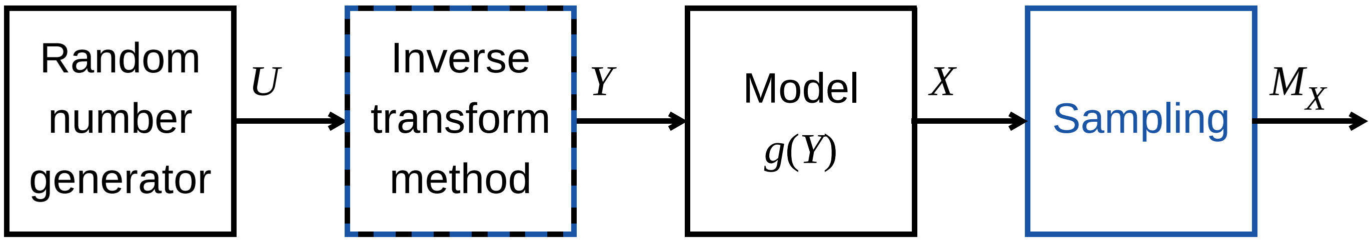
Importance Sampling
Stratified Sampling

Lecture 6, autumn 2016

Mikael Amelin

Introduction

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



Importance sampling

Consider 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.,

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

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

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

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

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

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

Importance sampling

- Expectation value:

$$\mathrm{E}[M_{X(Z)}] = \mathrm{E}[X(Z)] = \int_{\psi} f_Z(\psi) w(\psi) g(\psi) d\psi =$$

$$\int_{\psi} f_Z(\psi) \frac{f_Y(\psi)}{f_Z(\psi)} g(\psi) d\psi = \int_{\psi} f_Y(\psi) g(\psi) d\psi = \mathrm{E}[g(Y)],$$

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

Importance sampling

- Variance:

$$\mathrm{Var}[M_{X(Z)}] = \frac{\mathrm{Var}[X(Z)]}{n} = \frac{1}{n}(\mathrm{E}[(X(Z)^2)] - (\mathrm{E}[X(Z)])^2) =$$

$$= \frac{1}{n}\left(\int_{\psi} f_Z(\psi)(w(\psi)g(\psi))^2 d\psi - \mu_X^2\right) =$$

$$\frac{1}{n}\left(\int_{\psi} f_Z(\psi)\frac{f_Y^2(\psi)}{f_Z^2(\psi)} g^2(\psi) d\psi - \mu_X^2\right) =$$

$$= \frac{1}{n}\left(\int_{\psi} f_Y(\psi)w(\psi)g^2(\psi) d\psi - \mu_X^2\right).$$

Importance sampling

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

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

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

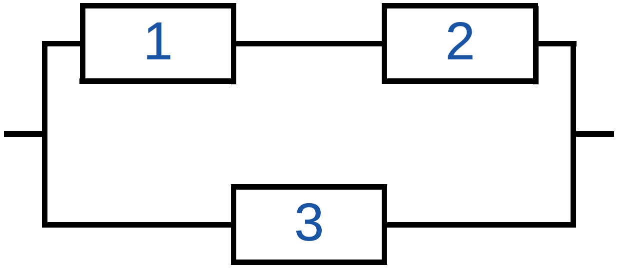
Excluding scenarios

- If $f_Z(\psi) = 0$ although $f_Y(\psi) > 0$ then some scenarios are in practice excluded from the simulation, i.e., we are introducing a systematic error.
- However, the systematic error can be acceptable if the excluded scenarios are not noticeable in the expectation value μ_X.

Example: Optimal importance sampling function

Problem

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



Example: Optimal importance sampling function

Solution

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

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

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

Multiple inputs

- 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.

Choice of importance sampling function

- We have seen that μ_X must be known in order to compute the optimal importance sampling function.
- 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.

