Design of Experiments

Budapest University of Technology and Economics
Fault Tolerant Systems Research Group

- Goal: parametrisation of the models
  - based on real-life data
- Experiment: collecting information
  - What information is to be collected?
  - What has to be observed? How many times?
  - To what can be concluded from the observations?
- (Statistical) Design of Experiment (DoE)
  - efficient method for designing/evaluating experiments
  - to found real and objective conclusions

Design of Experiment

- Experiment Design: Planning before executing
  - making experiments more robust
  - reducing variability
  - clear goals, unambiguous results
  - choosing between alternatives
  - small experiments, low cost
  - real information

Measuring numerical properties

- Let us measure the value of a property (1…5)

1 measurement – average: 4

5 measurements – average: 3,6

10 measurements – average: 3,3

15 measurements – average: 3,4

30 measurements – average: 3,3

(it seems to be rather constant)

50 measurements – average: 3,12

100 measurements – average: 2,88 **mean value:2,86**

Probability Variables

- Variables with multiple possible values (outcomes)
  - discrete or continuous values
  - probability distribution

- „Never show two thermometers at the same time.”

- If you measure it once, you know the value.
- If you measure it twice, you can never be sure.
  - Never? Sure?

Probability variable: X

- Statistics
  - mean value: $E(X)$ (average)
  - variance/deviation: $\sigma = \sqrt{E(X-\mu)^2}$ (spreading)

- Sampling: $x_1, x_2, \ldots x_t$ (measurements, observations)
  - empirical mean: $\bar{x} = \frac{x_1 + x_2 + \ldots x_t}{t}$
  - variance (incorrect): $\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots(x_t-\bar{x})^2}{t}}$
  - corrected emp. variance: $\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots(x_t-\bar{x})^2}{t-1}}$

Empirical Mean

- Method: computing mathematical average
  - repeated observations
  - independent of each other
  - under same circumstances
- Questions
  - How many observations are required?
  - By how far characterises the empirical mean the „real" expected value?
- The first thing to be clarified
  - distribution of empirical mean

Law of Large Numbers

- „The empirical mean converges to the mean value (with the probability 1), as the number of observations tend to infinity."

$$\frac{x_1 + x_2 + \ldots x_t}{t} \rightarrow \mu \qquad \text{if} \quad t \rightarrow \infty$$

Central Limit Theorem (CTL)

- The arithmetic mean of a sufficiently large number of iterates of independent random variables, …
  - each with a well-defined (finite) expected value: ***m***
  - and with a finite variance: ***s***
- … will be approximately normally distributed,
- regardless of the underlying distribution.

Central Limit Theorem (CTL)

