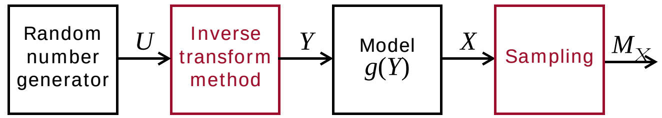
EG2080 Monte Carlo Methods
in Engineering

Theme 6

IMPORTANCE SAMPLING & STRATIFIED SAMPLING

INTRODUCTION

- All samples are treated equally in simple sampling.
- Sometimes it is possible to increase the accuracy by separating samples from different parts of a population.



IMPORTANCE SAMPLING

Assume an importance sampling function, $f_Z(\psi)$, with the following properties:

- The importance sampling function is a density function for a population $\mathcal{Z}$, i.e.,

$$\int_{-\infty}^{\infty} f_Z(\psi)d\psi = 1. \qquad (18)$$

- All values that appear in the population $\mathcal{Y}$ also appear in the population $\mathcal{Z}$:

$$f_Z(\psi) > 0 \ \forall \ \psi: f_Y(\psi) > 0. \qquad (19)$$

IMPORTANCE SAMPLING

- The probability of getting the outcome ψ from an observation of Y compared to the same probability for Z differs by a factor

$$w(\psi) = f_Y(\psi)/f_Z(\psi). \qquad (20)$$

- Study $X(Z) = w(Z) \cdot g(Z)$.

IMPORTANCE SAMPLING

- Expectation value:

$$E[X(Z)] = \int_{\psi \in \boldsymbol{R}} w(\psi)g(\psi)f_Z(\psi)d\psi =$$
$$= \{w(\psi) = f_Y(\psi)/f_Z(\psi)\} = \int_{\psi \in \boldsymbol{R}} g(\psi)f_Y(\psi)d\psi =$$
$$= E[g(Y)], \qquad (21)$$

i.e., sampling $X(Z)$ will produce an estimate of $E[X]$.

KTH Electrical Engineering

IMPORTANCE SAMPLING

- Variance:

$$Var[M_{X(Z)}] = \frac{Var[X(Z)]}{n} =$$
$$= \frac{1}{n}(E[(X(Z))^2] - (E[X(Z)])^2) =$$
$$= \frac{1}{n}\left(\int_{\psi \in \boldsymbol{R}} (w(\psi)g(\psi))^2 f_Z(\psi)d\psi - \mu_X^2\right) =$$
$$= \frac{1}{n}\left(\int_{\psi \in \boldsymbol{R}} w(\psi)g^2(\psi)f_Y(\psi)d\psi - \mu_X^2\right). \qquad (22)$$

KTH Electrical Engineering

IMPORTANCE SAMPLING

- If $f_Z(\psi) = f_Y(\psi)$ we get the same variance as for $Var[M_X]$ in simple sampling.
- If we choose

$$f_Z(\psi) = \frac{g(\psi)f_Y(\psi)}{\mu_X}, \qquad (23)$$

we get $Var[M_{X(Z)}] = 0$!
- If f_Z is close but not exactly equal to (23) we get $Var[M_{X(Z)}] > 0$, but $Var[M_{X(Z)}] < Var[M_X]$.
- Notice that a poor choice of f_Z can result in $Var[M_{X(Z)}] > Var[M_X]$!

KTH Electrical Engineering

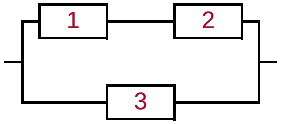
EXCLUDING SCENARIOS

- If the condition (19) is not fulfilled, we will exclude possible scenarios from the simulation, i.e., we are introducing a systematic error in the estimates M_X.
- However, the systematic error can be acceptable if the excluded scenarios are not noticeable in the expectation value μ_X.

KTH Electrical Engineering

EXAMPLE 25 - Optimal importance sampling function

Consider the system below, where each component has a reliability of 90%. Calculate the optimal importance sampling function.



Solution:

- Let $Y_i = 1$ if a component is functional and 0 otherwise.

EXAMPLE 25 - Optimal importance sampling function

Solution (cont.)

- Let $X = 1$ if the system is functional and 0 otherwise.
- It is easy to calculate that $\mu_X = 0.981$.

