



Theme 3

# SIMPLE SAMPLING

## LAW OF LARGE NUMBERS - History



- First described by the Swiss mathematician Jacob Bernouille (1654–1705).
- Further described by the French mathematician Siméon-Denis Poisson (1781-1840).
- Bernouille named it his “Golden Theorem”, but later it was referred to as “Bernouille’s Theorem”. Poisson coined the name “Law of Large Numbers”.

## LAW OF LARGE NUMBERS

*Theorem 9:* If  $x_1, \dots, x_n$  are independent observations of the random variable  $X$  then  $E[X]$  can be estimated by

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

Notice that  $m_X$  is actually the outcome of a random variable  $M_X$ !



## LAW OF LARGE NUMBERS - Proof

*Proof:* Let  $t_i$  denote the number of times that unit  $i$  appears in the samples; hence,  $t_i$  is an integer between 0 and  $n$ . The estimate of the expectation value can then be expressed as

$$M_X = \frac{1}{n} \sum_{i=1}^N t_i x_i.$$

The number of successful trials when  $n$  trials are performed and the probability of success is  $p$  in each trial is Bernouille-distributed. This means that  $t_i$  must be  $B(n, 1/N)$ -distributed.



## LAW OF LARGE NUMBERS - Proof



KTH Electrical Engineering

Proof (cont.)

$$E[M_X] = E\left[\frac{1}{n} \sum_{i=1}^N t_i x_i\right] = \sum_{i=1}^N \frac{1}{n} E[t_i] x_i = \{\text{the expectation value of a } B(n, p)\text{-distribution is } n \cdot p\} =$$

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

5

## SIMPLE SAMPLING



KTH Electrical Engineering

- Simple sampling means that completely random observations (**samples**) of a random variable,  $X$ , are collected.
- A sufficient number of samples will provide an estimate of  $E[X]$  according to the law of large numbers.
- Simple sampling can also be used to estimate other statistical properties (for example variance or probability distribution) of  $X$ .

6

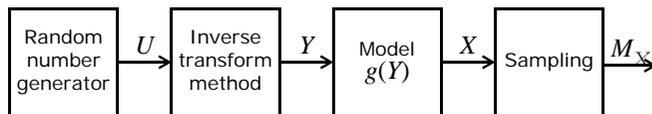
## SIMPLE SAMPLING - Computer simulation



KTH Electrical Engineering

In the computer simulation problem considered in this course, we generate samples by randomising the input values,  $y_i$ , and calculate the outcome  $x_i = g(y_i)$ .

A set of outcomes for the input variables will be referred to as a **scenario**.



7

## EXAMPLE 6 - Estimated mean

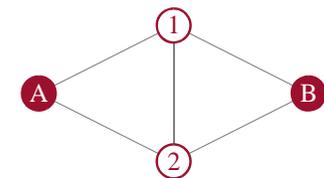
Table 1 Data for the system in example 6.

Leg	A ↔ 1	A ↔ 2	1 ↔ 2	1 ↔ B	2 ↔ B
Travelling time					
Short [min.]	6	10	3	12	14
Probability [%]	70	60	90	80	70
Long [min.]	9	12	5	17	18
Probability [%]	30	40	10	20	30



KTH Electrical Engineering

Estimate the expected travel time from A to B!



8

## EXAMPLE 6 - Estimated mean

Solution:

Table 2 Scenarios in the Monte Carlo simulation.

$U_1$	$U_2$	$U_3$	$U_4$	$U_5$	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$X$
0.95	0.76	0.62	0.41	0.06	9	12	3	12	14	21
0.23	0.46	0.79	0.94	0.35	6	10	3	17	14	23
0.61	0.02	0.92	0.92	0.81	6	10	5	17	18	23
0.49	0.82	0.74	0.41	0.01	6	12	3	12	14	18
0.89	0.44	0.18	0.89	0.14	9	12	3	17	14	26

$$m_X = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{21 + 23 + 23 + 18 + 26}{5} = 22.2.$$



KTH Electrical Engineering

9

## ESTIMATED VARIANCE

*Theorem 10:* If  $x_1, \dots, x_n$  are independent observations of the random variable  $X$  then  $\text{Var}[X]$  can be estimated by

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

For practical calculations,  $s_X^2$ , can be rewritten as

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



KTH Electrical Engineering

10

## EXAMPLE 7 - Estimated variance

Assume that 1 000 scenarios of the system in example 6 have been generated, and that the following results have been obtained:

$$\sum_{i=1}^{1000} x_i = 19\,857, \quad \sum_{i=1}^{1000} x_i^2 = 399\,651.$$

Estimate  $\text{Var}[X]$ !



