KTH Electrical Engineering

Exam in EG2080 Monte Carlo Methods in Engineering, 21 December 2011, 14:00–18:00, large conference 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)

Y_1 and Y_2 are two correlated random variables. The joint frequency function is given by

$$f_Y(x_1, x_2) = \begin{cases} 0.3 & \text{if } x_1 = 1, x_2 = 1, \\ 0.2 & \text{if } x_1 = 1, x_2 = 2, \\ 0.2 & \text{if } x_1 = 2, x_2 = 1, \\ 0.3 & \text{if } x_1 = 2, x_2 = 2, \\ 0 & \text{all other } (x_1, x_2). \end{cases}$$

Generate random values of Y_1 and Y_2. You may use any of the $U(0, 1)$-distributed random numbers provided in table 1.

Table 1 List of random values for problem 1.

Number	1	2	3	4
Value	0.82	0.13	0.63	0.28

Problem 2 (8 p)

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

$$\sum_{i=1}^{500} x_i = 248, \quad \sum_{i=1}^{500} x_i^2 = 252.$$

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 system $X = g(Y_1, Y_2)$. The inputs Y_1 and Y_2 are independent and there are six possible states for this system, as listed in the table below.

y_1	y_2	$f_{Y1}(x)$	$f_{Y2}(x)$	$g(y_1, y_2)$
1	0	0.2	0.01	5
1	2	0.2	0.99	1
2	0	0.6	0.01	10
2	2	0.6	0.99	2
3	0	0.2	0.01	15
3	2	0.2	0.99	7

The system is to be simulated using importance sampling. The following importance sampling functions have been suggested for the two inputs:

$$f_{Z1}(x) = \begin{cases} 0.1 & x = 1, \\ 0.5 & x = 2, \\ 0.4 & x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_{Z2}(x) = \begin{cases} 0.1 & x = 0, \\ 0.9 & x = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Which of the following three alternatives is most efficient: to apply only the importance sampling function for input 1, to apply only the importance sampling function for input 2, or to apply the importance sampling functions for both inputs?

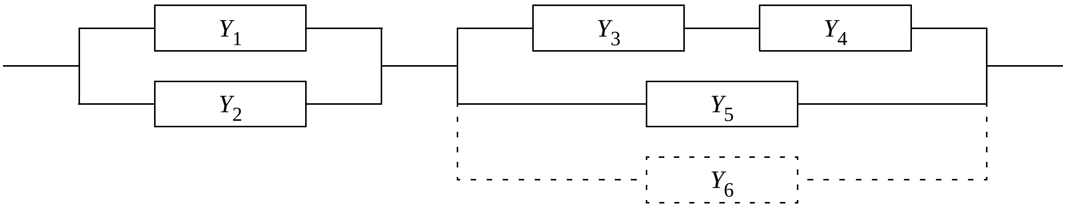
Problem 4 (8 p)

Consider the same system as in problem 3. Assume that this system is simulated using stratified sampling. Two strata have been defined so that all scenarios where $Y_2 = 0$ belong to stratum 1 and all scenarios where $Y_2 = 2$ belongs to stratum 2. Moreover, assume that $n/2$ scenarios will be generated in stratum 1 and $n/2$ scenarios in stratum 2.

a) (7 p) Is this application of stratified sampling more efficient than simple sampling with n samples?

b) (1 p) Can the efficiency of stratified sampling be improved further? If that is the case, what should be done? (A brief explanation is sufficient—no calculations are required.)

