Discrete Event Simulation

IN2045

Dr. Alexander Klein
Dr. Nils Kammenhuber
Prof. Dr.-Ing Georg Carle

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

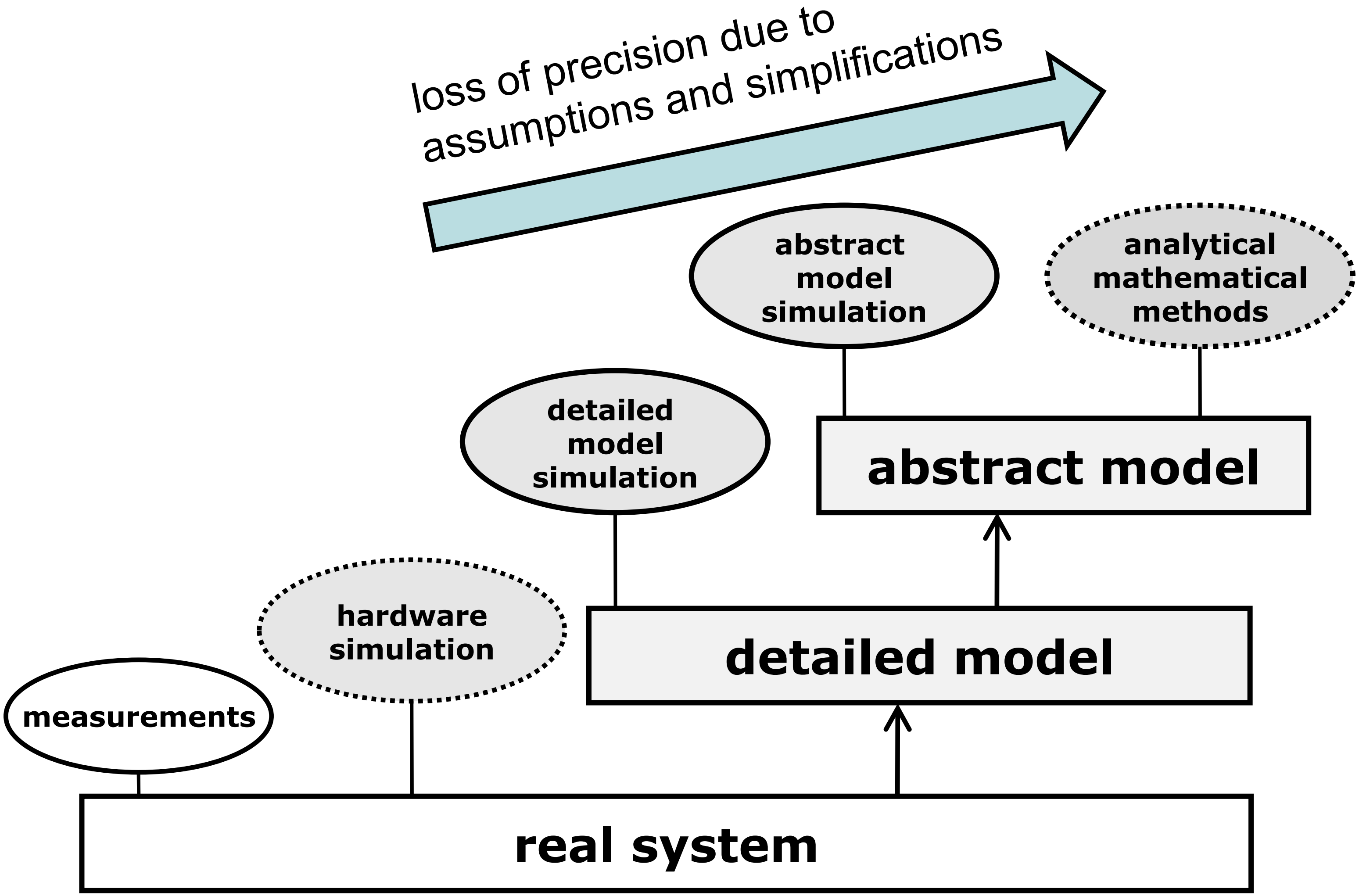
Topics

- Validation of Models:
  - Calibration
- Comparison of different systems
- Model Modifications:
  - Structural Change
  - Parameter Change
- Overfitting
- Comparison of Confidence Intervals:
  - Welch
  - Law and Kelton
  - Examples

