KTH Electrical Engineering

SIMULATION OF ELECTRICITY MARKETS

MONTE CARLO METHODS

Lectures 15-18
in
EG2050 System Planning

Mikael Amelin

COURSE OBJECTIVES

To pass the course, the students should show that they are able to

- apply both probabilistic production cost simulation and Monte Carlo simulation to calculate expected operation cost and risk of power deficit in an electricity market.

To receive a higher grade (A, B, C, D) the students should also show that they are able to

- create specialised models both for probabilistic production cost simulation and Monte Carlo simulation, and to use the results of an electricity market simulation to judge the consequences of various actions in the electricity market.

KTH Electrical Engineering

PROBABILITY DISTRIBUTIONS AND EXPECTATION VALUES

KTH Electrical Engineering

The probability distribution of a random variable can be described using the density function, $f_X(x)$.

The expectation value of a discrete random variable is then

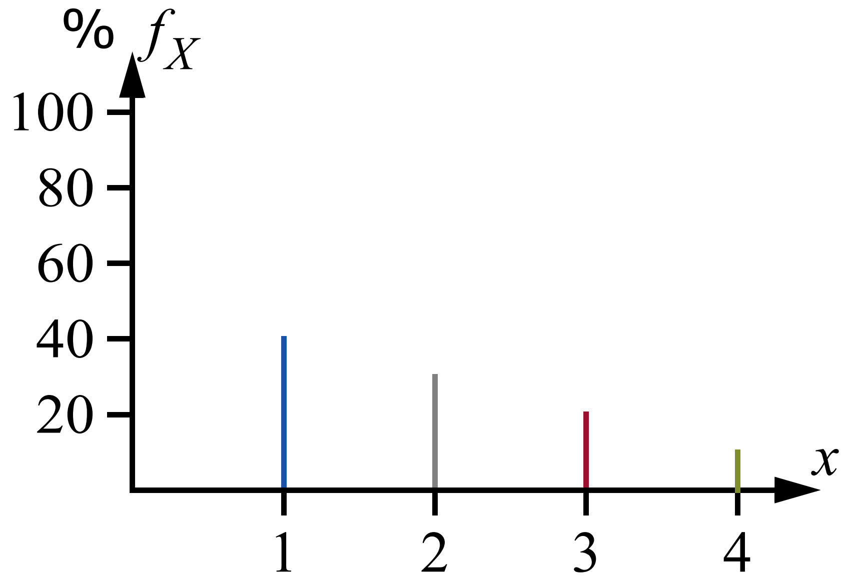
$$E[X] = \sum_x x f_X(x) .$$

PROBABILITY DISTRIBUTIONS AND EXPECTATION VALUES

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Example



$$E[X] = 0.4\cdot 1 + 0.3\cdot 2 + 0.2\cdot 3 + 0.1\cdot 4 = 2.$$

PROBABILITY DISTRIBUTIONS AND EXPECTATION VALUES

KTH Electrical Engineering

An alternative approach to describe probability distribution is to consider a random variable, X, as a population of individual units:

$$x_1, \ldots, x_N,$$

where

$$x_i = \text{outcome of } X \text{ for unit } i,$$
$$N = \text{number of units in the population.}$$

PROBABILITY DISTRIBUTIONS AND EXPECTATION VALUES

Using this alternative approach, the expectation value of a discrete random variable can be written as

$$E[X] = \frac{1}{N} \sum_{i=1}^{N} x_i .$$

KTH Electrical Engineering

PROBABILITY DISTRIBUTIONS AND EXPECTATION VALUES

KTH Electrical Engineering

Example

$$E[X] = \frac{1}{10}(1 + 1 + 1 + 1 + 2 + 2 + 2 + 3 + 3 + 4) = 2.$$

SIMPLE SAMPLING

Theorem 6.21. If there are n independent observations, $x_1, \ldots, x_n$, of the random variable X then the mean of these observations, i.e.,

$$m_X = \frac{1}{n} \sum_{i=1}^{n} x_i$$

is an estimate of $E[X]$.

KTH Electrical Engineering

SIMPLE SAMPLING

Compare $E[X] = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $m_X = \frac{1}{n}\sum_{i=1}^{n} x_i$.

- Simple sampling means that a limited number of random observations are evaluated instead of the whole population!

KTH Electrical Engineering

SIMPLE SAMPLING

- Notice that the estimate m_X is also a random variable!
- $E[m_X] = E[X]$
  (If not, the estimate would be biased.)
- $Var[m_X]$ is given by the following theorem:

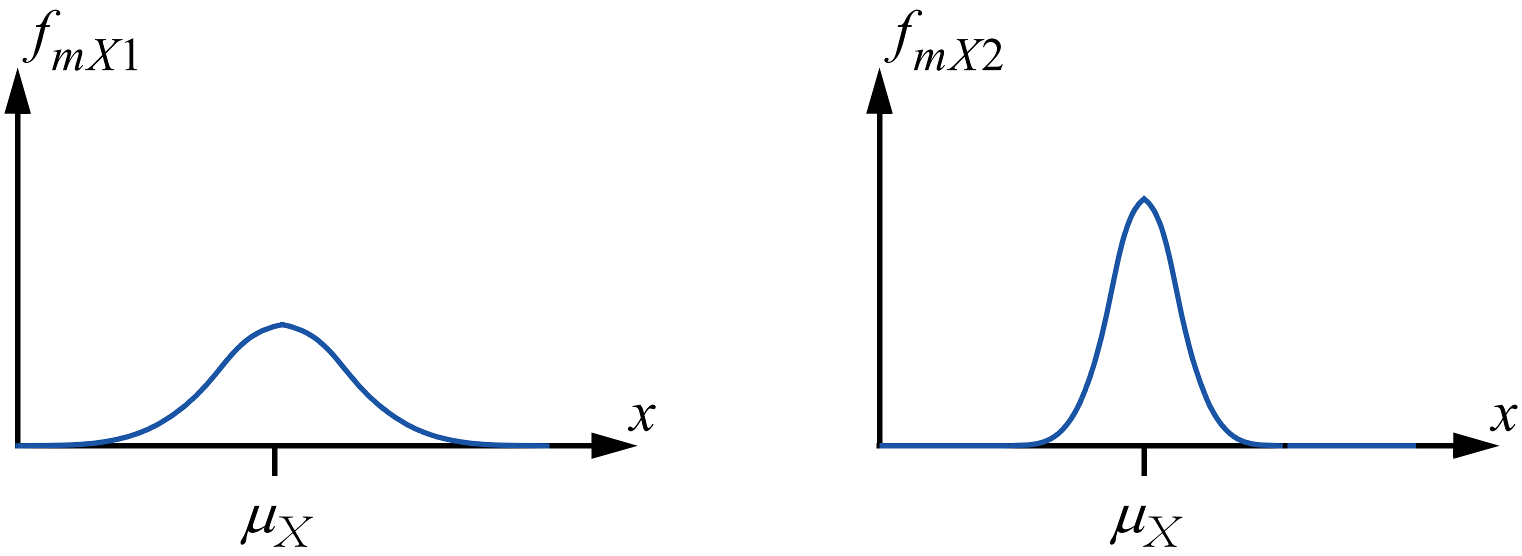
  Theorem 6.22. The variance of the estimate from simple sampling is

$$Var[m_X] = \frac{Var[X]}{n}.$$

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

The variance of the estimate, $Var[m_X]$, is interesting because it states how much an estimate might deviate from the true value.



Here m_{X1} is *likely* to be less accurate than m_{X2}.

KTH Electrical Engineering

SIMPLE SAMPLING - Precision

The practical conclusion of theorem 6.22 is that if the number of samples is increased, it is *likely* that we get a result close to the real value.

Example 6.20—Problem

Let C_i be the result of tossing a coin:

Heads $\Rightarrow C_i = 1$

Tails $\Rightarrow C_i = 0$

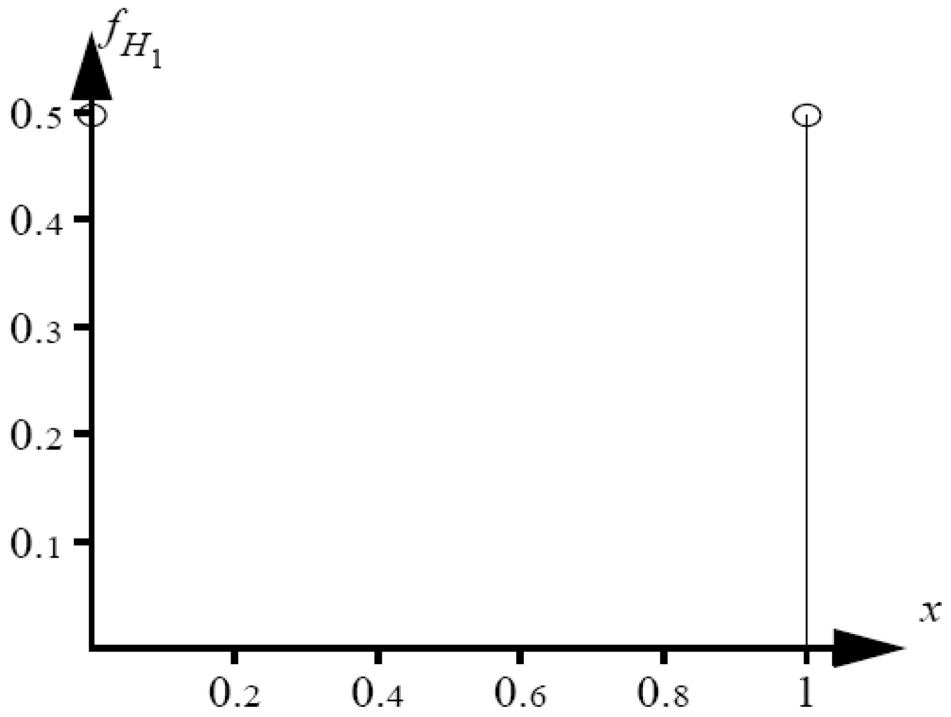
What is the probability distribution of

$$H_n = m_C = \frac{1}{n}\sum_{i=1}^{n} c_i?$$

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution

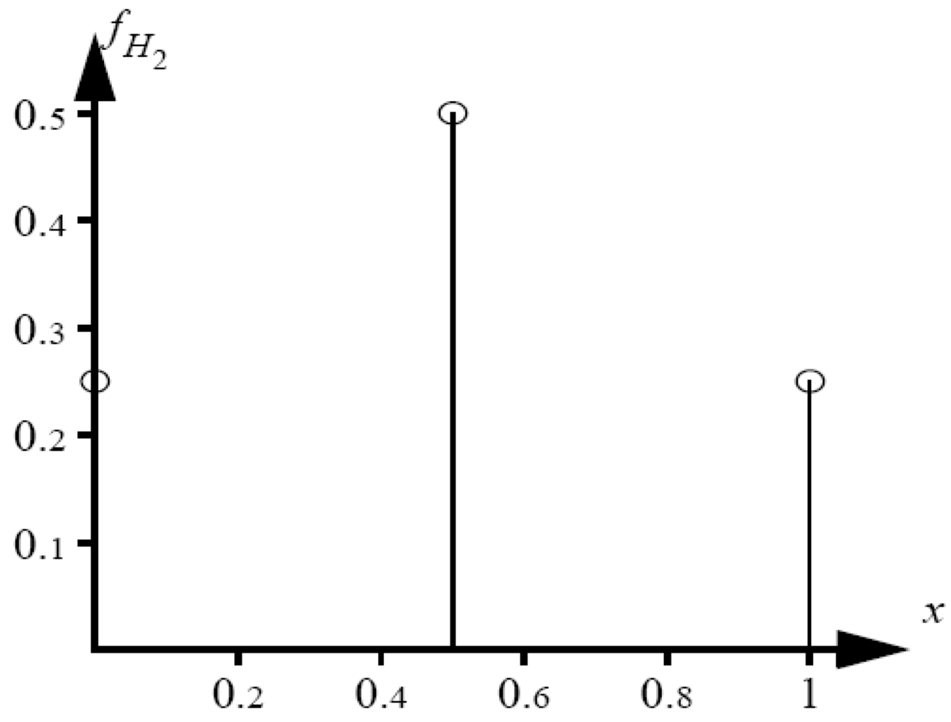


a) Density function after one trial.