Table 11 Analysis of the system in example 25.

y	0, 0, 0	0, 0, 1	0, 1, 0	0, 1, 1	1, 0, 0	1, 0, 1	1, 1, 0	1, 1, 1
$g(y)$	0	1	0	1	0	1	1	1
$f_Y(y)$	0.001	0.009	0.009	0.081	0.009	0.081	0.081	0.729
$f_Z(y) = g(y)f_Y(y)/\mu_X$	0	0.009	0	0.083	0	0.083	0.083	0.743

MULTIPLE INPUTS

- In general it is not practical to define a multivariate importance sampling function covering all possible scenarios, $\mathcal{Y}$.
- If there are K independent inputs to the system, Y_k, $k = 1, \ldots, K$, and f_{Zk} is the importance sampling function for the k:th input then the weight factor $w(Z)$ is given by

$$w_i = \prod_{k=1}^{K} \frac{f_Y(\psi_{k,i})}{f_Z(\psi_{k,i})}. \qquad (24)$$

MULTIPLE OUTPUTS

- The optimal importance sampling function depends on the statistical properties of the output X.
- If a system has multiple outputs, it is very likely that each output require different importance sampling functions.
- It might be acceptable to sacrifice some accuracy in one output if we gain accuracy in another.

DESIGNING THE IMPORTANCE SAMPLING FUNCTION

KTH Electrical Engineering

- We have seen that μ_X must be known in order to compute the importance sampling function according to (23).
- Finding a sufficiently good estimate of μ_X is essentially the same problem as to find a suitable approximative model $\tilde{g}$ to generate control variates; the expectation value of the control variate is a suitable approximation to μ_X.

EXAMPLE 26 - Importance sampling function for ICC

KTH Electrical Engineering

a) Use the simplified model in example 22 to suggest an f_Z which is suitable for estimating $E[NI]$. How efficient is the importance sampling function compared to simple sampling?

b) Same question as in part a, but suggest an f_Z which is suitable for estimating $E[MD]$.

c) Same question as in part a, but suggest an f_Z which is suitable for estimating both $E[NI]$ and $E[MD]$.

EXAMPLE 26 - Importance sampling function for ICC

KTH Electrical Engineering

Solution:

a) The simplified model does not take into account the available ingredients, A_i, and therefore we cannot approximate an importance sampling function for $A_i \Rightarrow$ use the real probability distribution for A_i.

Using the simplified model for NI in (23) gives the importance sampling function

$$f_{ZNI}(\psi) = \frac{2\psi \cdot f_{Dtot}(\psi)}{1\,050}.$$

EXAMPLE 26 - Importance sampling function for ICC

KTH Electrical Engineering

Solution (cont.)

b) Using the simplified model for MD in (23) gives the importance sampling function

$$f_{ZMD}(\psi) = \begin{cases} 0 & \text{if } \psi \leq 800, \\ \dfrac{f_{Dtot}(\psi)}{0.0032} & \text{if } \psi > 800. \end{cases}$$

c) Try using the mean of f_{ZNI} and f_{ZMD}.

EXAMPLE 26 - Importance sampling function for ICC

Table 12 *Results of ICC simulation in example 26.*

Simulation method	Expected net income [€/day]			Risk for missed delivery [%]			Average simulation time [h:min:s]
	Min	Av.	Max	Min	Av.	Max	
Enumeration		874			0.32		0:34:54
Simple samp.	829	872	909	0.00	0.27	1.00	0:02:25
Control var.	852	871	890	0.32	0.32	0.32	0:02:25
Imp. samp. a)	848	877	894	0.00	0.31	0.94	0:02:26
Imp. samp. b)	4.31	4.44	4.53	0.32	0.32	0.32	0:02:26
Imp. samp. c)	796	899	1 001	0.27	0.31	0.35	0:02:25

KTH Electrical Engineering

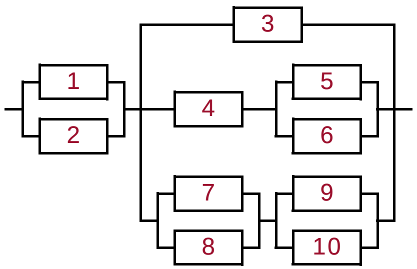
DUOGENEOUS POPULATIONS

- Importance sampling can still be valuable in those cases there it is hard to find approximations of g and μ_X.
- The importance sampling function allows us to force more important scenarios to appear more frequently during the sampling.
- Hence, we can increase the share of diverging units when sampling a duogeneous population.

KTH Electrical Engineering

Example 27 - Importance sampling of duogeneous population

Consider the system below, where each component has a reliability of 98%. Suggest an importance sampling function and test if it is better than simple sampling.



KTH Electrical Engineering

Example 27 - Importance sampling of duogeneous population

Solution:

- The probability that all components are working as they should is $0.98^{10} \approx 82\%$.
- However, it is the remaining 18% of the population that are interesting in the simulation, as the system cannot fail unless at least two components fail.
- Let us use importance sampling to reduce the probability of selecting a scenario where all components are working to say 20%.

KTH Electrical Engineering

Example 27 - Importance sampling of duogeneous population

KTH Electrical Engineering

Solution (cont.)

- Assume that all components have the reliability p_Z. We want p_Z^{10} to be approximately equal to $0.2 \Rightarrow$ choose $p_Z = 80\%$, i.e., use the importance sampling function

$$f_{Zi} = \begin{cases} 0.2 & \psi = 0, \\ 0.8 & \psi = 1, \\ 0 & \text{all other } \psi, \end{cases}$$

for each component.

