Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

inf Informatik
Computer Science

Wireless and Mobile Computing for Entertainment Applications

Appendix: Confidence intervals in stochastic simulation

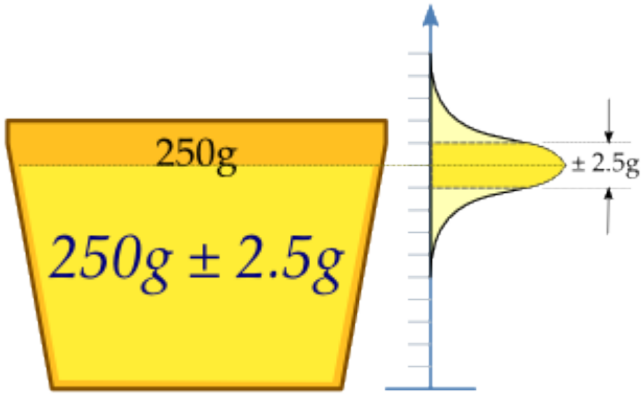
stefan@inf.ethz.ch

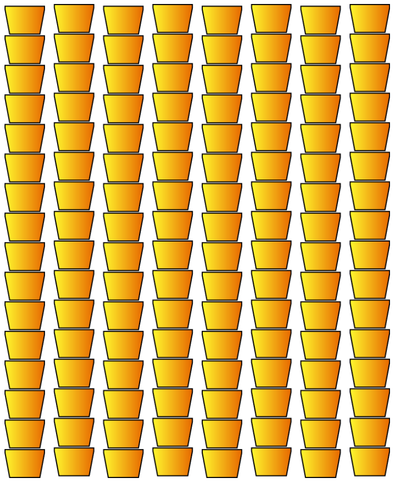
overview

- confidence interval
  - filler machine example
  - confidence level
  - Matlab code
- statistical hypothesis testing
  - magician example
  - statistical significance

filler machine example (1)

- machine that fills cups with Margarine
- content of each cup shows some variation: it is considered a random variable X
- variation assumed to be normally distributed around the desired mean of 250g
  - with standard deviation of 2.5g



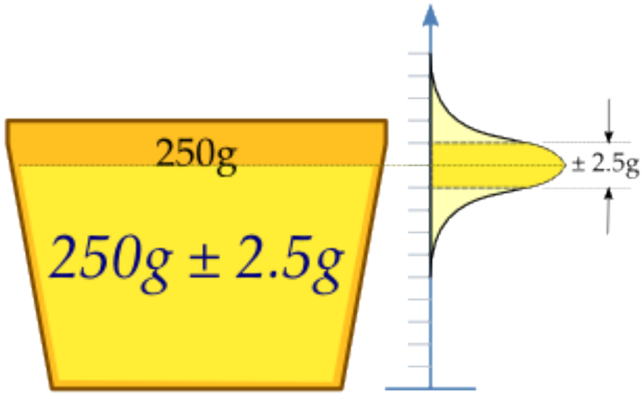


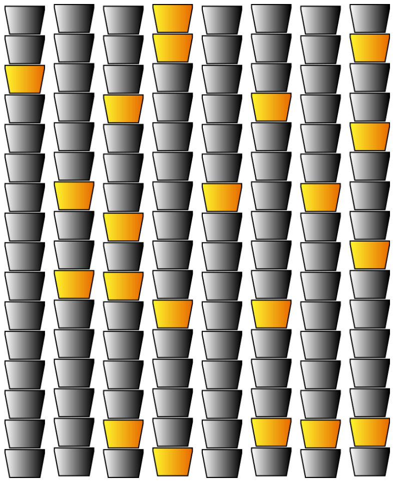
filler machine example (2)

- QUESTION: How well is the machine calibrated?
- take a sample of n = 25 cups randomly
- compute estimated mean value $\bar{X}$:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 250.2\,\mathrm{grams}.$$