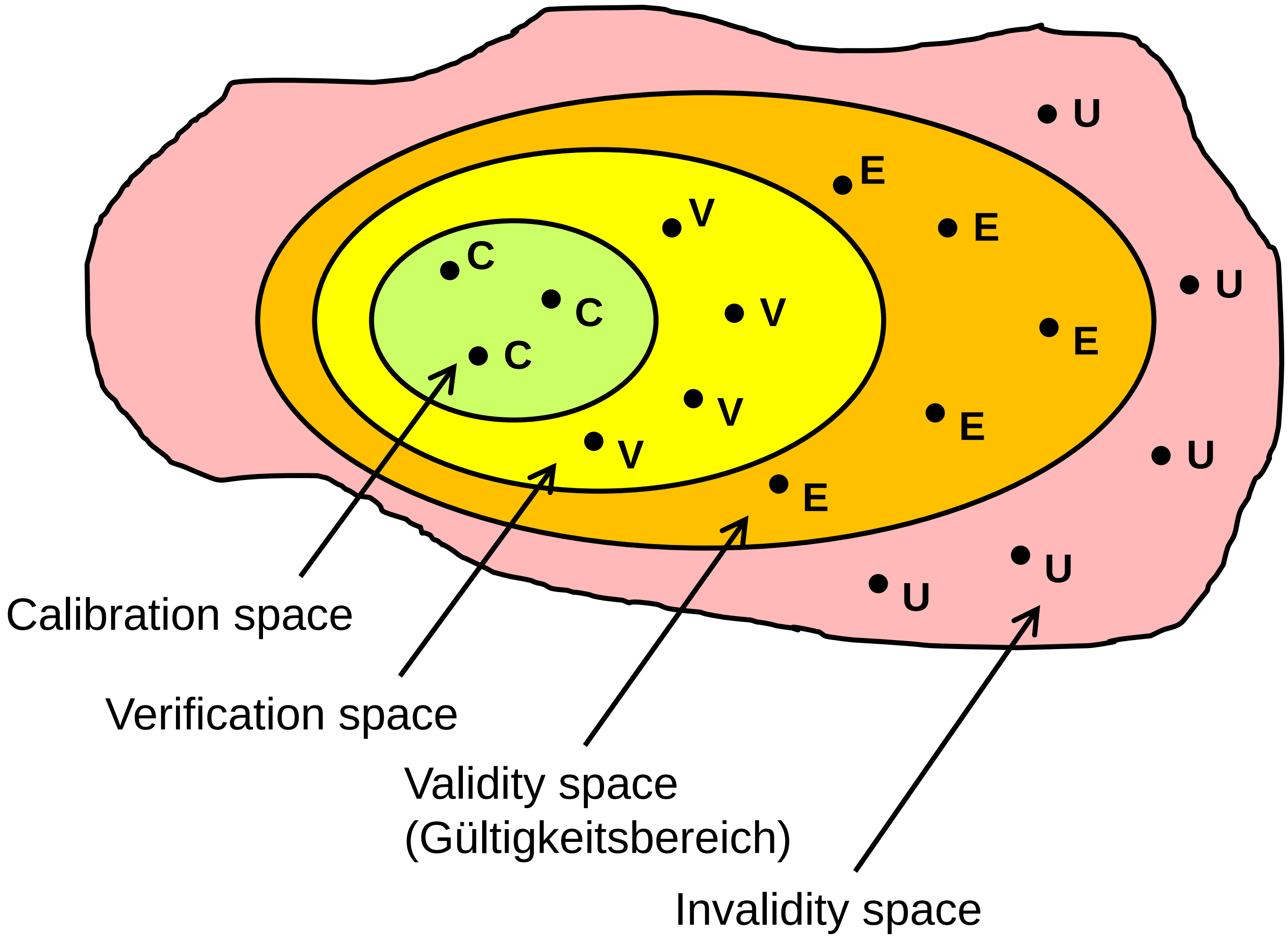
Validation of Models

Validation of Models

- Comparison of systems and models



- Calibration and Validation



Validation of Models

- **Calibration**
  - Identification and minimization/removal of differences between the real system and the model.
    - Structural change

      Modification of the program code in order to add new aspects or to modify existing processes.

