Network Analysis
IN2045
Evaluation of simulation and measurement results

Prof. Alexander von Bodisco
Stephan Günther
Dr. Heiko Niedermayer
Prof. Dr.-Ing. Georg Carle

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Note

- The slides may often say ***simulation***, yet the issues and methods also apply for ***measurements***, in particular if they come from observing variables over a certain set of
  - Times
  - Repetitions
  - Set-ups
  - …

Topics

- System Initialization
- Estimator
  - Consistent Estimator
  - Unbiased Estimator
  - Variance of an Estimator
  - Efficient Calculation of an Estimator
- Confidence Interval
  - Tschebyscheff Confidence Interval
  - Central Limit Theorem
  - t-Distribution Confidence Interval
- Evaluation of Simulation / Measurement Results
  - Replicate-Delete Method
  - Batch Means Method
  - Stationarity

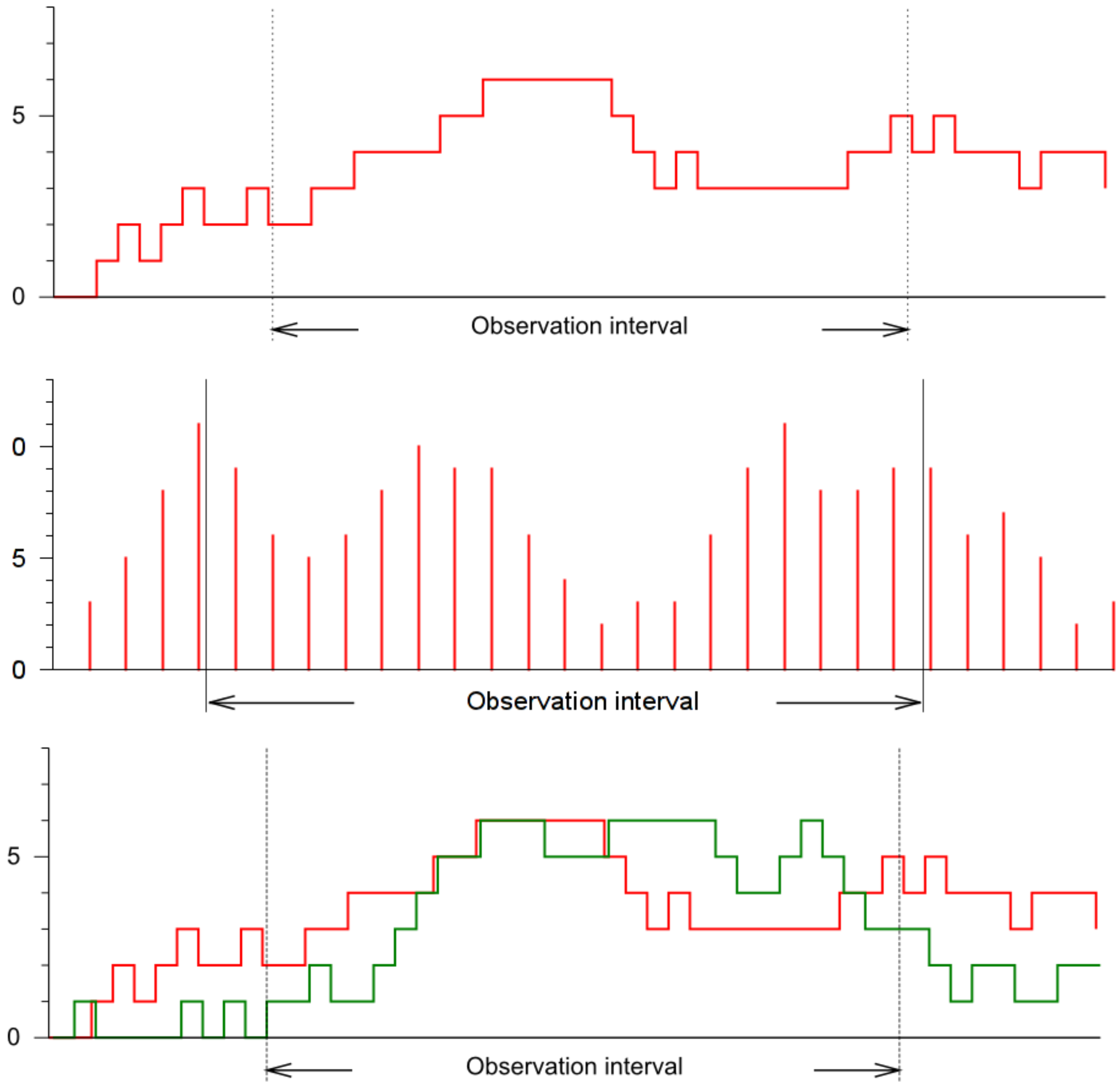
Evaluation of Simulation Results

Simulation study

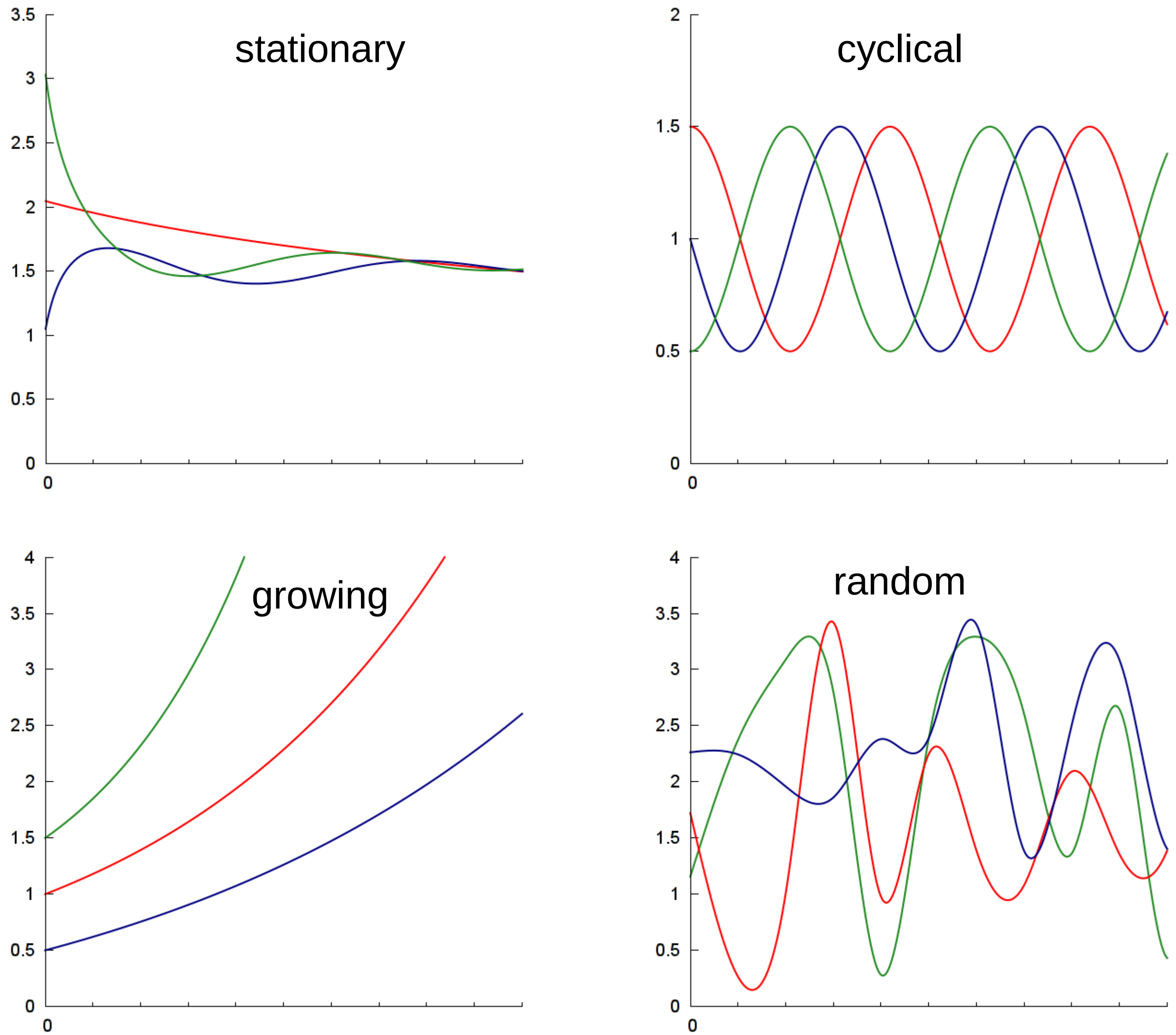
- Goals:
  - Evaluation of system S
  - Impact of (manageable) input variables C
  - Impact of (unmanageable) input variables U
  - Evaluation of outcome (result) P
  - Performance parameter Y

