KTH Electrical Engineering

Exam in EG2080 Monte Carlo Methods in Engineering, 18 December 2008, 13:00–17:00, 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.

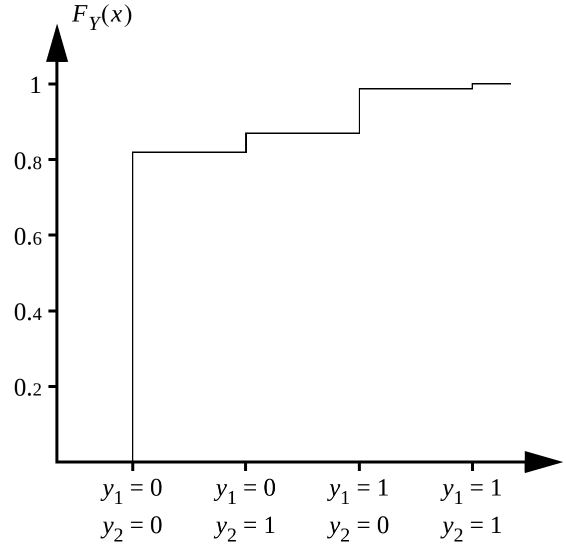
Allowed aids

The following aids are allowed in this exam:

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

Problem 1 (8 p)

A system has two binary inputs, Y_1 and Y_2, such that $P(Y_1 = 0) = 0.87$, $P(Y_1 = 1) = 0.13$, $P(Y_2 = 0) = 0.94$ and $P(Y_2 = 1) = 0.06$. The probability distributions of these two inputs have been combined into a single joint distribution as shown in the figure below.



a) (3 p) Apply the inverse transform method to convert the random value 0.46 from a $U(0, 1)$-distribution into values of Y_1 and Y_2.

b) (2 p) Assume that 20 scenarios are to be generated (i.e., we must generate 20 values of Y_1 and 20 values of Y_2). How many values from the random number generator are needed if complementary random numbers are applied? Do not forget to motivate your answer!

c) (1 p) Assume that 20 scenarios are to be generated. How many values from the random number generator if dagger sampling is applied? Assume that the dagger cycles for the two inputs are independent. Do not forget to motivate your answer!

d) (1 p) Assume that 20 scenarios are to be generated. How many values from the random number generator if dagger sampling is applied? Assume that the dagger cycles for the two inputs are reset at the end of the shorter cycle. Do not forget to motivate your answer!

e) (1 p) Assume that 20 scenarios are to be generated. How many values from the random number generator if dagger sampling is applied? Assume that the dagger cycles for the two inputs are reset at the end of the longer cycle. Do not forget to motivate your answer!

Problem 2 (8 p)

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

–14, –2, 11, –21, –9, –6, –6, 5, –4, –6.

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

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

c) (4 p) Assumed that M_X is normally distributed. How confident can one be that the true value of $E[X]$ is less than zero?

Hint: A table of the distribution function of the standardised normal distribution, i.e.,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

is given below. For $x < 0$ we can calculate $\Phi(x)$ using $\Phi(-x) = 1 - \Phi(x)$.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361

Problem 3 (8 p)

Consider a system $X = Y_1 \cdot Y_2$. There are nine possible scenarios for this system, as listed in the table below.

y_1	y_2	$g(y)$	$f_Y(y)$
–1	–1	1	0.04
–1	0	0	0.12
–1	1	–1	0.04
0	–1	0	0.12
0	0	0	0.36
0	1	0	0.12
1	–1	–1	0.04
1	0	0	0.12
1	1	1	0.04