      ⇨ The calibration and modeling of complex systems usually requires a large number of major structural changes.

    - Parameter change

      Modification of model parameters. No structural changes are required.

      ⇨ Parameter changes require less effort than structural changes.

      ⇨ Tune the parameters such that the parameter vector p minimizes the difference between the real system and the model

$$p_{opt} = \arg\min_p D(V_R, V_{S_{(p)}})$$

Validation of Models

Comparison of real system and model

- Definition of validity:
  - S Model
  - V_R Behavior/output of the real system
  - V_S Behavior/output of the model
  - $D(V_R, V_{S_K})$ Difference of behavior/output of the real system and the model.
  - D_{rs_max} Maximum allowed difference between the real system and the model.

⇨ Identify a maximum acceptance threshold D_{rs_max}

⇨ Find a model S such that $D(V_R, V_{S_K}) < D_{rs\_max}$

Validation of Models

- **Calibration**
  - Problems of calibration of statistic models:
    - Structural change

      Structural changes result in new models $S_1, S_2, \ldots, S_K$ which have to be verified and compared according to $D(V_R, V_{S_K})$.

    - Probabilistic results

      The behavior/output of the models depends on random variables. Thus, the results are probabilistic.

⇨ Is the difference between different models/configurations just the result of the random input variables or is it based on the structural changes/different setup?

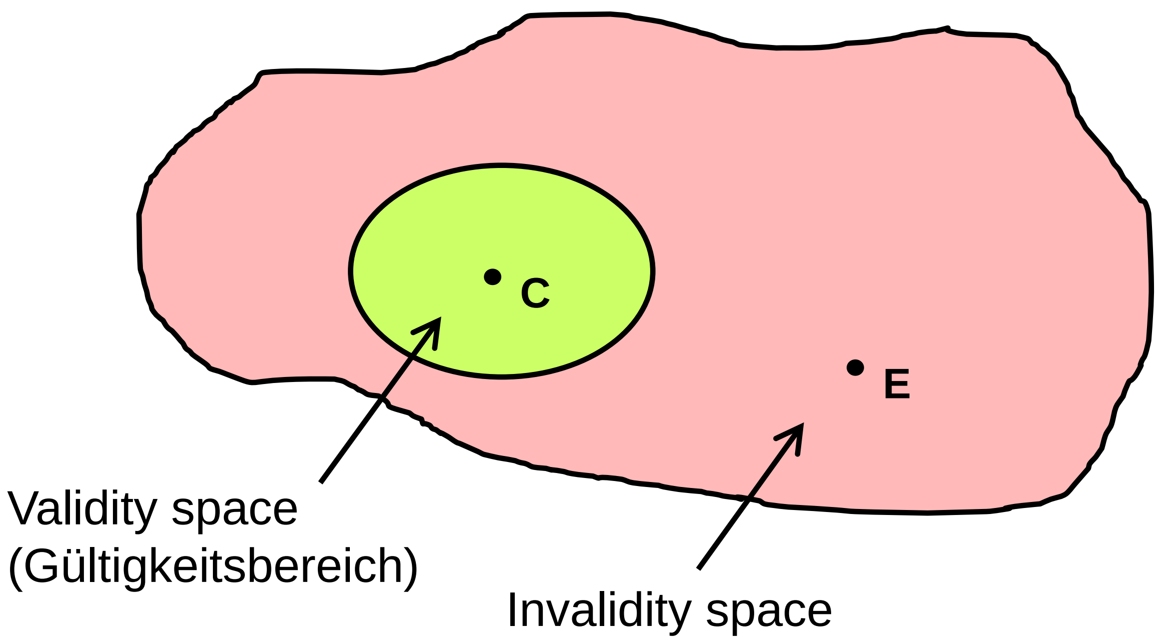
⇨ Define a maximum threshold D_{rs_max} which represents an acceptable difference between the real system and the model.