Example 27 - Importance sampling of duogeneous population

KTH Electrical Engineering

Solution (cont.)

Table 13 *Simulation results in example 27.*

Simulation method	$E[M_X]$*	$Var[M_X]$*	Average simulation time [s]
Enumeration	0.0004	0	0.20
Simple sampling	0.0004	$3.80 \cdot 10^{-7}$	0.75
Importance sampling	0.0004	$1.23 \cdot 10^{-8}$	3.70

* Estimated value based on 100 test simulations with 1 000 scenarios per simulation.

Example 27 - Importance sampling of duogeneous population

KTH Electrical Engineering

Solution (cont.)

- For simple sampling we get

$$T \cdot Var[M_X] \approx 2.85 \cdot 10^{-7},*$$

and for importance sampling we get

$$T \cdot Var[M_X] \approx 4.55 \cdot 10^{-8}.*$$

- Hence, importance sampling can be considered more efficient even though the simulation takes longer time.

* Cf. (4).

ESTIMATED VARIANCE

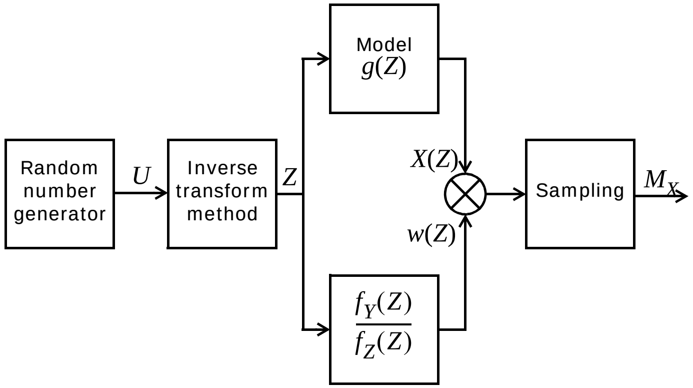
KTH Electrical Engineering

Theorem 13: In importance sampling $Var[X]$ can be estimated by

$$s^2_{X(Z)} = \frac{1}{n} \sum_{i=1}^{n} w(z_i) g^2(z_i) - m^2_{X(Z)}.$$

Notice that the estimate $s^2_{X(Z)}$ can be quite inaccurate if n is small—it is even possible that $s^2_{X(Z)}$ becomes negative!

SIMULATION PROCEDURE



SIMULATION PROCEDURE

- Step 1. Generate the first batch of scenarios, z_i, $i = 1, \ldots, n_b$, according to the importance sampling function f_Z.
- Step 2. Calculate $w_i = f_Y(z_i)/f_Z(z_i)$ and $x_i = g(z_i)$, $i = 1, \ldots, n_b$.
- Step 3. Calculate the sums $\sum_{i=1}^{n} w_i x_i$ and $\sum_{i=1}^{n} w_i x_i^2$.

SIMULATION PROCEDURE

- Step 4. Test stopping rule. If not fulfilled, repeat step 1 and 3 for the next batch.
- Step 5. Calculate estimates and present results.

STRATIFIED SAMPLING

Consider a population divided in groups (strata) with the following properties:

- $\mathcal{X}_h$ is the set of units belonging to stratum h.
- Each unit can only belong to one stratum:

$$\mathcal{X}_h \cap \mathcal{X}_k = \varnothing, h \neq k.$$

- Each unit must belong to one stratum:

$$\bigcup_h \mathcal{X}_h = \mathcal{X}.$$

STRATIFIED SAMPLING

- The stratum weight, ω_h, is the probability that a randomly selected unit belongs to stratum h, i.e.,

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

where N_h is the number of units in stratum h.

KTH Electrical Engineering

STRATIFIED SAMPLING

- Consider the estimate

$$m_X = \sum_{h=1}^{L} \omega_h m_{Xh},$$

where m_{Xh} are estimates of the expectation values of X_h, i.e., estimates of

$$\mu_{Xh} = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{h,i}.$$

KTH Electrical Engineering

STRATIFIED SAMPLING

- Expectation value:

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

$$= \sum_{h=1}^{L} \frac{N_h}{N} \cdot \frac{1}{N_h} \sum_{i=1}^{N_h} x_i = \frac{1}{N} \sum_{i=1}^{N} x_i = E[X], \qquad (25)$$

i.e., the weighted average of M_{Xh} is an estimate of $E[X]$.