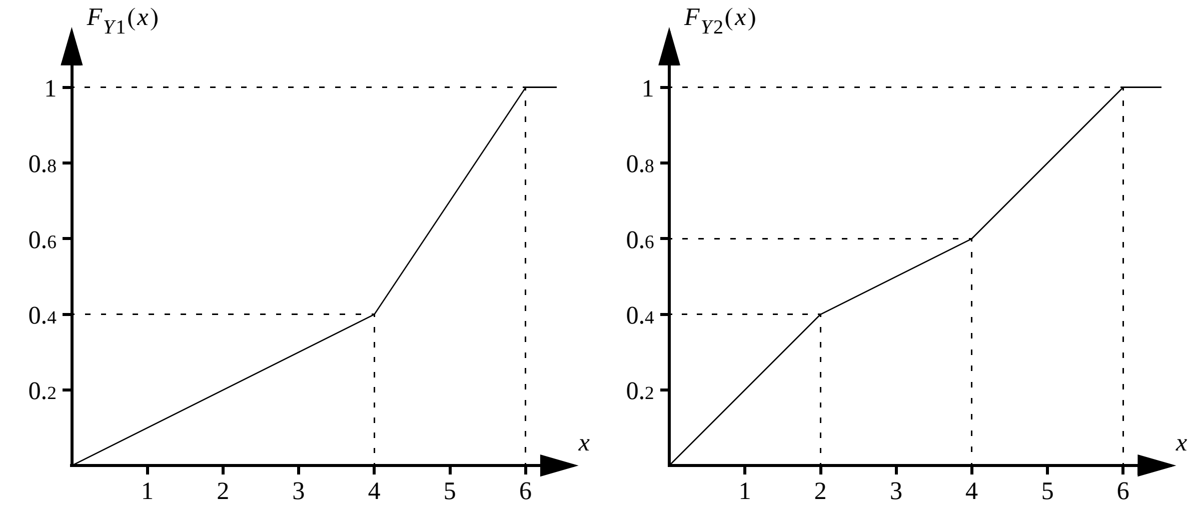
Determine the optimal importance sampling function for this system. How large is the variance of M_X when using the optimal importance sampling function?

Hint: The solution to the optimisation problem

$$\begin{aligned} &\text{minimise} && k\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right) \\ &\text{subject to} && a + b + c + d = 1, \\ &&& a > 0,\ b > 0,\ c > 0,\ d > 0, \end{aligned}$$

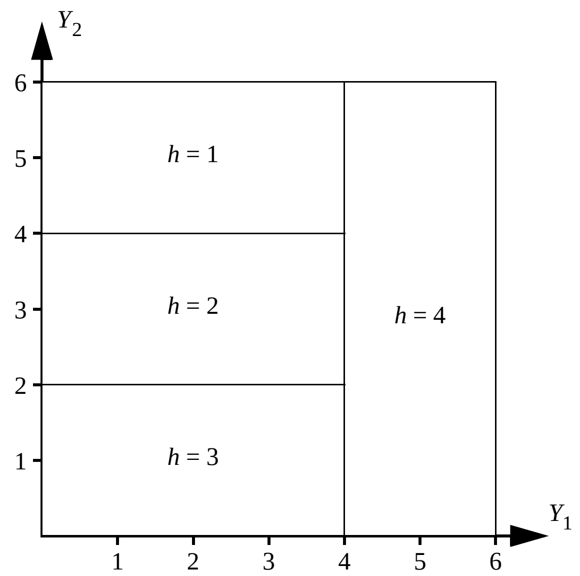
is $a = b = c = d = 0.25$.

Problem 4 (8 p)



Y_1 and Y_2 are two independent random variables with distribution functions $F_{Y1}(x)$ and $F_{Y2}(x)$ according to the figure above.

a) (3 p) The system $X = g(Y_1, Y_2)$ is simulated using stratified sampling. Strata have been defined according to the figure below. Calculate the stratum weights.



b) (5 p) Table 1 shows the results of a pilot study for a Monte Carlo simulation of the model $g(Y_1, Y_2)$. Suggest a sample allocation for the next batch of 100 samples.

Table 1 Results from the Monte Carlo simulation in problem 4 after a pilot study of 100 scenarios.