Problem 5 (8 p)



Consider the system above. In its original configuration it has five components. Now it is considered to add some extra redundancy to the system, by adding a backup component for components 3–5. Each component has a reliability of 90% and the state of each component is an input to the system. The inputs are equal to 1 if the corresponding component is operational, and 0 otherwise. The output is equal to 1 if the system is working, and 0 if it is not. Thus, for the original configuration we get the model

$$X_1 = g_1(Y) = \begin{cases} 1 & \text{if } (Y_1 + Y_2)(Y_3Y_4 + Y_5) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for the second model we get

$$X_2 = g_2(Y) = \begin{cases} 1 & \text{if } (Y_1 + Y_2)(Y_3Y_4 + Y_5 + Y_6) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The improved reliability of the system is to be estimated using correlated sampling and importance sampling. The importance sampling function is set up so that the probability of components 3–5 being operational is set to 50% each, whereas the other components still each have 90% probability of working, i.e.,

$$f_{Z1}(x) = f_{Y1}(x), f_{Z2}(x) = f_{Y2}(x), f_{Z3}(x) = f_{Z4}(x) = f_{Z5}(x) = \begin{cases} 0.5 & x = 0, \\ 0.5 & x = 1, \\ 0 & \text{otherwise,} \end{cases}$$

and $f_{Z6}(x) = f_{Y6}(x)$.

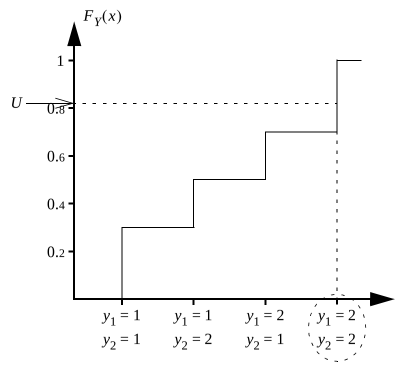
Table shows ten scenarios generated using these importance sampling functions. Based on these scenarios, what is the estimated increase in reliability by adding the sixth component?

Table 2 Ten scenarios from the Monte Carlo simulation in problem 5.

Scenario	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
1	1	1	0	1	1	1
2	1	1	0	1	1	1
3	1	0	1	1	1	1
4	1	1	1	0	1	0
5	1	0	0	1	1	1
6	1	0	0	0	1	1
7	1	1	0	1	1	1
8	1	1	1	0	0	1
9	1	1	1	0	1	1
10	1	1	1	1	1	1

Problem 1

The inverse transform method can be applied to the joint distribution function, as shown in the figure below. Using the first of the random values in table 1 gives the values $y_1 = 2$ and $y_2 = 2$.



Problem 2

a) $m_X = \frac{1}{500}\sum_{i=1}^{500} x_i = 0.496.$

b) $s_X^2 = \frac{1}{500}\sum_{i=1}^{500}(x_i - m_X)^2 = \frac{1}{500}\sum_{i=1}^{500} x_i^2 - m_X^2 = 0.258.$

$$a_X = \frac{s_X}{m_X\sqrt{n}} = \frac{\sqrt{0.258}}{0.496\sqrt{500}} \approx 0.0458.$$

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

Problem 3

The importance sampling function resulting in the lowest value of $Var[M_X]$ will be the most efficient (assuming that the differences in computation time are negligible). The variance of the estimate is given by

$$Var[M_X] = \frac{1}{n}\left(\sum_{\psi \in \Upsilon} g^2(\psi)\frac{f_Y^2(\psi)}{f_Z(\psi)} - \mu_X^2\right),$$

where Υ is the set of possible input values. Enumeration of Υ yields the following results:

ψ: y_1	y_2	$g(\psi)$	Real input probability distribution $f_Y(\psi) = f_{Y1}(y_1)f_{Y2}(y_2)$	Importance sampling applied to the first input $f_Z(\psi) = f_{Z1}(y_1)f_{Y2}(y_2)$	Importance sampling applied to the second input $f_Z(\psi) = f_{Y1}(y_1)f_{Z2}(y_2)$	Importance sampling applied to both inputs $f_Z(\psi) = f_{Z1}(y_1)f_{Z2}(y_2)$
1	0	5	0.002	0.001	0.020	0.010
1	3	1	0.198	0.099	0.180	0.090
2	0	10	0.006	0.005	0.060	0.050
2	3	2	0.594	0.495	0.540	0.450
3	0	15	0.002	0.004	0.020	0.040
3	3	7	0.198	0.396	0.180	0.360