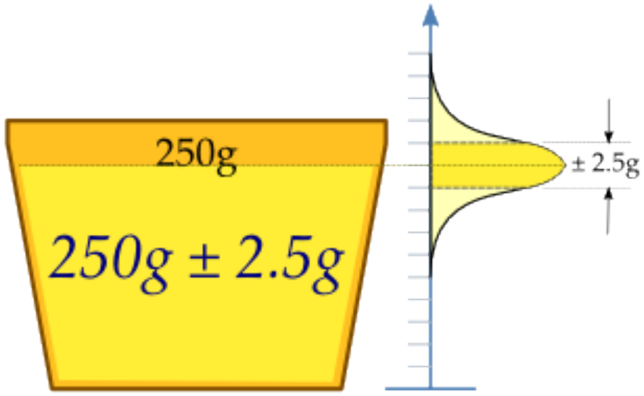


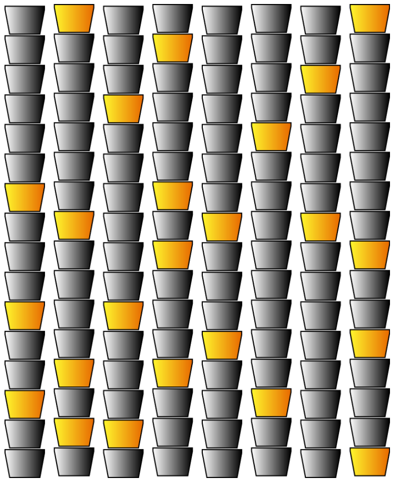


Random Sample 1 From Stock Of Cups

filler machine example (3)

- take a second sample randomly
- compute its mean value
- this time you estimate 250.3g

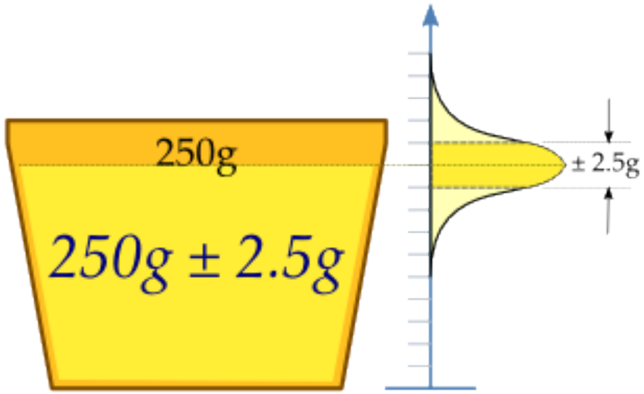


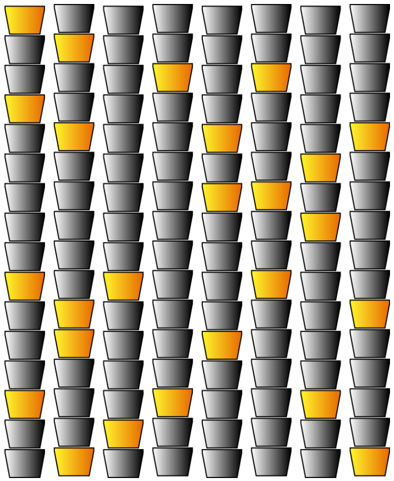


Random Sample 2 From Stock Of Cups

filler machine example (4)

- take a third sample randomly
- compute its mean value
- this time you estimate 249.8g