Stratum, h	Number of samples, n_h	Estimated stratum mean, m_{Xh}	Estimated stratum variance, s^2_{Xh}
1	25	60	1 024
2	25	65	529
3	25	90	1 936
4	25	110	3 600

Problem 5 (8 p)

A system is to be simualtaed using importance sampling and stratified sampling. Introduce the following symbols:

$f_{Zh}(x)$ = importance sampling function in stratum h,
$f_{Yh}(x)$ = density function of the input variables in stratum h,
$y_{h,i}$ = outcome of the input variables in the i:th scenario in stratum h,
w_i = weight factor of scenario y_i in stratum h, i.e., $f_{Yh}(y_{h,i})/f_{Zh}(y_{h,i})$,
$x_{h,i}$ = output for the i:th scenario in stratum h, i.e., $g_1(y_{h,i})$.

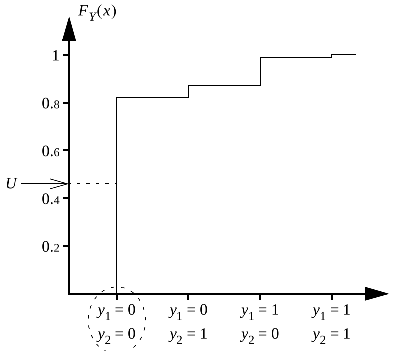
Table 2 shows the results of the simulation. Calculate the estimate of $E[X]$.

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

Stratum, h	Stratum weight, ω_h	Samples per stratum, n_h	$x_{h,1}$	$w_{h,1}$	$x_{h,2}$	$w_{h,2}$	$x_{h,3}$	$w_{h,3}$	$x_{h,4}$	$w_{h,4}$	$x_{h,5}$	$w_{h,5}$	$x_{h,6}$	$w_{h,6}$
1	0.5	5	20	0.50	8	1.60	5	1.20	12	1.50	13	1.40	–	–
2	0.3	3	25	0.30	12	1.70	18	0.95	–	–	–	–	–	–
3	0.2	6	22	1.04	30	0.64	36	0.52	18	1.08	27	0.64	48	0.51

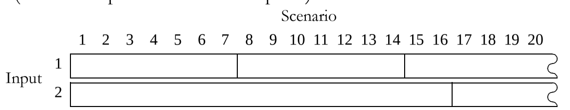
Problem 1

a) $F_Y^{-1}(0.46)$ can be calculated graphically as indicated in the figure below. The result is $y_1 = 0$ and $y_2 = 0$.

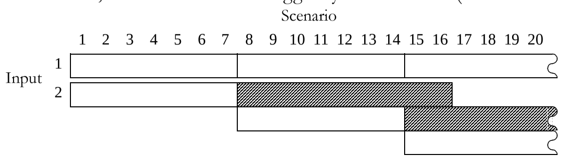


b) Let u_i, $i = 1, \ldots, 10$ be ten values from a pseudorandom number generator, and let $u_i^* = 1 - u_i$, $i = 1, \ldots, 10$. This gives us 20 values of both inputs, i.e., 20 scenarios. Hence, 10 values from the random number generator are needed.

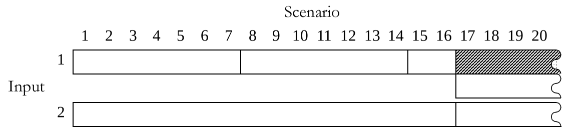
c) Each dagger cycle will require one value from the random number generator. The dagger cycles are represented by rectangles in the figure below. As can be seen, there is a need for five dagger cycles in total (three for input 1 and two for input 2).



d) The shaded areas below are the part of the longer dagger cycle that is ignored when the dagger cycle is reset. In this case, the total number of dagger cycles will be six (three for each input).



e) The shaded areas below are the part of the shorter dagger cycle that is ignored when the dagger cycle is reset. In this case, the total number of dagger cycles will be six (four for input 1 and two for input 2).