Validation of Models

- **Calibration**
  - Problems of calibration of statistic models:

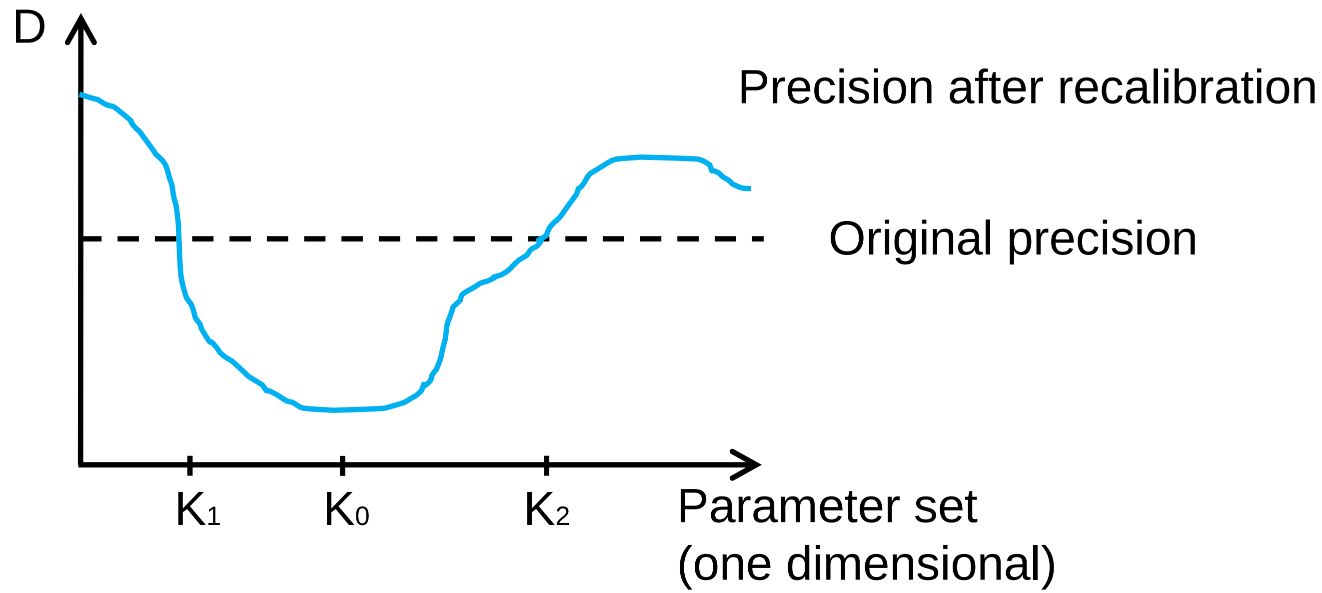
    ⇨ Assume that we have found a model S such that $D(V_R, V_{S_K}) < D_{rs\_max}$ for a certain scenario.

    • Can we assume that the condition holds for other scenarios?



Validation of Models

- **Calibration**
  - Problems of calibration of statistic models:
    - Overfitting
      – Structural and parameter changes often come with side-effects which are not obvious at the first glance.
      – The optimization for a certain parameter set may decrease the precision of the model for other parameter sets which results in an overall performance decrease.



Validation of Models

- **Measurement of the difference between system and model**
  - Measurement is essential for the determination of the behavior of the system and the model during the calibration.
  - Calibration and validation shall increase the trust in the model.
  - Questions
    - Which system(e.g. hardware, software, protocol stack) should be evaluated/simulated/compared with the simulator?
      - Is it possible to run the system with the target configuration?
      - Is it possible to use traces of the system as input of the simulation?
    - Which aspects/characteristics are chosen for the comparison?
      - What are the performance parameters(e.g. packet loss, jitter, delay, energy consumption, bandwidth)?
      - Which metric is of interest (average, minimum, maximum, median, x-quantile)?
      - Individual parameters or combined
    - What kind of methods should be used(analysis or simulation)?

