KTH Electrical Engineering

Exam in EG2080 Monte Carlo Methods in Engineering, 15 January 2014, 9:00–13:00, the seminar room

Instructions

The answer to each problem must begin on a new sheet, but answers to different parts of the same problem (a, b, c, etc.) can be written on the same sheet. The fields *Namn* (Name), *Blad nr* (Sheet number) and *Uppgift nr* (Problem number) must be filled out on every sheet.

Solutions should include sufficient detail that the argument and calculations can be easily followed.

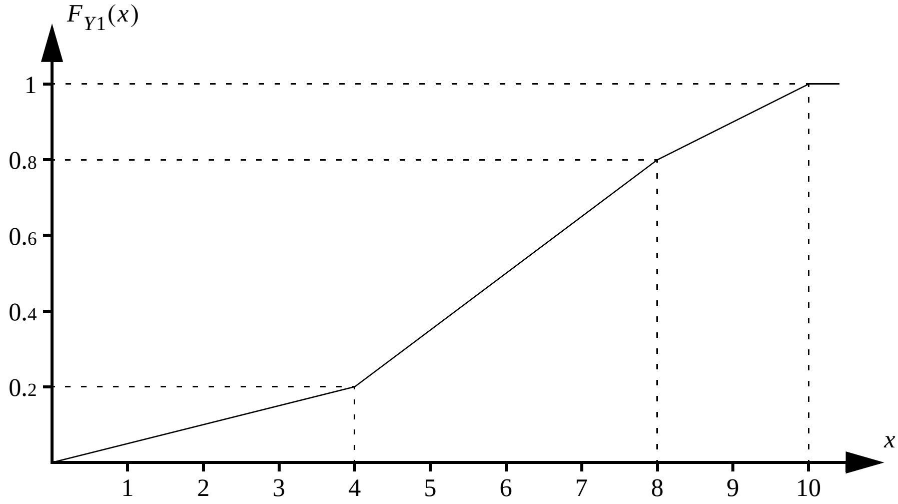
The exam can yield 40 points in total. The examinee is guaranteed to pass if the score is at least 33 points. There will be a possibility to complement for passing the exam if the result is at least 31 points.

Allowed aids

The following aids are allowed in this exam:

- Calculator without information relevant to the course.
- Formulae sheet.

Problem 1 (8 p)



a) (5 p) The figure above shows the distribution function of the continuous random variable Y_1. Which value of Y_1 is obtained if the random number 0.15 from a $U(0, 1)$-distribution is transformed according to the inverse transform method? What is the complementary random number of this value?

b) (3 p) The random variable Y_2 has the frequency function

$$f_{Y2}(x) = \begin{cases} 0.8 & x = 0, \\ 0.2 & x = 100, \\ 0 & \text{otherwise.} \end{cases}$$

Apply dagger sampling to generate a sequence of values of Y_2 based on the random number 0.15 from a $U(0, 1)$-distribution.

Problem 2 (8 p)

The following results have been obtained from a Monte Carlo simulation using simple sampling:

$$\sum_{i=1}^{900} x_i = 3\,780, \quad \sum_{i=1}^{900} x_i^2 = 19\,476.$$

a) (2 p) Calculate the estimated expectation value of X.

b) (4 p) Calculate the coefficient of variation.

c) (2 p) Assume that the coefficient of variation is used in the stopping criteria, and that the relative tolerance level is set to 0.01. Should more samples be generated or can the simulation be stopped at this point? Do not forget to motivate your answer!

Problem 3 (8 p)

Consider a simplified model of a system with one input:

$$\tilde{g}(Y) = \begin{cases} 10Y & \text{if } Y \le 50, \\ 15Y - 250 & \text{if } 50 < Y \le 80, \\ 25Y - 1\,050 & \text{if } 80 < Y. \end{cases}$$

The input Y is uniformly distributed between 0 and 100. Use the simplified model above to suggest an importance sampling function for this system!