Problem 2

a) $m_X = \frac{1}{10} \sum_{i=1}^{10} x_i = -5.2$.

b) $s_X^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - m_X)^2 = 72.16$.

c) If M_X is $N(-5.2, \sqrt{(72.16)/\ })$-distributed then $P(M_X < 0) = \Phi\left(\frac{0-(-5.2)}{\sqrt{7.216}}\right) \approx \Phi(1.94) \approx 97\%$.

Problem 3

The standard formula for the optimal importance sampling function cannot be applied here, because

$$\mu_X = \sum_\psi f_Y(\psi)g(\psi) = 0.04 \cdot 1 + 0 + 0.04 \cdot (-1) + 0 + 0 + 0 + 0.04 \cdot (-1) + 0 + 0.04 \cdot 1 = 0.$$

The variance of M_X when using importance sampling is given by

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

$$= \frac{1}{n}\left(1^2\frac{0.04^2}{f_Z(-1,-1)} + 0 + (-1)^2\frac{0.04^2}{f_Z(-1,1)} + 0 + 0 + 0 + (-1)^2\frac{0.04^2}{f_Z(1,-1)} + 0 + 1^2\frac{0.04^2}{f_Z(1,1)}\right).$$

According to the hint, the optimal importance sampling function must be to use the frequency function described in the table below. The resulting $Var[M_X]$ is then equal to $0.0256/n$.

y_1	y_2	$g(\psi)$	$f_Y(\psi)$	$f_Z(\psi)$
–1	–1	1	0.04	0.25
–1	0	0	0.12	0
–1	1	–1	0.04	0.25
0	–1	0	0.12	0
0	0	0	0.36	0
0	1	0	0.12	0
1	–1	–1	0.04	0.25
1	0	0	0.12	0
1	1	1	0.04	0.25

Problem 4

a) $\omega_1 = P(Y_1 \le 4) \cdot P(Y_2 \ge 4) = F_{Y1}(4) \cdot (1 - F_{Y2}(4)) = 0.16$.
Similarly, $\omega_2 = F_{Y1}(4) \cdot (F_{Y2}(4) - F_{Y2}(2)) = 0.08$, $\omega_3 = F_{Y1}(4) \cdot F_{Y2}(2) = 0.16$ and $\omega_4 = (1 - F_{Y1}(4)) = 0.6$.

b) According to the Neyman allocation, the target distribution of 2 000 samples is

$$n_h = n \frac{\omega_h s_{Xh}}{\sum_{k=1}^{L} \omega_k s_{Xk}} \approx \begin{cases} 20 & h = 1, \\ 7 & h = 2, \\ 28 & h = 3, \\ 144 & h = 4. \end{cases}$$

The first two strata have apparently been allocated too many samples already in the pilot study. Hence, we should use a compromise allocation that allocates the next 100 samples to strata 3 and 4. We would have 144 + 28 = 172 samples in strata 3 and 4 according to the target allocation, but we only have 150 samples available (25 + 25 from the pilot study and 100 from the next batch); thus, the compromise target after 200 scenarios could be 150/172 = 87% of the optimal allocation for these strata, i.e., 25 samples in stratum 3 and 125 samples in stratum 4. The resulting allocation for the 100 scenarios in the next batch is then to take all samples from stratum 4.

Problem 5

First we estimate the expectation value for each stratum:

$$m_{Xh} = \frac{1}{n_h} \sum_{i=1}^{n_h} w_{h,i} x_{h,i} = \begin{array}{ll} 65/5 = 13 & h = 1 \\ 45/3 = 15 & h = 2 \\ 120/6 = 20.3 & h = 3 \end{array}$$

The estimate for the entire population is obtained by

$$m_X = \sum_{h=1}^{3} \omega_h m_{Xh} = 0.5 \cdot 13 + 0.3 \cdot 15 + 0.2 \cdot 20.3 \approx 15.1.$$