Validation of Models

Comparison of samples

- The behavior of the real system and the model is in the following described by the random variables V_R and V_S, respectively.

- n measurements of the system $v_R = \{v_{R_1}, v_{R_2}, \ldots, v_{R_n}\}$
- m measurements of the model $v_S = \{v_{S_1}, v_{S_2}, \ldots, v_{S_m}\}$

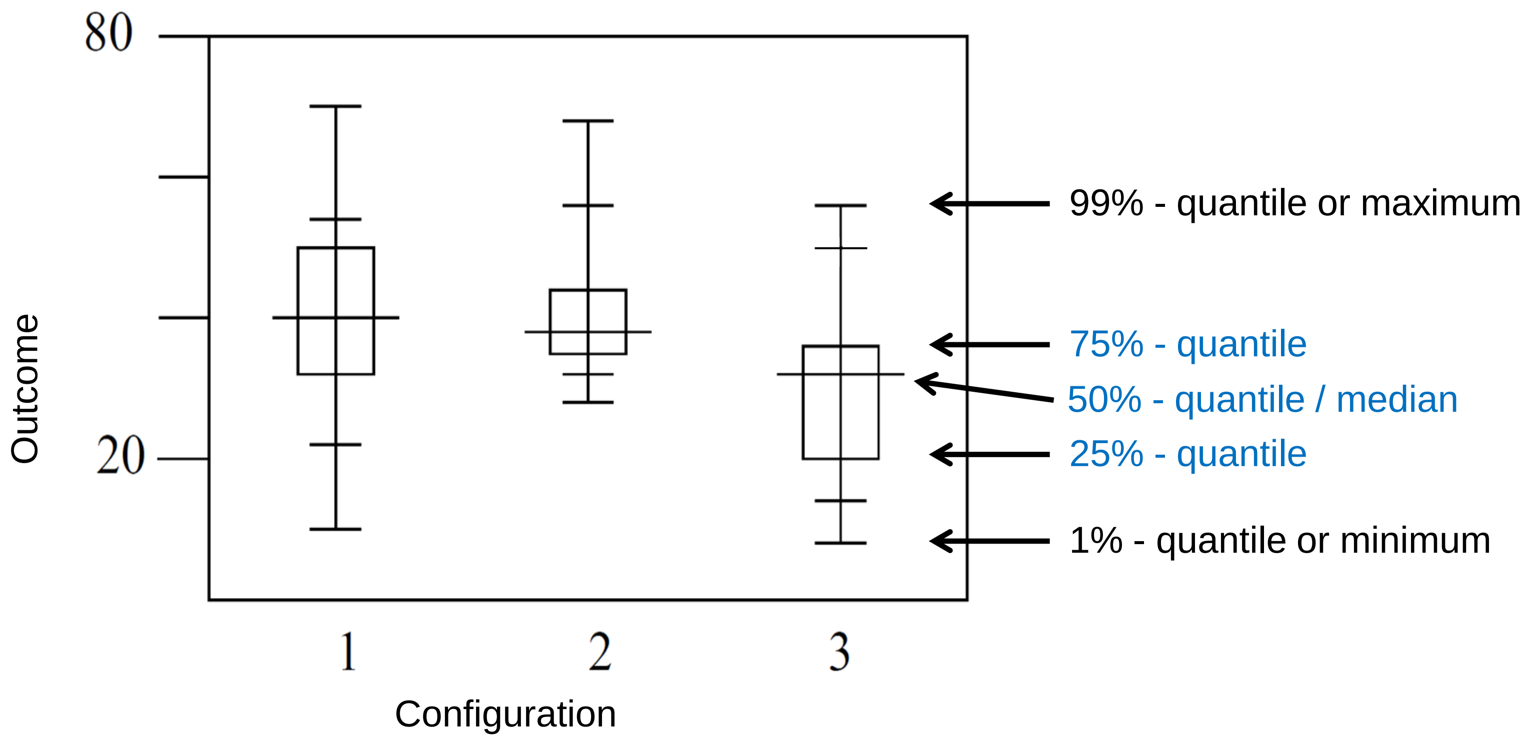
- Evaluation procedure:
  - Subjective comparison (inspection method) based on graphical representation of the samples.
    - Histograms, QQ-plots, PP-plots, box plots. (w/o outlier removal)
  - Objective comparison by using statistically firm evaluation procedures which use the samples as input. of the

⇨ The inspection method is an important decision criteria in spite of the subjective factors. However, it should be supplemented by statistically firm evaluation methods.