KTH Electrical Engineering

STRATIFIED SAMPLING

- Variance:

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

$$= \omega_1^2 Var[M_{X1}] + \ldots + \omega_L^2 Var[M_{XL}] +$$
$$+ 2\omega_1\omega_2 Cov[M_{X1}, M_{X2}] + \ldots$$
$$+ \ldots + 2\omega_{L-1}\omega_L Cov[M_{XL-1}, M_{XL}].$$

KTH Electrical Engineering

STRATIFIED SAMPLING

- Variance (cont.):

  If the estimates M_{Xh} are calculated separately then all covariance terms disappear and we get

$$Var\left[\sum_{h=1}^{L} \omega_h M_{Xh}\right] = \sum_{h=1}^{L} \omega_h^2 Var[M_{Xh}]. \quad (26)$$

KTH Electrical Engineering

STRATIFIED SAMPLING

- If there is only one stratum ($L = 1$, $\omega_1 = 1$) we get the same variance as for $Var[M_X]$ in simple sampling.
- If all strata are homogeneous, i.e., when $x_{h,i} = x_{h,j}\ \forall\ h, i, j$, we get $Var[\Sigma_h \omega_h M_{Xh}] = 0$!
- If strata are not strictly homogeneous we will get $Var[\Sigma_h \omega_h M_{Xh}] > 0$, but $Var[\Sigma_h \omega_h M_{Xh}] < Var[M_X]$.
- Notice that a poor choice of strata can result in $Var[\Sigma_h \omega_h M_{Xh}] > Var[M_X]$!

KTH Electrical Engineering

ESTIMATED VARIANCE

Theorem 14: In stratified sampling $Var[X]$ can be estimated by

$$s_X^2 = \sum_{h=1}^{L} \omega_h^2 s_{Xh}^2.$$

KTH Electrical Engineering

STRATUM PROPERTIES

- To apply stratified sampling, we need to estimate the expectation value and variance of each stratum.
- In some cases, we may calculate

$$\mu_{Xh} = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{h,i}, \quad \sigma_{Xh}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (x_{h,i} - \mu_{Xh})^2$$

  using analytical methods.

KTH Electrical Engineering

STRATUM PROPERTIES

- If an analytical solution is not possible, we can use simple sampling, i.e.,

$$m_{Xh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{h,i}, \quad (27)$$

$$s_{Xh}^2 = \frac{1}{n_h} \sum_{i=1}^{n_h} (x_{h,i} - m_{Xh})^2, \quad (28)$$

where n_h is the number of samples collected from stratum h.

SAMPLE ALLOCATION

Assume that n samples should be collected in a Monte Carlo simulation. How should the samples be distributed between the strata?

Theorem 15: $Var[\Sigma_h \omega_h M_{Xh}]$ is minimised if samples are distributed according to the Neyman allocation:

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

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

SAMPLE ALLOCATION

- The allocation according to theorem 15 is a flat optimum, which means that $Var[\Sigma_h \omega_h M_{Xh}]$ will not increase that much if we deviate from the optimal allocation.
- There is a number of possible practical problems to be addressed when looking for the Neyman allocation.

SAMPLE ALLOCATION
- Pilot study

- The standard deviation σ_{Xh} is generally not known and must be estimated using (28).
- However, it is not possible to estimate σ_{Xh} unless we have some samples, $x_{h,i}$, from stratum h.
- Thus we cannot apply the Neyman allocation until we have run a pilot study, where the number of samples per stratum is decided in advance.

SAMPLE ALLOCATION
- Pilot study

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

- The number of samples per stratum in the pilot study can be the same for all strata (a so-called proportional allocation).
- If we have some knowledge of the properties of the strata, we may concentrate the samples in the pilot study to selected strata.
  - Strata for which μ_{Xh} is known $\Rightarrow$ no samples.
  - Homogeneous strata $\Rightarrow$ few samples.
  - Heterogeneous or duogeneous strata $\Rightarrow$ many samples.

SAMPLE ALLOCATION
- Multiple outputs

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

- The optimal allocation depends on the statistical properties of the output X.
- If a system has multiple outputs, it is very likely that each output require different allocations.
- The solution is to calculate the optimal allocation for each output and then use a compromise of these.

SAMPLE ALLOCATION
- Multiple outputs

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

- The most straightforward compromise allocation is the mean of the optimal allocations for each output.
- However, sometimes it can be worthwhile to use a weighted mean of the optimal allocations for each output.

  For example, we may consider it more important to have an accurate estimate of one of the outputs.

SAMPLE ALLOCATION
- Multiple outputs

KTH VETENSKAP OCH KONST

KTH Electrical Engineering

- It is also possible to use dynamic weight factors when calculating the compromise allocation.
- For example, we may use the weight factor 1 for outputs which do not fulfil the stopping rule, and 0 for all other outputs.

EXAMPLE 28 - Compromise allocation

KTH Electrical Engineering

Use the results in table 14 to compute how the next 200 scenarios should be allocated.