- Mean value of the normal distribution

$$\mu = m$$

- Variance of the normal distribution

$$\sigma = s/\sqrt{t}$$

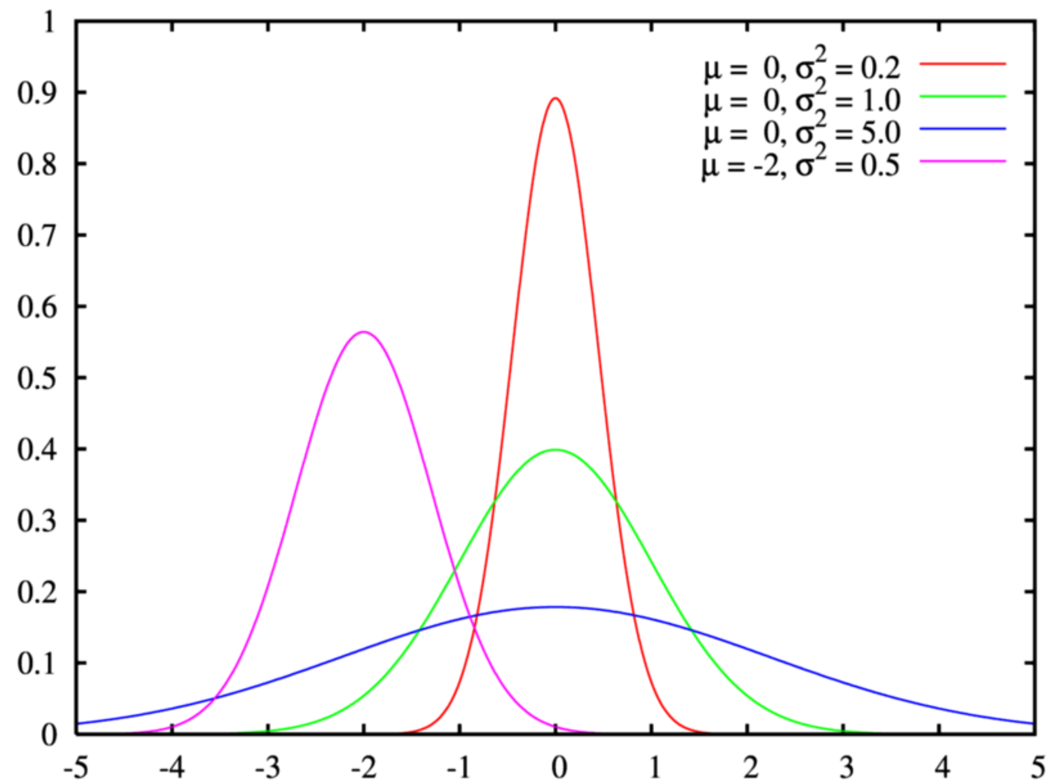
- Rules of thumb
  - for known s, the approximation is acceptable for t > 30
  - for unknown s, for t > 100

Normal (Gaussian) Distribution

- Probability density function: (not to be learnt)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Parameters
  - Mean μ
    - $\rightarrow$ **m**
  - Variance σ
    - $\rightarrow \mathbf{s}/\sqrt{t}$



Normal (Gaussian) Distribution

- The values concentrate around the mean

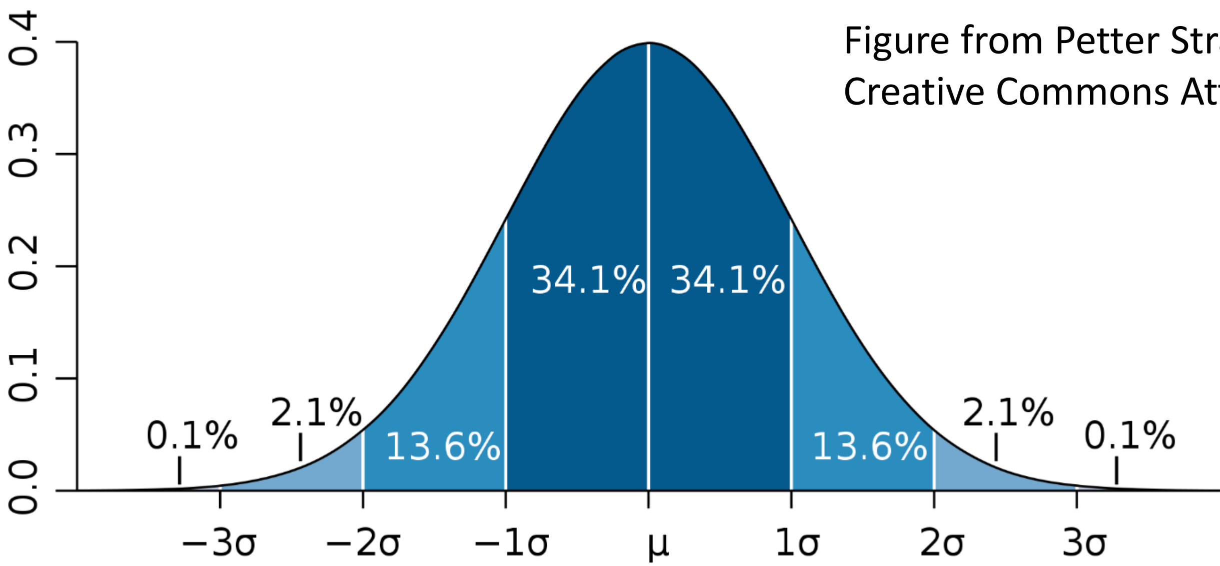


Figure from Petter Strandmark under the Creative Commons Attribution 2.5 Generic licence