KTH Electrical Engineering

11

## EXAMPLE 7 - Estimated variance

*Solution:* Start by estimating the expectation value:

$$m_X = \frac{1}{1\,000} \sum_{i=1}^{1\,000} x_i \approx 19.9.$$

The variance is then estimated by

$$s_X^2 = \frac{1}{1\,000} \sum_{i=1}^{1\,000} x_i^2 - m_X^2 = 399.651 - 19.857^2 \approx 5.3.$$



KTH Electrical Engineering

12

## ESTIMATED PROBABILITY DISTRIBUTION



KTH Electrical Engineering

- Sometimes we want to estimate the probability distribution of the output  $X$  and not just the expectation value and variance of  $X$ .
- Parameters of general probability distributions can usually be identified from the estimates  $m_X$  and  $s_X^2$ .

13

## EXAMPLE 8 - Estimated probability distribution



KTH Electrical Engineering

Consider a Monte Carlo simulation which has provided the results  $m_X = 89$  and  $s_X^2 = 8\,281$ . Estimate the probability distribution of  $X$  if  $X$  can be assumed to be

- a) normally distributed,
- b) exponentially distributed.

14

## EXAMPLE 8 - Estimated probability distribution



KTH Electrical Engineering

**Solution:**

- a) If  $Z$  is  $N(\mu, \sigma)$ -distributed then  $E[Z] = \mu$  and  $Var[Z] = \sigma^2$ . Here we have  $\mu = m_X$  and  $\sigma = \sqrt{s_X^2}$   
 $\Rightarrow X$  can be assumed to be  $N(89, 91)$ -distributed.
- b) If  $Z$  is  $E(\lambda)$ -distributed then  $E[Z] = 1/\lambda$  and  $Var[Z] = 1/\lambda^2$ . Here we have  $\lambda = 1/m_X$  and  $\lambda = \sqrt{1/s_X^2} \Rightarrow \{1 \text{ unknown, 2 equations—use least square solution}\} \Rightarrow X$  can be assumed to be  $E(1/90)$ -distributed.

15

## ESTIMATED PROBABILITY DISTRIBUTION



KTH Electrical Engineering

- If the probability distribution of  $X$  is completely unknown, it can be approximated by a discrete distribution based on the observations,  $x_1, \dots, x_n$ .
- Two alternatives:
  - Estimate the population  $X$  as the set  $\{x_1, \dots, x_n\}$ .
  - Estimate the frequency function  $f_X(x)$ .
- It is convenient to present the results graphically.

16

## ESTIMATED POPULATION

- **Step 1.** Sort the observations  $x_i, i = 1, \dots, n$ , in an increasing sequence.
- **Step 2.** Draw a "stair-case" where the height of each step is  $1/n$ .



KTH Electrical Engineering

17

## EXAMPLE 9 - Estimated population

A Monte Carlo simulation has resulted in the following samples of the output  $X$ :

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	54	26	62	58	23	48	60	65	46	66

Draw the estimated distribution function based on these samples.

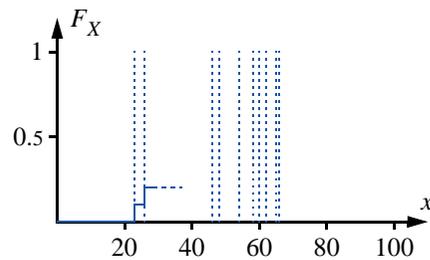


KTH Electrical Engineering

18

## EXAMPLE 9 - Estimated population

Solution:

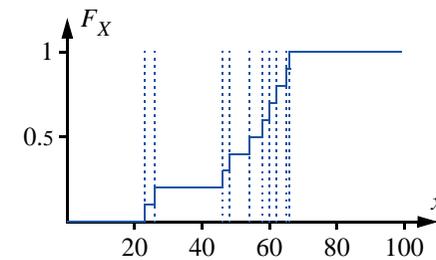


KTH Electrical Engineering

19

## EXAMPLE 9 - Estimated population

Solution (cont.)