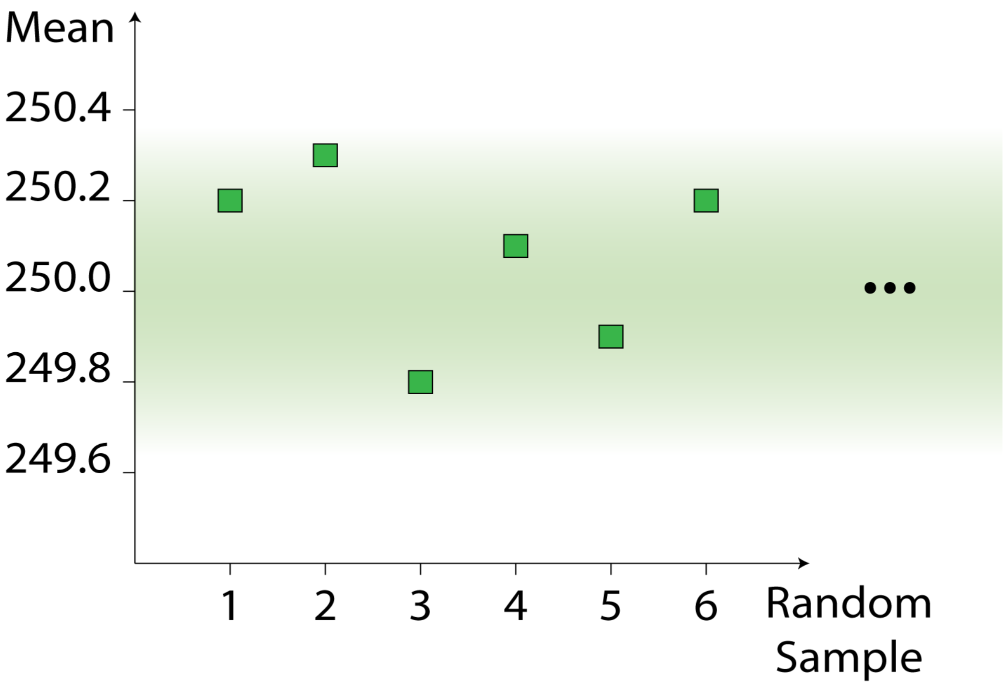




Random Sample 3 From Stock Of Cups

filler machine example (5)

- By repeating the process for many samples, …



Computing the means as single values does not solve our problem, but just postpone it to the point where you have multiple mean values (one per sample). You have to figure out what to do to make sense out of them.

… the obtained means appear to be distributed around the desired value! Now what?

confidence interval (1)

- for each computed mean we can compute a confidence interval
- the confidence interval is associated with a confidence level ($1-\alpha$)
  - typical values of the confidence level are 0.95 or 0.9
- interpretation:
  - tells how frequently, for repeated random samples, the corresponding confidence intervals cover the true parameter (real mean)
- although very close conceptually, it is NOT the probability that the true parameter is inside one specific confidence interval

confidence interval (2)

- how to compute it?
- a few facts: because the content of the cups is normally distributed, the means of the random samples are also normally distributed, with:
  - the same expected mean μ
    (the mean of the means, should be the mean ☺)
  - but different standard deviation, equal to $\sigma/\sqrt{n}$
  - in our example the standard deviation is 0.5g, since σ=2.5g and n=25

filler machine example (6)

- take the first random sample ($\bar{X}$ = 250.2g)
- standardize the random variable (enforcing the mean to zero and the variance to one) as:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{X} - \mu}{0.5}$$

- select a confidence level ($1-\alpha$) = 0.95
- denote confidence interval for Z as [-z; z]

filler machine example (7)

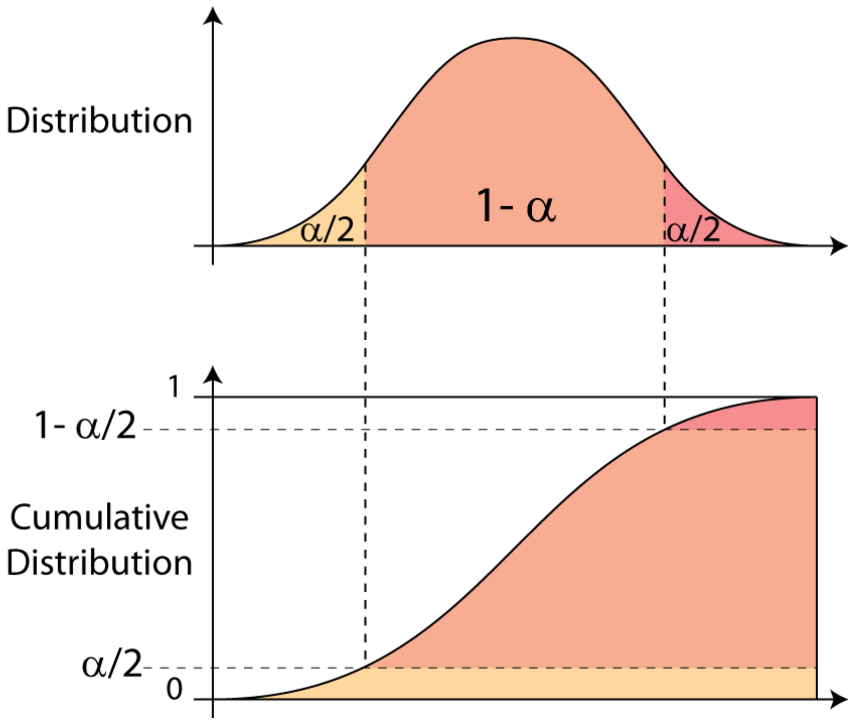
- set the probability of Z being inside the confidence interval equal to the conf. level:

$$P(-z \le Z \le z) = 1 - \alpha = 0.95.$$

- to find the values z, the cumulative distribution function Φ(z) can be used:

$$\Phi(z) = P(Z \le z) = 1 - \frac{\alpha}{2} = 0.975,$$

$$z = \Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.975) = 1.96,$$



filler machine example (8)