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution



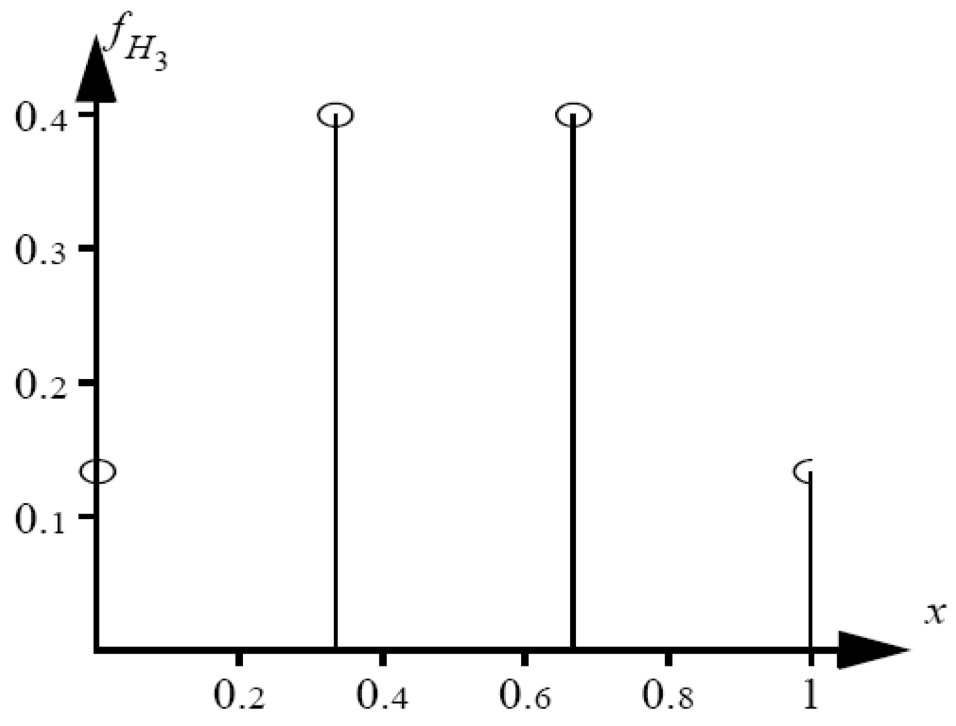
b) Density function after two trials.

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution



c) Density function after three trials.

KTH VETENSKAP OCH KONST

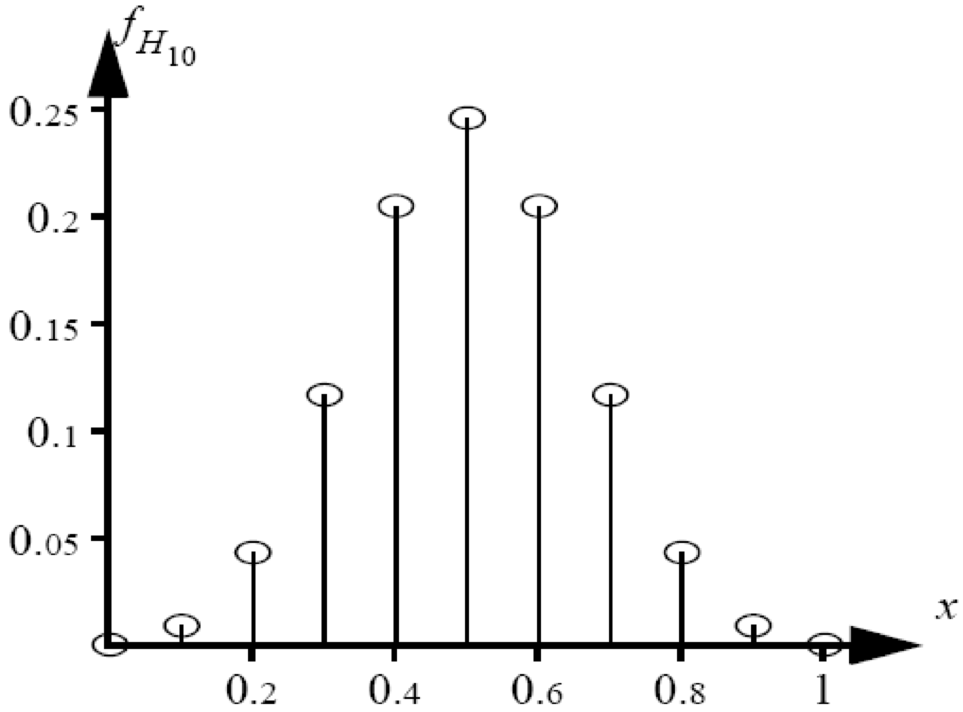
KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution

KTH VETENSKAP OCH KONST

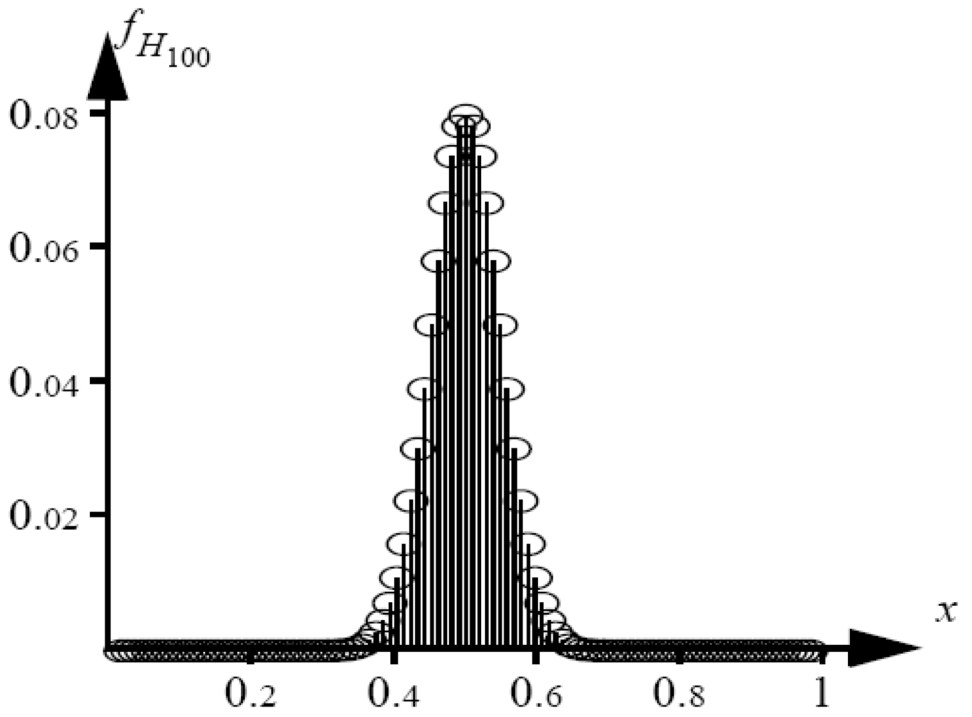
KTH Electrical Engineering



d) Density function after ten trials.

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution



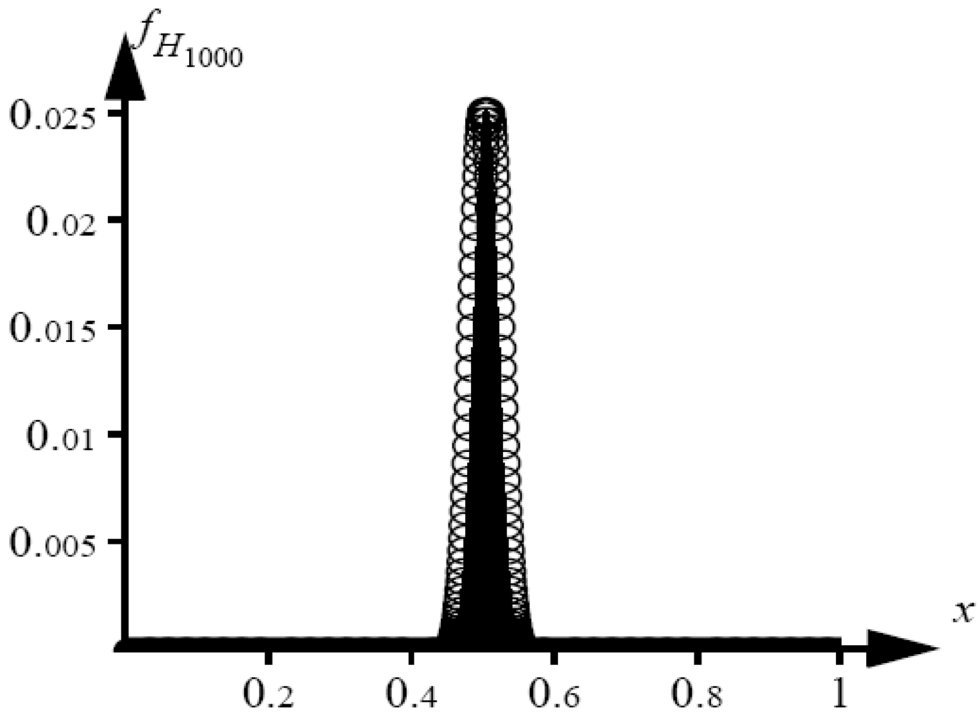
e) Density function after 100 trials.

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Solution



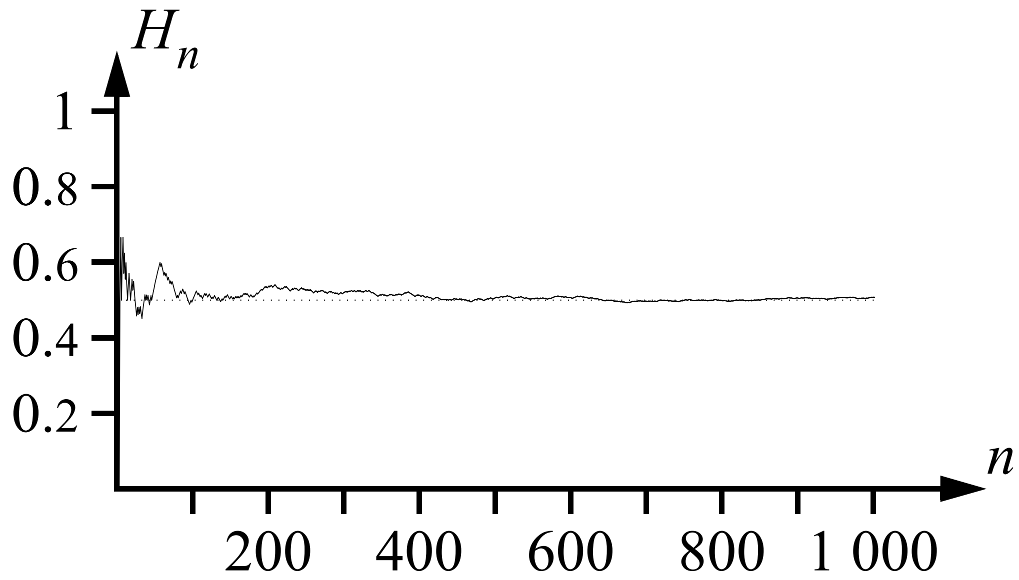
f) Density function after 1 000 trials.

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

SIMPLE SAMPLING - Accuracy

Example 6.20—Practical test



KTH Electrical Engineering

SIMPLE SAMPLING
- Convergence criteria

How do we know when to stop the sampling?

- Number of samples fixed in advance
- Estimate the precision, for example using the coefficient of variation

KTH Electrical Engineering

SIMPLE SAMPLING
- Convergence criteria

Example 6.21—Problem

- Probabilistic production simulation $\Rightarrow$ $LOLP_{PPC} = 1.0$%.
- Desired precision: 95% probability that the estimate is within ± 0.05% of the true value.

  This means that if the true $LOLP$ is 1.08% then we want the estimate to be in the interval 1.03% to 1.13%.
- The estimate m_{LOLO} is assumed to be normally distributed.

KTH Electrical Engineering

SIMPLE SAMPLING - Convergence criteria

KTH Electrical Engineering

Example 6.21—Solution

- The probability is 95% that an $N(\mu, \sigma)$-distributed random variable belongs to the interval $\mu \pm 1.96\sigma$.
- Here, we want the interval to be $\mu \pm 0.0005$
  $\Rightarrow$ The standard deviation of m_{LOLO} must be less than $0.0005/1.96 \approx 0.000255$
  $\Rightarrow$ The variance of m_{LOLO} must be less than $0.000255^2 \approx 6.5 \cdot 10^{-8}$.

SIMPLE SAMPLING - Convergence criteria

KTH Electrical Engineering

Example 6.21—Solution (cont.)

- The variance of m_{LOLO} depends on $Var[LOLO]$, which is unknown but can be estimated using the results from the PPC simulation:

$$Var[LOLO] \approx LOLP_{PPC}(1 - LOLP_{PPC}) =$$
$$= 0.01 \cdot 0.99 = 0.0099.$$

SIMPLE SAMPLING
- Convergence criteria

KTH Electrical Engineering

Example 6.21—Solution (cont.)

