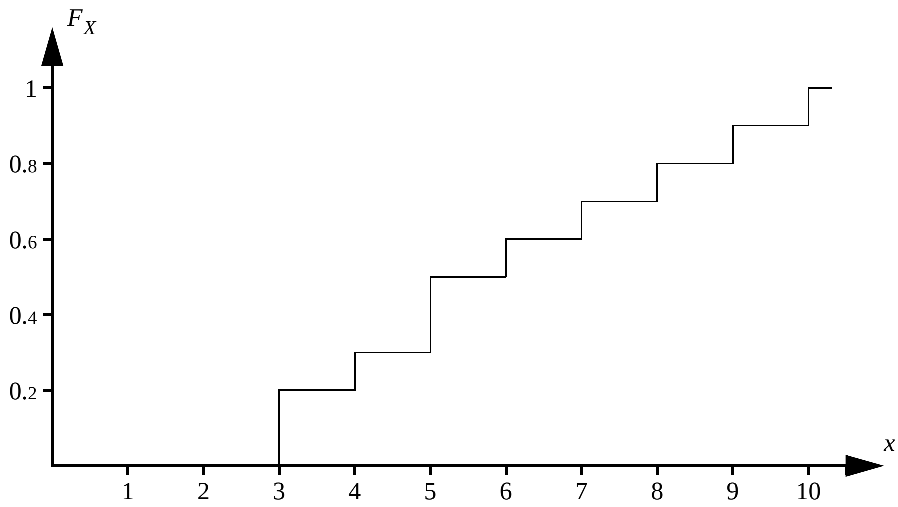
KTH Electrical Engineering

Electric Power Systems Lab
EG2080 MONTE CARLO METHODS IN ENGINEERING
Autumn 2014

Home assignments part I

Simple Discrete Probability Distribution

In these problems you will study a discrete random variable X with the following probability distribution:



Problem 1 — Basic Definitions

a) State the frequency function $f_X(x)$.

b) Enumerate the units x_i when X is considered as a population of N units, where the probability of selecting a specific unit is $1/N$.

c) Calculate $P(X \geq 5)$.

d) Calculate $E[X]$.

e) Calculate $Var[X]$.

Problem 2 — Confidence Intervals

The objective of this problem is to verify the formula for calculating confidence intervals. Write a program that performs the following tasks:

> ***Step 1.*** Generate 100 samples of the random variable X. Calculate the lower and upper bound of a 95% confidence interval for $E[X]$.
>
> ***Step 2.*** If the confidence interval includes the true expectation value, then record the simulation in step 1 as a successful simulation, otherwise record is as unsuccessful.
>
> ***Step 3.*** Repeat step 1 and 2 at least 1 000 times.

How many per cent of the simulations were recorded as successful? Does it seem reasonable to assume that the estimate M_X is normally distributed?

KTH Electrical Engineering

Home assignments part II

The Product Company

The Product Company is delivering its product to two different buyers. The company has three machines capable of manufacturing the product, and the machines have different production costs per unit and different capacities, as shown in table 1. There is also a risk that a machine fails, and cannot be used until it has been repaired (which takes one day).

Table 1 Machine data.

Machine, m	Production cost, c_m [€/unit]	Unavailability, q_m	Capacity, π_m
1	4	0.11	2 000
2	6	0.12	1 000
3	8	0.15	1 000

The company has no stock, i.e., the units must be produced on the same day as they are delivered. The units are delivered using lorries. Each lorry can carry 500 units and the cost of each lorry transport is 1 000 €. Notice that the transport cost is the same regardless of how many units the lorry is carrying.

The contract signed between the Product Company and its buyers states that the Product Company will pay a penalty fee of 25 € for each unit which cannot be delivered the day it has been ordered.

Inputs

The inputs to the simulation of the Product Company are the following:

D_b = demand of buyer b,
$\bar{P}_m$ = available generation capacity of machine m.

The demand is a truncated normal distribution rounded to integer values, i.e., let $\tilde{D}_1$ be N(800, 200)-distributed and let $\tilde{D}_2$ be N(600, 150)-distributed. The demand of buyer b is then 0 if $\tilde{D}_b \leq 0$; otherwise, it is equal to $\tilde{D}_b$ rounded to the closest integer.