- Problem:
  - Infinite number of different outcomes
  - Probability of a certain outcome cannot determined in advance
  - Parameter of interest can be regarded as random variable

Evaluation of Simulation Results



Evaluation of Simulation Results



Evaluation of Simulation Results

- **Evaluation:**
  - n simulation runs
  - m samples per simulation run
  - j_{th} sample of the corresponding simulation run
  - i – number of simulation run

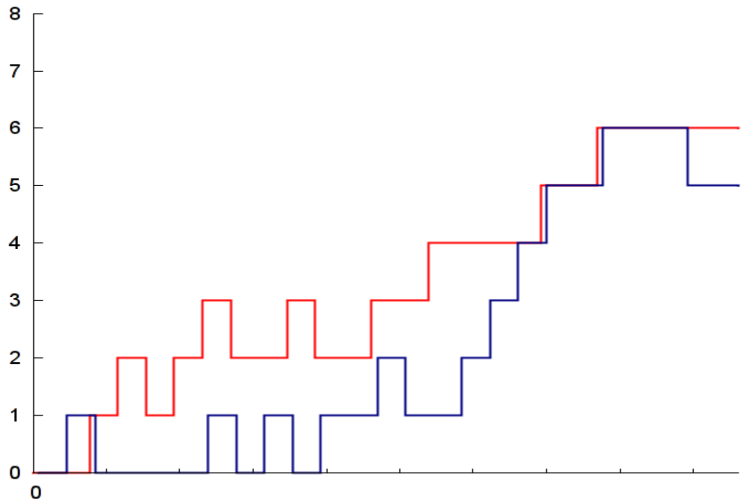
Trajectory ⇒

$$
\begin{matrix}
y_{11}, & \cdots & y_{1j}, & \cdots & y_{1m} \\
\vdots & & \vdots & & \\
y_{i1}, & \cdots & y_{ij}, & \cdots & y_{im} \\
\vdots & & \vdots & & \vdots \\
y_{n1}, & \cdots & y_{nj}, & \cdots & y_{nm}
\end{matrix}
$$

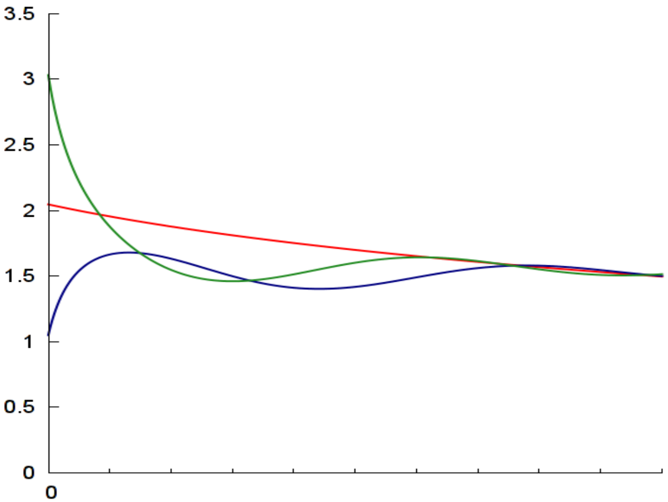
Simulation Matrix

Evaluation of Simulation Results

- **System Initialization**
  - Initialization state has to be chosen with respect to the expectation of the random variable
  - The transient phase of the system depends on the initialization state
  - The mean of the random variable (usually) converges to a certain level
  - Subsequent measurements are often correlated
    (e.g. waiting queue length)



Random variable Y for different simulation runs



Mean of random variable Y depending on the initialization state

Evaluation of Simulation Results

- **Estimator** (Schätzer)
  - Definition:

    An estimator is a statistic which is used to infer the value of an unknown parameter (estimand) in a statistic model.

  - Problem:
    - Estimate different characteristics (e.g. mean) of an observed parameter (e.g. delay or packet loss) with a small/certain number of samples.
    - Calculate the quality of the estimation

- **Consistent estimator** (konsistenter Schätzer)
  - Definition:

    An estimator is called consistent if its precision increases with the number of samples

$$\Rightarrow \quad \lim_{n\to\infty} P[|\tilde{Y}_j - E(Y_j)| > \varepsilon] = 0 \qquad \forall \varepsilon > 0$$

Evaluation of Simulation Results

- **Unbiased estimator** (erwartungstreuer Schätzer)
  - Definition:

    An estimator is called unbiased if its mean equals the true mean of the estimation parameter.