Problem 4 (8 p)

Alice is using her car to travel to work. She can choose between two alternative routes. Route A is the shorter one, but traffic might on the other hand be busy. Route B is longer, but there are hardly any unforeseen traffic jams. In order to decide which way to go, Alice listens to the traffic reports on the radio and estimates the travel time for each road, and then she chooses the route with the shortest forecasted travel time. If the forecasted travel time is the same for both routes, she will choose route A. The real travel time might of course be different from the forecast, and it might turn out that the route she has chosen is slower (but then there is no turning back).

The forecasted and real travel times on route A are correlated, but can be considered independent if we separate between low and busy traffic as well as accurate and inaccurate forecast, as described in the table below. The forecasted travel time for route B is always $U(45, 75)$-distributed and the real travel time is also $U(45, 75)$-distributed. The forecasted and real travel time on route B can be considered independent.

Table 1 Overview of forecasted and real travel times on route A.

Situation	Probability [%]	Forecasted travel time [min]	Real travel time [min]
Low traffic, accurate forecast	60	$U(30, 40)$	$U(30, 40)$
Busy traffic, inaccurate forecast	10	$U(30, 40)$	$U(60, 90)$
Low traffic, inaccurate forecast	2	$U(60, 90)$	$U(30, 40)$
Busy traffic, accurate forecast	28	$U(60, 90)$	$U(60, 90)$

a) (4 p) Suggest appropriate strata for a simulation of Alice's travelling, where the objective of the simulation is to estimate the expected travel time.

b) (4 p) Assume that the simulation should start with a pilot study and that 100 scenarios should be generated in the pilot study. How would it be appropriate to distribute the scenarios in the pilot study among the strata you suggested in part a? Do not forget to motivate your answer!

Problem 5 (8 p)

Table 2 shows the results from a Monte Carlo simulation using importance sampling and a control variate. The expectation value of the simplified model, $\tilde{g}(Y)$, is equal to 582.5. What is the estimated expectation value of the studied model, $X = g(Y)$?

Table 2 Results from the Monte Carlo simulation in problem 5.

i	$f_Y(y_i)$	$f_Z(y_i)$	$x_i = g(y_i)$	$z_i = \tilde{g}(y_i)$
1	0.0100	0.0060	429	351
2	0.0100	0.0123	857	716
3	0.0100	0.0061	411	353
4	0.0100	0.0149	1 105	869
5	0.0100	0.0037	250	216
6	0.0100	0.0242	1 754	1 409
7	0.0100	0.0154	1 090	899
8	0.0100	0.0160	1 108	934
9	0.0100	0.0182	1 299	1 063
10	0.0100	0.0060	437	350

Problem 1

a) The original random value is given by $y = F_{Y1}^{-1}(0.15)$ = {read from figure} = 3 and the complementary random number is then $y^* = F_{Y1}^{-1}(1 - 0.15)$ = 8.5.

b) Dagger cycle length S = floor $(1/p)$ = $\{p = 0.2\}$ = 5 ⇒

$$F^{\ddagger}_{Y2j}(x) = \begin{cases} 100 & \text{if } 0.2(j-1) \le x < 0.2j, \\ 0 & \text{otherwise,} \end{cases} \quad j = 1, \ldots, 5$$

$$y_j = F^{\ddagger}_{Y2j}(0.15) = \begin{cases} 100 & j = 1, \\ 0 & j = 2, 3, 4, 5. \end{cases}$$

Problem 2

a) $m_X = \frac{1}{900}\sum_{i=1}^{900} x_i = 4.2.$

b) $$s_X^2 = \frac{1}{900}\sum_{i=1}^{900}(x_i - m_X)^2 = \frac{1}{900}\sum_{i=1}^{900} x_i^2 - m_X^2 = 4.0.$$

$$a_X = \frac{s_X}{m_X\sqrt{n}} = \frac{\sqrt{4}}{4.2\sqrt{900}} \approx 0.0159.$$

c) $a_X > 0.01$ ⇒ The stopping criteria is not fulfilled and the simulation should continue.