- From theorem 6.22 we now have

$$Var[m_{LOLO}] = \frac{Var[LOLO]}{n}.$$

- $Var[m_{LOLO}] < 6.5 \cdot 10^{-8} \Rightarrow n > 152\ 127.$

SIMPLE SAMPLING
- Convergence criteria

KTH Electrical Engineering

Coefficient of variation

Definition: The coefficient of variation is defined as

$$a_X = \frac{\sqrt{Var[m_X]}}{m_X}.$$

SIMPLE SAMPLING
- Convergence criteria

Estimation of accuracy

- Select a few samples.
- Estimate $Var[X]$ by

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

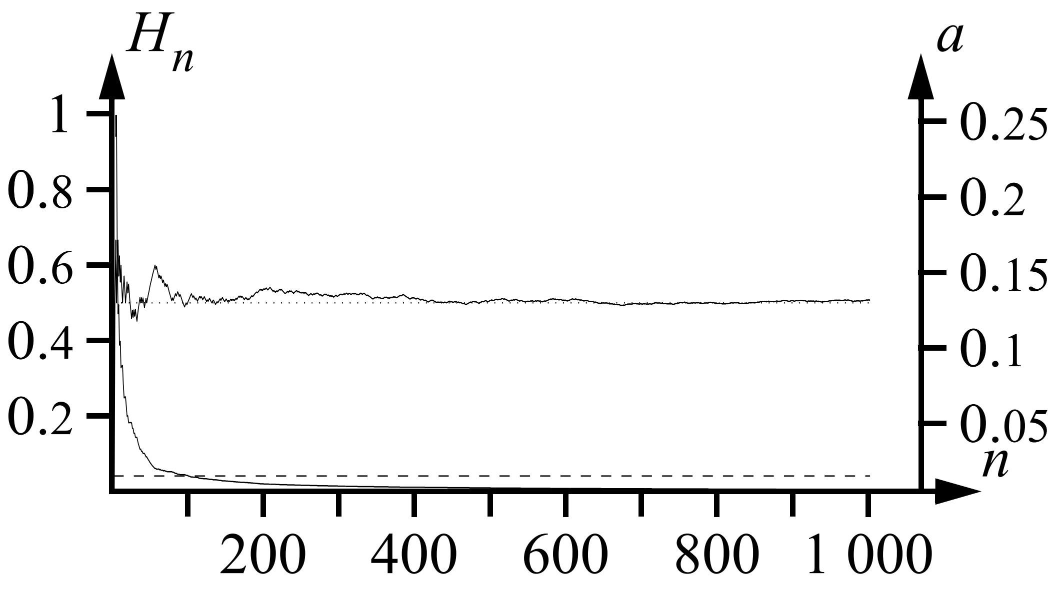
- Test if a_X is less than some tolerance limit, ρ. If yes, stop sampling, otherwise generate a few more samples, etc.

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

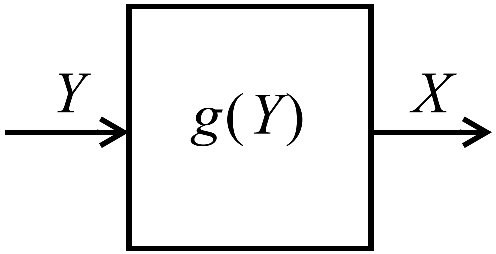
SIMPLE SAMPLING
- Convergence criteria

Example of using the coefficient of variation



KTH Electrical Engineering

SIMULATION OF ELECTRICITY MARKETS



KTH VETENSKAP OCH KONST

KTH Electrical Engineering

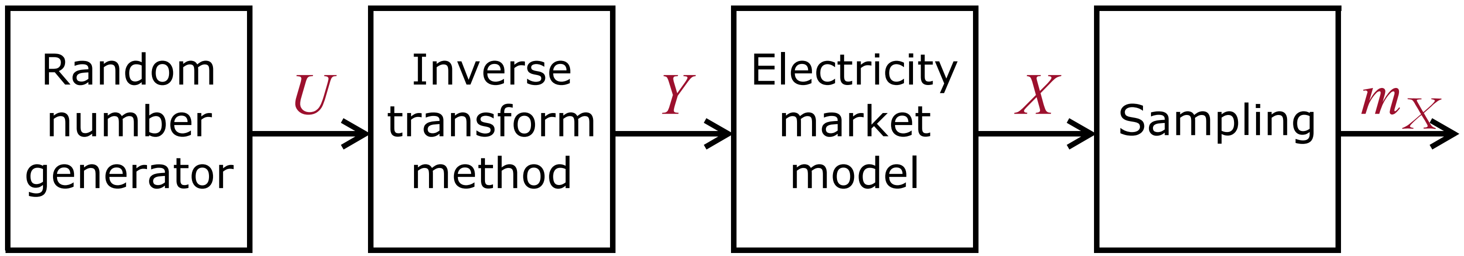
- The scenario parameters, Y, have known probability distributions.
- The result variables, X, have unknown probability distributions.
- We are primarily interested in system indices, which are defined as expectation values of some result variables.

SIMULATION OF ELECTRICITY MARKETS

Result variable	System index
TOC	$ETOC$
$LOLO$	$LOLP$
ENS	$EENS$

KTH Electrical Engineering

SIMPLE SAMPLING OF ELECTRICITY MARKETS



- Generate random numbers from uniform distribution
- Transform random numbers into appropriate probability distributions of the scenario parameters
- Determine how electricity market behaves in the scenario
- Sample the result variables

KTH Electrical Engineering

RANDOM NUMBER GENERATION

KTH Electrical Engineering

- Pseudorandom number generators are mathematical function which given a seed generates a sequence of numbers.
- A good pseudorandom number generator produces a sequence which closely mimics the properties of a $U(0, 1)$-distribution.

  Without knowledge of the pseudorandom number generator and the seed it is hardly possible to predict the next number in the sequence.
- Pseudorandom number generators are available in most programming languages.

TRANSFORMATION OF RANDOM NUMBER

KTH Electrical Engineering

It is not likely that the scenario parameters are $U(0, 1)$-distributed; hence, the output of the random number generator must be transformed to the appropriate probability distribution.

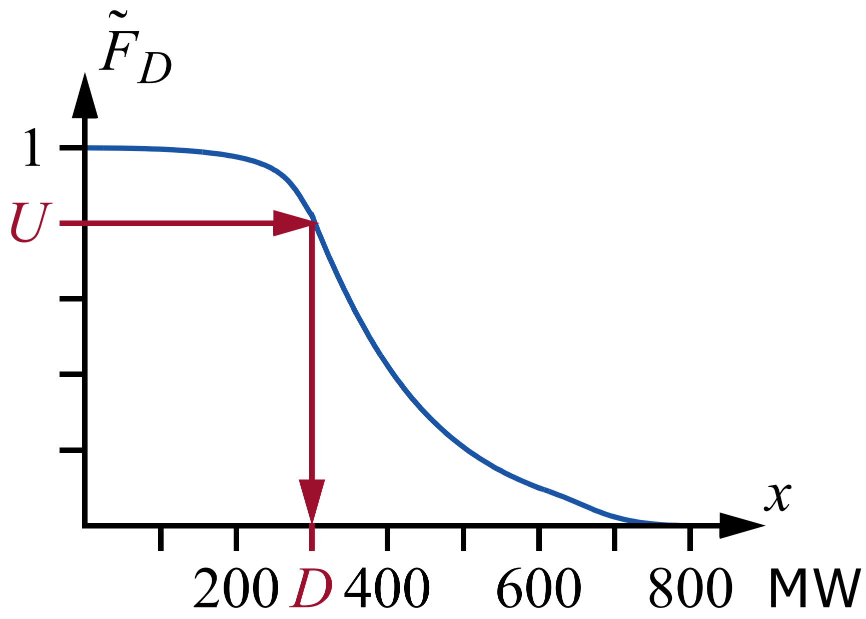
This can be done using the inverse transform method:

> *Theorem E.1.* If a random variable U is $U(0, 1)$-distributed then the random variable $Y = F_Y^{-1}(U)$ has the distribution function $F_Y(x)$.

TRANSFORMATION OF RANDOM NUMBER

Practical notice: We can use $\tilde{F}_Y$ instead of F_Y.

Example



KTH Electrical Engineering

TRANSFORMATION OF RANDOM NUMBER

The inverse of the distribution function of the normal distribution, $\Phi(x)$, does not exist!

$\Rightarrow$ Use an approximation of $\Phi^{-1}(x)$ instead.

This method is called the approximate inverse transform method and is described in theorem E.2 in the compendium.

KTH Electrical Engineering

ELECTRICITY MARKET MODEL

- A Monte Carlo simulation is not restricted to a specific electricity market model.
- The complexity of the electricity market model is only limited by the computation time.

KTH Electrical Engineering

ELECTRICITY MARKET MODEL

- In mathematical terms, the electricity market model is a function

$$x_i = g(y_i),$$

where

$$x_i = \text{result variables for scenario } i,$$

$$y_i = \text{scenario parameters of scenario } i,$$

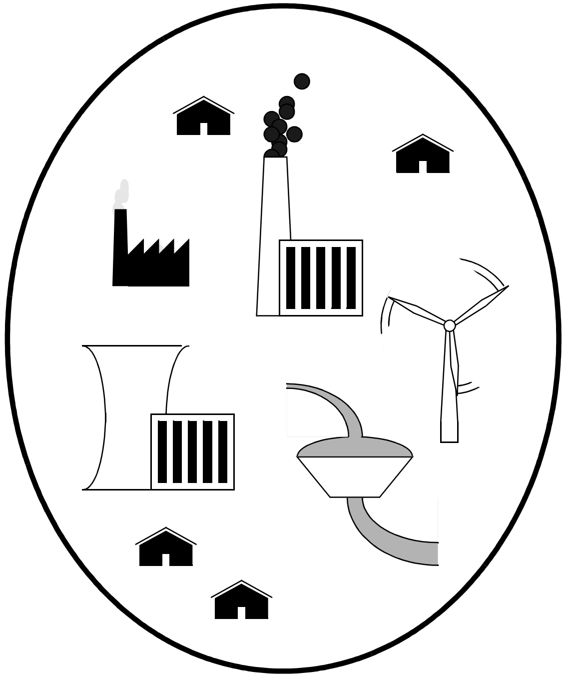
- In most cases the function g cannot be formulated explicitly, but must be indirectly defined from the solution to an optimisation problem.

KTH Electrical Engineering

ELECTRICITY MARKET MODEL - PPC model

Assume

- Perfect competition
- Perfect information
- Load is not price sensitive
- Neglect grid losses and limitations
- All scenario parameters can be treated as independent



KTH Electrical Engineering

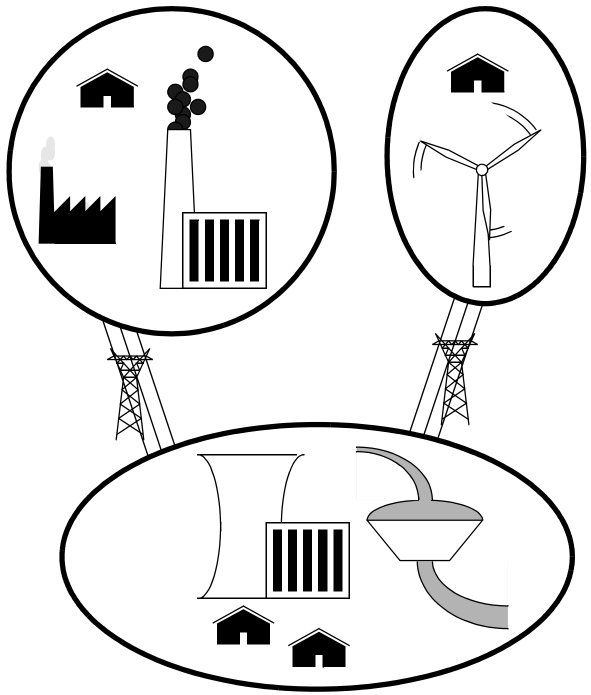
ELECTRICITY MARKET MODEL - Multi-area model

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Assume

- Perfect competition
- Perfect information
- Load is not price sensitive
- Transmission grid losses and limitations included
- Distribution grid losses and limitations neglected



MULTI-AREA MODEL - Example

System data

- Thermal units:
  - Oil condensing, 300 MW, 280 ¤/MWh, 95% availability, located in *North*
  - Nuclear, 1 000 MW, 100 ¤/MWh, 90% availability, located in *South*
  - Bio mass condensing, 400 MW, 180 ¤/MWh, 95% availability, located in *South*

KTH Electrical Engineering

MULTI-AREA MODEL - Example

System data (cont.)