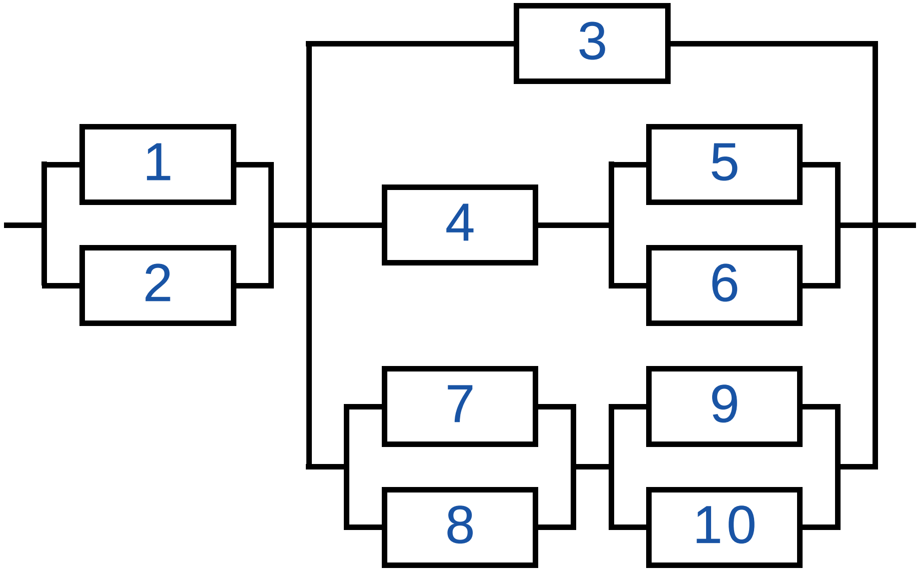
Duogeneous populations

- 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 the interesting scenarios to appear more frequently during the sampling.
- Hence, we can increase the share of diverging units when sampling a duogeneous population.

Example: Importance sampling of duogeneous population

Problem

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.



Example: 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%.

Example: Importance sampling of duogeneous population

Solution

- 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: Importance sampling of duogeneous population