Hence, the variances are given by

$$Var[M_X] \approx \frac{9.14 - \mu_X^2}{n} \text{ (when importance sampling is applied to the first input only)},$$

$$Var[M_X] \approx \frac{13.61 - \mu_X^2}{n} \text{ (when importance sampling is applied to the second input only)},$$

and

$$Var[M_X] \approx \frac{9.01 - \mu_X^2}{n} \text{ (when importance sampling is applied to both inputs)}.$$

The true expectation value μ_X does not depend on the importance sampling function; therefore, we can conclude that applying importance sampling to both inputs is the most efficient alternative.

Problem 4

a) Before we can investigate the variances of the two estimates, we need the properties of the variables to be sampled:

$$\mu_X = \sum_{x_1}\sum_{x_2} f_{Y1}(x_1)f_{Y2}(x_2)g(x_1, x_2) = 2.87,$$

$$\sigma_X^2 = \sum_{x_1}\sum_{x_2} f_{Y1}(x_1)f_{Y2}(x_2)(g(x_1, x_2) - \mu_X)^2 = 5.13,$$

$$\mu_{X1} = \sum_{x_1} f_{Y1}(x_1)g(x_1, 0) = 10,\ \sigma_{X1}^2 = \sum_{x_1} f_{Y1}(x_1)(g(x_1, 0) - \mu_{X1})^2 = 10,$$

$$\mu_{X2} = \sum_{x_1} f_{Y1}(x_1)g(x_1, 2) = 2.8,\ \sigma_{X2}^2 = \sum_{x_1} f_{Y1}(x_1)(g(x_1, 2) - \mu_{X2})^2 = 4.56.$$

We also need the stratum weights:

$$\omega_1 = P(Y_2 = 0) = f_{Y2}(0) = 0.01,\ \omega_2 = P(Y_2 = 2) = f_{Y2}(2) = 0.99.$$

The variance of the estimate when using simple sampling would be

$$Var[M_X] = \frac{\sigma_X^2}{n} = \frac{5.13}{n}.$$

This should be compared to the variance of the estimate when using stratified sampling, which is given by

$$Var[M_X] = \sum_{h=1}^{2} \omega_h^2 \frac{\sigma_{Xh}^2}{n/2} = \frac{2 \cdot 0.01^2 \cdot 10 + 2 \cdot 0.99^2 \cdot 4.56}{n} = \frac{8.94}{n}.$$

Assuming that the differences in computation time are negligible, the variance of the estimate is lower for simple sampling; hence, this application of stratified sampling was not efficient.

b) The efficiency of stratified sampling can be improved by changing the strata definition (so that strata are more homogenous) or by applying the Neyman allocation instead of proportional sampling.

Problem 5

The results for the ten scenarios are given in the table below.

Scenario	1	2	3	4	5	6	7	8	9	10
$g_1(z)$	1	1	1	1	1	1	1	0	1	1
$g_2(z)$	1	1	1	1	1	1	1	1	1	1

The expected difference between the two systems is estimated by

$$m_{(X2-X1)} = \frac{1}{n} \sum_{i=1}^{n} w(z_i)(g_2(z_i) - g_1(z_i)).$$

In this case, the difference $g_2(z_i) - g_1(z_i)$ is equal to zero for all scenarios except the eighth, i.e.,

$$m_{(X2-X1)} = \frac{1}{10} w(z_8) = \frac{1}{10}\left(\frac{0.9}{0.9} \cdot \frac{0.9}{0.9} \cdot \frac{0.9}{0.5} \cdot \frac{0.1}{0.5} \cdot \frac{0.1}{0.5} \cdot \frac{0.9}{0.9}\right) = 0.007.$$