KTH Electrical Engineering

20

## ESTIMATED APPROXIMATIVE FREQUENCY FUNCTION



KTH Electrical Engineering

- **Step 1.** Divide the  $x$ -axis in an arbitrary number of segments.
- **Step 2.** Count the number of samples  $x_i$  which belong to each interval,  $n_{seg}$ .
- **Step 3.** Draw a "stair-case" where each segment has a step, and where the height of each step is  $n_{seg}/n$ .

21

## EXAMPLE 10 - Approximative frequency function



KTH Electrical Engineering

Consider the same problem as in example 9.

**Solution:** Let us assume that the interesting range of outcomes for  $X$  is between 30 and 70.

a) This range is divided in 4 segments with a width of 10 each:

<i>seg</i>	1	2	3	4	5	6
Interval	$x \leq 30$	$30 < x, x \leq 40$	$40 < x, x \leq 50$	$50 < x, x \leq 60$	$60 < x, x \leq 70$	$70 < x$
$n_{seg}$	2	0	2	2	4	0

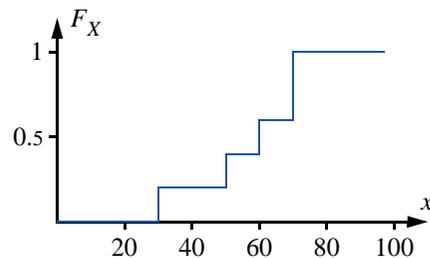
22

## EXAMPLE 10 - Approximative frequency function

Solution (cont.)



KTH Electrical Engineering



23

## ESTIMATED PROBABILITY DISTRIBUTION - Comparison



KTH Electrical Engineering

- An estimated distribution function based on a population will provide a smoother curve.
- The smoothness of an estimated distribution function based on an approximative frequency function will depend on the number of segments.
- All samples  $x_i, i = 1, \dots, n$ , must be stored to estimate the population  $\mathcal{X}$ , whereas it is only necessary to store all counters  $n_{seg}$  to approximate a frequency function.

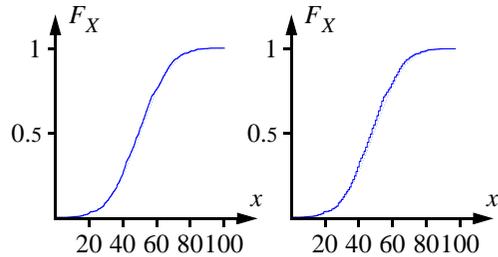
24

## ESTIMATED PROBABILITY DISTRIBUTION - Comparison

Example:  $X \in N(50, 15)$ ,  $n = 1\,000$ , 102 segments in approximative frequency function.



KTH Electrical Engineering



25

## ACCURACY

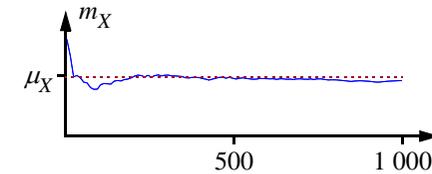
- A Monte Carlo simulation does not converge towards the true expectation value in the same sense as the geometric series

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots$$

is converging to 1.



KTH Electrical Engineering



26

## ACCURACY

- As we collect more samples, the *probability* of getting an inaccurate estimate is decreasing, but there is no guarantee that we get a better estimate if we generate another sample.
- It is more or less inevitable that the result of a Monte Carlo simulation is inaccurate—the *question is how inaccurate it is!*



KTH Electrical Engineering

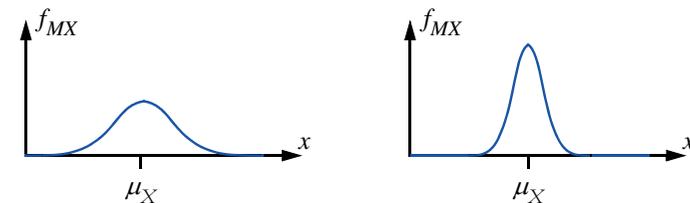
27

## ACCURACY

- Remember that  $M_X$  is a random variable. According to theorem 9 we have  $E[M_X] = E[X]$ .
- Hence,  $Var[M_X]$  is an indicator of the accuracy of the simulation method.