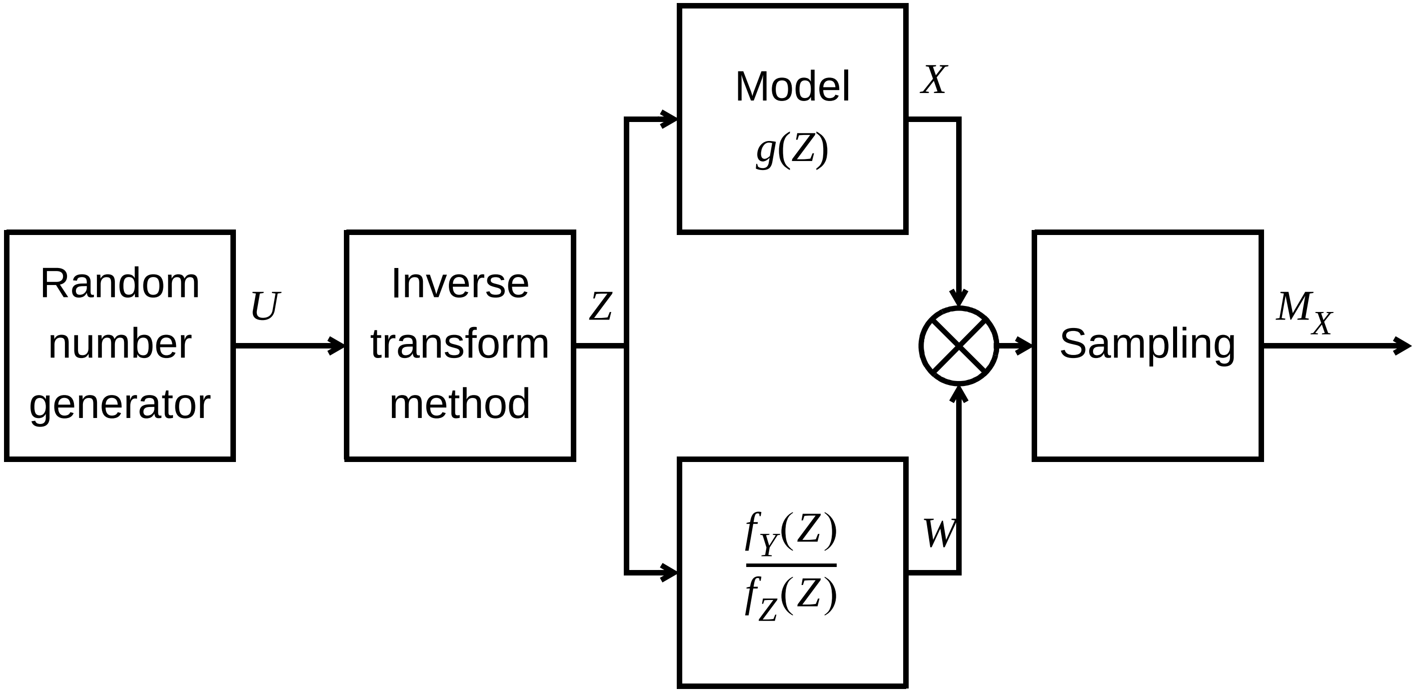
Simulation results

Simulation method	$E[M_X]$*	$\mathrm{Var}[M_X]$*	Average simulation time [s]	Efficiency
Enumeration	0.0004	0	0.20	0
Simple sampling	0.0004	$3.80 \cdot 10^{-7}$	0.75	$2.85 \cdot 10^{-7}$
Importance sampling	0.0004	$1.23 \cdot 10^{-8}$	3.70	$4.55 \cdot 10^{-8}$

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

Simulation procedure

Importance sampling



Simulation procedure

Importance sampling

- **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

$$\sum_{i=1}^{n} w_i x_i \text{ and } \sum_{i=1}^{n} w_i x_i^2.$$

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

Example: Importance sampling of Akabuga District

Importance sampling functions:

- **Large generator:** Increase unavailability to 20%.
- **Small generator:** No change.
- **Transmission line:** increase unavailability to 2%.
- **Total load:** 20% probability of each state (200, 300, 400, 500 and 600 kW respectively).
- **Share of total load in Akabuga:** No change.

Example: Importance sampling of Akabuga District

Simulation method	Average simulation time [ms]	M_{TOC}					M_{LOLO}				
		Min.	Mean	Max.	Var.	Eff.	Min.	Mean	Max.	Var.	Eff.
Enumera-tion	2		443.4		0	0		0.0331		0	0
Simple sampling	8	366.1	436.0	488.8	480.7	3 758	0.0200	0.0326	0.0480	0.000034	0.000263
Compl. r.n.	11	391.9	444.0	484.8	311.2	3 303	0.0160	0.0332	0.0520	0.000036	0.000384
Dagger sampling	8	408.3	444.7	509.0	438.7	4 995	0.0210	0.0334	0.0480	0.000029	0.000327
Control variates	13	424.7	442.2	467.4	53.4	688	0.0250	0.0327	0.0410	0.000012	0.000151
Imp. sam-pling	8	416.1	440.1	474.6	155.6	1 292	0.0262	0.0331	0.0435	0.000013	0.000104

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:

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

- 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.,

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

where N_h is the number of units in stratum h.

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

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

Stratified sampling

- Expectation value:

$$\mathrm{E}\left[\sum_{h=1}^{L}\omega_h M_{Xh}\right] = \sum_{h=1}^{L}\omega_h \mu_{Xh} = \sum_{h=1}^{L}\frac{N_h}{N}\cdot\frac{1}{N_h}\sum_{i=1}^{N_h}x_i = \frac{1}{N}\sum_{i=1}^{N}x_i =$$
$$= \mathrm{E}[X],$$