- Given the definition of Z, and the values of z, σ and n, we solve for μ:

$$0.95 = 1 - \alpha = P(-z \leq Z \leq z) = P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq 1.96\right)$$

$$= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{X} - 1.96 \times 0.5 \leq \mu \leq \bar{X} + 1.96 \times 0.5\right)$$

$$= P\left(\bar{X} - 0.98 \leq \mu \leq \bar{X} + 0.98\right).$$

- Inserting the value of $\bar{X}$ = 250.2g, we obtain:

$$(\bar{x} - 0.98; \bar{x} + 0.98) = (250.2 - 0.98; 250.2 + 0.98) = (249.22; 251.18).$$

filler machine example (9)

- in general, we found that for a sample $\bar{X}$, it holds:

$$0.95 = P\left(\bar{X} - 0.98 \leq \mu \leq \bar{X} + 0.98\right)$$

- each random sample has it's own confidence interval
- each confidence interval is either hit or not by the mean
  - therefore the probability of μ being in that interval is either 0 or 1
- but overall, 95% of the confidence intervals contain the true value μ = 250g
- confidence level tells what the probability is that the true mean μ=250g lies inside one confidence interval ONLY FOR REPEATED random samples

filler machine example (10)

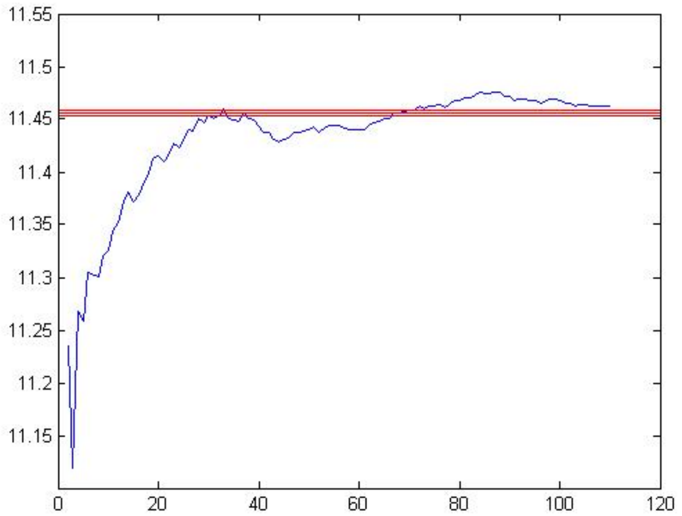
- to answer the original question …
- since the desired value 250g is inside the confidence interval with confidence level of 95% generated from the first random sample, there is no reason to believe the machine is not sufficiently calibrated.

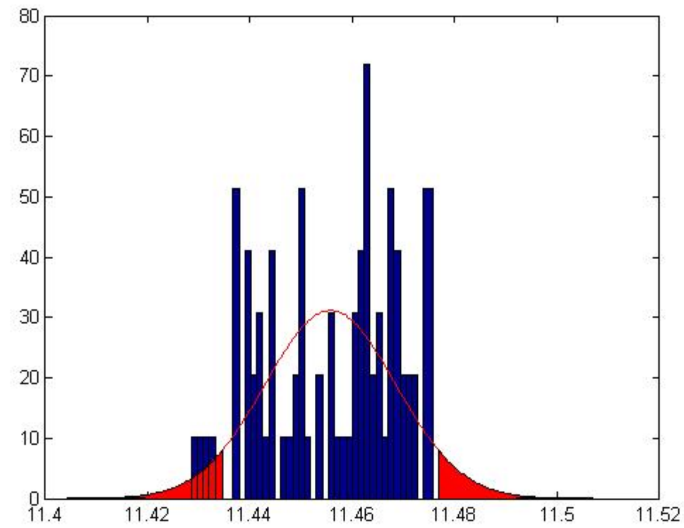
Matlab code (1)

- In Matlab
  - function normfit() (Statistics Toolbox) does the work for you!
- input:
  - Data
  - Alpha (1-a = conf level)
- output:
  - Mean
  - Standard Deviation
  - Mean confidence interval
  - Standard Deviation confidence interval