The available generation capacities are two-state random variables, i.e.,

$$f_{\bar{P}_m}(x) = \begin{cases} q_m & x = 0, \\ 1 - q_m & x = \pi_m \\ 0 & \text{all other } x, \end{cases} \tag{1}$$

where the values of q_m and π_m are given in table 1.

Model parameters

The following parameters are used in the model of the company:

c_m = production cost per unit for machine m,
d = transport cost per lorry,
f = penalty cost for each unit that is not delivered on time to the buyer,
τ = transport capacity per lorry,

The values of these parameters are given above.

Outputs

The objective of the simulation is to estimate the expected operation cost of the Product Company, i.e.,

C_{tot} = total operation cost of the product company.

Mathematical model

The operation of the Product Company can be described as an optimisation problem, where the objective is to minimise the costs (production, transport and penalty fees for units that are not delivered on time), and the constraints are that the production should equal the demand (unless the production capacity is insufficient), and that the capacity of the lorries must be sufficient to transport the produced units to the buyer. However, this optimisation problem is quite simple, and the solution can be found without using an optimisation algorithm.

Sort the machines according to increasing production cost, i.e., $c_1 \le c_2 \le c_3$. The production is given by

$$P = \begin{cases} D_1 + D_2 & \text{if } D_1 + D_2 \le \bar{P}_1 + \bar{P}_2 + \bar{P}_3, \\ \bar{P}_1 + \bar{P}_2 + \bar{P}_3 & \text{if } D_1 + D_2 > \bar{P}_1 + \bar{P}_2 + \bar{P}_3. \end{cases} \tag{2}$$

Consequently, the production cost is given by

$$C_P = \begin{cases} c_1 P & \text{if } P \le \bar{P}_1, \\ c_1\bar{P}_1 + c_2(P - \bar{P}_1) & \text{if } \bar{P}_1 < P \le \bar{P}_1 + \bar{P}_2, \\ c_1\bar{P}_1 + c_2\bar{P}_2 + c_3(P - \bar{P}_1 - \bar{P}_2) & \text{if } \bar{P}_1 + \bar{P}_2 < P, \end{cases} \tag{3}$$

and the total penalty cost is given by

$$C_U = f(D_1 + D_2 - P). \tag{4}$$

The number of trucks needed when the production capacity is sufficient can be calculated from

$$T_b = \text{ceil}\,(D_b/\tau),\ b = 1, 2, \tag{5}$$

where $x = \text{ceil}\,(y)$ is the smallest integer x such that $x \ge y$.

However, if the production capacity is not sufficient, the transport cost is more complicated to determine, as there are several possibilities to distribute the available units between the two buyers. There are five possible solutions. The first one is

$$T_1 + T_2 = \text{ceil}\,(P/\tau). \tag{6a}$$

This solution is definitely the best if the difference between the total demand and the total production corresponds to at least one lorry, i.e., if $D_1 + D_2 - P > \tau$. If this condition is not fulfilled, we should also investigate the following four solutions:

$$T_1 = \text{floor}\,(D_1/\tau),\ T_2 = \text{floor}\,(D_2/\tau), \tag{6b}$$

$$T_1 = \text{floor}\,(D_1/\tau),\ T_2 = \text{ceil}\,(D_2/\tau), \tag{6c}$$

$$T_1 = \text{ceil}\,(D_1/\tau),\ T_2 = \text{floor}\,(D_2/\tau), \tag{6d}$$

$$T_1 = \text{ceil}\,(D_1/\tau),\ T_2 = \text{ceil}\,(D_2/\tau), \tag{6e}$$

where $x = \text{floor}\,(y)$ is the largest integer x such that $x \le y$. For each of the solutions (6b)–(6e), we need to check that the minimum number of undelivered units does not exceed the difference between the demand and the production. The minimum undelivered units to a buyer is given by

$$U_b = \max\,(0, D_b - \tau T_b). \tag{7}$$

To calculate the transport cost, we need to investigate all four alternatives, check if they are acceptable (i.e., if $U_1 + U_2 \le D_1 + D_2 - P$), and then select the acceptable solution using the least number of trucks. The total transport cost is then given by