  - Example:
    - Assume a very large population of elements with a different characteristic (e.g. height of individuals) and μ being the mean of the characteristic
    - Let $E[\hat{Y}_i]$ be the mean of n collected sample values and $\overline{X}$ the random variable which consists of these mean values.

$$\Rightarrow \quad \overline{X} = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$$

$$\Rightarrow \quad E(\overline{X}) = E\left[\frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)\right] = \frac{1}{n} E\left[X_1 + X_2 + \ldots + X_n\right]$$

$$\Rightarrow \quad E(X_1) = E(X_2) = \ldots = E(X_n) = \mu$$

$$\Rightarrow \quad E(\overline{X}) = \mu$$

❑ Estimation of E(Y)

- Estimator $\tilde{Y}$
- Estimation (value) $\hat{Y}$
- Point Estimator of E(Y) $\hat{Y}_j = \frac{1}{n} \cdot \sum_{i=1}^{n} y_{ij}$
- Random Variable $\tilde{Y}_j$
- Outcome of $RV\ \tilde{Y}_j$ ⇨ $\hat{Y}_j$
- Consistency of $\tilde{Y}$ ⇨ $\lim_{n \to \infty} P[|\tilde{Y}_j - E(Y_j)| > \varepsilon] = 0 \qquad \forall \varepsilon > 0$

Quality of the estimator still unknown

Evaluation of Simulation Results

- Variance of the estimator $\tilde{Y}$ represents a first indicator of its quality
  - Calculation of the variance of the estimator after n samples

$$\sigma^2(\tilde{Y}_j) = E\left(\left(\tilde{Y}_j - E(\tilde{Y}_j)\right)^2\right)$$

$$= E\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_{ij} - E(Y_j)\right)^2\right)$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left((Y_{ij} - E(Y_j))^2\right) + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n} E\left((Y_{ij} - E(Y_j))\cdot(Y_{kj} - E(Y_j))\right)\right)$$

Double sum represents the correlated part which is 0 if y_{ij} are uncorrelated

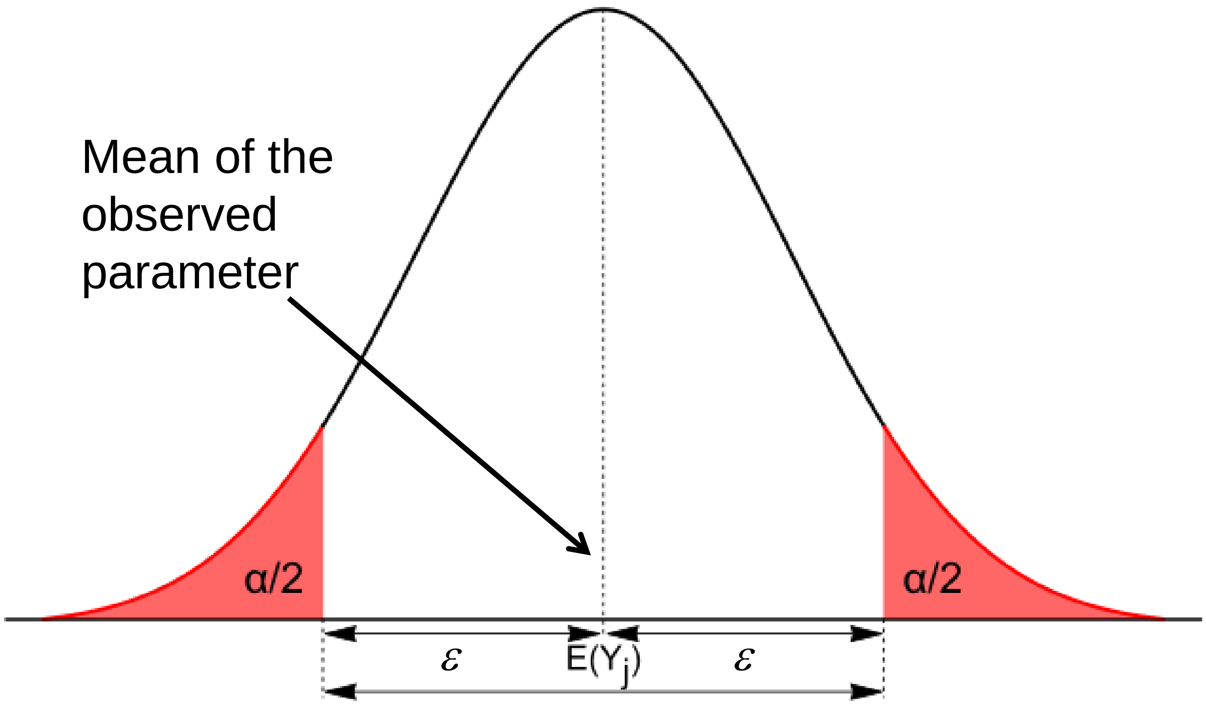
$$\sigma^2(\tilde{Y}_j) = \frac{1}{n^2}\sum_{i=1}^{n} E\left((Y_{ij} - E(Y_j))^2\right) = \frac{1}{n^2}\cdot n\cdot\sigma^2(Y_j) = \frac{\sigma^2(Y_j)}{n}$$

- Variances of the estimator increases
  - with the variance of the estimand
  - if the number of samples is reduced

Evaluation of Simulation Results

- **Unbiased estimator for** $\sigma^2(Y_j)$

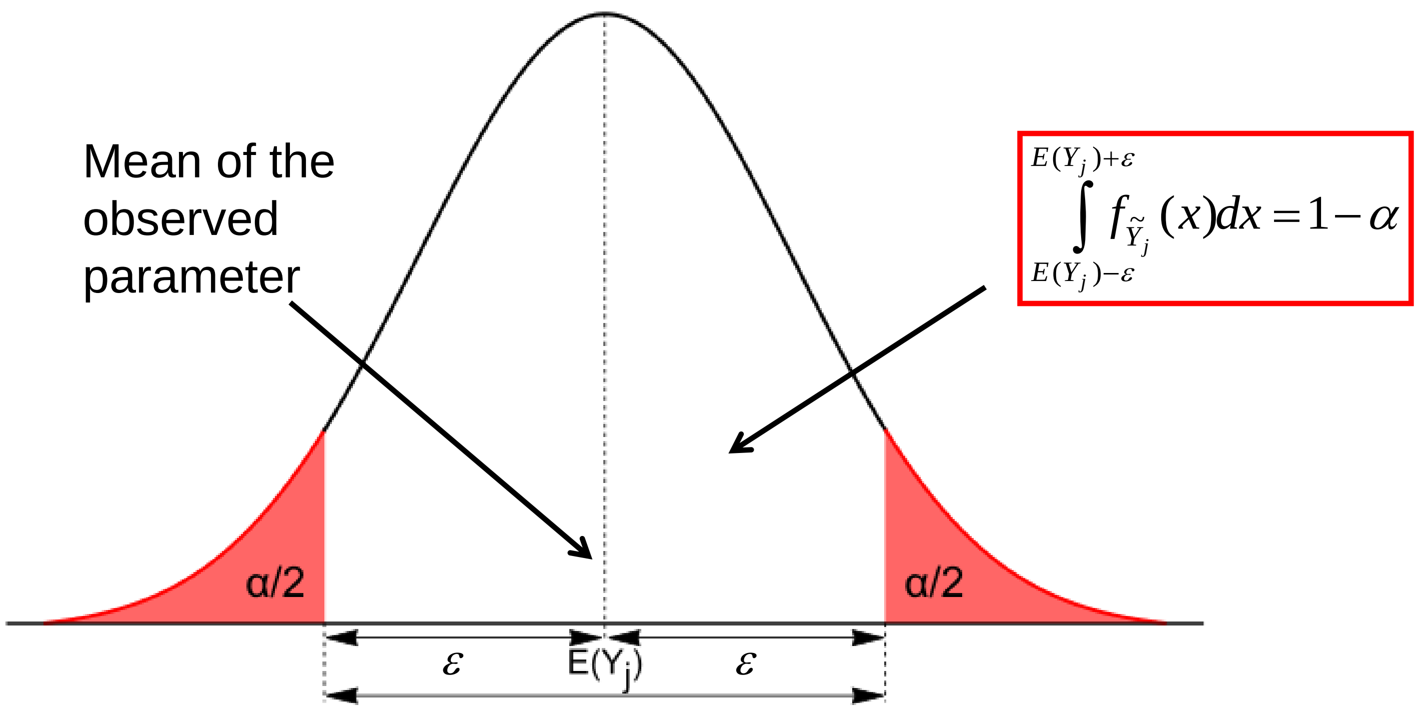
$$\tilde{S}_j^2(Y_j) = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_{ij} - \tilde{Y}_j\right)^2$$