Matlab code (2)

- a few plot helpers ...
  - confidenceIntervalAnalysis()
- input:
  - Data, Alpha (1- α = conf level)
- output:
  - mean + confidence interval
  - plot of the mean and confidence interval on the data
  - plot of the confidence level on the distribution





overview

- confidence interval
  - filler machine example
  - confidence level
  - Matlab code
- statistical hypothesis testing
  - magician example
  - statistical significance

magician example (1)

- A man claims he has “clairvoyance“ (has extra-sensory perception).
  - We want to test him!
- We show him the reverse of a randomly chosen play cards 25 times and ask which suit it belongs to.
- Interpretation of the number of hits X:
  - With 5 or 6, he is clearly an impostor.
  - With 23 or 24 he truly is a magician!
  - But what about values in-between? From what value on is he gifted?

Statistical significance is a concept that allows us to treat the last question.

statistical significance

- a result is **statistically significant** if it is unlikely to have occurred by chance
  - this does not mean that the result has a real impact in the real world, but just that is statistically surprising

- associated with a **significance level α**
  - typical values: 5%, 1%, 0.1%
- interpretation: α = 0.1% → "There is only one chance over a thousand that this could have happened by chance"

magician example (2)

- denote with ***p*** the probability of a correct guess
- two hypotheses:
  - H0: $p = \frac{1}{4}$ (random guess)
  - H1: $p > \frac{1}{4}$ (magic powers)
- only one hypothesis is true
- With ***c***, we denote how many correct guesses we require to believe that the guy is magical.
  - The higher c, the more strict and critical we are.

The first hypothesis is also called null hypothesis, as it is what we believe is the default state of things.

magician example (3)

- our test will accept one of the two hypotheses
- one way to consider our test is to look at false positive, when a random guesser is accepted as magician
- the value *c* influence such probability:
  - *With c* = 25, there is a very low probability that the guy is guessing randomly

$$P(\text{reject } H_0|H_0 \text{ is valid}) = P(X \geq 25|p = \tfrac{1}{4}) = \left(\tfrac{1}{4}\right)^{25} \approx 10^{-15},$$

  - With *c* = 10, there is a higher probability to have a false positive

$$P(\text{reject } H_0|H_0 \text{ is valid}) = P(X \geq 10|p = \tfrac{1}{4}) \approx 0.07.$$

magician example (4)

- in practice, one selects a significance level α, and sets the probability of a false positive to it
  - Example: For α = 1%

$$P(\text{reject } H_0|H_0 \text{ is valid}) = P(X \geq c|p = \tfrac{1}{4}) \leq 0.01.$$

- then one picks the smallest ***c*** for which the inequality holds.
  - In our example ***c*** = 12.
- the general formula is:

$$P(reject\,H_0|H_0\,is\,valid) = (\tfrac{1}{4})^c(\tfrac{3}{4})^{25-c}\frac{25!}{c!(25-c)!} \leq \alpha$$

How do I solve it? For instance using a brute force with growing ***c*** from 0 to 25.

magician example (5)

- notice that a very bad result, can be statistically significant!
- imagine that the guy guessed all 25 cards wrong
  - the probability of guessing a card wrong is p' = ¾
- the probability of guessing 0 card correctly is:

$$P(\text{reject } H_0 | H_0 \text{ is valid}) = P(X \geq 25 | p' = \tfrac{3}{4}) = \left(\tfrac{3}{4}\right)^{25} \approx 0.00075.$$

- this is very unlikely, and as well magical as high scores!

magician example (6)

- to answer the original question …
- 1) decide how strict you want to be!
  - for instance: very strict, with a significance level α of 1%!
- 2) compute the corresponding c
  - for α = 1% we obtain c = 12
- 3) test the candidate and compare the number of correct guesses with c:
  - with 12 or more correct guesses the guy has magic powers
  - with less than 12 correct guesses they guy is an impostor

further reading

- http://en.wikipedia.org/wiki/Confidence_interval
- http://www.stat.yale.edu/Courses/1997-98/101/confint.htm
- http://en.wikipedia.org/wiki/Statistical_hypothesis_testing

Wireless and Mobile Computing for Entertainment Applications

Confidence intervals in stochastic simulation – END OF APPENDIX –

stefan@inf.ethz.ch