$$C_T = d(T_1 + T_2). \tag{8}$$

Example 1. Assume that $D_1 = 210$, $D_2 = 220$, $P = 250$ and $\tau = 100$. Since the condition $D_1 + D_2 - P > \tau$ is fulfilled, we may use (6a) to determine the total number of lorries needed:

$$T_1 + T_2 = \text{ceil}(P/\tau) = 3.$$

Example 2. Assume that $D_1 = 210$, $D_2 = 220$, $P = 400$ and $\tau = 100$. Since (6b) requires the least number of trucks, we start by checking if this solution is acceptable:

$$T_1 = \text{floor}\,(210/100) = 2 \Rightarrow U_1 = \max\,(0, 210 - 100 \cdot 2) = 10,$$
$$T_2 = \text{floor}\,(220/100) = 2 \Rightarrow U_2 = \max\,(0, 220 - 100 \cdot 2) = 20.$$

This solution is acceptable, because $U_1 + U_2 = 30$ is equal to $D_1 + D_2 - P = 210 + 220 - 400 = 30$. Hence, 200 units should be transported to buyer 1 and 200 units should be transported to buyer 2.

Example 3. Assume that $D_1 = 210$, $D_2 = 220$, $P = 420$ and $\tau = 100$. Since (6b) requires the least number of trucks, we start by checking if this solution is acceptable:

$$T_1 = \text{floor}\,(210/100) = 2 \Rightarrow U_1 = \max\,(0, 210 - 100 \cdot 2) = 10,$$
$$T_2 = \text{floor}\,(220/100) = 2 \Rightarrow U_2 = \max\,(0, 220 - 100 \cdot 2) = 20.$$

This solution is however not acceptable, because $U_1 + U_2 = 30$, and that is larger than $D_1 + D_2 - P = 210 + 220 - 420 = 10$. Next, we try the solution according to (6c):

$$T_1 = \text{floor}\,(210/100) = 2 \Rightarrow U_1 = \max\,(0, 210 - 100 \cdot 2) = 10,$$
$$T_2 = \text{ceil}\,(220/100) = 3 \Rightarrow U_2 = \max\,(0, 220 - 100 \cdot 3) = 0.$$

Now we have an acceptable solution, because $U_1 + U_2 = 10$ which is equal to $D_1 + D_2 - P$. The conclusion is that 200 units should be transported to buyer 1 and 220 units should be transported to buyer 2.

Example 4. Assume that $D_1 = 190$, $D_2 = 220$, $P = 400$ and $\tau = 100$. Since (6b) requires the least number of trucks, we start by checking if this solution is acceptable:

$$T_1 = \text{floor}\,(190/100) = 1 \Rightarrow U_1 = \max\,(0, 190 - 100 \cdot 1) = 90,$$
$$T_2 = \text{floor}\,(220/100) = 2 \Rightarrow U_2 = \max\,(0, 220 - 100 \cdot 2) = 20.$$

This solution is however not acceptable, because $U_1 + U_2 = 110$, and that is larger than $D_1 + D_2 - P = 190 + 220 - 400 = 10$. Trying (6c) will not help, because we will still have

U_1 = 90. Therefore, we go on and test (6d):

T_1 = ceil (190/100) = 2 ⇒ U_1 = max (0, 190 – 100 · 2) = 0,

T_2 = floor (220/100) = 2 ⇒ U_1 = max (0, 220 – 100 · 2) = 20.

This solution is not acceptable neither, because $U_1 + U_2 = 20 > 10$; we must use the final alternative according to (6e):

T_1 = ceil (190/100) = 2 ⇒ U_1 = max (0, 190 – 100 · 2) = 0,

T_2 = ceil (220/100) = 3 ⇒ U_1 = max (0, 220 – 100 · 3) = 0.

It will be necessary to use five trucks in this case, regardless of if we choose to send 190 units to buyer 1 and 210 units to buyer 2, or if we choose to send 180 units to buyer 1 and 220 to buyer 2, etc.