Distribution of the estimator $\tilde{Y}_j$

Evaluation of Simulation Results

- Assume that the probability density function of the estimator $f_{\tilde{Y}_j}(y)$ is known in advance
  - The probability that $\tilde{Y}_j$ lies within the interval is $1-\alpha$



Distribution of the estimator $\tilde{Y}_j$

Evaluation of Simulation Results

- **Biased estimator for the sample variance** $\sigma^2(Y_j)$

$$\tilde{S}_{nj}^2(Y_j) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_{ij} - \tilde{Y}_j\right)^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \frac{\left(\sum_{i=1}^{n} x_i^2\right)}{n^2}$$

$\tilde{S}_{nj}^2$ is a biased estimator of the sample variance since it systematically underestimates it.

- **Bessel's correction**

The biased estimator can be transformed in an unbiased estimator of the sample variance by applying Bessels's correction.

$$\frac{n}{n-1}$$

- **Unbiased estimator for the sample variance** $\sigma^2(Y_j)$

$$\tilde{S}_j^2(Y_j) = \frac{n}{n-1}\cdot\tilde{S}_{nj}^2(Y_j) = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_{ij} - \tilde{Y}_j\right)^2$$

Evaluation of Simulation Results

Efficient calculation of the estimator

- 1. Problem

  Number of samples y_{ij} can become high which results in high memory consumption

  - Solution

    Recursion $\hat{Y}_j(k) = \frac{k-1}{k}\hat{Y}_j(k-1) + \frac{y_{kj}}{k}$

- 2. Problem

  The calculation of the sample variance requires the direct evaluation of the estimation of the variance

  $$\hat{S}_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{ij} - \hat{Y}_j\right)^2$$

  ⇒ Every sample y_{ij} has to be stored

  - Solution
    - Store the sums of y_{ij} and y_{ij}^2
    - Calculate $\hat{S}_j^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(y_{ij}\right)^2 - n \cdot \left(\hat{Y}_j\right)^2\right)$

The size of n can be reduced if means are used instead of single sample values.

Evaluation of Simulation Results

Confidence interval

- Definition: $P[|\tilde{Y}_j - E(Y_j)| \geq \varepsilon] = \alpha$

  Calculate an interval 2ε around the $E(Y_j)$ such that a sample of $\tilde{Y}_j$ lies in the interval with a probability of $1-\alpha$

  Larger α ⇒ Smaller ε

  Smaller α ⇒ Larger ε

- $\hat{Y}_j \pm \varepsilon$ is called $(1-\alpha)\cdot 100\%$ confidence interval of $E(Y_j)$
- $[\hat{Y}_j - \varepsilon, \hat{Y}_j + \varepsilon]$ is called interval estimator

⇒ Interval estimators are more important than point estimators since they make probability based assumptions which consider the variance of the estimator.

Evaluation of Simulation Results

Confidence interval according to Tschebyscheff

- Let X be a random variable with mean $E(X)$ and variance $\sigma^2(X)$

$$P[|X - E(X)| \geq c] \leq \frac{\sigma^2(X)}{c^2} \qquad \forall c > 0$$

- Replace c by ε and X by $\tilde{Y}_j$. Variance of $\tilde{Y}_j$ is $\frac{\sigma^2(Y_j)}{n}$

$$\Rightarrow \quad P[|\tilde{Y}_j - E(Y_j)| \geq c] \leq \frac{\sigma^2(X)}{n \cdot \varepsilon^2}$$

Evaluation of Simulation Results

Confidence interval according to Tschebyscheff

- Example:
  - Calculate the 90% confidence interval of

$$\Rightarrow \quad \alpha = \frac{\sigma^2(Y_j)}{n \cdot \varepsilon^2} = 0.1$$

  - Now assume a sample size n = 10

$$\Rightarrow \quad \frac{\sigma^2(Y_j)}{10 \cdot \varepsilon^2} = 0.1 \quad \Leftrightarrow \quad \varepsilon^2 = \sigma^2(Y_j) \quad \Leftrightarrow \quad \varepsilon = \sigma(Y_j)$$

  - Thus, a confidence interval of half size requires four times the number of samples

$$\Rightarrow \quad \frac{\sigma^2(Y)}{n \cdot \varepsilon^2} = \frac{\sigma^2(Y)}{m \cdot (\varepsilon/2)^2} \quad \Leftrightarrow \quad m = 4 \cdot n$$

Evaluation of Simulation Results

Confidence interval according to Tschebyscheff

- Disadvantages:
  - The calculation requires knowledge of the variance $\sigma^2(Y)$ of the estimand which is typically unknown and must thus be replaced by the estimator $\hat{S}_j^2$.

Note that this breaks the pre-condition of Tschebyscheff which makes the calculated bounds invalid.

  - Tschebyscheff intervals are very large / pessimistic since they are valid for any given distribution.

The pessimistic characteristic of Tschebyscheff is often used as justification for replacing the variance $\sigma^2(Y)$ with the estimator $\hat{S}_j^2$.

Evaluation of Simulation Results

❑ Confidence interval according to Tschebyscheff

- Example:
  
  - Flipping a coin. RV Y => Y e {0-head, 1-tail}
  - $E(Y) = 0.5$, $\sigma^2(Y) = 0.25$
  - Flipping the coin 10 times after another ⇒ n = 10 (samples)
  - Calculate 90% confidence interval ⇒ $\alpha = 0.1$
  - Experiment 1: 0000101001

    ⇒ $\hat{Y} = 0.3 \quad \wedge \quad \hat{S}^2 = 0.23333$

    ⇒ [-0.183, 0.783] ⇒ [0, 0.783]
  - Experiment 2: 0110111001

    ⇒ $\hat{Y} = 0.6 \quad \wedge \quad \hat{S}^2 = 0.26667$

    ⇒ [0.084, 1.116] ⇒ [0.084, 1.000]

Evaluation of Simulation Results

Confidence interval according to Tschebyscheff

- Example:
  
  - Flipping a coin. RV Y => $Y \in$ {0-head, 1-tail}
  - $E(Y) = 0.5$, $\sigma^2(Y) = 0.25$
  - Flipping the coin 20 times after another ⇒ n = **20** (samples)
  - Calculate 90% confidence interval ⇒ $\alpha = 0.1$
  - Concatenation of Experiment 1 and 2: 00001 01001 01101 11001

    ⇒ $\hat{Y} = 0.3 \quad \wedge \quad \hat{S}^2 = 0.23333$

    ⇒ [0.089, 0.811] ⇒

  - Summary:
    - The true mean lies within the interval with a probability of 90%
    - 1 of 10 experiments will not contain the true mean

Evaluation of Simulation Results

Central limit theorem

The distribution of the (normalized and centralized) sum of a large number of independent and identical distributed random variables can be approximated by the (standard) normal distribution.