- Non-dispatchable units:
  - Run-of-the-river hydro, 2 000 MW (80%), 1 900 MW (10%), 1 800 MW (10%), negligible operation cost, located in *North*
  - Wind farm, 100 MW (10%), 80 MW (5%), 60 MW (10%), 40 MW (15%), 20 MW (25%), 0 MW (35%), negligible operation cost, available capacity 0 MW (35%), located in *Isle*

KTH Electrical Engineering

MULTI-AREA MODEL - Example

System data (cont.)

- Transmission lines:
  - AC line between *North* and *South*, 1 200 MW, 4% losses, 100% availability
  - HVDC link from *South* to *Isle* (one direction only!), 200 MW, 2% losses, 100% availability
- Load:
  - *North*: $N(600,100)$
  - *South*: $N(2000,300)$
  - *Isle*: $N(100,20)$

(No correlations, no price sensitivity, no compensation paid for disconnected load.)

KTH Electrical Engineering

MULTI-AREA MODEL - Example

KTH Electrical Engineering

Problem

Formulate a multi-area model for the system and show how the result variables TOC, $LOLO$ and ENS are calculated.

Solution

Parameters

D_n = load in area n (scenario parameter—will be randomised during simulation),

MULTI-AREA MODEL - Example

Solution (cont.)

$\overline{G}_g$ = available generation capacity in thermal unit g (scenario parameter—will be randomised during simulation),

$\overline{P}_{n,m}$ = maximal transmission from area n to area m =

$$= \begin{cases} 1\,200 & n = 1, m = 2, \\ 1\,200 & n = 2, m = 1, \\ 200 & n = 2, m = 3, \end{cases}$$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

KTH Electrical Engineering

Solution (cont.)

$\overline{W}_n =$ available non-dispatchable generation capacity in area n (scenario parameter—will be randomised during simulation),

$\beta_{Gg} =$ operation cost in thermal unit $g =$

$$= \begin{cases} 280 & g = 1, \\ 100 & g = 2, \\ 180 & g = 3, \end{cases}$$

MULTI-AREA MODEL - Example

Solution (cont.)

$\beta_{Ln,\,m}$ = loss coefficient for transmission from area n to area m =

$$= \begin{cases} 0.04 & n = 1, m = 2, \\ 0.04 & n = 2, m = 1, \\ 0.02 & n = 2, m = 3, \end{cases}$$

β_{Un} = penalty for unserved load in area n = $= 500,\ n = 1, 2, 3.$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

Optimisation variables

G_g = generation in thermal unit g, $g = 1, 2, 3$,

$P_{n, m}$ = transmission from area n to area m, $(n, m) = (1, 2), (2, 1), (2, 3)$,

U_n = unserved load in area n, $n = 1, 2, 3$,

W_n = generation in non-dispatchable unit n, $n = 1, 3$.

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

Objective function

$$\text{minimise} \sum_{g=1}^{3} \beta_{Gg} G_g + \sum_{n=1}^{3} \beta_{Un} U_n .$$

Constraints

Load balance in *North*:

$$G_1 + W_1 + 0.96 P_{2,1} = D_1 - U_1 + P_{1,2} .$$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

Load balance in *South*:

$$G_2 + G_3 + 0.96P_{1,2} = D_2 - U_2 + P_{2,1} + P_{2,3}.$$

Load balance in *Isle*:

$$W_3 + 0.98P_{2,3} = D_3 - U_3.$$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

Variable limits

$$0 \leq G_g \leq \overline{G}_g, \qquad g = 1, 2, 3,$$

$$0 \leq P_{n,m} \leq \overline{P}_{n,m}, (n, m) = (1, 2), (2, 1), (2, 3),$$

$$0 \leq U_n \leq D_n, \qquad n = 1, 2, 3,$$

$$0 \leq W_n \leq \overline{W}_n, \qquad n = 1, 3.$$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

The result variables are calculated by solving the optimisation problem for the specific values of the scenario parameters and then calculate

$$TOC = \sum_{g=1}^{3} \beta_{Gg} G_g,$$

$$ENS = \sum_{n=1}^{3} U_n,$$

KTH Electrical Engineering

MULTI-AREA MODEL - Example

Solution (cont.)

$$LOLO = \begin{cases} 0 & \text{if } ENS = 0, \\ 1 & \text{if } ENS > 0. \end{cases}$$

KTH Electrical Engineering

HOME ASSIGNMENTS PART IV - Hints

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Problem 22

Define multi-area model.

- State probability distribution of scenario parameters.
- State value of model constants.
- Formulate optimisation problem.
- Show how the result variables TOC and $LOLO$ are calculated from the solution to the optimisation problem.

MONTE CARLO SIMULATION - Example system

KTH Electrical Engineering

- Run-of-the-river hydro, 150 kW, 100% availability, negligible operation cost
- Diesel generator set, 40–100 kW, 100% availability, 1 ¤/kWh
- Diesel generator set, 0–50 kW, 100% availability, 2 ¤/kWh
- Load $N(180, 40)$-distributed [kW]
- Dummy load (water heater) can absorb surplus generation

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

SIMPLE SAMPLING - Example

Example 6.22 (simple sampling)

In our example system, we need to consider

- One scenario parameter, D (the load), i.e., $Y = [D]$.
- One result variable, TOC (operation cost), i.e., $X = [TOC]$.

SIMPLE SAMPLING - Example

KTH Electrical Engineering

Example 6.22 (cont.)

- The electricity market model is the explicit function $X = g(Y)$, where

$$g(Y) = \begin{cases} 0 & Y \leq 150, \\ 2(Y - 150) & 150 < Y \leq 170, \\ 40 & 170 < Y \leq 190, \\ Y - 150 & 190 < Y \leq 250, \\ 100 + 2(Y - 250) & 250 < Y \leq 300, \\ 200 & 300 < Y. \end{cases}$$

SIMPLE SAMPLING - Example

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Example 6.22 (cont.)

To randomise a scenario, we generate a $U(0, 1)$-distributed random number and transform it to an $N(180, 40)$-distribution.

Scenario, i	1	2	3	4	5	6	7	8	9	10
D [kWh/h]	165	273	144	147	185	147	225	120	147	152
TOC [¤/h]	30	146	0	0	40	0	75	0	0	4

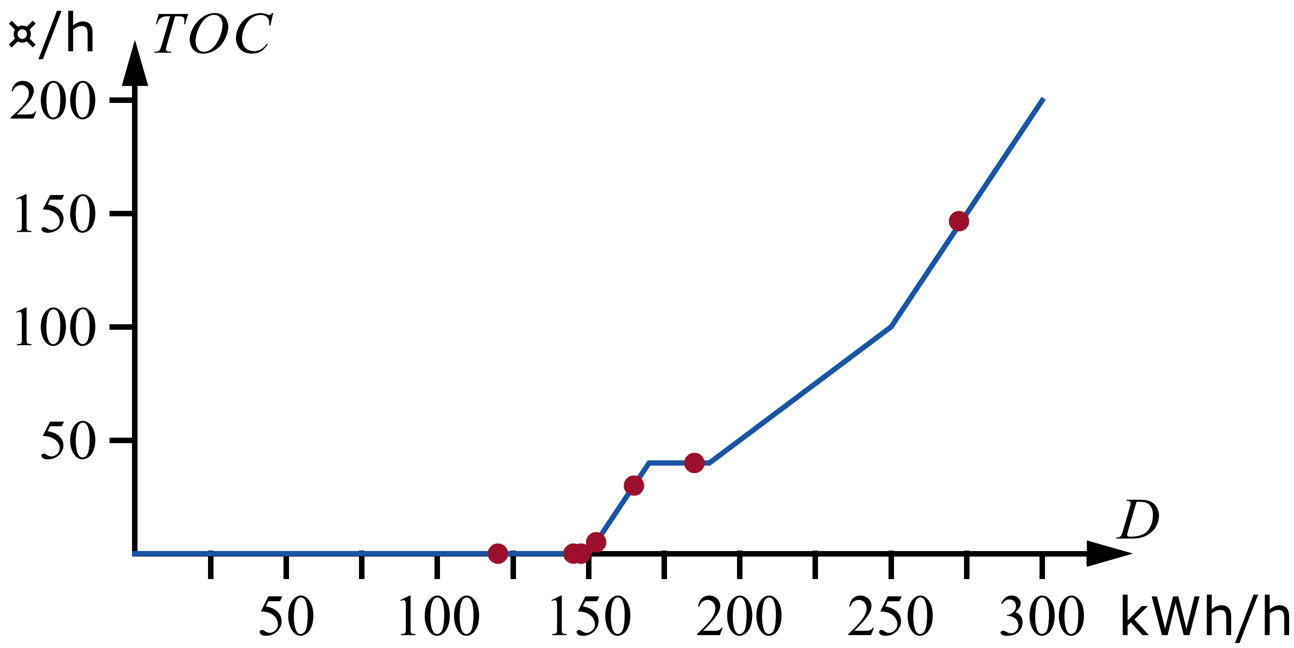
$$m_{TOC} = \frac{1}{10} \sum_{i=1}^{10} toc_i = \ldots = 29.50.$$

SIMPLE SAMPLING - Example

Example 6.22 (cont.)

KTH VETENSKAP OCH KONST

KTH Electrical Engineering



SIMPLE SAMPLING - Example

Example 6.22 (cont.)

- True value: $ETOC = 39.66$ ¤/h.
- Estimate from simple sampling: $m_{TOC} = 29.50$ ¤/h.

KTH Electrical Engineering

HOME ASSIGNMENTS PART IV - Hints

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Problem 23

Apply simple sampling.

- Analyse scenarios using the multi-area model from problem 22.
- Estimate $ETOC$ and $LOLP$.

MONTE CARLO SIMULATION - Efficiency

- A huge number of samples might be necessary to obtain a reasonable accuracy $\Rightarrow$ long computation time.
- However, we might have some information about the simulation results already before we start sampling.
- Sometimes the information can be used to improve the accuracy (i.e., reduce $Var[m_X]$).

KTH Electrical Engineering

MONTE CARLO SIMULATION - Efficiency

- Methods to reduce $Var[m_X]$ are referred to as *variance reduction techniques*.

  In this course we will consider three variance reduction techniques:
  - Complementary random numbers
  - Control variates
  - Stratified sampling

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Theory

KTH Electrical Engineering

- Assume that m_{X1} and m_{X2} are two separate estimates of μ_X, i.e., $E[m_{X1}] = E[m_{X2}] = \mu_X$.
- The mean of these two estimates, i.e., $(m_{X1} + m_{X2})/2$, is also an estimate of μ_X, because

$$E\left[\frac{m_{X1} + m_{X2}}{2}\right] = \frac{1}{2}(E[m_{X1}] + E[m_{X2}]) =$$

$$= \frac{1}{2}(\mu_X + \mu_X) = \mu_X.$$

COMPLEMENTARY RANDOM NUMBERS - Theory

- How good is the new estimate? Study

$$Var\left[\frac{m_{X1} + m_{X2}}{2}\right] = \frac{1}{4}Var[m_{X1} + m_{X2}] =$$

$$= \frac{1}{4}(Var[m_{X1}] + Var[m_{X2}] + 2Cov[m_{X1}, m_{X2}]).$$

KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Theory

KTH Electrical Engineering

- If m_{X1} and m_{X2} are obtained from two separate simulations using simple sampling with n samples in each simulation, then $Var[m_{X1}] = Var[m_{X2}]$ and $Cov[m_{X1}, m_{X2}] = 0$ $\Rightarrow$

$$Var\left[\frac{m_{X1} + m_{X2}}{2}\right] = \ldots = \frac{Var[m_{X1}]}{2}.$$

Cf. theorem 6.22: Twice as many samples should cut the variance of the estimate in half.