Finally, the total operation cost is the sum of the production cost, the penalty cost and the transport cost, i.e.,

$$C_{tot} = C_P + C_U + C_T. \quad (9)$$

Problem 3 — Complementary Random Numbers & Dagger Sampling

a) What are the probability distributions of $\overline{P}_{tot} = \overline{P}_1 + \overline{P}_2 + \overline{P}_3$ and $D_{tot} = D_1 + D_2$ respectively?

b) Generate four scenarios to a Monte Carlo simulation of the Product Company using complementary random numbers for $\overline{P}_{tot}$ and D_{tot}. Present the results in a table showing the pseudorandom numbers as well as the resulting values for the inputs (cf. table 2). Please indicate which pseudo-random numbers you have used to generate which input value.

Table 2 Scenarios generated using complementary random numbers.

Scenario	Pseudo-random numbers			Production capacity [units/day]				Demand [units/day]		
	U_1	U_2	…	$\overline{P}_1$	$\overline{P}_2$	$\overline{P}_3$	$\overline{P}_{tot}$	D_1	D_2	D_{tot}
1										
2										
3										
4										

c) Generate ten scenarios to a Monte Carlo simulation of the Product Company using dagger sampling for $\overline{P}_1$, $\overline{P}_2$ and $\overline{P}_3$. Present the results in a table showing the pseudorandom numbers as well as the resulting values for the inputs (cf. table 2). Please indicate which pseudo-random numbers you have used to generate which input value.

Problem 4 — Control Variate

a) Suggest a simplified model of the Product Company and calculate the expected operation cost according to the simplified model.

Hint: You can try to neglect one or more of the following features of the model: the penalty fee, the transport cost, the differences in production cost in the machines and the capacity limitations of the machines.

b) Generate ten scenarios for a Monte Carlo simulation of the Product Company. Calculate the results of the simplified model as well as the detailed model (described in the introduction to part II). Present the results in a table showing the input and output values for each scenario (cf. table 3). What is the estimated expected operation cost for the Product Company?

Table 3 Scenarios generated using stratified sampling.

Scenario	Production capacity [units/day]				Demand [units/day]			Operation cost [€/day]	
	$\bar{P}_1$	$\bar{P}_2$	$\bar{P}_3$	$\bar{P}_{tot}$	D_1	D_2	D_{tot}	Simplified model	Detailed model
1									
2									
3									
…									
10									

Problem 5 — Correlated Sampling

The Product Company wants to investigate if the transport costs will be reduced if they switch to larger lorries, which have 50% higher transport capacity and 20% higher transport cost. Generate ten scenarios and calculate the operation cost for the system with smaller lorries and the system with larger lorries respectively. Present the results in a table showing the input and output values for each scenario (cf. table 3). What is the estimated expected difference in operation cost between smaller and larger lorries?

Problem 6 — Importance Sampling

a) Use the simplified model from problem 4a to suggest an importance sampling function.

b) Use your importance sampling function from problem 6a to generate five scenarios. Present the results in a table showing the pseudorandom numbers as well as the resulting values for the inputs (cf. table 2). Please indicate which pseudorandom numbers you have used to generate which input value.

c) What is the estimated expected operation cost based on the five scenarios from problem 6b?

Problem 7 — Stratified Sampling

a) Use a strata tree to suggest a stratification for a simulation of the Product Company.

b) Calculate the mean for each stratum either by using two samples or—if possible—by analytical calculations. Present the generated scenarios in a table, showing the input values and the corresponding output value (cf.table 4). What is the estimated mean?

Table 4 Scenarios generated using stratified sampling.

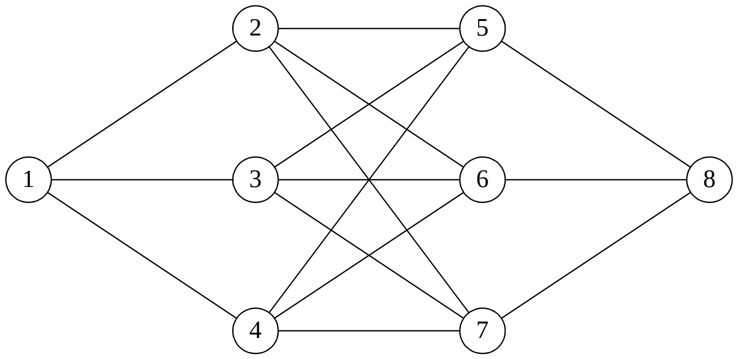
Stratum	Scenario	Production capacity [units/day]				Demand [units/day]			Operation cost [€/day]
		$\bar{P}_1$	$\bar{P}_2$	$\bar{P}_3$	$\bar{P}_{tot}$	D_1	D_2	D_{tot}	
1	1 2								
2	1 2								
…									