Lindeberg-Lévy theorem

Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables within the same probability space which are independent and follow the same distribution. The mean of each random variable is μ and the standard variation is σ.

In the following we take a closer look at the nth sum of the sequence.

$$S_n = X_1 + X_2 + \ldots + X_n \Rightarrow E[S_n] = n\mu \Rightarrow \sigma^2[S_n] = n\sigma$$

Introduce a new standardized random variable Z_n

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Evaluation of Simulation Results

- **Central limit theorem**

  The distribution of the random variable Z_n converges against the (standard) normal distribution according to the central limit theorem if the number of summands n increases.

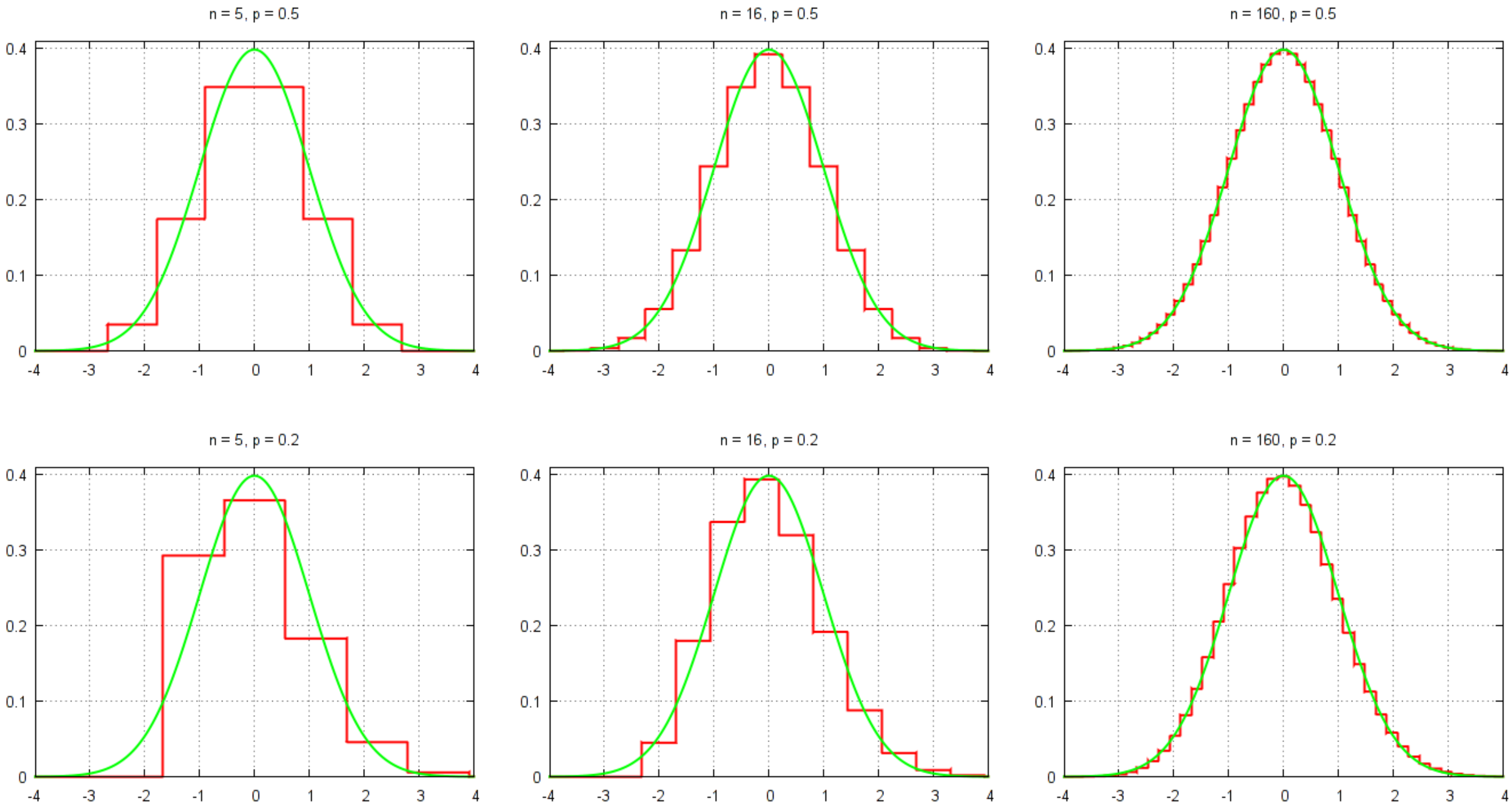
$$\Rightarrow \quad \lim_{n\to\infty} P(Z_n \le z) = \Phi(z) \qquad z \in R$$

With $\Phi(z)$ representing the (standard) normal distribution $N(0,1)$

$$\Rightarrow \quad \lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \le z\right) = \Phi(z)$$

$$\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$$

- Central limit theorem



Sum of binomial distributed random variables

Confidence interval according to the central limit theorem

- Idea: The central limit theorem is still valid if σ^2 is replaced by $\tilde{S}^2$. Thus, it is possible to calculate the critical values out of the normal distribution.

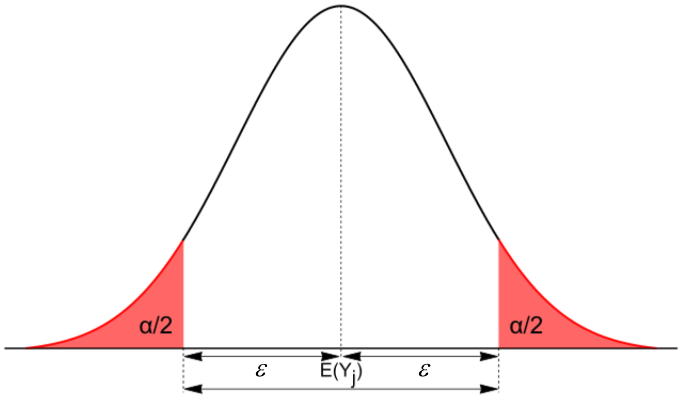
- Recapitulate the “flipping of a coin example” with $\tilde{Y}$ representing the distribution of the estimator and Y being the distribution of the estimand. Then we can calculate the confidence interval as follows:

$$\Rightarrow \quad P[|Z| \geq \varepsilon] = P\left[\left|\frac{\tilde{Y} - E(Y)}{\tilde{S}/\sqrt{n}}\right| \geq \varepsilon\right] = \alpha$$

$$\Rightarrow \quad P[|\tilde{Y} - E(Y)| \geq \varepsilon \cdot \tilde{S}/\sqrt{n}] = \alpha$$

$$\Rightarrow \quad \boxed{\tilde{Y} \pm z_\alpha \cdot \tilde{S}/\sqrt{n}}$$

$\Rightarrow$ z_α is the $\alpha/2$ percentile of $N(0,1)$



Evaluation of Simulation Results

Confidence interval according to the central limit theorem

The central limit theorem generates smaller confidence intervals

⇨ Tschebyscheff [0.089, 0.811]