KTH Electrical Engineering



28

## ACCURACY

*Theorem 11:* In simple sampling, the variance of the estimated expectation value is

$$\text{Var}[M_X] = \frac{\text{Var}[X]}{n} \cdot \frac{N-n}{N}.$$

- The factor  $(N-n)/N$  is called **fpc** (*finite population correction*).
- For infinite populations we get

$$\text{Var}[M_X] = \frac{\text{Var}[X]}{n}. \quad (3)$$



KTH Electrical Engineering

29

## REPLACEMENT

- Sampling *without* replacement means that a sampled unit cannot appear again in the selection.
- Sampling *with* replacement means that a sampled unit can appear more than once in the selection.  
This is equivalent to having an infinite number of copies of each unit in the population; hence, the population is infinite.



KTH Electrical Engineering

30

## EXAMPLE 11 - Replacement

Consider an urn with two red balls and two green balls.



- Sampling without replacement  $\Rightarrow$  pick up a ball, check its colour, put the ball aside.
- Sampling with replacement  $\Rightarrow$  pick up a ball, check its colour, put the ball back in the urn.



KTH Electrical Engineering

31

## EXAMPLE 11 - Replacement

Let  $R$  be equal to 1 if a red ball has been selected and 0 if the ball was green.

$E[R]$  = share of red balls; in this case 0.5.

Study the probability distribution of  $M_R$  if one to four samples are drawn with and without replacement respectively.



KTH Electrical Engineering

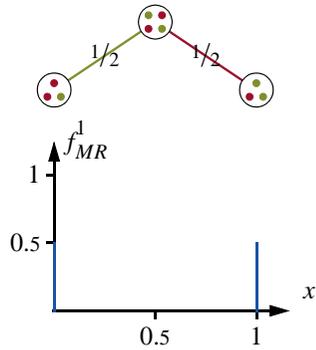
32

## EXAMPLE 11 - Replacement

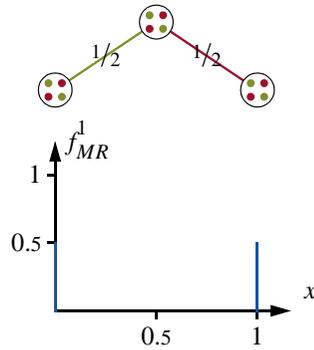
Frequency function for  $m_R$  after 1 trial:



Without replacement



With replacement



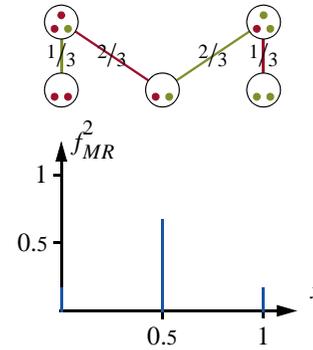
33

## EXAMPLE 11 - Replacement

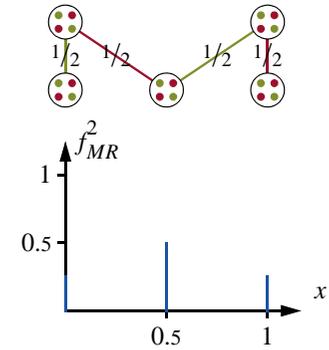
Frequency function for  $M_R$  after 2 trials:



Without replacement



With replacement



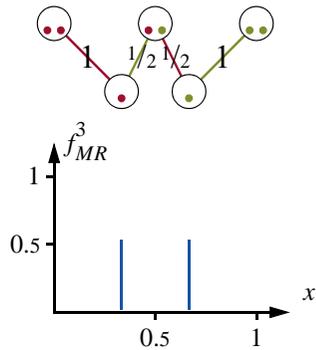
34

## EXAMPLE 11 - Replacement

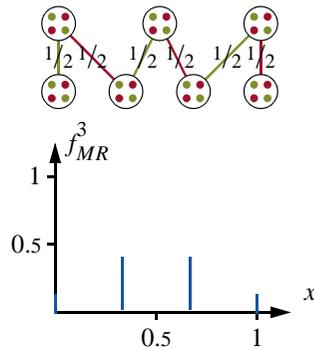
Frequency function for  $M_R$  after 3 trials:



Without replacement



With replacement



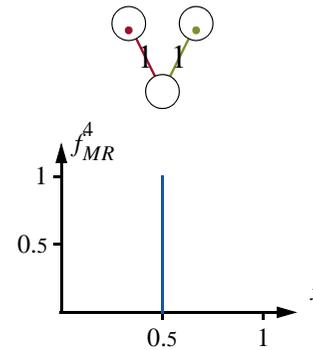
35

## EXAMPLE 11 - Replacement

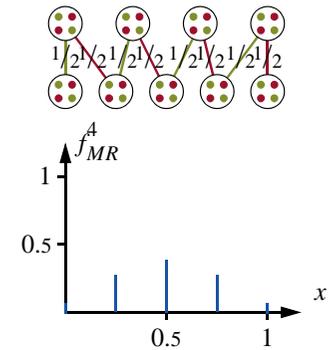
Frequency function for  $M_R$  after 4 trials:



Without replacement



With replacement



36

## COMPUTER SIMULATION WITHOUT REPLACEMENT



KTH Electrical Engineering

- Apparently, we will get better estimates if we use sampling without replacement.
- However, there are some practical problems to sample without replacement in a computer simulation: Each time we generate a value of the inputs,  $Y$ , the probability distribution  $F_Y(x)$  must be recalculated to exclude  $y_i$ .

37

## COMPUTER SIMULATION WITHOUT REPLACEMENT



KTH Electrical Engineering

- An alternative approach to exclude input values that have already been sampled would be to store all values of  $y_i$ . A new scenario,  $y_j$ , is then only analysed if  $y_j \neq y_i \forall i = 1, \dots, j-1$ . If that is not the case, the scenario is rejected and new values for  $y_j$  are generated.
- Is the reduction of  $Var[M_X]$  worth the extra work to achieve sampling without replacement, or is it better to simply collect some more samples?

38

## COMPUTER SIMULATION WITHOUT REPLACEMENT



KTH Electrical Engineering

In the continuation of this course, we will assume that sampling with replacement is used, and ignore the fpc terms in the formulae!

39

## ACCURACY



KTH Electrical Engineering

- Before a simulation is started, we should decide which accuracy is necessary, i.e., to define a **tolerance** level for the simulation.
- During the simulation, we must have some possibility to see if the result is within the tolerance level, i.e., to define a **stopping rule** for the simulation.
- After the simulation, we want to know the size of the random error, i.e., to determine the **confidence interval** of the simulation.

40

## TOLERANCE



KTH Electrical Engineering

- The tolerance towards estimation errors is of course depending on the application.
- As the true value is unknown, it is inevitable that we can only estimate the uncertainty of the Monte Carlo simulation results. In other words, we have to accept that there always is a risk that the result is outside the tolerance limits.
- The desired accuracy can either be expressed as an absolute value, for example  $\pm 1$ , or as a relative error, for example  $\pm 10\%$ .

41

## TOLERANCE - Duogeneous populations



KTH Electrical Engineering

*Definition 10:* A duogeneous population consists of a majority of homogeneous units (referred to as **conformist** units) and a minority of **diverging** units.

- All conformist units have the same value,  $x_C$ .
- The diverging units may either be homogeneous (i.e., all diverging units have the same value,  $x_D$ ) or heterogeneous (i.e., different diverging units have different values).

42

## TOLERANCE - Duogeneous populations



KTH Electrical Engineering

- The diverging units can be of large importance to the statistical properties of the random variable  $X$ .
- The smaller the share of diverging units, the harder it is to estimate.
  - It takes more samples before we have obtained a sufficient number of diverging units.
- Incorrect estimates of the share of diverging units may cause large relative errors.