Validation of Models

- Subjective comparison

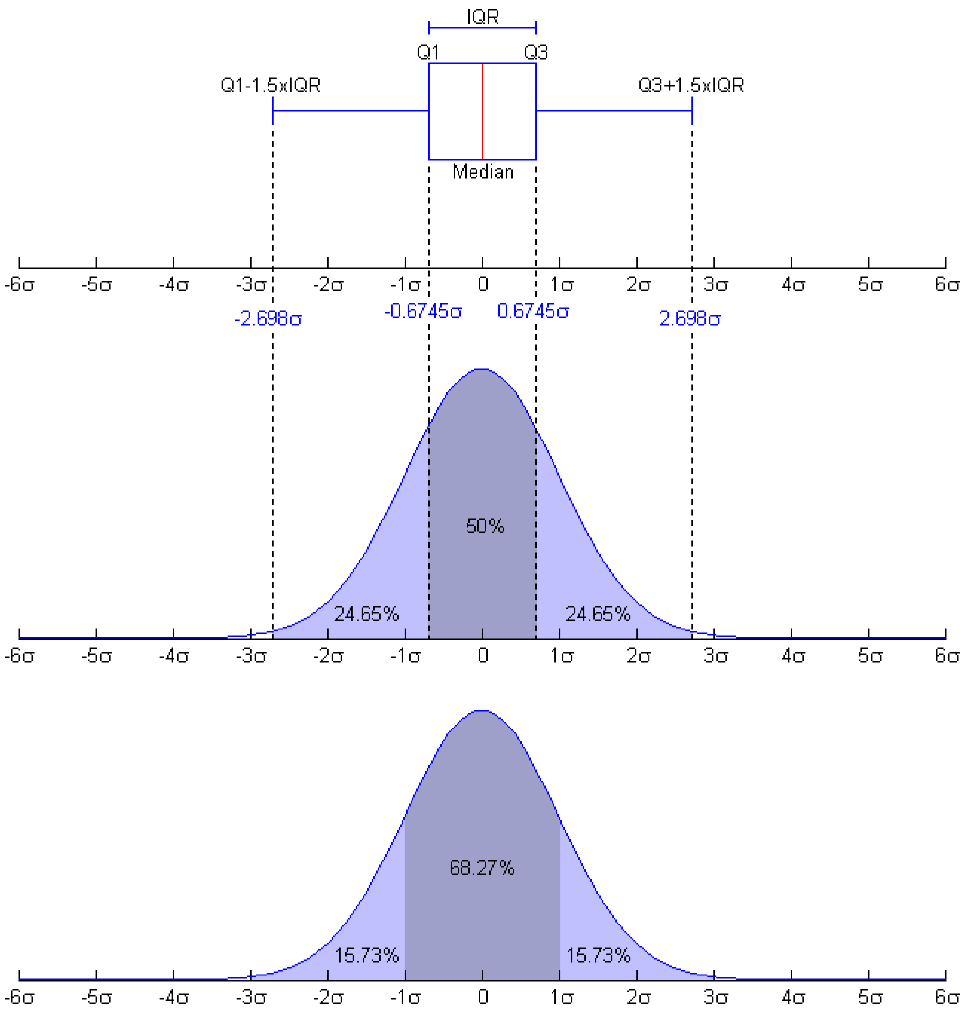
  Box plot / Whisker plot



The inner box always represents the 25%, 50% and the 75% - quantile. The other values are chosen with respect to the samples. (c.f. How to lie with statistics)

Validation of Models

- Subjective comparison
  - Box plot / Whisker plot



Picture taken from Wikipedia

Validation of Models

- **Comparison of confidence intervals**
  - Confidence intervals are used to evaluate the mean of two samples.