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

Stratified sampling

- Variance:

$$\mathrm{Var}\left[\sum_{h=1}^{L}\omega_h M_{Xh}\right] = \omega_1^2\mathrm{Var}[M_{X1}] + \ldots + \omega_L^2\mathrm{Var}[M_{XL}] +$$

$$+ 2\omega_1\omega_2\mathrm{Cov}[M_{X1}, M_{X2}] + \ldots$$

$$+ 2\omega_{L-1}\omega_L\mathrm{Cov}[M_{XL-1}, M_{XL}].$$

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

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

Stratified sampling

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

Stratum properties

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

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

- Otherwise we can use simple sampling, i.e.,

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

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

Applying stratified sampling

- The precision of stratified sampling depends on $\sum_h \omega_h^2 \mathrm{Var}[M_{Xh}]$, where ω_h and $\mathrm{Var}[M_{Xh}]$ both depend on the definition of strata and where $\mathrm{Var}[M_{Xh}]$ also depend on the number of samples we collect from stratum h.
- An effective application of stratified sampling thus requires that we manage to
  - define good strata (stratification),
  - distribute the total number of samples over the strata in a good way (sample allocation).
- Both these problems can be addressed in several ways.

Sample allocation

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

Theorem 5.2 (Neyman allocation): For a given stratification, the variance of the estimate from stratified sampling, i.e., $Var[\Sigma_h \omega_h M_{Xh}]$, is minimised if samples are distributed according to

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

Sample allocation

- The Neyman allocation is a flat optimum, which means that $Var[\Sigma_h \omega_h M_{Xh}]$ will not increase that much if we have an allocation which deviates some from the optimal allocation.
- There is a number of possible practical problems to be addressed if we wish to apply the Neyman allocation.

Sample allocation

Pilot study

- The standard deviation σ_{Xh} is generally not known and must be estimated:

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

- 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

- 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

- 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.
- 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. (We may consider it more important to have an accurate estimate of one of the outputs.)

Sample allocation

Multiple outputs

- 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: Compromise allocation

Problem

Use the results in table 1 to compute how the next 200 scenarios should be allocated.

Table 1 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Estimated 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: Compromise allocation

Solution

Applying the Neyman allocation to both outputs yields the following results:

Table 2 Compromise allocation.

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

- 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.
  - This might not be a problem (as in the previous example).
  - The number of samples in some strata could however become negative if we just subtract the number of samples collected so far from the target compromise allocation.

Sample allocation

Batch allocation

- 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: Missing samples 1

Problem

Use the results in table 3 to compute how the next 200 scenarios should be allocated.

Table 3 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Estimated 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

Example: Missing samples 1

Solution

We get the same compromise allocation as in the previous example:

Table 4 Compromise allocation.

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

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).
- **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.

Sample allocation

Batch allocation

- **Step 3.** Let $\mathcal{H}^+$ be the index set of strata which should be allocated more samples, i.e.,

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

- **Step 4.** Calculate the total number of samples needed to collect at least as many samples as in the target allocation for each stratum:

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

Sample allocation

Batch allocation

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

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

- **Step 6.** Calculate the total number of additional samples compared to the target allocation for each stratum:

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

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: Missing samples 2

Problem

Use the results in table 5 to compute how the next 200 scenarios should be allocated.

Table 5 Results from a Monte Carlo simulation.

Stratum, h	Stratum weight, ω_h	Number of samples, n_h	Estimated 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: Missing samples 2

Solution

We get the same compromise allocation as in the previous examples:

Table 6 Compromise allocation.