43

## EXAMPLE 12 - Duogeneous populations



KTH Electrical Engineering

Consider the random variables  $X_1$  and  $X_2$ , which have the frequency functions on the form:

$$f(x) = \begin{cases} 1-p & x = 0, \\ p & x = 1, \\ 0 & \text{all other } x, \end{cases}$$

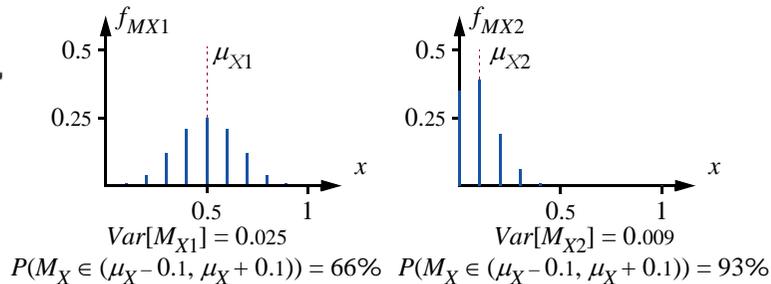
where  $p = 0.5$  for  $X_1$  and  $p = 0.1$  for  $X_2$ .

How accurate are the estimates  $M_{X_1}$  and  $M_{X_2}$  after 10 samples?

44

## EXAMPLE 12 - Duogeneous populations

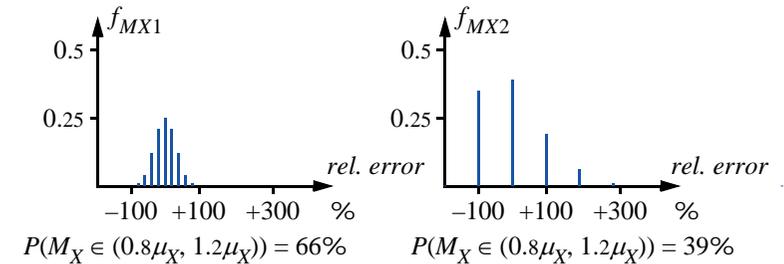
The frequency functions of  $m_{X1}$  and  $m_{X2}$  respectively looks like this after 10 samples:



45

## EXAMPLE 12 - Duogeneous populations

The frequency functions of the *relative errors* of  $M_{X1}$  and  $M_{X2}$  respectively looks like this after 10 samples:



46

## STOPPING RULE

A stopping rule is a test that can be performed after collecting  $n$  samples to decide if more samples should be collected or if the results are reasonably accurate.

Stopping rules can be designed arbitrarily. Here we will study three possible options:

- Fixed number of samples.
- Relative tolerance.
- Rejection of impossible results.



47

## FIXED NUMBER OF SAMPLES

This stopping rule simply means that we decide before the simulation how many samples we should collect.

The number of samples can be decided based on either **prior experience** of simulating the system or from **rough estimations**.



48

## EXAMPLE 13 - Estimated number of samples



KTH Electrical Engineering

It is known that the reliability of a system is approximately 99%, and Monte Carlo simulation is used to provide a more accurate estimate.

The acceptable tolerance of the estimate is that there should be 95% probability that  $|M_X - \mu_X| \leq 0.0005$ .

How many samples should be collected in the simulation?

49

## EXAMPLE 13 - Estimated number of samples



KTH Electrical Engineering

**Solution:** Assume that  $M_X$  is normally distributed\* with the mean  $\mu_X$  and the standard deviation  $\sigma_{mX}$ . The probability is 95% that  $M_X$  is in the interval  $\mu_X \pm 1.96\sigma_{mX}$ . Hence, the standard deviation of  $M_X$  may not be larger than  $0.0005/1.96 \approx 0.000255$ .

The standard deviation  $\sigma_{MX}$  can be calculated if we estimate the variance of  $X$ .

\* Which seems reasonable, considering that  $M_X$  is proportional to the sum of a large number of independent random variables

50

## EXAMPLE 13 - Estimated number of samples



KTH Electrical Engineering

**Solution (cont.)** Assuming the probability distribution

$$f_X(x) = \begin{cases} 0.01 & x = 0, \\ 0.99 & x = 1, \\ 0 & \text{all other } x, \end{cases}$$

we get

$$\text{Var}[X] = 0.99(1 - 0.99)^2 + 0.01(0.01 - 0.99)^2 \approx 0.0099.$$

51

## EXAMPLE 13 - Estimated number of samples



KTH Electrical Engineering

**Solution (cont.)** According to theorem 11 we get

$$\sigma_{MX}^2 \approx \frac{\text{Var}[X]}{n} \leq 0.000255^2$$

$\Rightarrow$

$$n \geq 152\,127.$$

To be on the safe side, we should round this results upwards and collect for example 153 000 samples.