Table 14 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samp-les, n_h	Estimated stratum standard deviations	
			Output 1, s_{X1h}	Output 2, s_{X2h}
1	0.1	300	0	0.080
2	0.2	100	750	0.007
3	0.3	200	1 500	0.002
4	0.4	200	1 000	0

EXAMPLE 28 - Compromise allocation

KTH Electrical Engineering

Solution: Applying theorem 15 to both outputs yields the following results:

Table 15 Compromise allocation in example 28.

Stratum, h	Optimal allocation		Compromise allocation	Sample allocation in next batch
	Output 1	Output 2		
1	0	800	400	100
2	150	140	145	45
3	450	60	255	55
4	400	0	200	0

SAMPLE ALLOCATION - Batch allocation

KTH Electrical Engineering

- The compromise allocation gives the total number of samples per stratum.
- However, we have already collected a number of samples per stratum (in the pilot study and in previous batches) $\Rightarrow$ we need to calculate the sample allocation for the next batch only.
- The number of samples in some strata could become negative if we just subtract the number of samples collected so far from the target compromise allocation.

SAMPLE ALLOCATION - Batch allocation

KTH Electrical Engineering

- It may not be possible to achieve the target compromise allocation. We should then choose an allocation for the next batch which brings us as close as possible.
- However, there is no self-evident definition of what “as close as possible” means $\Rightarrow$ there might be many possible solutions to this problem.

EXAMPLE 29 - Missing samples in one stratum

Use the results in table 16 to compute how the next 200 scenarios should be allocated.

Table 16 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Estimated stratum standard deviations	
			Output 1, s_{X1h}	Output 2, s_{X2h}
1	0.1	100	0	0.080
2	0.2	150	750	0.007
3	0.3	300	1 500	0.002
4	0.4	250	1 000	0

KTH Electrical Engineering

EXAMPLE 29 - Missing samples in one stratum

Solution: We get the same compromise allocation as in example 28:

Table 17 Compromise allocation in example 29.

Stratum, h	Compromise allocation	Number of samples so far	Best possible allocation for next batch
1	400	100	200
2	145	150	0
3	255	300	0
4	200	250	0

KTH Electrical Engineering

SAMPLE ALLOCATION - Batch allocation

- A straightforward idea is that all the number of samples in the next batch is reduced by the same share for all strata where the number of samples cannot reach the target value.

Algorithm

- Step 1. Calculate a compromise allocation, $n_h^{\circledcirc}$ for $n + n_b$ samples (where n is the total number of samples collected so far and n_b is the number of samples to be collected in batch b).

KTH Electrical Engineering

SAMPLE ALLOCATION - Batch allocation

Algorithm (cont.)

- Step 2. Calculate a preliminary batch allocation according to $n'_{h,b} = n_h^{\circledcirc} - n_h$, where n_h is the number of samples collected so far from stratum h.
- Step 3. Let $\mathcal{H}^+$ be the index set of strata which should be allocated more samples, i.e.,

$$\mathcal{H}^+ = \{h: n'_{h,b} > 0\}.$$

KTH Electrical Engineering

SAMPLE ALLOCATION - Batch allocation

Algorithm (cont.)

- Step 4. Calculate the number of wanted samples according to

$$n^+ = \sum_{h \in \mathcal{H}^+} n'_{h,b}.$$

- Step 5. Let $\mathcal{H}^-$ be the index set of strata which have received to many samples, i.e.,

$$\mathcal{H}^- = \{h: n'_{h,b} < 0\}.$$

SAMPLE ALLOCATION - Batch allocation

Algorithm (cont.)

- Step 6. Calculate the number of unwanted samples according to

$$n^- = -\sum_{h \in \mathcal{H}^-} n'_{h,b}.$$

SAMPLE ALLOCATION - Batch allocation

- Step 7. Calculate the batch allocation according to

$$n_{h,b} = \begin{cases} 0 & \forall\ h \in \mathcal{H}^-, \\ (1 - n^-/n^+)n'_{h,b} & \forall\ h \in \mathcal{H}^+. \end{cases}$$

EXAMPLE 30 - Missing samples in several strata

Use the results in table 18 to compute how the next 200 scenarios should be allocated.

Table 18 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Estimated stratum standard deviations	
			Output 1, s_{X1h}	Output 2, s_{X2h}
1	0.1	250	0	0.080
2	0.2	175	750	0.007
3	0.3	275	1 500	0.002
4	0.4	100	1 000	0

EXAMPLE 30 - Missing samples in several strata

Solution: We get the same compromise allocation as in example 28:

Table 19 Compromise allocation in example 30.

Stratum, h	Compro-mise allo-cation	Number of samples so far	Best possible allocation for next batch	Surplus/ deficit
1	400	250	$0.8 \cdot 150 = 120$	–8.5%
2	145	175	0	+38%
3	255	275	0	+15%
4	200	100	$0.8 \cdot 100 = 80$	–10%

THE CARDINAL ERROR

- A Neyman allocation based on a pilot study introduces a risk for a sampling error which is independent of the number of samples.
- This cardinal error can only be avoided by careful design of strata.