⇨ Central Limit [0.262, 0.638]

- Question: What is the minimum value for n to allow the assumption that the estimator is normally distributed?
  - The minimum value of n depends on the distribution of the estimand Y. In worst case scenarios, the mean of the estimand $E[Y]$ may be outside the confidence interval with a probability which is significantly higher than α.

Evaluation of Simulation Results

- **Empirical evaluation of the confidence interval calculation according to the (student) t-distribution**

  - Problem:
    Only a view results for different distributions are known.

  - In the following we assume that Y is already normal distributed

$$\Rightarrow \quad \frac{\tilde{Y} - E[Y]}{\tilde{S} / \sqrt{n}}$$ follows a t-distribution with n-1 degrees of freedom

  - Critical / popular values of the t-distribution can be taken from tables

$$\Rightarrow \quad \hat{Y} \pm t_{n-1,1-\alpha/2} \cdot \frac{\widehat{S}}{\sqrt{n}}$$

Evaluation of Simulation Results

- **Empirical evaluation of the confidence interval calculation according to the (student) t-distribution**

  - Idea:

    Apply a known distribution and calculate the confidence interval. Then repeat the experiment k times and estimate the probability with which the outcome of the experiment remains within the calculated boundaries.

  - Example: 90% confidence interval, k = 500 repetitions

Verteilung	n=5	n=10	n=20	n=40
Normal	0.910	0.902	0.898	0.900
Exponential	0.854	0.878	0.870	0.890
Lognormal	0.758	0.768	0.842	0.852
Hyperexp.	0.584	0.586	0.682	0.774

Table taken from Law: Simulation Modeling and Analysis, 4th Edition

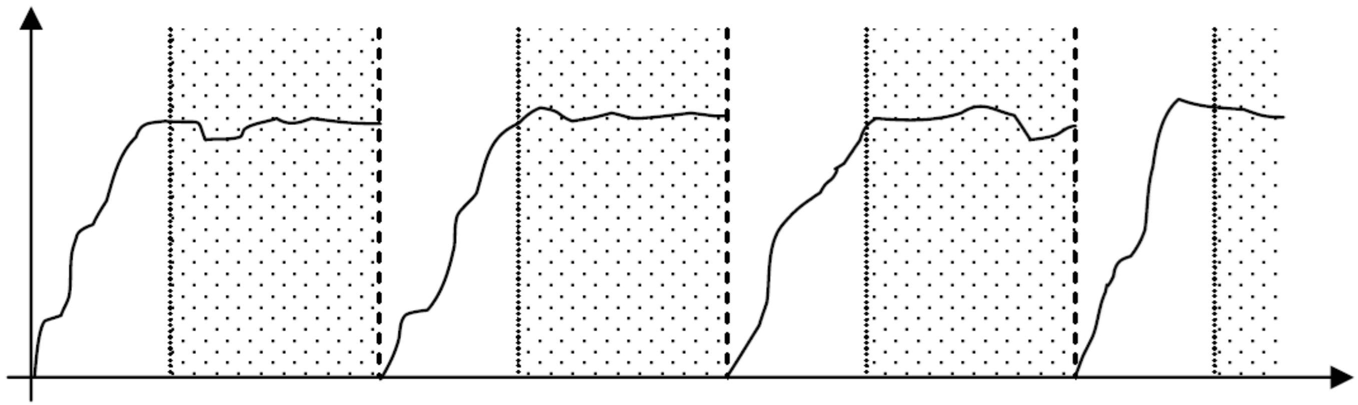
Evaluation of Simulation Results

How to get useful simulation results out of a simulation (measurement)

Evaluation of Simulation Results

- **Replicate-Delete Method (LK 9.5.2)**
  - Estimate the duration of the transient phase
  - Replicate – Simulate a large number of runs
  - Delete – Remove the transient phase since it does not contain meaningful results
  - The duration of the simulation has to be a much longer than the duration of the transient phase

  - Calculate the confidence intervals by using the mean values of the individual simulation runs



Evaluation of Simulation Results

Replicate-Delete Method

- Advantage:
  - Most simple approach
  - Less affected by correlation
  - Typically supported by all simulation tools
- Disadvantage:
  - Requires correct estimation of the duration of the transient phase.
  - Underestimation of the duration of transient phase results in falsified simulation results.
  - Requires more time compared to Batch-Means since the transient phase has to be simulated several times.

Covariance

Covariance is a measure which describes how two variables change together

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X] \cdot E[Y]$$

- Special Case: $Cov(X,X) = VAR[X]$
- Other Characteristics:
  - $Cov(X,a) = 0$
  - $Cov(X,Y) = Cov(Y,X)$
  - $Cov(aX,bY) = abCov(X,Y)$
  - $Cov(X+a,Y+b) = Cov(X,Y)$

Correlation function

Correlation function describes how two random variable tend to derivate from their expectation

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{VAR(X) \cdot VAR(Y)}}$$

- Characteristics:
  - $Y = X$ ⟹ $Cor(X,Y) = 1$ (Maximum positive)
  - $Y = -X$ ⟹ $Cor(X,Y) = -1$ (Maximum negative)
  - $Cor(X,Y) > 0$ Both random variable tend to have either high or low values (difference to their expectation)
  - $Cor(X,Y) < 0$ The random variables differ from each other such that one has high values while the other has low values and vice versa (difference to their expectation)

Autocorrelation (LK 4.9)

- Autocorrelation is the cross-correlation of a signal with itself. In the context of statistics it represents a metric for the similarity between observations of a stochastic process. From a mathematical point of view, autocorrelation can be regarded as a tool for finding repeating patterns of a stochastic process.