Problem 3

The expectation value of the simplified model is given by

$$\mu_Z = E[\tilde{g}(Y)] = \int_{-\infty}^{\infty} f_Y(x)\tilde{g}(x)dx = \{f_Y(x) = 0 \text{ for } x < 0 \text{ and } x > 100\} = \int_0^{100} f_Y(x)\tilde{g}(x)dx =$$

$$= \int_0^{50} \frac{10x}{100}dx + \int_{50}^{80} \frac{15x - 250}{100}dx + \int_{80}^{100} \frac{25x - 1\,050}{100}dx = \ldots = 582.5.$$

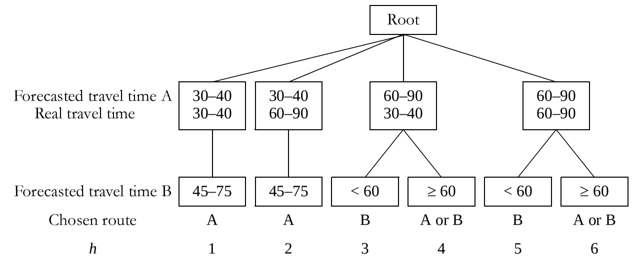
The optimal importance sampling function based on the simplified model is then given by

$$f_Z(x) = \frac{f_Y(x)\tilde{g}(x)}{\mu_Z} = \left\{ \frac{f_Y(x)}{\mu_Z} = \frac{\frac{1}{100}}{582.5} = \frac{1}{58\,250} \text{ for } 0 < x \le 100 \right\} =$$

$$= \begin{cases} 10x/58\,250 & \text{if } 0 < x \le 50, \\ (15x - 250)/58\,250 & \text{if } 50 < x \le 80, \\ (25x - 1\,050)/58\,250 & \text{if } 80 < x \le 100, \\ 0 & \text{all other } x. \end{cases}$$

Problem 4

a) We can use a strata tree to identify scenarios where Alice's choice of route is easily predicted. We put the situation (i.e., each row in table 1) on the first level below the root and the forecasted travel time of route B on the second level, as in the figure below. In the first stratum, the forecasted travel time of route A is always smaller than the forecasted travel time of route B; hence, Alice will choose route A. This is also the case in the second stratum, but the difference is that in this case the forecast is highly inaccurate and the resulting travel time will be much longer compared to the scenarios in the first stratum. In the third and fifth strata, the forecasted travel time of route B is always smaller then the forecast for route A, and Alice will go route B. In the remaining strata, Alice's choice cannot be predicted.[1].



b) The expectation values of strata 1, 2, 3 and 5 can be computed analytically. For example, in stratum 1 we know that Alice will choose route A and that the expected real travel time is 35 minutes. Similarly, in stratum 3 Alice will go route B and the expected real travel time will then be 60 minutes (the real travel time on route B is independent of the forecast and will consequently be $U(45, 75)$-distributed regardless of whether or not the forecasted travel time is less than 60 minutes).

This means that there are only two strata, where the expectation value need to be estimated. A reasonable solution is to allocate the same amount of scenarios in the pilot study to each stratum, i.e., 50 scenarios should be generated in stratum 4 and another 50 scenarios should be generated in stratum 6.

Problem 5

We want to estimate the difference between the detailed and simplified models, i.e., $X - Z$ (where $Z = \tilde{g}(Y)$). However, since importance sampling is used, each sample should be multiplied by a

1. It could be noted that strata 4 and 6 could be further divided to identify scenarios where the forecasted travel time on route A is larger than 75, in which case we know that Alice will choose route B. However, this strata tree is sufficient for this problem.

weight factor. Thus, we get the estimated difference

$$m_{(X-Z)} = \sum_{i=1}^{n} \frac{f_Y(y_i)}{f_Z(y_i)}(x_i - z_i) = \frac{1}{10}\left(\frac{0.0100}{0.0060}(429-351) + \ldots + \frac{0.0100}{0.0060}(437-350)\right) \approx 124.0.$$

Finally, we add the expectation value of the control variate to get the estimated expectation value of the detailed model:

$$m_X = m_{(X-Z)} + \mu_Z = 706.5.$$