COMPLEMENTARY RANDOM NUMBERS - Theory

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

- However, if m_{X1} and m_{X2} are negatively correlated, the variance of the estimate can be lower than for simple sampling. If we have n samples in each simulation, then $Var[m_{X1}] = Var[m_{X2}]$ and $Cov[m_{X1}, m_{X2}] < 0 \Rightarrow$

$$Var\left[\frac{m_{X1} + m_{X2}}{2}\right] = \ldots =$$

$$= \frac{Var[m_{X1}]}{2} + \frac{1}{2}Cov[m_{X1}, m_{X2}].$$

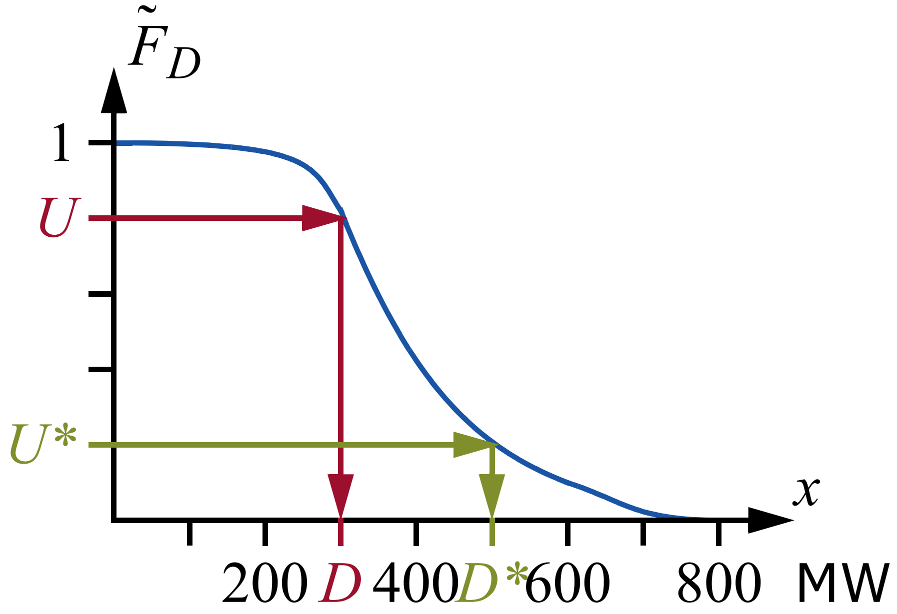
COMPLEMENTARY RANDOM NUMBERS - Theory

KTH Electrical Engineering

How do we get negatively correlated estimates?

- Let U (the original random number) be $U(0, 1)$-distributed.
- Then $U^* = 1 - U$ (the complementary random number) is also $U(0, 1)$-distributed.
- U and U^* are negatively correlated ($\rho_{U,\, U^*} = -1$).
- $Y = F_Y^{-1}(U)$ and $Y^* = F_Y^{-1}(U^*)$ will also be negatively correlated ($\rho_{Y,\, Y^*} \geq \rho_{U,\, U^*}$).

COMPLEMENTARY RANDOM NUMBERS - Theory



KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Theory

KTH Electrical Engineering

- $X = g(Y)$ and $X^* = g(Y^*)$ will also be negatively correlated ($\rho_{X,\,X^*} \geq \rho_{Y,\,Y^*} \geq \rho_{U,\,U^*}$).
- If m_{X1} is based on observations of X and m_{X2} is based on observations of X^* then m_{X1} and m_{X2} will also be negatively correlated.

COMPLEMENTARY RANDOM NUMBERS - Implementation

KTH Electrical Engineering

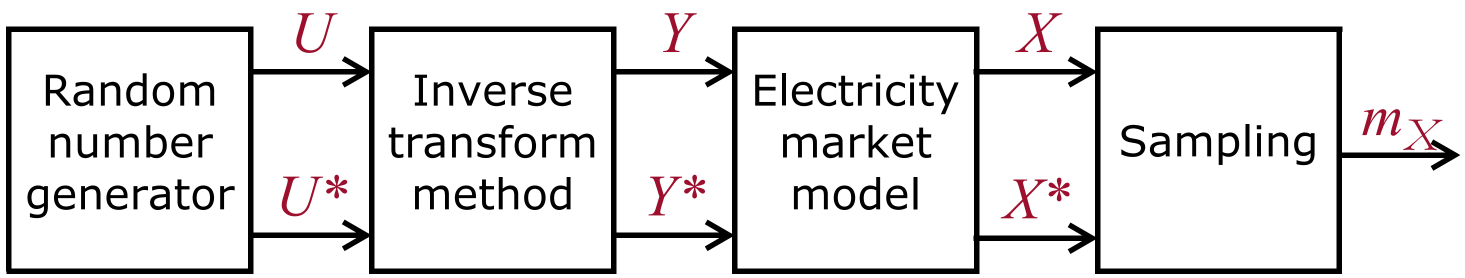
Practical observation:

$$m_{X1} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_{X2} = \frac{1}{n}\sum_{i=1}^{n} x_i^*$$

$$\Rightarrow \frac{m_{X1} + m_{X2}}{2} = \frac{1}{2n}\sum_{i=1}^{n}(x_i + x_i^*).$$

Hence, there is no need to differentiate between samples based on original and complementary random numbers respectively.

COMPLEMENTARY RANDOM NUMBERS - Implementation



- Generate random numbers from uniform distribution (original and complementary)
- Transform all random numbers into appropriate probability distributions of the scenario parameters
- Determine how electricity market behaves in original and complementary scenarios
- Sample the result variables

KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Implementation

If there are S scenario parameters then we can create in total 2^S scenarios on various combinations of original and complementary random numbers.

Example

Two scenario parameters, $\overline{G}$ and D:

- Original scenario: $\overline{G}, D$
- Complementary scenarios: $\overline{G}, D^*, \overline{G}^*, D, \overline{G}^*, D^*$

KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Implementation

KTH Electrical Engineering

- Generating too many complementary scenarios might be inefficient.
- Hence, we should only generate complementary random numbers for those scenario parameters where the negative correlation between Y and Y^* will be detectable in the result variables!

COMPLEMENTARY RANDOM NUMBERS - Implementation

KTH Electrical Engineering

Example: Complementary random numbers for multi-area model

- The total load, $D_{tot} = \Sigma D_n$, has a stronger correlation to TOC than the individual area loads, D_n.
- Randomise the total load D_{tot} and its complementary random number D^*_{tot}.
- Randomise two sets of preliminary loads in the areas, D^i_n and D^{ii}_n respectively.

COMPLEMENTARY RANDOM NUMBERS - Implementation

Example (cont.)

- Finally, scale the preliminary area loads so that they match the total load, i.e.,

$$\text{Scenario 1: } D_n = \frac{D_{tot}}{\sum_{m \in \mathcal{N}} D_m^i} D_n^i$$

$$\text{Scenario 2: } D_n = \frac{D_{tot}^*}{\sum_{m \in \mathcal{N}} D_m^{ii}} D_n^{ii}$$

KTH Electrical Engineering

COMPLEMENTARY RANDOM NUMBERS - Example

KTH Electrical Engineering

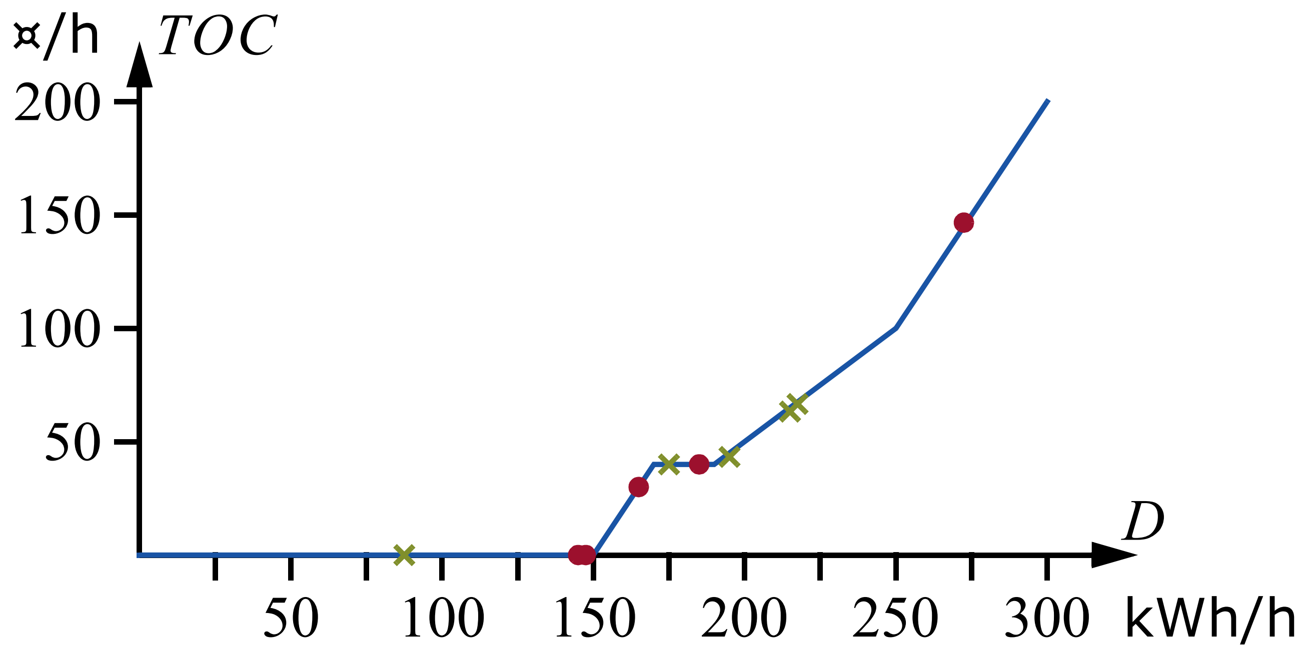
Example 6.26

Scenario, i	1	2	3	4	5
D [kWh/h]	165	273	144	147	185
TOC [¤/h]	30	146	0	0	40
Scenario, i	6	7	8	9	10
D^* [kWh/h]	195	87	216	213	175
TOC^* [¤/h]	45	0	66	63	40

$$m_{TOC} = \frac{1}{10} \sum_{i=1}^{10} toc_i = \ldots = 43.00.$$

COMPLEMENTARY RANDOM NUMBERS - Example

Example 6.26 (cont.)



COMPLEMENTARY RANDOM NUMBERS - Example

KTH Electrical Engineering

Example 6.26 (cont.)

- True value: $ETOC = 39.66$ ¤/h.
- Estimate from simple sampling: $m_{TOC} = 29.50$ ¤/h.
- Estimate using complementary random numbers: $m_{TOC} = 43.00$ ¤/h.

CONTROL VARIATES - Theory

- Assume that we have two electricity market models: a detailed model $X = g(Y)$ and a simplified model $Z = \tilde{g}(Y)$.

  The results of the simplified model are referred to as control variates.

- Assume that we want to calculate the system indices for the detailed model (i.e., estimate $E[g(Y)]$) and that we already know the system indices for the simplified model, $E[\tilde{g}(Y)] = \mu_Z$.
- Sample the difference between the result variables and the control variate, i.e., $X - Z$!

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

CONTROL VARIATES - Theory

- An estimate of the system indices for the detailed model is obtained by adding the system indices of the simplified model to the estimated difference between the two models, because

$$E[m_{(X-Z)} + \mu_Z] = E[X - Z] + \mu_Z =$$
$$= E[X] - E[Z] + \mu_Z = E[X].$$

KTH Electrical Engineering

CONTROL VARIATES - Theory

- How good is the new estimate?

$$Var[m_{(X-Z)} + \mu_Z] = \frac{1}{n}Var[X - Z] + 0 =$$

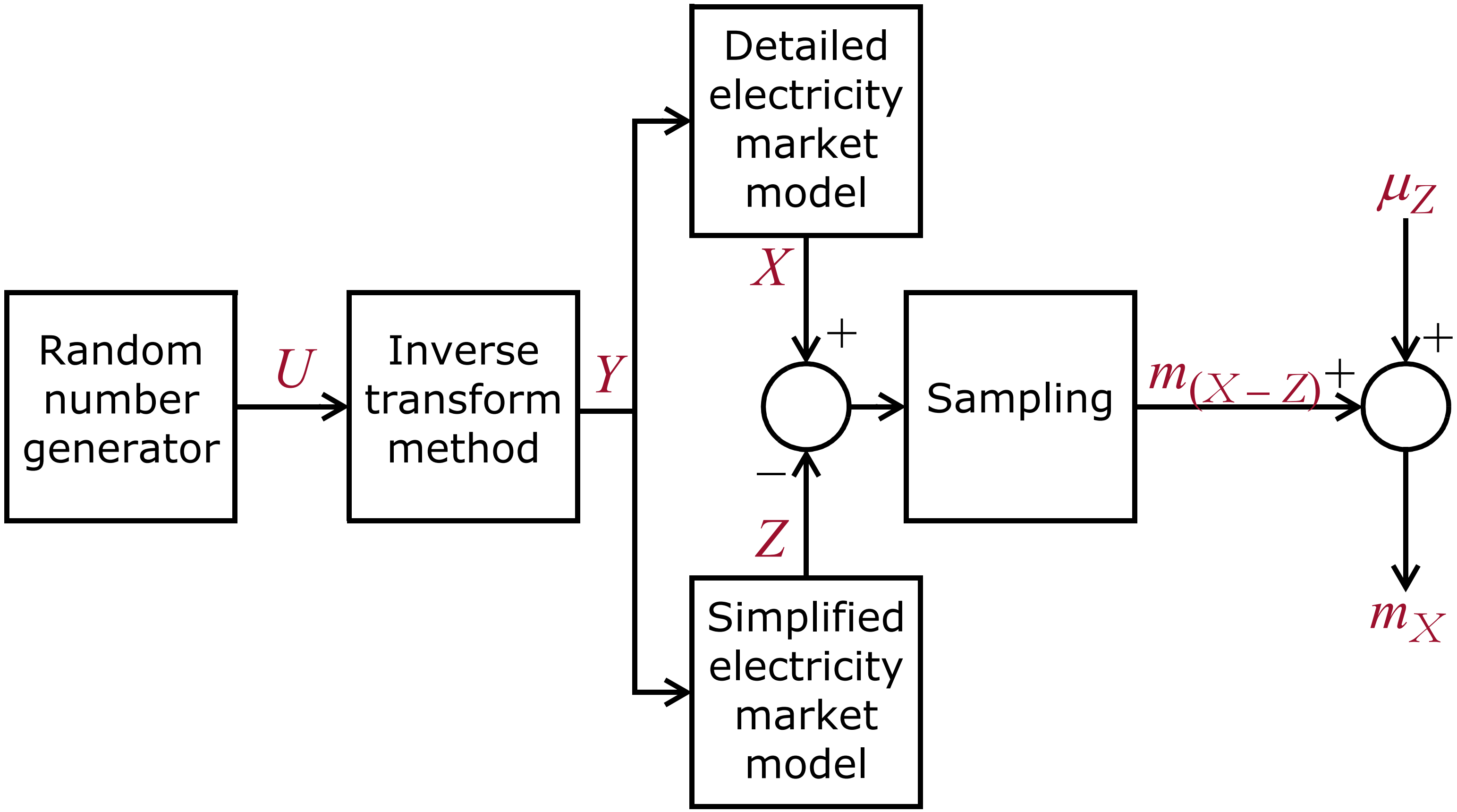
$$= \frac{1}{n}(Var[X] + Var[Z] - 2Cov[X, Z]).$$