Definition:

- Correlation of two samples with distance k from a stochastic process X is given by:

$\Rightarrow$ $Cor(X,Y)$ with $Y_i = X_{i+j}$

Use case:

- Test of random number generators (remember spectral test)
- Evaluation of simulation results (c.f. Batch-Means)

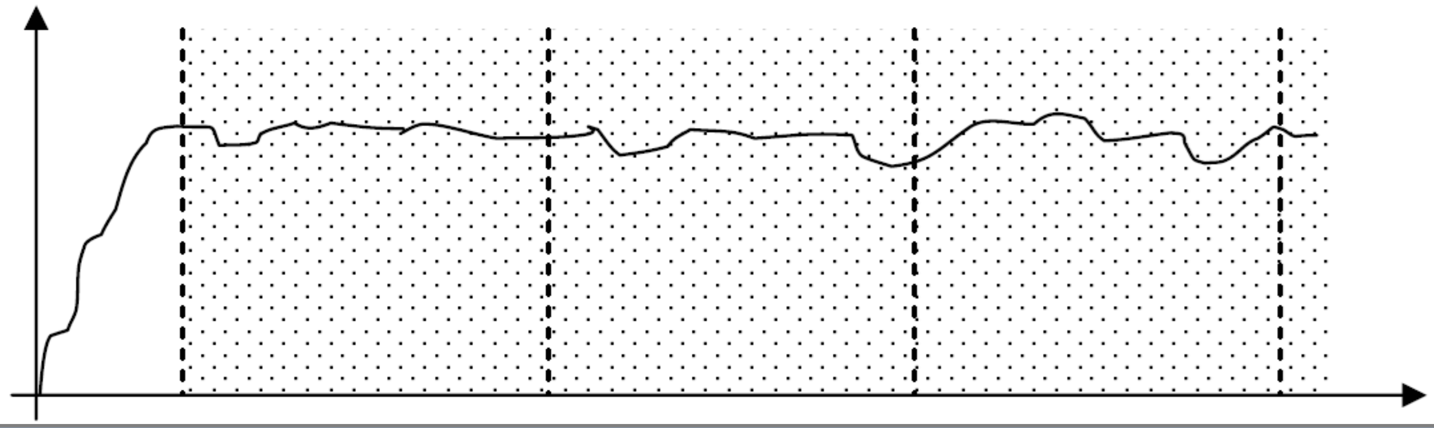
Evaluation of Simulation Results

Batch-Means Method (LK 9.5.3)

- Estimate the duration of the transient phase
- Perform a long simulation run
- Remove the transient phase
- Divide the gathered results in n intervals of equal length (Batches) which hold m samples

⇒ Assure that the mean of subsequent batches is uncorrelated (calculate the empirical autocorrelation)

⇒ Number of batches $n \geq 10$ Batch size $m \geq 10 \cdot x$



Evaluation of Simulation Results

Batch-Means Method

- Calculate the confidence intervals by using the mean values of the batches
- Minimize the absolute and relative error by increasing the number of batches (longer simulation run)

Optional approach:

- Estimate the duration of the transient phase
- Choose a sufficient value for m
- Simulate until the confidence interval has the desired size

Evaluation of Simulation Results

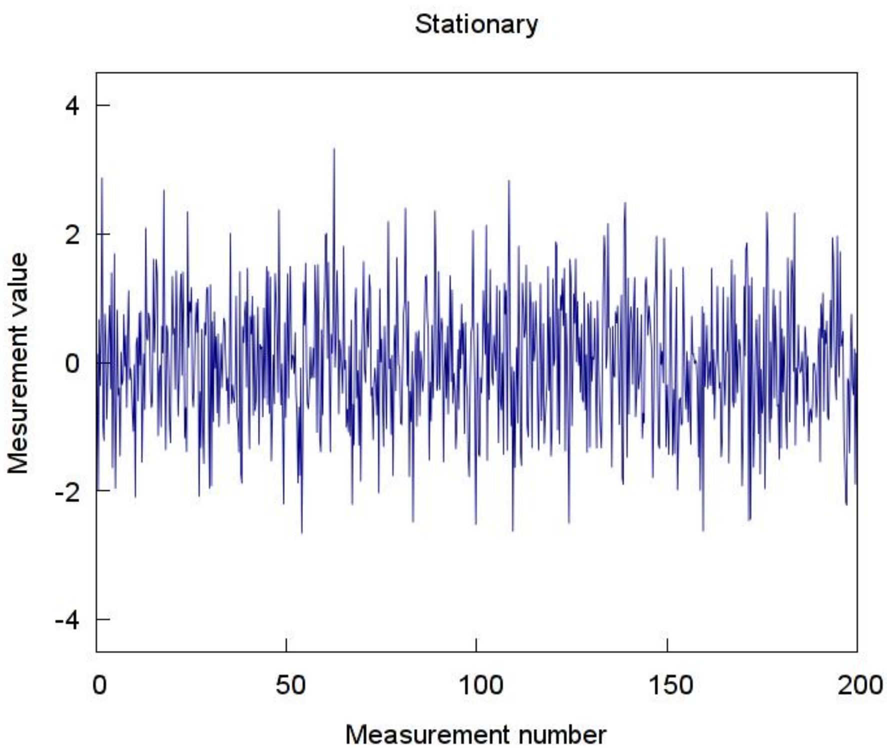
Batch-Means Method

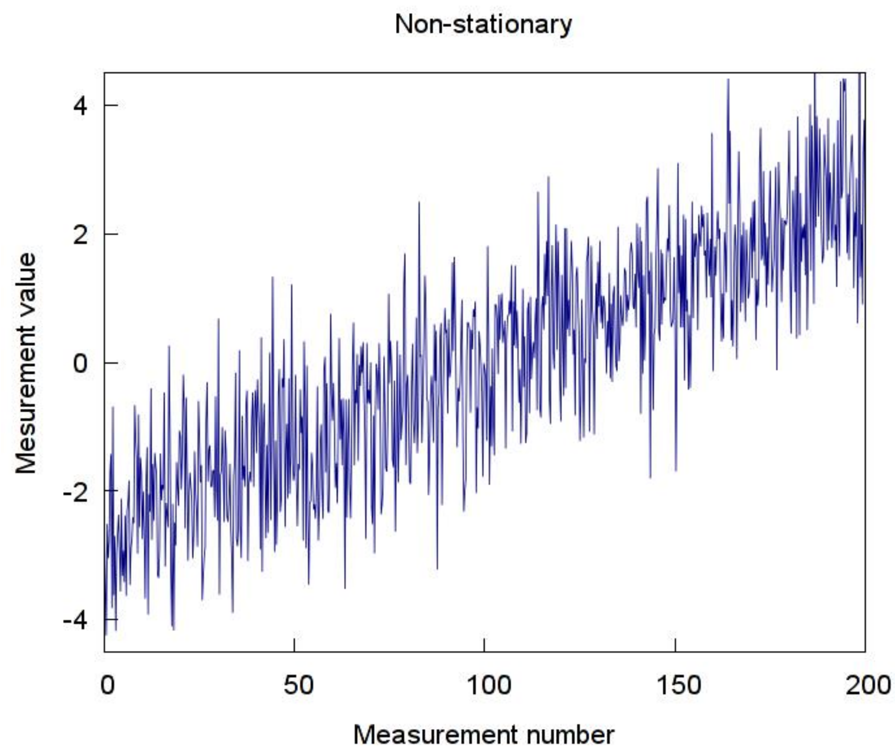
- Advantage:
  - Minimizes the time to get meaningful results since only a single transient phase has to be simulated
  - Errors of the estimation of the duration of the transient phase decrease with increasing number of batches