Stratum, h	Compromise allocation	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

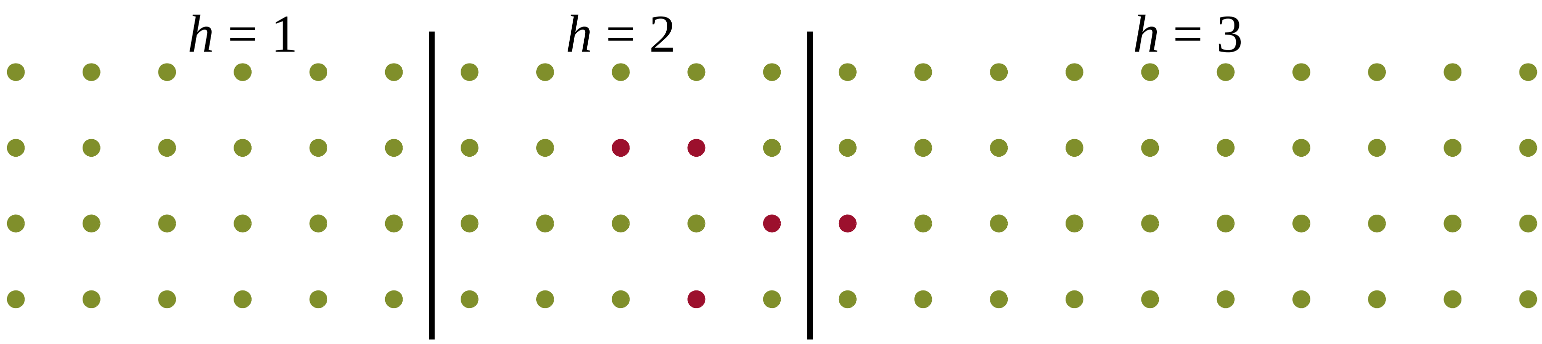
- The cause of the cardinal error is that the pilot study might incorrectly identify a stratum as homogeneous, i.e., $s_h = 0$ although $\sigma_h > 0$.
- No more samples will then be collected from stratum $h \Rightarrow$ impossible to detect that the stratum is actually heterogeneous!
- There will be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

Example: Cardinal error for a duogeneous population

Problem

The population 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: Cardinal error for a duogeneous population

Solution

- Stratum 1 is really homogeneous and the risk 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 is approximately 11%.
- 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 is approximately 78%.

Stratification

A good stratification should

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

  i.e., strata with high stratum weights should be as homogeneous as possible (low $\mathrm{Var}[M_{Xh}]$),
- avoid the cardinal error.

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 $\text{cum}\ \sqrt{f_X(x)}$ 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, $f_{\tilde{X}}(x)$.

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 duoge-neous population.
- The strata tree method requires that we have some knowledge of how the system is responding to some key inputs.

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.

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 Υ should be represented in the strata tree.

The strata tree

Definition

- Each branch will constitute of subset of the possible scenarios, Υ, 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 advisable 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.

Example: Simple strata tree

Problem

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 $\bar{P}$ are given in tables 7 and 8.

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

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

Use a strata tree to suggest an appropriate stratification.

Example: Simple strata tree

Problem

Table 7 Probability distribution of the number of passengers.

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 8 Probability distribution of the transport capacity.

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.