Idea:

- Calculate a confidence interval based on the difference of the mean of each sample.
- It is assumed that both samples have the same mean if the calculated confidence interval includes 0.
- The range of the confidence interval indicates whether the differences of both samples are within tolerable limits.

In the following it is assumed that the samples are statistically independent.

❑ Comparison of confidence intervals

Welch

- Estimators

$$\Rightarrow \quad \tilde{\mu}_R = \frac{1}{n} \cdot \sum_{i=1}^{n} V_{R_i} \qquad \tilde{S}_R^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(V_{R_i} - \tilde{\mu}_R \right)^2$$

$$\Rightarrow \quad \tilde{\mu}_S = \frac{1}{m} \cdot \sum_{i=1}^{m} V_{S_i} \qquad \tilde{S}_S^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(V_{S_i} - \tilde{\mu}_S \right)^2$$

- Difference of both samples are defined as follows:
  - $v_{RS\,i} = v_{R\,i} - v_{S\,i}$
  - $v_{RS} = \left\{ v_{RS_1}, v_{RS_2}, \ldots, v_{RS_n} \right\}$
  - $\tilde{\mu}_{RS} = \frac{1}{n} \cdot \sum_{i=1}^{n} V_{RS_i}$
  - $\tilde{S}_R^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(V_{R_i} - \tilde{\mu}_R \right)^2$

$$\Rightarrow \quad \boxed{\hat{\mu}_{RS} \pm t_{n-1,1-\alpha/2} \cdot \tilde{S}_{RS} / \sqrt{n}}$$

⚠ Both samples have to be statistically independent.

⚠ The samples must be of the same size. $m = n$

⚠ The variance of both samples must be equal. $Var(V_R) = Var(V_S)$

Validation of Models

- **Comparison of confidence intervals**

  Welch

  - Problem:
    - Sample size is typically different.
    - Additional measurements are often not possible.
    - A reduction of the sample size results in loss of information.
    - The variance of the system is not known in advance.
    - The variance of the system usually not equals the variance of the model.

⇨ The described method leads to acceptable results if the sample are of the same size and their variance only slightly differs.

⇨ A different method should be used if the sample size is different.

Validation of Models

- **Comparison of confidence intervals**

Law and Kelton

The following approximation of the Welsh's method was introduced by Law and Kelton. It generates acceptable results for samples of different size $m \neq n$ and variance $Var(V_R) \neq Var(V_S)$ in a wide range of experiments.

– Estimated sample mean $\hat{\mu}_{RS}$ of the difference:

$$\Rightarrow \quad \hat{\mu}_{RS} = \hat{\mu}_R - \hat{\mu}_S$$

– Estimated variance $\hat{S}^2_{RS}$ of the difference:

$$\Rightarrow \quad \hat{S}^2_{RS} = \frac{\hat{S}^2_R}{n} + \frac{\hat{S}^2_S}{m}$$

– Approximated degrees of freedom $\hat{f}$ of the difference:

$$\Rightarrow \quad \hat{f} = \frac{\left(\hat{S}^2_{RS}\right)^2}{\frac{1}{n-1}\left(\frac{\hat{S}^2_R}{n}\right)^2 + \frac{1}{m-1}\left(\frac{\hat{S}^2_S}{m}\right)^2}$$

Validation of Models

- **Comparison of confidence intervals**

  Law and Kelton

  Since $\widehat{f}$ is typically no integer it has to be rounded or interpolated as follows.

$$\Rightarrow \quad t_{\hat{f},1-\alpha/2} = \left(\widehat{f} - \left\lfloor \widehat{f} \right\rfloor\right) \cdot t_{\left\lfloor \hat{f} \right\rfloor,1-\alpha/2} + \left(\left\lceil \widehat{f} \right\rceil - \widehat{f}\right) \cdot t_{\left\lceil \hat{f} \right\rceil,1-\alpha/2}$$

  The confidence interval of V_{RS} can then be calculated as follows:

$$\Rightarrow \quad \hat{\mu}_{RS} \pm t_{\hat{f},1-\alpha/2} \cdot \widehat{S}_{RS}$$

Validation of Models

- **Comparison of confidence intervals**

  Example

  Law and Kelton [2000, Kap. 10.2]

i	v_{Ri}	v_{Si}	$v_{Ri} - v_{Si}$
1	126.97	118.21	8.76
2	124.31	120.22	4.09
3	126.68	122.45	4.23
4	122.66	122.68	0.02
5	127.23	119.40	7.83

⇨ Are the samples based on the same distribution?