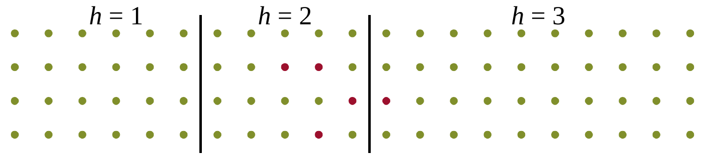
THE CARDINAL ERROR

- The cause of the cardinal error is that the pilot study might incorrectly identify a stratum as homogeneous, i.e., $s_h = 0$ although $\sigma_h > 0$.
- No more samples will then be collected from stratum $h \Rightarrow$ impossible to detect that the stratum is actually heterogeneous!
- There will be a risk for cardinal error in all simulations, unless if all strata really are homogeneous.
- However, the risk will be particularly noticeable for duogeneous populations.

EXAMPLE 31 - Cardinal error in duogeneous population

The population $\mathcal{X}$ consists of 95 conformist units and 5 diverging units. Assume that there are three strata as shown in the figure below.

Ten samples are collected from each stratum in the pilot study. What is the probability of cardinal error?



EXAMPLE 31 - Cardinal error in duogeneous population

Solution:

- Stratum 1 is in reality homogeneous, which means that the risk for cardinal error is 0%.
- In stratum 2 we have 4 diverging units and 16 conformist units. We will get the estimate $s_2 = 0$ either if we only sample conformist units (probability $0.8^{10} \approx 11\%$) or diverging units (probability $0.2^{10} \approx 10^{-7}$), i.e., the risk of cardinal error is approximately 11%.

KTH Electrical Engineering

EXAMPLE 31 - Cardinal error in duogeneous population

Solution (cont.)

- In stratum 3 we have 1 diverging unit and 39 conformist units. We will get the estimate $s_3 = 0$ either if we only sample conformist units (probability $0.975^{10} \approx 78\%$) or diverging units (probability $0.025^{10} \approx 10^{-18}$), i.e., the risk of cardinal error is approximately 78%.

KTH Electrical Engineering

STRATIFICATION

- We have seen how samples should be allocated to minimise $Var[\Sigma_h \omega_h M_{Xh}]$.
- However, from (26) we get that

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

$$= \sum_{h=1}^{L} \omega_h^2 \frac{Var[X_h]}{n_h},$$

i.e., $Var[\Sigma_h \omega_h M_{Xh}]$ is also depending on how we define strata.

KTH Electrical Engineering

THE CUM $\sqrt{f}$-RULE

- It can be shown that for a single output X, the variance of the estimate, $Var[\Sigma_h \omega_h M_{Xh}]$, is approximately minimised if strata are chosen so that they create equal intervals on the cum $\sqrt{f_X(x)}$ scale, where $f_X(x)$ is the density function of X.
- In practice, $f_X(x)$ is not known, but if we have a control variate, then we can use the density function of the control variate, $f_{\tilde{X}}(x)$.

KTH Electrical Engineering

EXAMPLE 32 - The cum $\sqrt{f}$-rule in the ICC simulation

Apply the cum $\sqrt{f}$ –rule to define 8 strata for a simulation of the system in example 15.

Solution: The output MD is a duogeneous population and the only cum $\sqrt{f}$ –rule is therefore not applicable to MD.

Concerning the net income, we can use the control variate $\tilde{NI}$ to define strata. From example 22 we have $\tilde{NI} = 2D_{tot}$; hence, we have $f_{\tilde{NI}}(x) = f_{Dtot}(x/2)$.

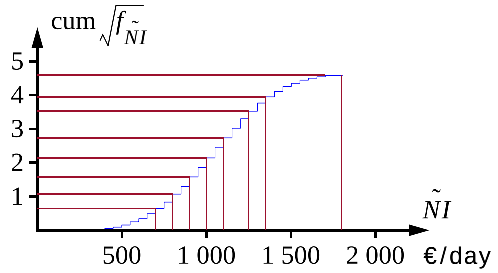
EXAMPLE 32 - The cum $\sqrt{f}$-rule in the ICC simulation

Solution (cont.)

$\tilde{NI}$	$f_{Dtot}\left(\frac{\tilde{NI}}{2}\right)$	cum $\sqrt{f_{Dtot}\left(\frac{\tilde{NI}}{2}\right)}$
300	0.0002	0.0156
400	0.0015	0.0539
450	0.0015	0.0922
…	…	…
1 800	0.0002	4.5921

EXAMPLE 32 - The cum $\sqrt{f}$-rule in the ICC simulation

Solution (cont.)



THE STRATA TREE

- The strata tree is a tool to organise the inputs to the system in such a way that the output values are more easily predictable.
- A strata tree can for example be used to identify strata where we can expect to find diverging units in a duogeneous population.
- The strata tree method requires that we have some knowledge of how the system is responding to some key inputs.