- X and Z are results from models of the same system $\Rightarrow$ X and Z should be positively correlated $\Rightarrow$ $Var[m_{(X-Z)} + \mu_Z] < Var[m_X]$, i.e., sampling using a control variate can be more efficient than using simple sampling!

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

CONTROL VARIATES - Theory



KTH Electrical Engineering

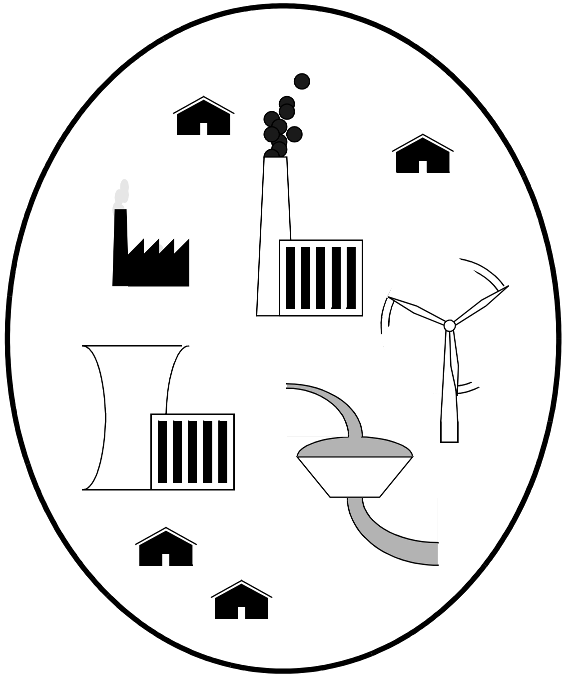
CONTROL VARIATES
- Simplified model

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Assume

- Perfect competition
- Perfect information
- Load is not price sensitive
- Neglect grid losses and limitations
- All scenario parameters can be treated as independent



CONTROL VARIATES - Simplified model

KTH Electrical Engineering

Parameters

D_{tot} = total load (scenario parameter—will be randomised during simulation),

$\overline{G}_g$ = available generation capacity in thermal unit g (scenario parameter—will be randomised during simulation),

β_{Gg} = operation cost in thermal unit g,

β_{Un} = penalty for unserved load.

CONTROL VARIATES - Simplified model

Optimisation variables

$$\tilde{G}_g = \text{generation in thermal unit } g,$$
$$\tilde{U} = \text{unserved load.}$$

Objective function

$$\sum_{g \in G} \beta_{Gg}\tilde{G}_g + \beta_U\tilde{U}.$$

KTH Electrical Engineering

CONTROL VARIATES - Simplified model

KTH Electrical Engineering

Load balance constraint

$$\sum_{g \in G} \tilde{G}_g = D_{tot} - \tilde{U}.$$

Variable limits

$$0 \leq \tilde{G}_g \leq \overline{G}_g,$$

$$0 \leq \tilde{U} \leq D_{tot}.$$

CONTROL VARIATES - Simplified model

The control variates are calculated by solving the optimisation problem for the specific values of the scenario parameters and then calculate

$$\tilde{TOC} = \sum_{g \in G} \beta_{Gg} \tilde{G}_g,$$

$$\tilde{ENS} = \tilde{U},$$

$$\tilde{LOLO} = \begin{cases} 0 & \text{if } \tilde{ENS} = 0, \\ 1 & \text{if } \tilde{ENS} > 0. \end{cases}$$

KTH Electrical Engineering

KTH Electrical Engineering

CONTROL VARIATES - Simplified model

The expectation values of the control variates are calculated by running a probabilistic production cost simulation, i.e.,

$$\mu_{\tilde{TOC}} = ETOC_{PPC},$$

$$\mu_{\tilde{ENS}} = EENS_{PPC},$$

$$\mu_{\tilde{LOLO}} = LOLP_{PPC}.$$

CONTROL VARIATES - Example

Example 6.27

In the simplified model we ignore the lower generation limit on the larger diesel generator set, i.e.,

$$\tilde{g}(Y) = \begin{cases} 0 & Y \leq 150, \\ Y - 150 & 150 < Y \leq 250, \\ 100 + 2(Y - 250) & 250 < Y \leq 300, \\ 200 & 300 < Y. \end{cases}$$

KTH Electrical Engineering

CONTROL VARIATES - Example

Example 6.27 (cont.)

Scenario, i	1	2	3	4	5	6	7	8	9	10
D [kWh/h]	165	273	144	147	185	147	225	120	147	152
TOC [¤/h]	30	146	0	0	40	0	75	0	0	4
$\tilde{TOC}$ [¤/h]	15	146	0	0	35	0	75	0	0	2

KTH Electrical Engineering

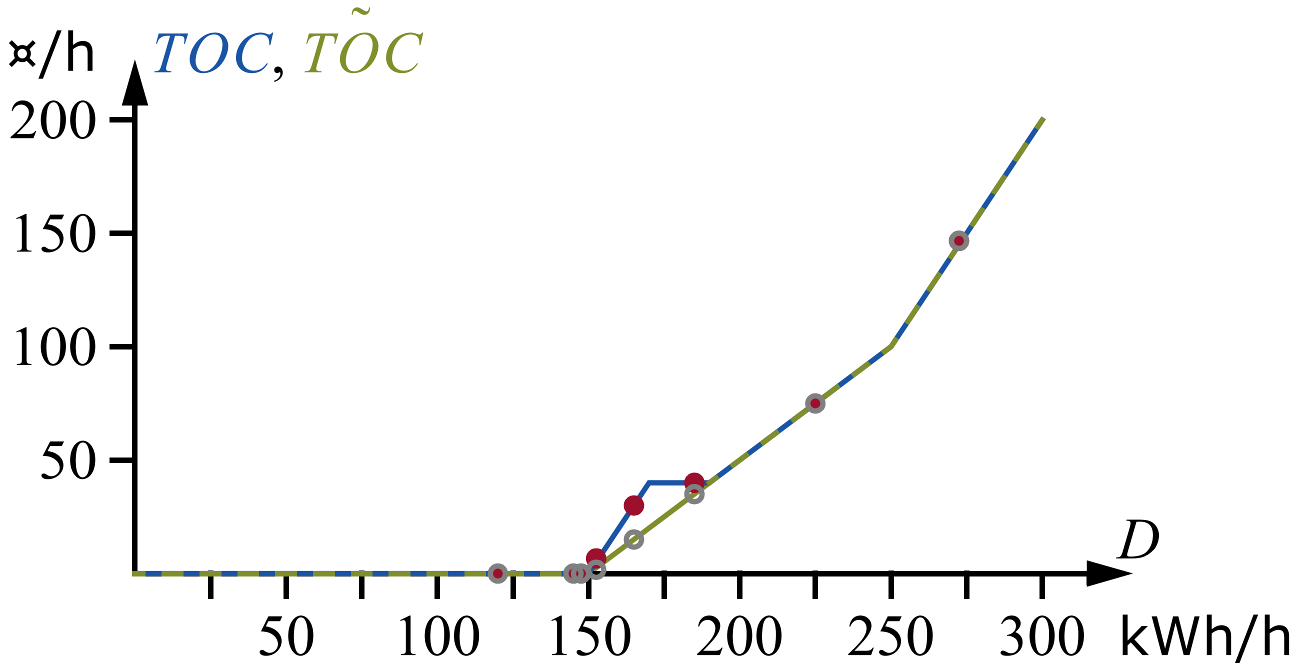
$$ETOC_{PPC} = 36.27.$$

$$m_{TOC} = \frac{1}{10} \sum_{i=1}^{10} (toc_i - \tilde{toc}_i) + \mu_{\tilde{TOC}} = \ldots =$$

$$= 2.20 + 36.27 = 38.47.$$

CONTROL VARIATES - Example

Example 6.27 (cont.)



CONTROL VARIATES - Example

Example 6.27 (cont.)

- True value: $ETOC = 39.66$ ¤/h.
- Estimate from simple sampling: $m_{TOC} = 29.50$ ¤/h.
- Estimate using complementary random numbers: $m_{TOC} = 43.00$ ¤/h.
- Estimate using a control variate: $m_{TOC} = 38.47$ ¤/h.

KTH Electrical Engineering

STRATIFIED SAMPLING
- Theory

KTH Electrical Engineering

- Assume that a population is divided in separate parts, strata, so that each unit belongs to exactly one stratum.

  If X_h is the set of units belonging to stratum h then we must have

  - $X_h \cap X_k = \varnothing, h \neq k$ (no overlapping strata)
  - $\bigcup_h X_h = X$ (the strata should include the entire population)

STRATIFIED SAMPLING - Theory

KTH Electrical Engineering

- Each stratum is assigned a stratum weight corresponding to the size of the stratum, i.e.,

$$\omega_h = \frac{N_h}{N} = P(X \in \mathcal{X}_h).$$

N_h is the number of units in stratum h and N is the number of units in the population.

STRATIFIED SAMPLING - Theory

- Assume that we have determined estimates of $E[X_h]$ for each stratum.
  - Estimate using simple sampling:

$$m_{Xh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{h,i}.$$

  - Analytical value: $m_{Xh} = \mu_{Xh}$.