KTH Electrical Engineering

Electric Power Systems Lab
EG2080 MONTE CARLO METHODS IN ENGINEERING
Autumn 2014

Home assignments part III

Travelling Path



Consider a traveller who want to go from node 1 to node 8 in the shortest possible time. Let $T_{i,j}$ be the time needed to go from node i to node j. $T_{i,j}$ are two-state random variables, with the probability distribution

$$f_{Ti,j}(x) = \begin{cases} 1 - p_{i,j} & x = a_{i,j}, \\ p_{i,j} & x = b_{i,j}, \\ 0 & \text{all other } x, \end{cases} \tag{10}$$

where the values of $p_{i,j}$, $a_{i,j}$ and $b_{i,j}$ are given in table 5.

The best path can be found using the following algorithm:

$$T_{i,8} = \min_{j \in \{5, 6, 7\}} (T_{i,j} + T_{j,8}), \; i = 2, \dots 4, \tag{11a}$$

$$T_{1,8} = \min_{j \in \{2, 3, 4\}} (T_{1,j} + T_{j,8}). \tag{11b}$$

Problem 8 — Efficiency of Dagger Sampling

a) The inputs to your system have different dagger cycle lengths. Describe how this can be managed when the system is simulated using dagger sampling.

Table 5 Path data.

(i, j)	$p_{i,j}$	$a_{i,j}$	$b_{i,j}$	(i, j)	$p_{i,j}$	$a_{i,j}$	$b_{i,j}$	(i, j)	$p_{i,j}$	$a_{i,j}$	$b_{i,j}$
(1, 2)	0.06	10	40	(2, 7)	0.03	25	55	(4, 6)	0.03	15	45
(1, 3)	0.03	15	45	(3, 5)	0.07	10	40	(4, 7)	0.03	20	50
(1, 4)	0.03	20	50	(3, 6)	0.03	20	50	(5, 8)	0.06	10	40
(2, 5)	0.06	10	40	(3, 7)	0.03	25	55	(6, 8)	0.03	15	45
(2, 6)	0.03	15	45	(4, 5)	0.07	10	40	(7, 8)	0.03	20	50

b) Investigate the efficiency of dagger sampling on this system by running 20 simulations with 500 samples per simulation. Let m_{Ts} be the estimate from simulation s. What is the lowest, average and highest values of m_{Ts}? What is the estimated variance of M_T, i.e.,

$$\frac{1}{20}\sum_{s=1}^{20} m_{Ts}^2 - \left(\frac{1}{20}\sum_{s=1}^{20} m_{Ts}\right)^2? \tag{12}$$

Problem 9 — Efficiency of Importance Sampling

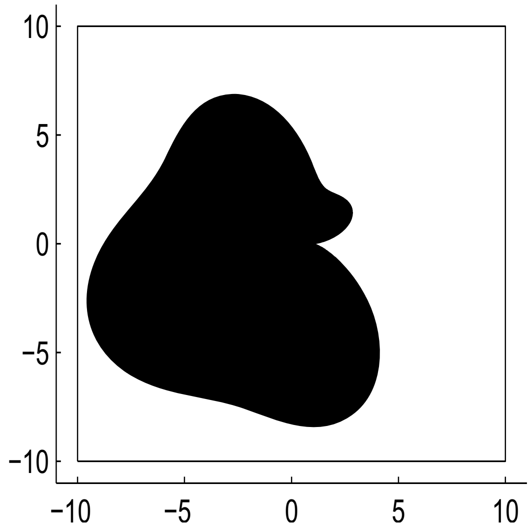
Suggest an importance sampling function for this system, and test the efficiency in the same way as for dagger sampling (i.e., 20 simulations with 500 samples each).