KTH Electrical Engineering

THE STRATA TREE - Definition

- Each node in the strata tree except the root specifies a subset of the population, $\mathcal{Y}_j$, of input j.
- Each node is assigned a node weight, which is equal to the probability that the outcome for the input Y_j is within the subset $\mathcal{Y}_j$, given that the outcome of the inputs in the nodes above belongs to their specified subsets, i.e.,

  $P(Y_j \in \mathcal{Y}_j | Y_k \in \mathcal{Y}_k, k = 1, \ldots, j - 1)$.
- The root holds no information and has the node weight 1.

KTH Electrical Engineering

THE STRATA TREE - Definition

- The nodes along a branch of the strata tree must specify subsets for all inputs of the system.

  If the system has J inputs then there should be at least $J + 1$ levels in the strata tree.
- It is allowed to add “dummy nodes” which do not specify a subset for an input.

  The dummy nodes can be used to simplify the calculation of the node weights.
- All units of the input population $\mathcal{Y}$ should be represented in the strata tree.

KTH Electrical Engineering

THE STRATA TREE - Definition

- Each branch will constitute of subpopulation of the possible scenarios, $\mathcal{Y}$, i.e., a branch constitutes a stratum.
- The stratum weight is the product of the node weights along the branch.
- The number of branches can be quite high and it is not practical to have too many strata; therefore, it is advicable combine several branches where the resulting output values are similar.

  The strata weight is then the sum of the products of the node weights along each branch.

KTH Electrical Engineering

EXAMPLE 33 - Simple strata tree

Consider a bus line, where the number of passengers is P and the transport capacity of the buses is $\bar{P}$. The probability distributions of P and $\bar{P}$ is given in tables 20 and 21.

The objective is to study if the transport capacity is sufficient, i.e., the model is

$$X = \begin{cases} 0 & \text{if } P \le \bar{P}, \\ 1 & \text{if } P > \bar{P}. \end{cases}$$

Use a strata tree to suggest an appropriate stratification.

EXAMPLE 33 - Simple strata tree

KTH Electrical Engineering

Table 20 Data for the bus line in example 33.

Time period	$f_P(x)$				
	$x = 25$	$x = 50$	$x = 75$	$x = 100$	$x = 125$
Day time*	0.01	0.20	0.58	0.20	0.01
Night time*	0.50	0.49	0.01	0	0

Table 21 Data for the bus line in example 33.

Time period	$f_{\overline{P}}(x)$		
	$x = 0$	$x = 50$	$x = 100$
Day time*	0.0001	0.0198	0.9801
Night time*	0.0100	0.9900	0

* Day time: 7 am to 7 pm. Night time: 7 pm to 7 am.

EXAMPLE 33 - Simple strata tree

KTH Electrical Engineering

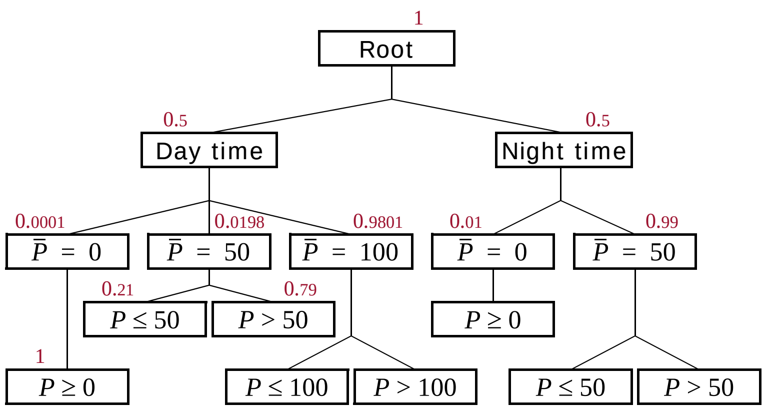
Solution: The output X can be predicted by comparing the capacity of the bus line, $\overline{P}$, to the number of passengers, P. Hence, we should have one level in the tree representing $\overline{P}$ and one level representing P.

There is a correlation between P and $\overline{P}$, which makes it difficult to calculate the node weights. However, the correlation can be managed by including the time of the day in the strata tree.

EXAMPLE 33 - Simple strata tree

KTH Electrical Engineering

Solution (cont.)



EXAMPLE 34 - Effectiveness of ICC simulation

KTH Electrical Engineering

Consider the same system as in example 15. Compare the true expectation values to the probability distribution of estimates from a stratified sampling simulations using approximately 200 samples.

a) Define strata using the cum $\sqrt{f}$ –rule.

b) Define strata using a strata tree.