52

## RELATIVE TOLERANCE

*Definition 11:* The coefficient of variation for the estimate  $m_X$  is given by

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

Using theorem 10 and theorem 11 we get the following estimate of  $a_X$ :

$$a_X \approx \frac{\sqrt{s_X^2/n}}{m_X} = \frac{s_X}{m_X \cdot \sqrt{n}}.$$



KTH Electrical Engineering

53

## RELATIVE TOLERANCE

- Before the simulation, we define an acceptable relative tolerance level,  $\rho_X$ .
- During the simulation we estimate the coefficient of variation; if it is less than the tolerance level then the simulation is stopped, otherwise more samples are needed.



KTH Electrical Engineering

54

## REJECTION OF IMPOSSIBLE RESULTS

- In some simulations an insufficient number of samples can result in estimates which are obviously incorrect.

**Example:** Duogeneous population where the estimate  $s_X^2$  is equal to 0, which implies that only conformist units have been sampled so far.

- Tests if the results are reasonable or not can be combined with other stopping rules.



KTH Electrical Engineering

55

## CONFIDENCE INTERVAL

*Definition 12:* A confidence interval is an interval which has some specific probability (**confidence level**) to include the true value.

- The confidence level can be chosen arbitrarily. Common choices are 95%, 99% or 99.9%.
- To calculate the confidence interval, we need to make an assumption about the probability distribution of  $M_X$ .



KTH Electrical Engineering

56

## CONFIDENCE INTERVAL



KTH Electrical Engineering

*Theorem 12:* Let  $t_\alpha$  be the value such that  $\Phi(t_\alpha) = 1 - \alpha/2$ . \* If  $M_X$  is normally distributed then  $m_X \pm t_\alpha \cdot s_X/\sqrt{n}$  is a confidence interval with confidence level  $1 - \alpha$ .

Table 3 Values of  $t_\alpha$

Confidence level, $1 - \alpha$	95%	99%	99.9%
$t_\alpha$	1.9600	2.5758	3.2905

\*  $\Phi(x)$  is the distribution function of a  $N(0, 1)$ -distributed random variable.

57

## CONFIDENCE INTERVAL



KTH Electrical Engineering

*Proof:* We are looking for a value  $\delta$  such that the probability of  $\mu_X$  is belonging to the interval  $m_X \pm \delta$  is equal to  $1 - \alpha$ , i.e.,

$$P(\mu_X \leq M_X + \delta) - P(\mu_X \leq M_X - \delta) = 1 - \alpha.$$

This equation can be rewritten as

$$P(M_X \leq \mu_X + \delta) - P(M_X \leq \mu_X - \delta) = 1 - \alpha.$$

Assuming that  $M_X$  is normally distributed, the mean is  $\mu_X$ , because  $E[M_X] = E[X]$  (theorem 9) and the standard distribution is  $s_X/\sqrt{n}$ , because  $\text{Var}[M_X] = \text{Var}[X]/n$  (theorem 11).

58

## CONFIDENCE INTERVAL



KTH Electrical Engineering

*Proof (cont.)* If  $X$  is  $N(\mu, \sigma)$ -distributed then

$$P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Hence we get

$$\Phi\left(\frac{\mu_X + \delta - \mu_X}{s_X/\sqrt{n}}\right) - \Phi\left(\frac{\mu_X - \delta - \mu_X}{s_X/\sqrt{n}}\right) = 1 - \alpha.$$

Now, due to the symmetry,  $\Phi(1 - x) = 1 - \Phi(x)$ ; hence,

$$2\Phi\left(\frac{\delta}{s_X/\sqrt{n}}\right) - 1 = 1 - \alpha \Rightarrow \Phi\left(\frac{\delta}{s_X/\sqrt{n}}\right) = 1 - \frac{\alpha}{2}.$$

59

## CONFIDENCE INTERVAL



KTH Electrical Engineering

*Proof (cont.)* As  $t_\alpha$  fulfils the condition  $\Phi(t_\alpha) = 1 - \alpha/2$  we get that

$$\frac{\delta}{s_X/\sqrt{n}} = t_\alpha.$$

This gives the confidence interval

$$m_X \pm \delta = m_X \pm t_\alpha \cdot s_X/\sqrt{n}. \blacksquare$$

60

## EXAMPLE 14 - Confidence interval



KTH Electrical Engineering

Assume that 1 000 scenarios of the system in example 6 have been generated, and that the following results have been obtained:

$$\sum_{i=1}^{1\ 000} x_i = 19\ 857, \quad \sum_{i=1}^{1\ 000} x_i^2 = 399\ 651.$$

Calculate a 95% confidence interval for  $E[X]$ !