KTH Electrical Engineering

Electric Power Systems Lab
EG2080 MONTE CARLO METHODS IN ENGINEERING
Autumn 2014

Home assignments part IV

Area Calculation



Consider the black shape within the square above. If a point within the square is selected randomly, i.e., let Y_1 and Y_2 be two independent $U(-10, 10)$-distributed random numbers, then the probability of selecting a point within the black shape must be equal to the area of the shape divided by the area of the square, i.e., $A/400$. Now, consider the system

$$X = g(Y) = \begin{cases} 0 & \text{if } (Y_1, Y_2) \text{ is outside the shape,} \\ 400 & \text{if } (Y_1, Y_2) \text{ is inside the shape.} \end{cases} \tag{13}$$

The expectation value $E[X]$ must then be

$$E[X] = f_X(0) \cdot 0 + f_X(400) \cdot 400 = \frac{A}{400} \cdot 400 = A; \tag{14}$$

hence, the area of the black shape can be determined by estimating $E[X]$ using Monte Carlo methods.

To test if the selected point is within the shape, calculate the distance between the point (y_1, y_2) and the origin, i.e.,

$$r_Y = \sqrt{y_1^2 + y_2^2}, \tag{15}$$

where y_1 is the horizontal position and y_2 is the vertical position of the selected point. Then check if r_Y is less than or equal to the outside border of the shape, which is given by the following function:

$$r(\vartheta) = \sin(4\vartheta) - \frac{1}{8}\vartheta^3 + \frac{\pi}{8}\vartheta^2 + \frac{\pi^2}{4}\vartheta + 1, 0 \le \vartheta \le 2\pi, \qquad 0 \le \vartheta \le 2\pi. \tag{16}$$

In (16), the angle ϑ must be in the interval [0, 2π] and can be calculated according to

$$\vartheta = \begin{cases} \arctan\left(\frac{y_2}{y_1}\right) & \text{if } y_1 > 0 \text{ and } y_2 \ge 0, \\ \arctan\left(\frac{y_2}{y_1}\right) + 2\pi & \text{if } y_1 > 0 \text{ and } y_2 < 0, \\ \arctan\left(\frac{y_2}{y_1}\right) + \pi & \text{if } y_1 < 0, \\ \frac{\pi}{2} & \text{if } y_1 = 0 \text{ and } y_2 \ge 0, \\ \frac{3\pi}{2} & \text{if } y_1 = 0 \text{ and } y_2 < 0, \end{cases} \tag{17}$$

Problem 10 — Efficiency of Complementary Random Numbers

a) Since the system has two inputs, there are several possible combinations of original and complementary random numbers. Describe how this can be managed when the system is simulated using complementary random numbers.

b) Investigate the efficiency of complementary random numbers on this system by running 20 simulations with 500 samples per simulation. Let m_{Xs} be the estimate from simulation s. What is the lowest, average and highest values of m_{Xs}? What is the estimated variance of M_X, i.e.,

$$\frac{1}{20}\sum_{s=1}^{20} m_{Xs}^2 - \left(\frac{1}{20}\sum_{s=1}^{20} m_{Xs}\right)^2? \tag{18}$$

Problem 11 — Efficiency of Control Variates

Choose a shape with a known area (for example a circle, triangle or square), which can be used as a control variate. Then test the efficiency in the same way as for complementary random numbers (i.e., 20 simulations with 500 samples each).

Problem 12 — Efficiency of Stratified Sampling

a) Define strata for this system, by dividing the square in smaller rectangles (the strata must be rectangular, because otherwise Y_1 and Y_2 cannot be randomised independently within the stratum). Try to create strata which are either homogeneous or duogeneous with approximately 50% conformist units and 50% diverging units—avoid strata with a small share of diverging units. Assume that 100 samples are to be generated in the pilot study. Which sample allocation do you think is appropriate for the pilot study?

b) Test the efficiency in the same way as for complementary random numbers (i.e., 20 simulations with 500 samples each). Apply the Neyman allocation to calculate the sample allocation for the remaining samples after the pilot study.