- The value of the normally distributed probability variable…
  - is not further than **1*σ*** from *μ* in **68%** of the cases
  - is not further than **2*σ*** from *μ* in **95%** of the cases
  - is not further than **3*σ*** from *μ* in **68%** of the cases
  - …

Confidence Intervals

- We observe t (>30) times a property (with an arbitrary distribution and with variance ***s***)
- Its empirical mean …
  - … estimates the value ***m*** with max. $\boldsymbol{s}/\sqrt{\boldsymbol{t}}$ inaccuracy with **68%** assurance,
  - … falls in a $\boldsymbol{2s}/\sqrt{\boldsymbol{t}}$ radius interval around ***m*** with **95%** assurance,
  - … falls in a $\boldsymbol{3s}/\sqrt{\boldsymbol{t}}$ radius interval around ***m*** with **99,7%** assurance.
- For an increasing ***t*** the interval shrinks by $\sqrt{\boldsymbol{t}}$

DoE Example

- 30 observation for the mean
  - empirical mean: 2,3 s (More observations required?)
  - empirical variance: ***s*** = 1,1 s
- Goal
  - the 99,7% confidence interval should be 0,6 s wide
- Design of Experiment
  - Required radius (half of width) = 3σ = $\mathbf{3s}/\sqrt{\mathbf{t}}$ < 0,3 s
    - (σ is the variance of the average, not that of the property!)
  - Therefore *t* = 121 observations are required
- Where am I cheating?

Correction

- The parameters of the real distribution are usually *a priori* unknown. (Why else would we measure?)
- Therefore the variance of the property cannot be used
- Only empirical variation can be used → instead of *Gaussian/normal* we have a *Student t* distribution
  - (other confidence intervals)
- In case of $t \to \infty$: Student → normal
- Rule of thumb: for $t > 100$ Gauss distribution can be assumed

Practice 1

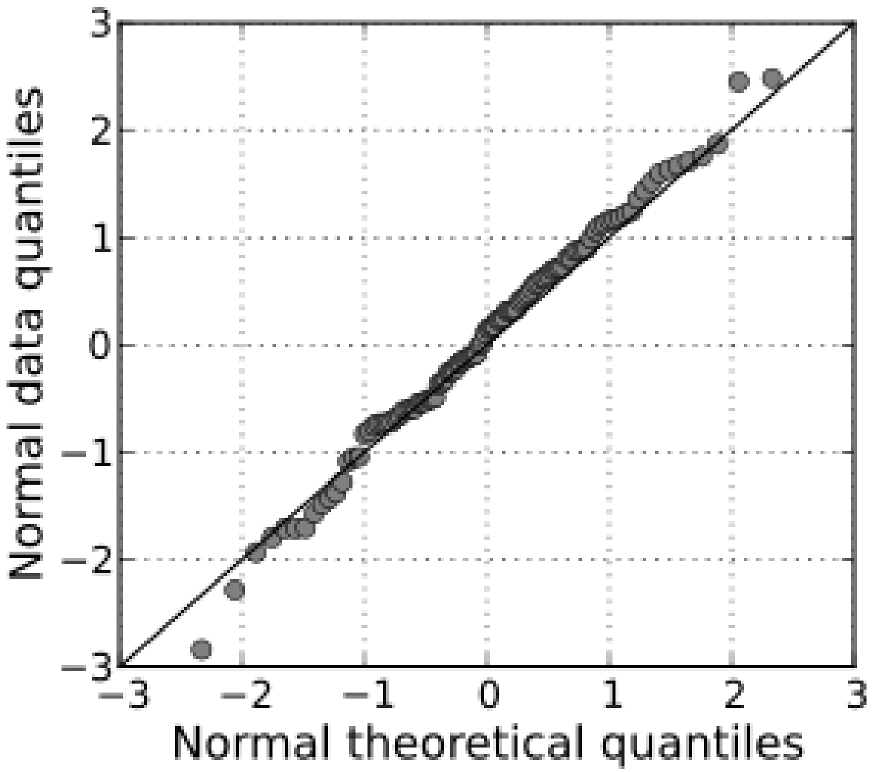
We have an IT system. The execution time of a given service varies (e.g. because of paging, garbage collection and varying cache hits in the memory). We have set up a „near real life" benchmark to measure the execution time.

a) The results of the first 10 runs are:
   37s, 34s, 35s, 39s, 57s, 41s, 36s, 35s, 61s, 35s
   What are the empirical mean and variance of this short experiment?