61

## EXAMPLE 14 - Confidence interval



KTH Electrical Engineering

**Solution:** From the solution to example 7 we have  $m_X \approx 19.86$  and  $s_X \approx \sqrt{5.3}$ . Hence, we can conclude that there is a 95% probability that the expected least travel time is

$$m_X \pm \frac{t_\alpha \cdot s_X}{\sqrt{n}} = \frac{1.96 \cdot \sqrt{5.3}}{\sqrt{1\ 000}} \approx 19.86 \pm 0.14.$$

62

## SIMULATION PROCEDURE

It does not seem practical to test the stopping rule after each generated sample. Hence, the simulation can be divided in batches, where  $n_b$  samples are collected in each batch.

- **Step 1.** Generate the first batch of scenarios,  $y_i, i = 1, \dots, n_b$ .
- **Step 2.** Calculate  $x_i = g(y_i), i = 1, \dots, n_b$ .



KTH Electrical Engineering

63

## SIMULATION PROCEDURE

- Simple sampling

- **Step 3.** Calculate the sums  $\sum_{i=1}^n x_i$  and  $\sum_{i=1}^n x_i^2$

or store all samples  $x_i$  (depending on the objective of the simulation).

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



KTH Electrical Engineering

64

## EFFECTIVENESS



KTH Electrical Engineering

- If we have two methods to estimate the same expectation value, how do we determine which one is the more efficient?
- Assume method 1 provides the estimate  $M_{X1}$  and method 2 provides the estimate  $M_{X2}$ .

$$T_1 \cdot \text{Var}[M_{X1}] < T_2 \cdot \text{Var}[M_{X2}]^* \quad (4)$$

⇒ Method 1 is more effective than method 2.

\*  $T_i$  is the computation time of method  $i$ . In many cases,  $T_i$  can be assumed to be directly proportional to the number of generated samples,  $n_i$ .

65

## EFFECTIVENESS



KTH Electrical Engineering

- To use (4) we need to estimate  $\text{Var}[M_{X1}]$  and  $\text{Var}[M_{X2}]$ , but if a simulation method is inaccurate, we may get poor estimates.
- Another possibility is to use simple sampling to estimate the probability distributions of  $M_{X1}$  and  $M_{X2}$  respectively.

66

## EXAMPLE 15 - Effectiveness of ICC simulation

Data for the Ice-Cream Company:

Table 4 Ice-cream flavours

Flavour	Critical ingredients	Demand [litre/day]				
		0	50	100	125	200
Blueberry	Milk & Cream	25%	50%	25%		
Chocolate	Eggs, Milk & Cream		25%		50%	25%
Lemon	Eggs	25%	50%	25%		
Pineapple	None	25%	50%	25%		
Strawberry	Milk & Cream		25%		50%	25%
Vanilla	Eggs, Milk & Cream		25%		50%	25%



KTH Electrical Engineering

67

## EXAMPLE 15 - Effectiveness of ICC simulation

- Four small ice-cream machines (operation cost 15 €, capacity 100 litres) and two large ice-cream machines (operation cost 25 €, capacity 200 litres).
- The ingredient cost of ice-cream is 1 €/litre.
- The restaurants pay 3 €/litre.
- The availability of the ingredients are 95%.

Compare the true expectation values to the probability distribution of estimates from a simple sampling simulation using 200 samples.



KTH Electrical Engineering

68

## EXAMPLE 15 - Effectiveness of ICC simulation

Table 5 Results of ICC simulation in example 15.



KTH Electrical Engineering

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 sampling	829	872	909	0.00	0.27	1.00	0:02:25

69

## VARIANCE REDUCTION TECHNIQUES



KTH Electrical Engineering

- We have in many cases some knowledge about the system to be simulated.
- This knowledge can be used to improve the accuracy of the simulation.
- Methods based on some knowledge of the system are called **variance reduction techniques**, since improving the accuracy is equivalent to reducing  $Var[M_X]$ .

70