⇨ What is the level of significance (Signifikanzniveau)?

Validation of Models

- Comparison of confidence intervals

  Example

  Law and Kelton [2000, Kap. 10.2]

⇒ $\widehat{\mu}_S = 120.59$ $\widehat{S}_S^2 = 3.76$

⇒ $\widehat{\mu}_R = 125.57$ $\widehat{S}_R^2 = 4.00$

⇒ $\widehat{\mu}_{RS} = 4.98$ $\widehat{S}_{RS}^2 = 1.55$

⇒ $\hat{f} = 7.99$

⇒ Level of significance $\alpha = 0.1$, conf. interval $[2.67, 7.29]$

⇒ Level of significance $\alpha = 0.01$, conf. interval $[0.81, 9.15]$

A significant difference between the real system and the simulation can be assumed since 0 lies not within the confidence interval.

Validation of Models

- **Comparison of confidence intervals**

  Example

  Welch (taken from Law and Kelton [2000, Kap. 10.2])

i	v_{Ri}	v_{Si}	v_{Ri} - v_{Si}
1	126.97	118.21	8.76
2	124.31	120.22	4.09
3	126.68	122.45	4.23
4	122.66	122.68	0.02
5	127.23	119.40	7.83

Now we assume that both samples are based on distributions V_S and V_R which have the same variance and are statistically independent.

$$\Rightarrow \quad \hat{S}_{RS}^2 = \frac{\sum_{i=1}^{n} \left((v_{R_i} - v_{S_i}) - \hat{\mu}_{RS} \right)^2}{n \cdot (n-1)}$$

$$\Rightarrow \quad \hat{\mu}_{RS} \pm t_{n-1,1-\alpha/2} \cdot \tilde{S}_{RS} / \sqrt{n}$$

Validation of Models

- **Comparison of confidence intervals**

  Example

  Welch (taken from Law and Kelton [2000, Kap. 10.2])

⇨ $\widehat{\mu}_{RS} = 4.98$ $\widehat{S}^2_{RS} = 1.56$

⇨ Level of significance $\alpha = 0.1$, conf. interval $[1.66, 8.30]$

⇨ Level of significance $\alpha = 0.01$, conf. interval $[-2.20, 12.16]$

According to the definition of the confidence interval, the true value lies with a probability of $1-\alpha$ within the confidence interval.

- We can only assure with a probability of 10% that the mean of the distributions of both samples are not equal.
- A sample size of 5 is too small to allow a meaningful comparison.

Inverse student-t distribution

One Sided	75%	80%	85%	90%	95%	97.50%	99%	99.50%	99.75%	99.90%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.50%	99.80%	99.90%
1	1	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.92	4.303	6.965	9.925	14.09	22.33	31.6
3	0.765	0.978	1.25	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.19	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5	0.727	0.92	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.86	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.1	1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10	0.7	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.93	4.318
13	0.694	0.87	1.079	1.35	1.771	2.16	2.65	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.14
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.69	0.865	1.071	1.337	1.746	2.12	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.74	2.11	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.33	1.734	2.101	2.552	2.878	3.197	3.61	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.86	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.85
21	0.686	0.859	1.063	1.323	1.721	2.08	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.06	1.319	1.714	2.069	2.5	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.06	2.485	2.787	3.078	3.45	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.69
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.31	1.697	2.042	2.457	2.75	3.03	3.385	3.646
40	0.681	0.851	1.05	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2	2.39	2.66	2.915	3.232	3.46
80	0.678	0.846	1.043	1.292	1.664	1.99	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.29	1.66	1.984	2.364	2.626	2.871	3.174	3.39
120	0.677	0.845	1.041	1.289	1.658	1.98	2.358	2.617	2.86	3.16	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.96	2.326	2.576	2.807	3.09	3.291