KTH Electrical Engineering

Exam in EG2080 Monte Carlo Methods in Engineering, 4 February 2008, 14:00–19:00, the 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 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) (5 p) The random variable Y_1 has the frequency function

$$f_{Y1}(x) = \begin{cases} 0.5 & x = 1, \\ 0.25 & x = 2, \\ 0.25 & x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

Which value of Y_1 is obtained if the random number 0.55 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

$$f_{Y2}(x) = \begin{cases} 0.7 & x = 1, \\ 0.3 & x = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Apply dagger sampling to generate a sequence of values of Y_2 based on the random number 0.55 from a $U(0, 1)$-distribution.

Problem 2 (8 p)

The following samples have been obtained in a Monte Carlo simulation using simple sampling:

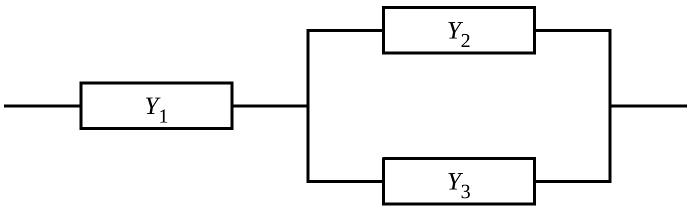
300, 26, 93, 67, 222, 144, 61, 2, 172, 59.

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

b) (2 p) Calculate the estimated variance of X.

c) (4 p) Calculate a 99% confidence interval for the expectation value of X. It can be assumed that m_X is normally distributed.

Problem 3 (8 p)



Consider the system above, where each component has a reliability of 90%. The inputs Y_1, Y_2 and Y_3 are equal to 1 if a component is operational, and 0 otherwise. The output X is equal to 1 if the system is working, and 0 if it is not, i.e.,

$$X = \begin{cases} 1 & \text{if } Y_1 = 1 \text{ and } Y_2 + Y_3 \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The system is to be simulated using importance sampling. Is it more efficient to use the importance sampling function

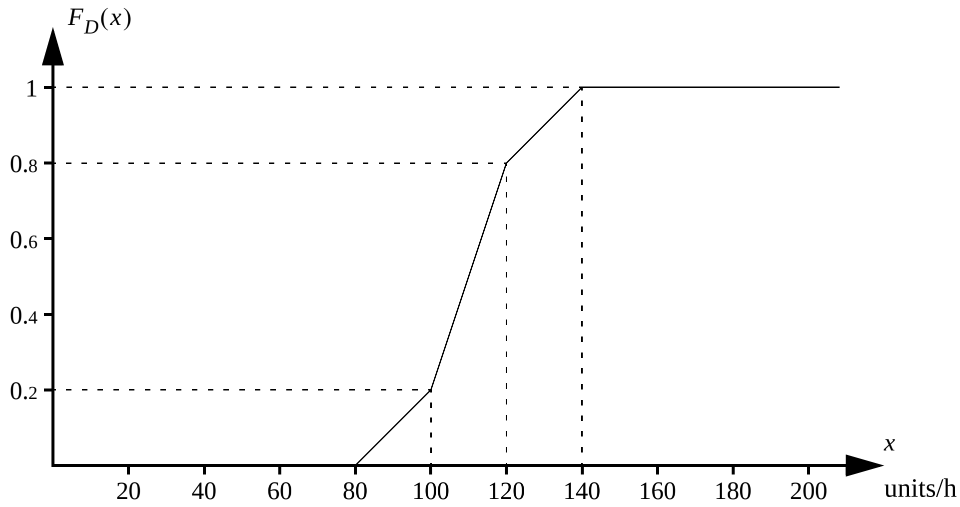
$$f_{Z1}(y) = \begin{cases} 0.5 & y = 0, \\ 0.5 & y = 1, \\ 0 & \text{all other } y, \end{cases}$$

for each component, or to use the importance sampling function

$$f_{Z2}(y) = \begin{cases} 0.2 & y = 0, \\ 0.8 & y = 1, \\ 0 & \text{all other } y, \end{cases}$$

for each component? The computation time is the same for both importance sampling functions.

Problem 4 (8 p)



The figure above shows the distribution function of the hourly demand, D, for The Company. The Company uses two machines to produce enough units to cover the demand. Normally, the availability of the machines is 90%, but whenever one machine is down, the heavier loading of the other machine decreases the availability; hence, there is a correlation between outages in the two machines. The probability of both machines being available is 81%, the probability that only one machine is available is 16% and the probability that none machine is available is 3%. When available, each machine can produce 100 units per hour.

a) (4 p) Suggest strata which differentiates scenarios where the available production capacity is larger than the demand from scenarios where the available production capacity is insufficient.
Hint: A strata tree might be useful.

b) (4 p) Calculate the stratum weights for the stratification you suggested in part a.

Problem 5 (8 p)

Monte Carlo simulation has been used to compare the two systems $X_1 = g_1(Y)$ and $X_2 = g_2(Y)$. Some results of the simulation are shown in table 1. Calculate the estimated difference between $E[X_1]$ and $E[X_2]$. If the table does not contain all the information you need to solve the problem, then you may assume an arbitrary value for the missing information.

