same shape, where they start and stop, at almost the same level. However, t-test has a higher power for all μ, and hence is the superior test in this simulation.

Table 2: Power table for different μ with SIGN.test.

μ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Power	0.092	0.179	0.288	0.416	0.558	0.691	0.797	0.869	0.948	0.972	0.99

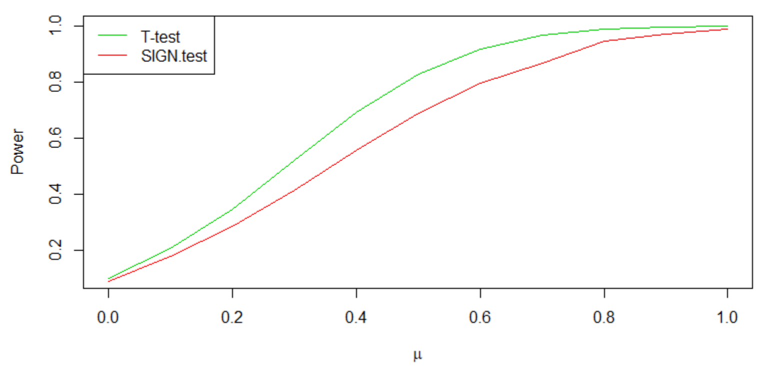

Figure 6: Power curves for t-test and SIGN.test, over different μ.