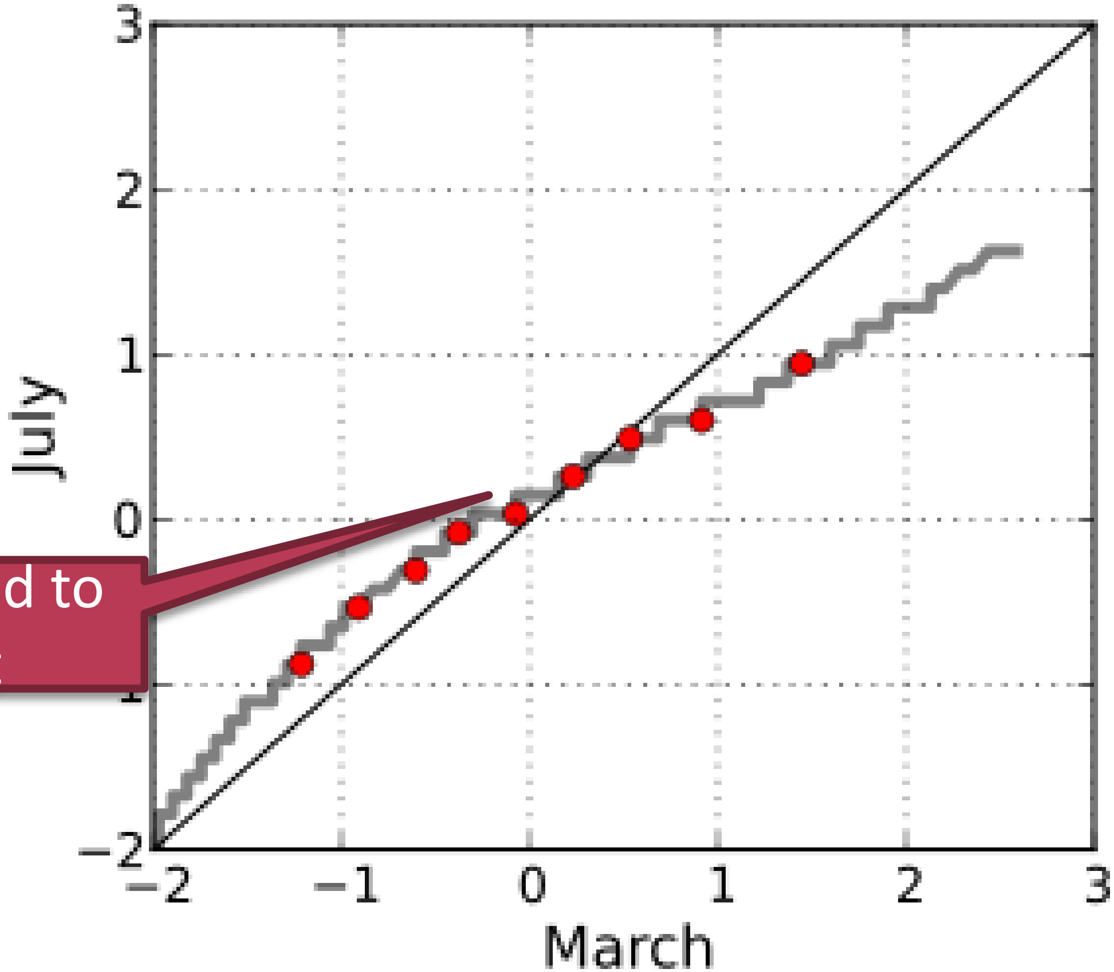
b) After 10.000 runs we see an empirical mean of 44,3s and an emp. variance 11,6s.
   How precise is this value?

Example: Q-Q plot

- Quantile – Quantile plot
- Similarity of distributions
- Example: normal distributions



Q-Q plot



July: skewed to the left

Skewness