Table 1 Results from the Monte Carlo simulation in problem 5.

Stratum, h	Stratum weight, ω_h	Samples per stratum, n_h	Results from system 1 $\sum_{i=1}^{n_h} g_1(y_{h,i})$	Results from system 2l, $\sum_{i=1}^{n_h} g_2(y_{h,i})$
1	0.5	200	1 600	1 500
2	0.3	400	4 800	3 800
3	0.2	400	8 000	6 000

Problem 1

a) The distribution function of Y_1 is given by

$$F_{Y1}(x) = \sum_{t \le x} f_{Y1}(t) = \begin{cases} 0 & x < 1, \\ 0.5 & 1 \le x < 2, \\ 0.75 & 2 \le x < 3, \\ 1 & 3 \le x. \end{cases}$$

The original random value is given by $y = F_{Y1}^{-1}(0.55) = 2$ and the complementary random number is then $y^* = F_{Y1}^{-1}(1 - 0.55) = 1$.

b) Dagger cycle length S = floor $(1/p) = \{p = 0.3\} = 3 \Rightarrow$

$$F^{\ddagger}_{Yj}(x) = \begin{cases} 2 & \text{if } 0.3(j-1) \le x < 0.3j, \\ 1 & \text{otherwise,} \end{cases} \quad j = 1, 2$$

$$y_j = F^{\ddagger}_{Yj}(0.55) = \begin{cases} 1 & j = 1, \\ 2 & j = 2, \\ 1 & j = 3. \end{cases}$$

Problem 2

a) $m_X = \frac{1}{10} \sum_{i=1}^{10} x_i = 114.6.$

b) $s_X^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - m_X)^2 \approx 7\,929.2.$

c) $\delta = \frac{t_\alpha \cdot s_X}{\sqrt{n}} \approx \{t_{0.99} = 2.5758\} \approx 72.5 \Rightarrow 114.6 \pm 72.5$ is a 99% confidence interval for $E[X]$.

Problem 3

The importance sampling function resulting in the lowest value of $Var[M_X]$ will be the most efficient, as the computation time is the same for both methods. The variance of the estimate is given by

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

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

ψ			$g(\psi)$	$f_Y(\psi) = f_{Y1}(y_1)f_{Y2}(y_2)f_{Y3}(y_3)$	$f_Z(\psi) = f_{Z1}(y_1)f_{Z1}(y_2)f_{Z1}(y_3)$	$f_Z(\psi) = f_{Z2}(y_1)f_{Z2}(y_2)f_{Z2}(y_3)$
y_1	y_2	y_3				
0	0	0	0	0.001	0.125	0.008
0	0	1	0	0.009	0.125	0.032
0	1	0	0	0.009	0.125	0.032
0	1	1	0	0.081	0.125	0.128
1	0	0	0	0.009	0.125	0.032
1	0	1	1	0.081	0.125	0.128
1	1	0	1	0.081	0.125	0.128
1	1	1	1	0.729	0.125	0.512

Hence, the variances are given by

$$Var[M_X] \approx \frac{4.36 - \mu_X^2}{n} \text{ (for the first method)}$$

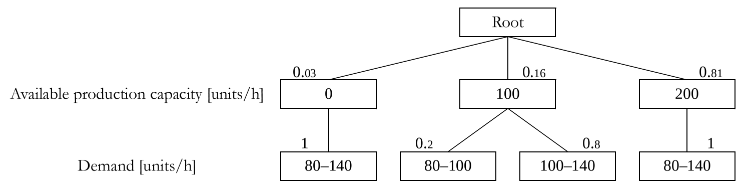
and

$$Var[M_X] \approx \frac{1.14 - \mu_X^2}{n} \text{ (for the second method).}$$

The true expectation value μ_X does not depend on the importance sampling function; therefore, we can conclude that the second method (i.e., using the importance sampling function f_{Z2}) is more efficient.

Problem 4

a) We may use a strata tree to compare the available production capacity to the demand. Put the three possible states of the available production capacity on the first level below the root and then the demand on the next level.



We have four branches in the strata tree, but since we only want to differentiate between sufficient or insufficient capacity, we do not need to introduce more than two strata; the first and third branch all represent scenarios where the production capacity is sufficient, and the second and fourth then represent scenarios where the capacity is not enough to cover the demand.

b) The resulting stratum weights are the sum of the products of the node weights along each branch, i.e.,

$$\omega_1 = 0.81 \cdot 1 + 0.16 \cdot 0.2 = 0.842,$$

$$\omega_1 = 0.16 \cdot 0.8 + 0.03 \cdot 1 = 0.158.$$

Problem 5

The information is sufficient. We start by calculating the estimated difference between the two systems for each stratum:

$$m_{(X1-X2)h} = \frac{1}{n_h} \sum_{i=1}^{h} (g_1(y_{h,i}) - g_2(y_{h,i})) = \begin{cases} 0.5 & h = 1, \\ 2.5 & h = 2, \\ 5 & h = 3. \end{cases}$$

Then we calculate the estimated difference between the detailed model and the control variate for the entire population:

$$m_{(X1-X2)} = \sum_{=1}^{3} \omega_h m_{(X1-X2)i} = 2.$$