- Disadvantage:
  - Calculation of n and m is complicated and usually requires detailed knowledge of the simulation
  - Calculation of the autocorrelation of the intervals have to be calculated in order to assure that the corresponding means are not correlated

Stationarity

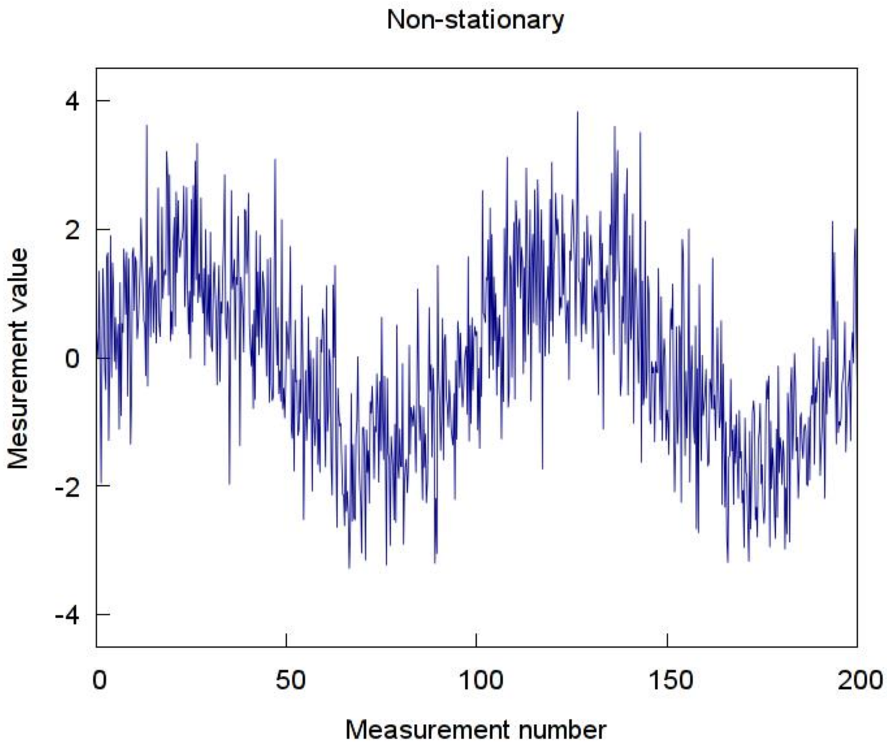
- What’s stationarity? – An intuitive graphical explanation:

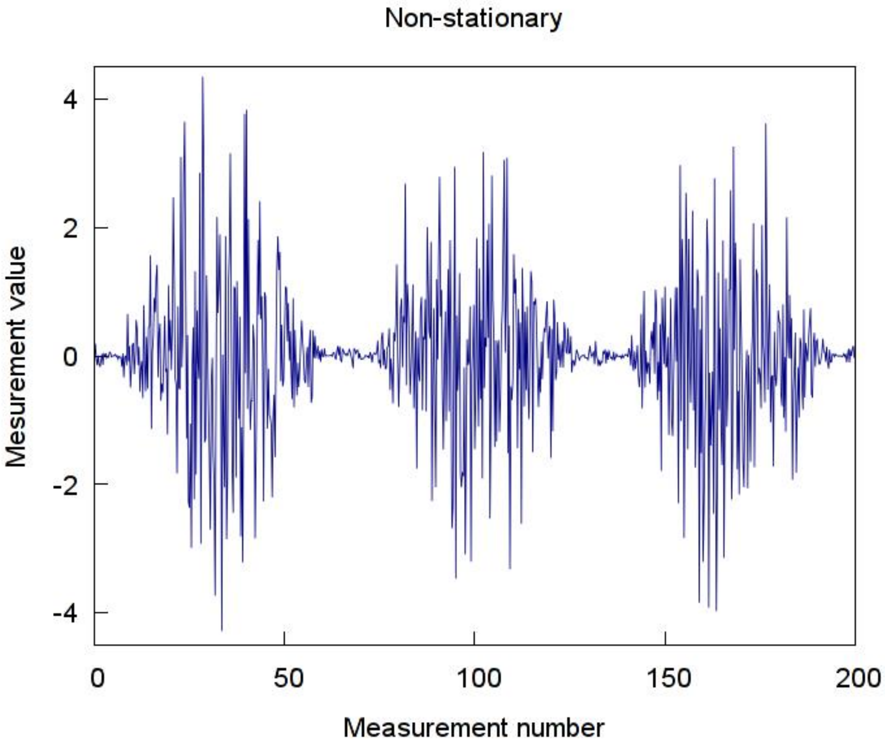




Stationarity

- Its not just trends





Stationarity

- Non-stationarity
  - May happen if you observe a phenomenon on the Internet.
    - Number of Attacks seen by IDS
    - Average Number of Facebook friends of a Facebook user
    - Currently popular movie
    - Changing network graphs due to some process of adding and deleting edges
    - Measurement time exceeds the time where assumptions hold
    - …
  - Non-stationarity may be overcome when changing the observed variables relevant for your actual question (given that the non-stationary variable is not your main objective)

Mathematical definitions of stationarity

- Strong stationarity:
  - All samples Xi are drawn from exactly the same underlying distribution
  - In practice, this is hard or impossible to prove
- Other types of stationarity:
  - Mean stationary: $\mu(X_i) = \text{const}\ \forall i$
  - Variance stationary: $\mathrm{Var}(X_i) = \text{const}\ \forall i$
  - Covariance stationary: $\mathrm{Cov}(X_i, X_{i+k}) = \text{const}(k)\ \forall i$ (only dependent on lag)
  - Weakly stationary: The Xi are mean stationary and covariance stationary
  - In practice, weak stationarity is most commonly used

Arithmetics

Assume Xi and Yi are [weakly] stationary processes. Then…:

- You can shift a stationary process:
  α + Xi is stationary, too
- You can scale a stationary process:
  $\beta \cdot$ Xi is stationary, too
- You can add stationary processes together:
  Xi + Yi is stationary, too

Why…?

- Important term in statistics
  - Many methods, algorithms, mechanisms assume that all samples come from the same distribution
    - Warning: We experience phenomena such as convergence phases at the beginning of simulations, etc. – this means it's not stationary [yet]!
    - Often would need strong stationarity, but often weak can do the trick
  - May be interesting to analyse if the output of a simulator / experiment / … is [weakly] stationary or not
- How to test for [weak] stationarity?
  - Tests usually built into statistics packages
  - Parametric tests for stationarity
    - Make assumptions about underlying data (e.g., normally distributed)
  - Nonparametric tests for stationarity
    - Need more measurements (usually 5%–35% more samples)

Example

- Calculating confidence intervals
  - Assumption: All samples are drawn from the same population
  - But what if you take measurements from a process that has not converged yet?
  - Solution: Check the time series of measurements for stationarity