KTH Electrical Engineering

STRATIFIED SAMPLING - Theory

KTH Electrical Engineering

- The weighted mean of the estimated stratum expectation values, m_{Xh}, is an estimate of $E[X]$, because

$$E\left[\sum_{h=1}^{L} \omega_h m_{Xh}\right] = \sum_{h=1}^{L} \omega_h \mu_{Xh} =$$

$$= \sum_{h=1}^{L} \frac{N_h}{N} \cdot \frac{1}{N_h} \sum_{i=1}^{N_h} x_i = \frac{1}{N} \sum_{i=1}^{N} x_i = E[X].$$

STRATIFIED SAMPLING - Theory

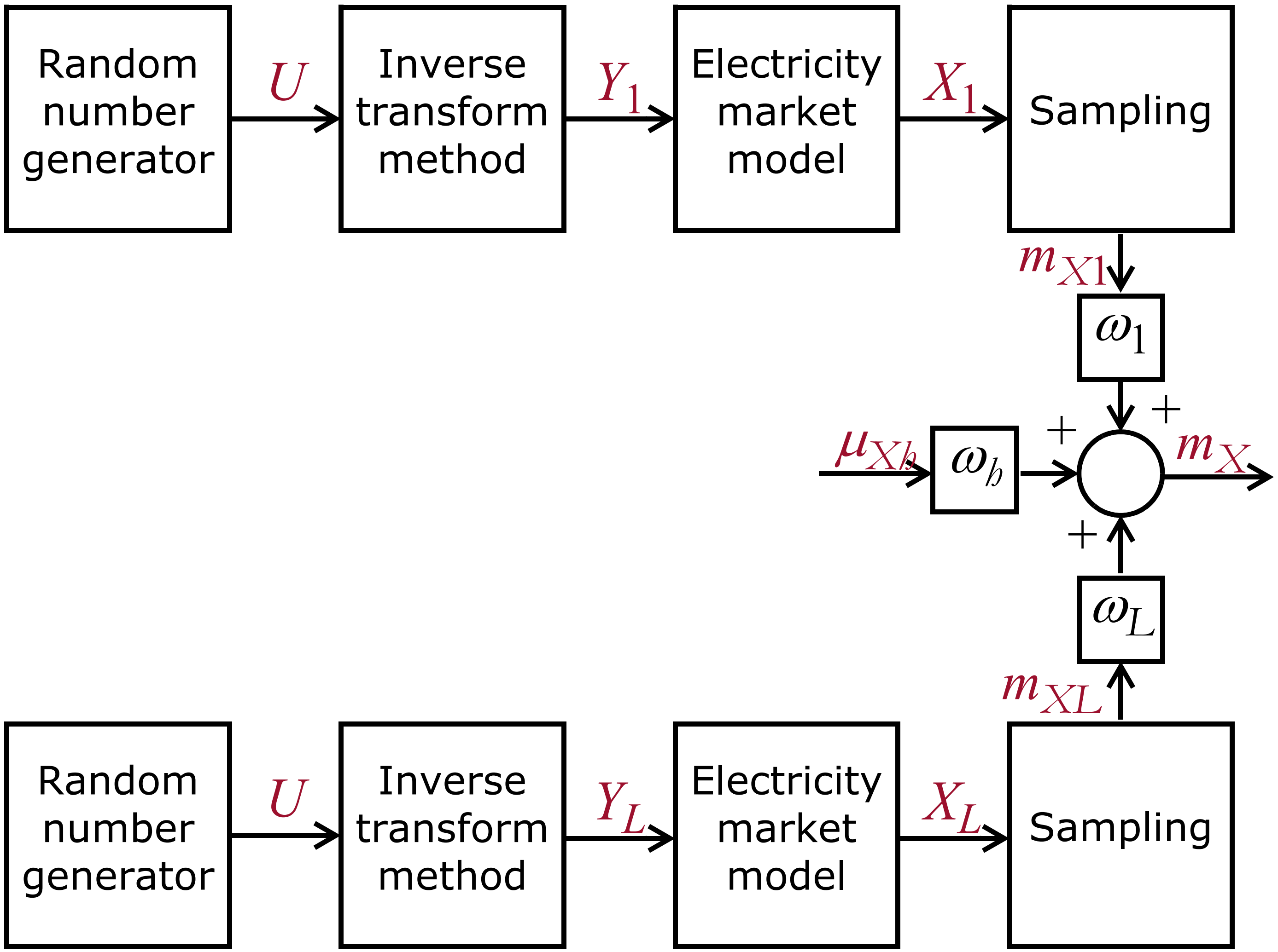
KTH Electrical Engineering

- How good is the new estimate? It can be shown that

$$Var\left[\sum_{h=1}^{L} \omega_h m_{Xh}\right] = \sum_{h=1}^{L} \omega_h^2 Var[m_{Xh}].$$

- Well chosen strata can result in a lesser variance than for simple sampling!
  However, the opposite is also possible!

STRATIFIED SAMPLING - Theory



KTH Electrical Engineering

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

STRATIFIED SAMPLING - Implementation

- Strata cannot be defined based on the values of the result variables, as these values are unknown until we calculate $x = g(y)$.
- A stratum must therefore be defined in terms of possible values for the scenario parameters.
- Hence, we must be able to generate random numbers which belong to a specific part of a probability distribution.

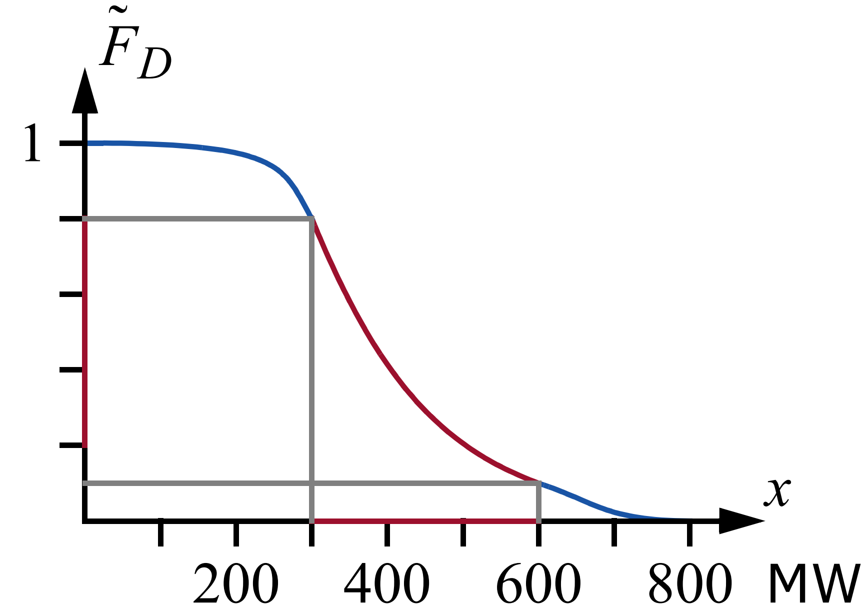
STRATIFIED SAMPLING - Implementation

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Example:

Generate $D \in (300, 600)$.



STRATIFIED SAMPLING - Example

KTH Electrical Engineering

Example 6.28

Introduce the following strata:

1. All scenarios such that $D \leq 150$
2. All scenarios such that $150 < D \leq 250$
3. All scenarios such that $250 < D$

STRATIFIED SAMPLING - Example

KTH Electrical Engineering

Example 6.28 (cont.)

Calculate the stratum weights:

$$\omega_1 = P(D \leq 150) = \Phi(-0.75) \approx 0.23,$$

$$\omega_2 = P(150 < D \leq 250) = \\ = \Phi(1.75) - \Phi(-0.75) \approx 0.73,$$

$$\omega_3 = P(250 < D) = 1 - \Phi(1.75) \approx 0.04.$$

STRATIFIED SAMPLING - Example

Example 6.28 (cont.)

KTH Electrical Engineering

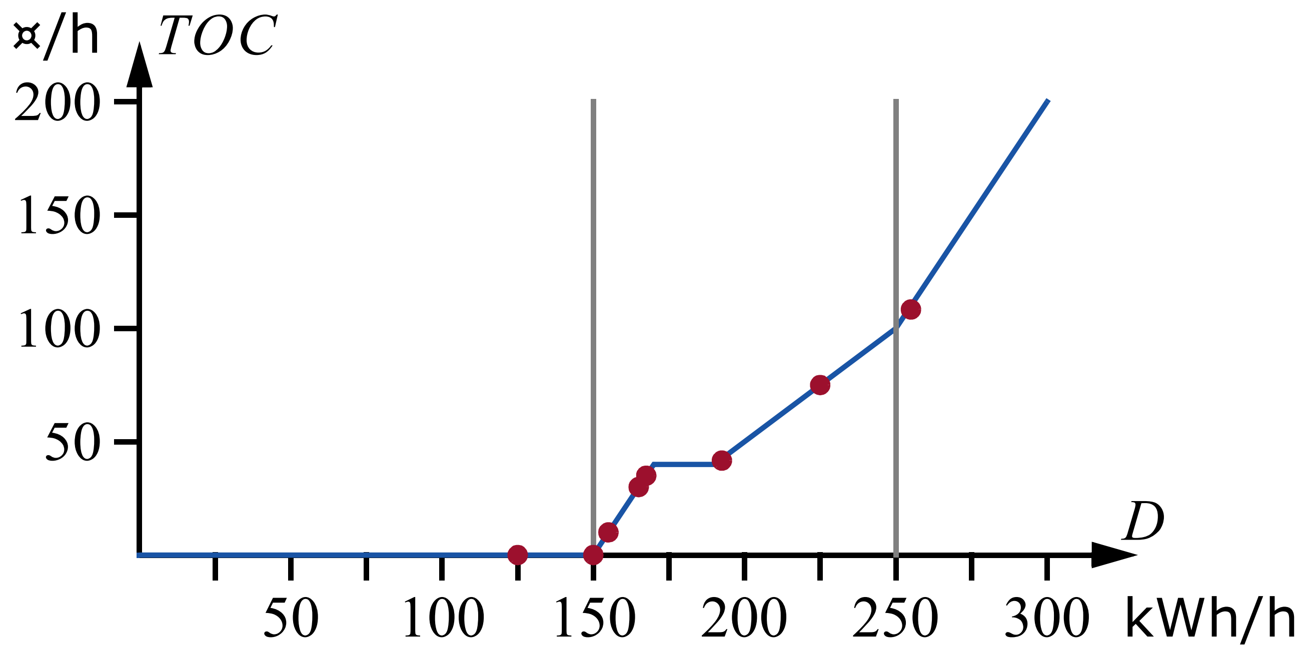
Stratum, h	1		2						3	
Scenario, i	1	2	1	2	3	4	5	6	1	2
D [kWh/h]	124	150	166	168	193	167	224	156	254	255
TOC [¤/h]	0	0	33	36	43	34	74	12	108	110

$$m_{TOC} = \omega_1 \frac{1}{2} \sum_{i=1}^{2} x_{1,i} + \omega_2 \frac{1}{6} \sum_{i=1}^{6} x_{2,i} + \omega_3 \frac{1}{2} \sum_{i=1}^{2} x_{3,i}$$

$$= \ldots = 32.72.$$

STRATIFIED SAMPLING - Example

Example 6.28 (cont.)



KTH VETENSKAP OCH KONST

KTH Electrical Engineering

STRATIFIED SAMPLING - Example

KTH Electrical Engineering

Example 6.28 (cont.)

- True value: $ETOC = 39.66$ ¤/h.
- Simple sampling: $m_{TOC} = 29.50$ ¤/h.
- Complementary random numbers: $m_{TOC} = 43.00$ ¤/h.
- Control variate: $m_{TOC} = 38.47$ ¤/h.
- Stratified sampling: $m_{TOC} = 32.72$ ¤/h.

COMBINED VARIANCE REDUCTION TECHNIQUES

Example 6.29

KTH Electrical Engineering

Stratum, h	1	2			3
Scenario, i	1	1	2	3	1
D [kWh/h]	124	166	168	193	254
$\tilde{TOC}$ [¤/h]	0	16	18	43	108
TOC [¤/h]	0	32	36	43	108
Scenario, i	2	4	5	6	2
D^* [kWh/h]	138	217	215	185	276
$\tilde{TOC}^*$ [¤/h]	0	67	65	35	152
TOC^* [¤/h]	0	67	65	40	152

COMBINED VARIANCE REDUCTION TECHNIQUES

Example 6.29 (cont.)

$$m_{TOC} = \mu_{T\tilde{O}C} + \omega_1 \frac{1}{2} \sum_{i=1}^{2} (toc_{1,i} - t\tilde{o}c_{1,i}) +$$

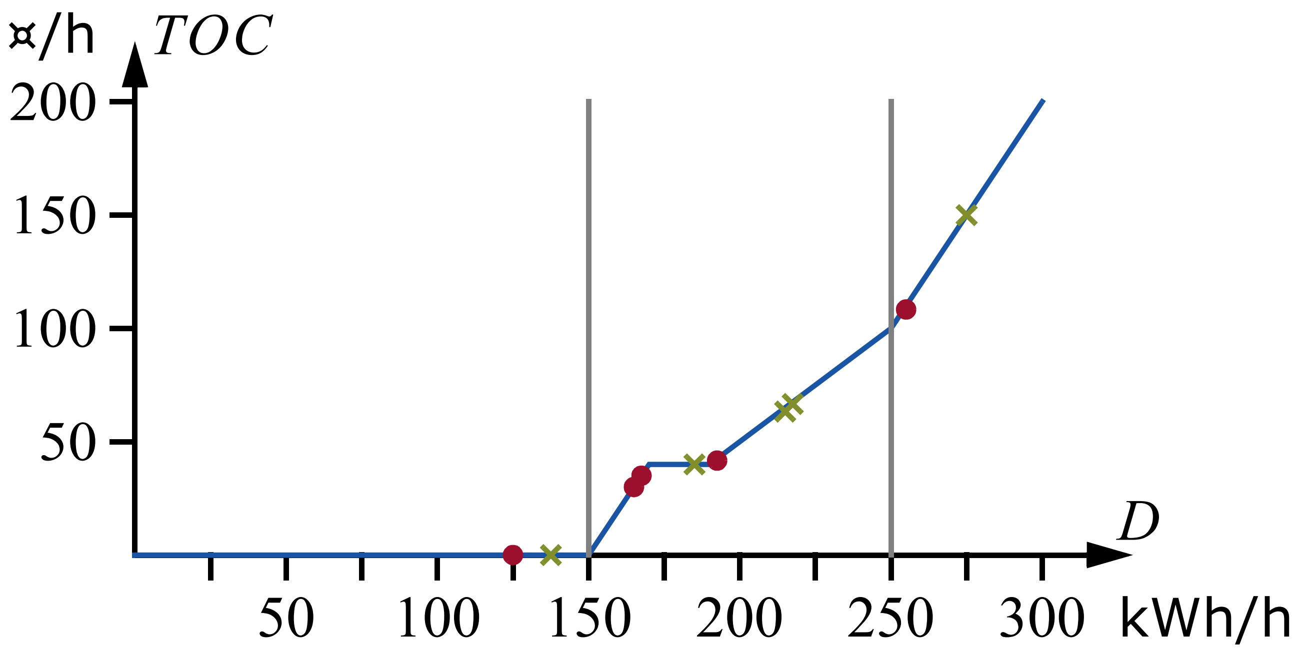
$$+ \omega_2 \frac{1}{6} \sum_{i=1}^{6} (toc_{2,i} - t\tilde{o}c_{2,i}) +$$

$$+ \omega_3 \frac{1}{2} \sum_{i=1}^{2} (toc_{3,i} - t\tilde{o}c_{3,i}) + = \ldots = 43.44.$$

KTH Electrical Engineering

COMBINED VARIANCE REDUCTION TECHNIQUES

Example 6.29 (cont.)