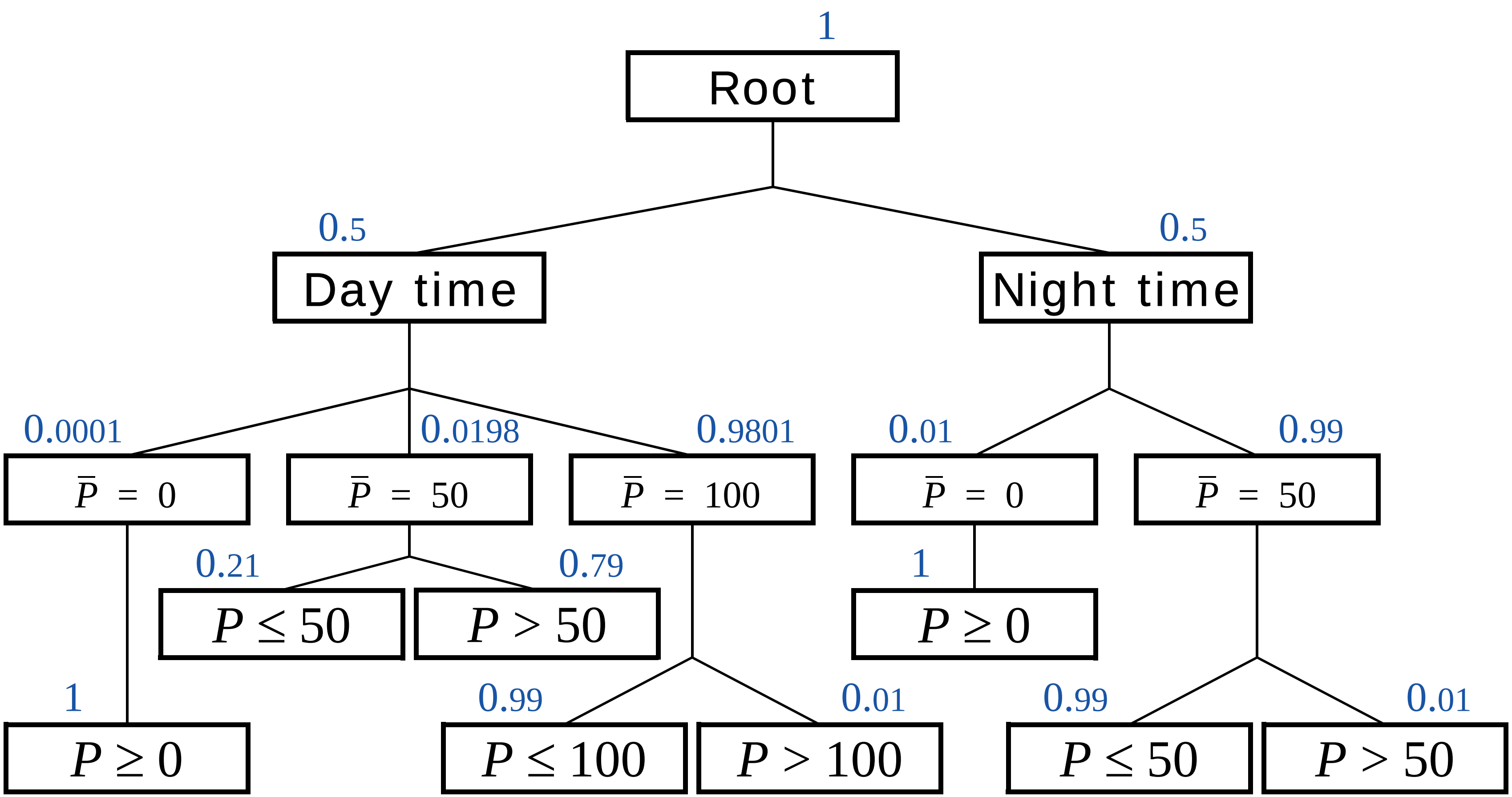
Example: Simple strata tree

Solution

- The output X can be predicted by comparing the capacity of the bus line, $\bar{P}$, to the number of passengers, P. Hence, we should have one level in the tree representing $\bar{P}$ and one level representing P.
- There is a correlation between P and $\bar{P}$, 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: Simple strata tree

Solution



Example: Stratified sampling of Akabuga District

Define a stratum tree for the system with the following levels:

- Available transmission capacity.
- Total available capacity in diesel generator sets.
- Total load.

h	$\bar{P}$	$\bar{G}_{tot}$	D_{tot}	TOC	$LOLO$	ω_h
1	0	0 + 0	200–600	0	1	0.0002
2		0 + 150	200–600	1 800	1	0.0008
3		200 + 0	200	2 000	0 or 1	0.0004
4		200 + 0	300–600	2 000	1	0.0014
5		200 + 150	200–300	≥ 2 000	0	0.0043
6		200 + 150	400	≥ 3 680	0 or 1	0.0018
7		200 + 150	500–600	3 800	1	0.0011

Example: Stratified sampling of Akabuga District

Define a stratum tree for the system with the following levels:

- Available transmission capacity.
- Total available capacity in diesel generator sets.
- Total load.

h	$\bar{P}$	$\bar{G}_{tot}$	D_{tot}	TOC	$LOLO$	ω_h
8	300	0–350	200–300	0	0	0.5940
9		0 + 0	400–600	0	1	0.0079
10		0 + 150	400	≥ 650	0	0.0198
11		0 + 150	500–600	1 800	1	0.0119
12		200 + 0	400–500	≥ 542	0	0.0624
13		200 + 0	600	2 000	1	0.0089
14		200 + 150	400–600	≥ 542	0	0.2851

Example: Stratified sampling of Akabuga District

Simulation method	Average simulation time [ms]	M_{TOC}					M_{LOLO}				
		Min.	Mean	Max.	Var.	Eff.	Min.	Mean	Max.	Var.	Eff.
Enumera-tion	2		443.4		0	0		0.0331		0	0
Simple	8	366.1	436.0	488.8	480.7	3 758	0.0200	0.0326	0.0480	0.000034	0.000263
Compl. r.n.	11	391.9	444.0	484.8	311.2	3 303	0.0160	0.0332	0.0520	0.000036	0.000384
Dagger	8	408.3	444.7	509.0	438.7	4 995	0.0210	0.0334	0.0480	0.000029	0.000327
Ctrl var.	13	424.7	442.2	467.4	53.4	688	0.0250	0.0327	0.0410	0.000012	0.000151
Imp.	8	416.1	440.1	474.6	155.6	1 292	0.0262	0.0331	0.0435	0.000013	0.000104
Stratified sampling	12	417.9	443.2	476.1	134.2	1 578	0.0330	0.0331	0.0332	0.000000	0.000000

Random number generation for strata

- 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 number generation for strata

