KTH Electrical Engineering

Exam in EG2080 Monte Carlo Methods in Engineering, 20 December 2007, 8:00–13:00, 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.

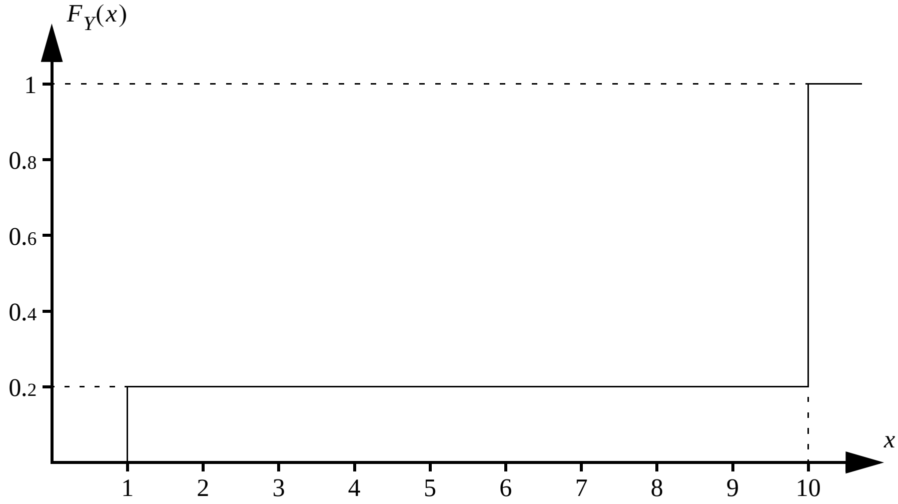
The exam can yield 40 points in total. The examinee is guaranteed to pass if the score is at least 33 points. There will be a possibility to complement for passing the exam with the grade E if the result is at least 31 points.

Allowed aids

The following aids are allowed in this exam:

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

Problem 1 (8 p)



a) (2 p) The figure above shows the distribution function of the discrete random variable Y. Apply the inverse transform method to generate a value of Y based on the random number 0.65 from a $U(0, 1)$-distribution.

b) (3 p) Assume that the value $y = 1$ has been randomised according to the probability distribution above. Do you have sufficient information to determine the complementary random value y^*? If yes, what is the value of y^*? If no, which information would be needed?

c) (3 p) Apply dagger sampling to generate a sequence of values of Y based on the random number 0.65 from a $U(0, 1)$-distribution.

Problem 2 (8 p)

The expectation value of the random variable X has been estimated using two different simulation methods. Each method was tested five times, with different seeds for the random number generator. The results are presented in table 1. It can be assumed that there is no bias in the estimates, i.e., $E[m_X] = E[X]$ for both simulation methods. Which method seems most efficient, according to these test runs? Do not forget to motivate your answer!

Table 1 Results of testing two different simulation methods.

Method 1		Method 2	
Estimate, m_X	Simulation time, T [s]	Estimate, m_X	Simulation time, T [s]
113	0.96	103	1.06
110	0.77	113	0.98
96	0.92	92	1.04
122	1.03	113	1.15
119	0.92	99	1.02

Problem 3 (8 p)

Consider a system $X = g(Y_1, Y_2)$, where

$$g(Y_1, Y_2) = \begin{cases} Y_1 + Y_2 & \text{if } Y_1 + Y_2 < 10, \\ 2(Y_1 + Y_2) & \text{if } 10 \le Y_1 + Y_2 < 20, \\ 3.25(Y_1 + Y_2) & \text{if } 20 \le Y_1 + Y_2. \end{cases}$$

The inputs are discrete random variables, and each input is either equal to 0, 5 or 10. The probability of each state is recorded in table 2.

Table 2 Probability distribution of the inputs in problem 3.

State, x	$f_{Y1}(x)$	$f_{Y2}(x)$
0	0.25	0.20
5	0.50	0.60
10	0.25	0.20

What is the optimal importance sampling function for this system?

Problem 4 (8 p)

Table 3 shows the results of the first 1 000 samples from a Monte Carlo simulation.

Table 3 Results from the Monte Carlo simulation in problem 4.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Simulation results	
			$\sum_{h=1}^{n_h} x_{h,i}$	$\sum_{h=1}^{n_h} x_{h,i}^2$
1	0.4	300	60 000	312 000 000
2	0.3	300	150 000	750 000 000
3	0.2	300	165 000	259 500 500
4	0.1	100	100 000	100 000 000

a) (2 p) Estimate the expectation value of the output variable.

b) (6 p) Suggest an appropriate sample allocation for the next 1 000 scenarios to be generated.

Problem 5 (8 p)

Consider a Monte Carlo simulation, where the objective is to estimate the difference between the two systems $X_1 = g_1(Y)$ and $X_2 = g_2(Y)$. There is also a simplified model $Z = \tilde{g}(Y)$; this model does not recognise the difference between the systems g_1 and g_2, but the expectation value of the simplified model can be calculated using an analytical method. Hence, it is known that $E[Z] = 500$.

Complementary random numbers are applied to one of the inputs of the systems, i.e., for each scenario y_i that is generated, we also get a complementary scenario y_i*. The results of the analysing 1 000 scenarios as well as their complementary scenarios are shown in table 4. What is the resulting estimate of $E[X_1 - X_2]$?

Table 4 Results of the Monte Carlo simulation in problem 5.