COMBINED VARIANCE REDUCTION TECHNIQUES

Example 6.29 (cont.)

- True value: $ETOC = 39.66$ ¤/h.
- Simple sampling: $m_{TOC} = 29.50$ ¤/h.
- Complementary random numbers: $m_{TOC} = 43.00$ ¤/h.
- Control variate: $m_{TOC} = 38.47$ ¤/h.
- Stratified sampling: $m_{TOC} = 32.72$ ¤/h.
- Combined method: $m_{TOC} = 43.44$ ¤/h.

COMPARISON OF SIMULATION METHODS

Results from 1 000 simulations with different seeds for the random number generator.

Simulation method	Lowest *ETOC* estimate	Average *ETOC* estimate	Highest *ETOC* estimate
Simple sampling	6.33	39.78	81.65
Complementary random numbers	31.48	39.85	64.22
Control variate	36.27	40.03	46.08
Stratified sampling	19.44	39.78	59.74
Combination	36.95	40.03	43.30

KTH Electrical Engineering

STRATIFIED SAMPLING - Strata tree

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

How should strata be defined in order to achieve the largest variance reduction?

Consider that

$$Var[m_X] = \sum_{h=1}^{L} \omega_h^2 Var[m_{Xh}].$$

$\Rightarrow$ If all $Var[m_{Xh}]$ are small then $Var[m_X]$ will be small.

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

- It is preferable if all scenarios belonging to a stratum give the same or very similar values for the result variables.
- To define efficient strata, we must be able to predict the results of the scenarios (without actually calculating the result variables).
- The strata tree is a tool to systematically categorise the scenarios.

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

A strata tree is a tree structure with the following properties:

- The root of the tree contains no information.
- All other nodes specify a number of possible outcomes for one or more scenario parameters.
- Each node has a node weight, which is given by the probability to get the specified outcome for the scenario parameters.

KTH Electrical Engineering

STRATIFIED SAMPLING - Strata tree

- The node weight of the root is 1.
- The scenario parameters along a branch of the tree should be independent of each other.
- Each branch will include a part of the total population, i.e., each branch corresponds to a stratum. The stratum weight is obtained by multiplying the node weights along the branch.

Strata with similar properties can be merged, i.e., one stratum may consist of several branches.

STRATIFIED SAMPLING - Strata tree

The strata tree should include all possible scenarios. This requirement is guaranteed to be fulfilled if

- all children of a certain node specify outcomes for the same scenario parameters,
- the sum of the node weights of the children is equal to 1.

KTH Electrical Engineering

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

The possible values of TOC and $LOLO$ can be predicted if we know the total available generation capacity, $\overline{G}$ (thermal) and $\overline{W}$ (non-dispatchable), as well as the total load, D.

- One node for each possible state of the available generation capacity should be in an upper level of the strata tree.
  Requires discrete probability distributions!
- One node for each interesting load interval in a lower level of the strata tree.

STRATIFIED SAMPLING - Strata tree

- In a multi-area model, we need to consider that there are some load intervals for which it is more difficult to predict TOC and $LOLO$.
- Assume that we know the maximal losses, $\bar{L}$, and the maximal unused generation capacity due to transmission congestion, $\overline{U}_W$ (renewable) and $\overline{U}_{WG}$ (total).

KTH Electrical Engineering

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

- $D \leq \overline{W} - \overline{U}_W \Rightarrow TOC = 0, LOLO = 0$

  The load can be covered using only non-dispatchable units.

- $\overline{W} - \overline{U}_W < D \leq \overline{W} - \overline{L} \Rightarrow TOC \geq 0, LOLO = 0$

  It is possible that the load can be covered using only non-dispatchable units, but other units might have to be dispatched due to transmission congestion.

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

- $\overline{W} - \overline{L} < D \leq \overline{W} \Rightarrow TOC \geq 0, LOLO = 0$

  It is possible that the load can be covered using only non-dispatchable units, but other units might have to be dispatched due to transmission losses.

- $\overline{W} < D \leq \overline{W} + \overline{G} - \overline{U}_{WG} \Rightarrow TOC > 0, LOLO = 0$

  The load cannot be covered using only non-dispatchable units, but the generation capacity is sufficient thanks to the other units.

STRATIFIED SAMPLING - Strata tree

- $\overline{W} + \overline{G} - \overline{U}_{WG} < D \leq \overline{W} + \overline{G} - \overline{L} \Rightarrow TOC > 0$, $LOLO = 0$ or 1

  It is possible that the generation capacity is sufficient, but load shedding might become necessary due to transmission congestion.

- $\overline{W} + \overline{G} - \overline{L} < D \leq \overline{W} + \overline{G} \Rightarrow TOC > 0$, $LOLO = 0$ or 1

  It is possible that the generation capacity is sufficient, but load shedding might become necessary due to transmission losses.

STRATIFIED SAMPLING - Strata tree

- $\overline{W} + \overline{G} < D \Rightarrow TOC > 0, LOLO = 1$

  Load shedding is unavoidable.

Notice that other combinations of TOC and $LOLO$ are also possible, for example if $\overline{G} = 0$.

KTH Electrical Engineering

STRATIFIED SAMPLING - Strata tree

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Example 6.30 — System data

Generation

- Wind power, available capacity 0 kW (50%) or 150 kW (50%), negligible operation cost.
- Diesel generator set, 250 kW, 80% availability, operation cost 10 ¤/kWh.

Load

- Evenings: Mji $N(175,48)$, Kijiji $N(75,20)$.
- Other time: Mji $N(120,24)$, Kijiji $N(30,7)$.

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Example 6.30 — System data

Transmission

- The maximal losses on the line between Mji and Kijiji are 3 kW.

Example 6.30 — Problem

Suggest an appropriate strata tree and calculate the stratum weights!

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Example 6.30 — Solution

Suitable strata tree

- Level 0: Root
- Level 1: Time of day
- Level 2: Available generation capacity
- Level 3: Load

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	D	Node weight				
Day/ Night	0.75	0	0	0.1	≥ 0	1	0.075	0	1	*
		150	0	0.1	≤ 147	0.452	0.034	0	0	I
					147–150	0.048	0.004	0	0/1	*
					> 150	0.5	0.038	0	1	*

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	D	Node weight				
Day/ Night	0.75	0	250	0.4	≤ 247	0.452	0.034	> 0	0	IV
					247–250	0.048	0.004	> 0	0/1	VI
					> 250	≈ 0	≈ 0	> 0	1	VII

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	*D*	Node weight				
Day/ Night	0.75	150	250	0.4	≤ 147	0.452	0.136	0	0	I
					147–150	0.048	0.014	≥ 0	0	III
					150–397	0.5	0.15	> 0	0	IV
					397–400	≈ 0	≈ 0	> 0	0/1	VI
					> 400	≈ 0	≈ 0	> 0	1	VII

STRATIFIED SAMPLING - Strata tree

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	*D*	Node weight				
Evening	0.25	0	0	0.1	≥ 0	1	0.025	0	1	*
		150	0	0.1	≤ 147	0.024	0.001	0	0	I
					147–150	0.003	≈ 0	0	0/1	*
					> 150	0.973	0.243	0	1	*

STRATIFIED SAMPLING - Strata tree

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	D	Node weight				
Evening	0.25	0	250	0.4	≤ 247	0.452	0.034	> 0	0	IV
					247–250	0.048	0.004	> 0	0/1	VI
					> 250	0.5	0.05	> 0	1	VII

STRATIFIED SAMPLING - Strata tree

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Time level		Generation level			Load level		Stratum weight	*TOC*	*LOLO*	Type
Period	Node weight	$\overline{W}$	$\overline{G}$	Node weight	*D*	Node weight				
Evening	0.25	150	250	0.4	≤ 147	0.024	0.002	0	0	I
					147–150	0.003	≈ 0	≥ 0	0	III
					150–397	0.970	0.097	> 0	0	IV
					397–400	≈ 0	≈ 0	> 0	0/1	VI
					> 400	0.002	≈ 0	> 0	1	VII

HOME ASSIGNMENTS PART IV - Hints

KTH Electrical Engineering

Problem 24

Define strata using a strata tree.

- Calculate the maximal losses.
- Identify interesting load intervals.
- Calculate node and stratum weights.

Recommended exercise: 6.16a

HOME ASSIGNMENTS PART IV - Hints

KTH Electrical Engineering

Problem 25

Perform a small Monte Carlo simulation to estimate $ETOC$ and $LOLP$.

- Randomise a scenario (`norminterval` will generate both an original and complementary random number from the normal distribution of your choice).
- Analyse the scenario using the multi-area model (problem 22) and a PPC model.
- Calculate estimates according to (6.46).

Recommended exercises: 6.14, 6.15, 6.16b

HOME ASSIGNMENTS PART IV - Hints

Problem 26

Compare probabilistic production cost simulation and Monte Carlo simulation.

KTH Electrical Engineering

STRATIFIED SAMPLING
- Sample allocation

Theorem (Neyman allocation): $Var[m_X]$ for n samples is minimised if the sample are distributed between the strata according to

$$n_h = \frac{\omega_h \sigma_{Xh}}{\sum_{k=1}^{L} \omega_k \sigma_{Xk}} n,$$

where $\sigma_{Xh} = \sqrt{Var[X_h]}$.

KTH Electrical Engineering

STRATIFIED SAMPLING
- Sample allocation

The Neyman allocation corresponds to a flat optimum, i.e., it is possible that we get a $Var[m_X]$ which is close to the optimal value, even if we do not use the best sample allocation.

KTH Electrical Engineering

STRATIFIED SAMPLING
- Sample allocation

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Problem 1:

$Var[X_h]$ are unknown.

- Estimate σ_{Xh} by

$$s_{Xh} = \sqrt{\frac{1}{n_h}\sum_{i=1}^{n_h}(x_{h,i} - m_{Xh})^2}.$$

- Notice that σ_{Xh} cannot be estimated unless $n_h > 0$!

STRATIFIED SAMPLING - Sample allocation

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Procedure

- Run a pilot study where the number of scenarios per stratum is determined in advance.
- Calculate an appropriate allocation.
- Run a batch of scenarios.
- Test convergence criteria.
- If more scenarios are needed, update the sample allocation, run the next batch, etc.

STRATIFIED SAMPLING - Sample allocation

KTH Electrical Engineering

Problem 2:

We are simultaneously sampling several result variables and a sample allocation that is optimal for one result variable might not be optimal for another.

- Calculate the optimal sample allocation with respect to each result variable.
- Use a compromise allocation (for example the mean of the allocations).

STRATIFIED SAMPLING
- Sample allocation

Example:

Stratum	Optimal allocation for *TOC*	Optimal allocation for *LOLO*	Compromise allocation
1	0	0	0
2	1 028	0	514
3	388	0	194
4	4	1 420	712
5	0	0	0
Σ	1 420	1 420	1 420

KTH Electrical Engineering

STRATIFIED SAMPLING
- Sample allocation

KTH Electrical Engineering

Problem 3:

It might not be possible to achieve the target sample allocation.

- Try to get as close as possible!
  (Cf. algorithm described in the compendium, page 144.)

STRATIFIED SAMPLING - Sample allocation

KTH Electrical Engineering

Example:

Stratum	Allocation			
	Compromise	So far	Next batch	Total
1	0	94	0	94
2	514	530	0	530
3	194	68	77	145
4	712	512	123	635
5	0	16	0	16
Σ	1 420	1 220	200	1 420