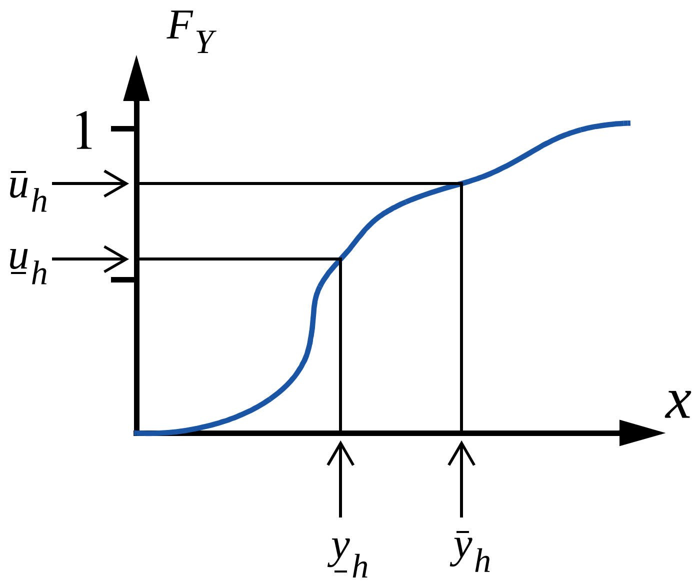
Algorithm

$$\underline{u}_h = F_Y(\underline{y}_h)$$

$$\bar{u}_h = F_Y(\bar{y}_h)$$

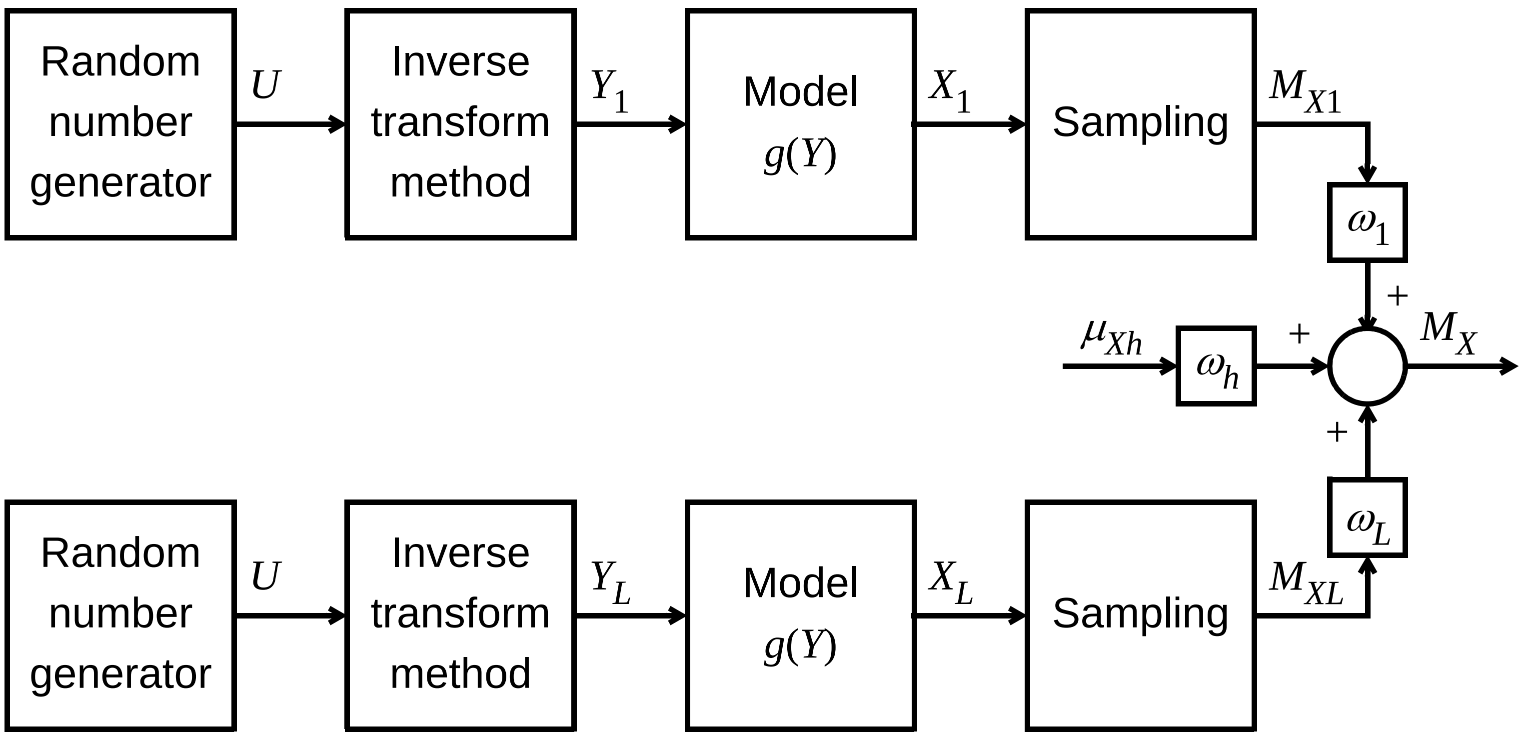
$$U_h = \underline{u}_h + (\bar{u}_h - \underline{u}_h)U$$

$$Y_h = F_Y^{-1}(U_h)$$



Simulation procedure

Stratified sampling



Simulation procedure

Stratified sampling

- **Step 1.** Calculate μ_{Xh} for those strata where it is possible.
- **Step 2.** Generate the first batch of scenarios (the pilot study), $y_{h,i}$, $i = 1, \ldots, n_h$, and calculate the sample allocation for the next batch.
- **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 $\sum_{i=1}^{n} x_{h,i}^2$.
- **Step 5.** Test the stopping rule. If not fulfilled, calculate the sample allocation for the next batch and repeat steps 3 to 4.
- **Step 6.** Calculate estimates and present results.

Variance reduction

- To achieve a variance reduction, we need to have some information of the simulated system.
- For importance sampling, we need to find an importance sampling function which directs the sampling to the 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.