$\sum_{i=1}^{1000} g_1(y_i)$	$\sum_{i=1}^{1000} g_1(y_i^*)$	$\sum_{i=1}^{1000} g_2(y_i)$	$\sum_{i=1}^{1000} g_2(y_i^*)$	$\sum_{i=1}^{1000} \tilde{g}(y_i)$	$\sum_{i=1}^{1000} \tilde{g}(y_i^*)$
540 000	550 000	520 000	540 000	490 000	520 000

Problem 1

a) $y = F_Y^{-1}(u) = 10.$

b) If $y = 1$ then $u \le 0.2$. Hence, $u^* \ge 0.8 \Rightarrow y^* = 10$.

c) Dagger cycle length $S = \text{floor}(1/p) = \{p = 0.2\} = 5 \Rightarrow$

$$F^{\ddagger}_{Yj}(x) = \begin{cases} 1 & \text{if } 0.2(j-1) \le x < 0.2j, \\ 10 & \text{otherwise,} \end{cases} \quad j = 1, \ldots, 5$$

$$y_j = F^{\ddagger}_{Yj}(0.65) = \begin{cases} 10 & j = 1, \\ 10 & j = 2, \\ 10 & j = 3, \\ 1 & j = 4, \\ 10 & j = 5. \end{cases}$$

Problem 2

The simulation method having the least value of the product of the computation time and the variance of the estimate is considered to be most efficient. Hence, we need to estimate the expected computation time and the variance of the estimates:

$$E[T_k] \approx m_{Tk} = \frac{1}{5}\sum_{i=1}^{5} t_{k,i} = \begin{cases} 0.92 & k = 1, \\ 1.05 & k = 2, \end{cases}$$

$$E[M_{Xk}] \approx m_{MXk} = \frac{1}{5}\sum_{i=1}^{5} m_{Xk,i} = \begin{cases} 112 & k = 1, \\ 104 & k = 2, \end{cases}$$

$$Var[M_{Xk}] \approx s^2_{MXk} = \frac{1}{5}\sum_{i=1}^{5} (m_{Xk,i} - m_{MXk})^2 = \begin{cases} 82 & k = 1, \\ 66.4 & k = 2. \end{cases}$$

Since $m_{T1} \cdot s^2_{MX1} = 75.44 > m_{T2} \cdot s^2_{MX2} = 69.72$, we may conclude that method 2 is more efficient.

Problem 3

The inputs have nine possible states, and the expectation value $E[X] = \mu_X$ can be calculated using enumeration:

$$\mu_X = \sum_{i=1}^{9} f_Y(\psi_i) g(\psi_i) = \ldots = 20,$$

where f_Y is the combined probability distribution of both inputs. Then, the importance sampling function should be chosen as

$$f_Z(\psi) = \frac{g(\psi) f_Y(\psi)}{\mu_X},$$

More details are provided in the table below:

Y_1	Y_2	f_Y	$X = g(Y_1, Y_2)$	f_Z
0	0	0.05	0	0
0	5	0.15	5	0.0375
0	10	0.05	20	0.0500
5	0	0.10	5	0.0250
5	5	0.30	20	0.3000
5	10	0.10	30	0.1500
10	0	0.05	20	0.0500
10	5	0.15	30	0.2250
10	10	0.05	65	0.1625

Problem 4

a) First we estimate the expectation value for each stratum:

$$m_{Xh} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{h,i} = \begin{cases} 200 & h = 1, \\ 500 & h = 2, \\ 550 & h = 3, \\ 1\,000 & h = 4. \end{cases}$$

The estimate for the entire population is obtained by

$$m_X = \sum_{h=1}^{4} \omega_h m_{Xh} = 440.$$

b) The estimated variance in each stratum is given by

$$s^2_{Xh} = \frac{1}{n_h}\sum_{i=1}^{n_h} (x_{h,i} - m_{Xh})^2 = \frac{1}{n_h}\sum_{i=1}^{n_h} x^2_{h,i} - m^2_{Xh} = \begin{cases} 1\,000\,000 & h = 1, \\ 2\,250\,000 & h = 2, \\ 562\,500 & h = 3, \\ 0 & h = 4. \end{cases}$$

According to the Neyman allocation, the target distribution of 2 000 samples should be

$$n_h = n\frac{\omega_h s_{Xh}}{\sum_{k=1}^{L} \omega_k s_{Xk}} = \begin{cases} 800 & h = 1, \\ 900 & h = 2, \\ 300 & h = 3, \\ 0 & h = 4. \end{cases}$$

Apparently, we already have enough scenarios in strata 3 and 4; the next 1 000 scenarios should therefore be allocated to strata 1 and 2. One possible compromise is to generate 455 scenarios in stratum 1 and 545 scenarios in stratum 2.

Problem 5

The control variate is not relevant for the difference between X_1 and X_2. Based on the original scenarios, the estimated difference between the systems is

$$\frac{1}{1\,000}\sum_{i=1}^{1000}(x_{1,i}-x_{2,i}) = \frac{1}{1\,000}\left(\sum_{i=1}^{1000}x_{1,i}-\sum_{i=1}^{1000}x_{2,i}\right)$$

and the estimate based on the complementary scenarios is

$$\frac{1}{1\,000}\left(\sum_{i=1}^{1000}x_{1,i}^{*}-\sum_{i=1}^{1000}x_{2,i}^{*}\right).$$

The final estimate is the average of these two, i.e.,

$$m_{(X1-X2)} = \frac{1}{2\,000}\left(\sum_{i=1}^{1000}x_{1,i}-\sum_{i=1}^{1000}x_{2,i}+\sum_{i=1}^{1000}x_{1,i}^{*}-\sum_{i=1}^{1000}x_{2,i}^{*}\right) = 15.$$