EXAMPLE 34 - Effectiveness of ICC simulation

Solution:

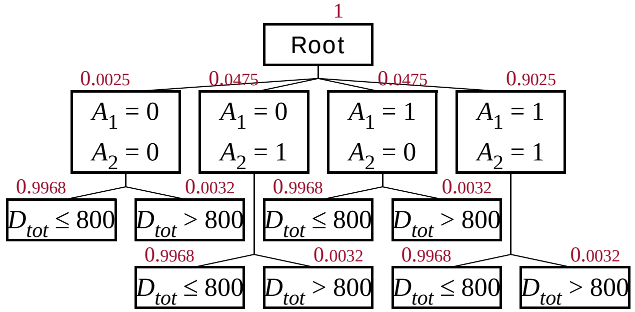
a) Use the stratification from example 32.

b) The availability of ingredients have an impact on how much ice-cream the company has to sell for half price. Hence, put all possible combinations of A_i in one level of the strata tree.

The company will miss a delivery if the demand exceeds the capacity of the machines. Therefore, let us put D_{tot} in one level of the strata tree.

EXAMPLE 34 - Effectiveness of ICC simulation

Solution (cont.)



EXAMPLE 34 - Effectiveness of ICC simulation

Table 22 Results of ICC simulation in example 34.

Simulation method	Expected net income [€/day]			Risk for missed delivery [%]			Average simulation time [h:min:s]
	Min	Av.	Max	Min	Av.	Max	
Enumeration		874			0.32		0:34:54
Simple samp.	829	872	909	0.00	0.27	1.00	0:02:25
Control var.	852	871	890	0.32	0.32	0.32	0:02:25
Stratified sampling a)	862	893	910	0.00	0.10	0.44	0:02:20
Stratified sampling b)	849	874	902	0.32	0.32	0.32	0:02:23

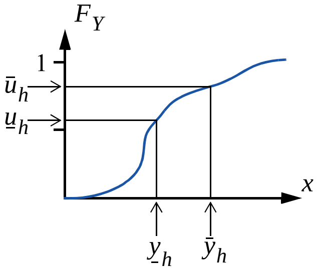
RANDOM NUMBERS FOR STRATIFIED INPUTS

- Each stratum has a specific set of possible values for each input.
- The transformation of pseudorandom values, U, into input values Y may need to be modified.
- Assume that the input values for stratum h should be in the range $\underline{y}_h$ to $\bar{y}_h$.

RANDOM NUMBERS FOR STRATIFIED INPUTS

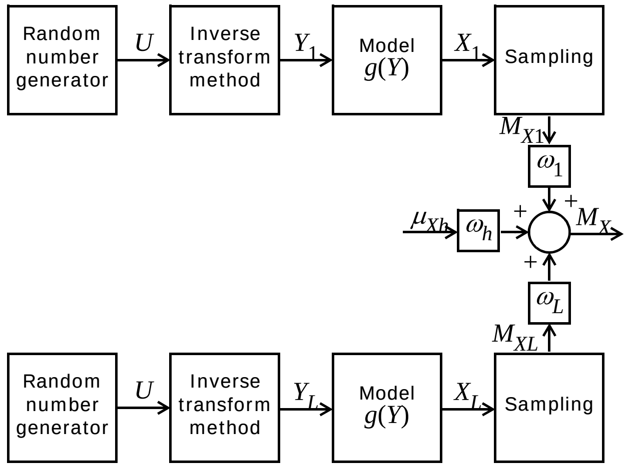
Algorithm

$$\underline{u}_h = F_Y(\underline{y}_h)$$
$$\bar{u}_h = F_Y(\bar{y}_h)$$
$$U_h = \underline{u}_h + (\bar{u}_h - \underline{u}_h)U$$
$$Y_h = F_Y^{-1}(U_h)$$



KTH Electrical Engineering

SIMULATION PROCEDURE



KTH Electrical Engineering

SIMULATION PROCEDURE

- Step 1. Calculate μ_{Xh} and σ_{Xh} for those strata where it is possible.
- Step 2. Generate the first batch of scenarios (the pilot study), $y_{h,i}$, $i = 1, \ldots, n_h$.
- Step 3. Calculate $x_{h,i} = g(y_{h,i})$, $i = 1, \ldots, n_h$.
- Step 4. Update the sums $\sum_{i=1}^{n} x_{h,i}$ and $\sum_{i=1}^{n} x_{h,i}^2$.

KTH Electrical Engineering

SIMULATION PROCEDURE

- Step 5. Test the stopping rule. If not fulfilled, calculate the sample allocation for the next batch and repeat step 2 to 4.
- Step 6. Calculate estimates and present results.

KTH Electrical Engineering

VARIANCE REDUCTION

KTH Electrical Engineering

- To achieve a variance reduction, we need to have some information of the simulated system.
- For importance sampling, we need to know which units of the population are the most important for the expectation value.
- For stratified sampling, we must be able to divide the population in strata such that the total time to obtain accurate estimates for each stratum is less than the time to obtain accurate estimates for the entire population.