Output Analysis

The simulation results

- The simulation
  - receives as inputs parameters represented by random variables
  - produces as outputs observations of the random variables we are interested in
- For each random variable we are interested in computing
  - an estimation of the *mean value* for the variable
  - an estimation of the error committed providing that mean value for the variable

Types of simulations

- Steady-state simulation:
  - The system runs continuously (or at least over a very long period of time)
  - We want to study long-run, or steady-state, properties of the system
- Terminating simulation:
  - The system under study evolves from time 0 to a well defined ending time, either known in advance or related to a specific event
  - It depends on the initial conditions

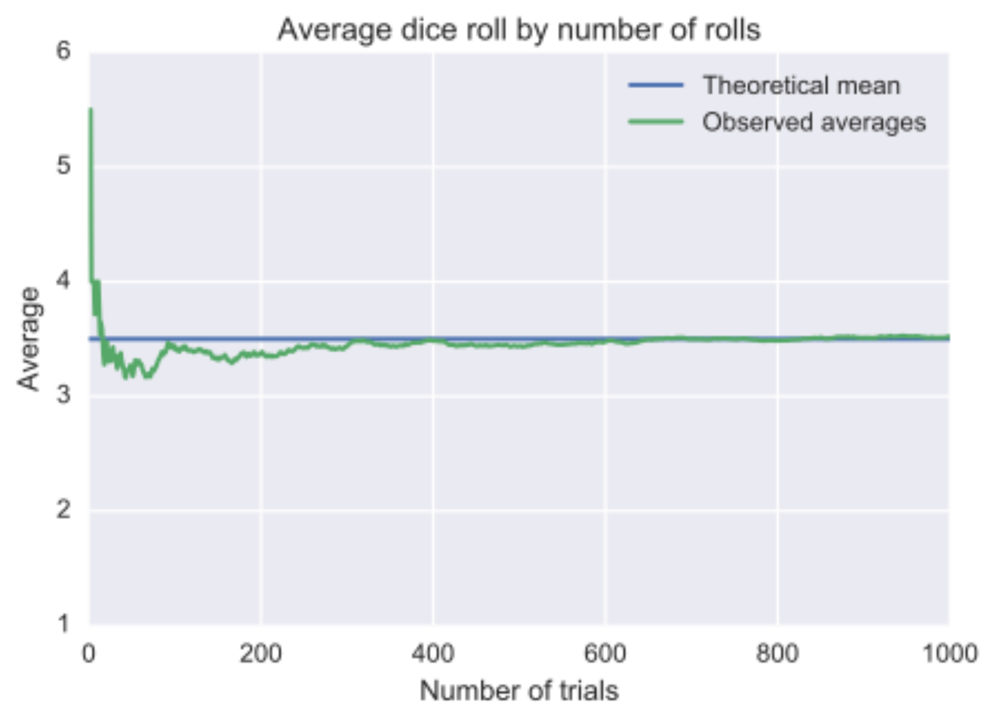
Steady-state simulation

- A system continuously running can be studied estimating its performance parameters over long time intervals
  - X is the metric under study (the average of a random quantity)
  - Being x_1, x_2, … x_i, … independent istances of X collected observing the running system over long periods, by the LLN it holds that

$$X = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

Steady-state simulation

"Law of large numbers: [...] as the number of identically distributed, randomly generated variables increases, their sample mean (average) approaches their theoretical mean."



Steady-state simulation

- Each estimate of X, x_i, is obtained by collecting during the simulation several instances of the observed quantity
- For example
  - During a simulation run we observe N_i packets
  - For each packet *j*, we measure the delay d_j
  - We compute the average delay

$$x_i = \frac{1}{N_i} \sum_{j=1}^{N_i} d_j$$

Confidence intervals

- A confidence interval is an *estimation of the error* we make when we estimate the average value X of a random quantity using an averaging process over a limited number of samples and replications (e.g., X is the average delay, the average throughput,…)
- Assumptions:
  - The observed process is stationary
  - X has average *m* and variance s^2
  - The *n* observations of X ($x_1, x_2, \dots x_i, \dots, x_n$) are independent, i.e., they are iid variables

Confidence intervals

- The average estimation of X, m, is given from the arithmetic mean of the observations

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- $E[\hat{x}]=m$, $Var(\hat{x})=s^2/n$
- From central limit theorem, for large n, the estimate can be approximated by a normally distributed RV $\hat{x} \approx N(m, s/\sqrt{n})$

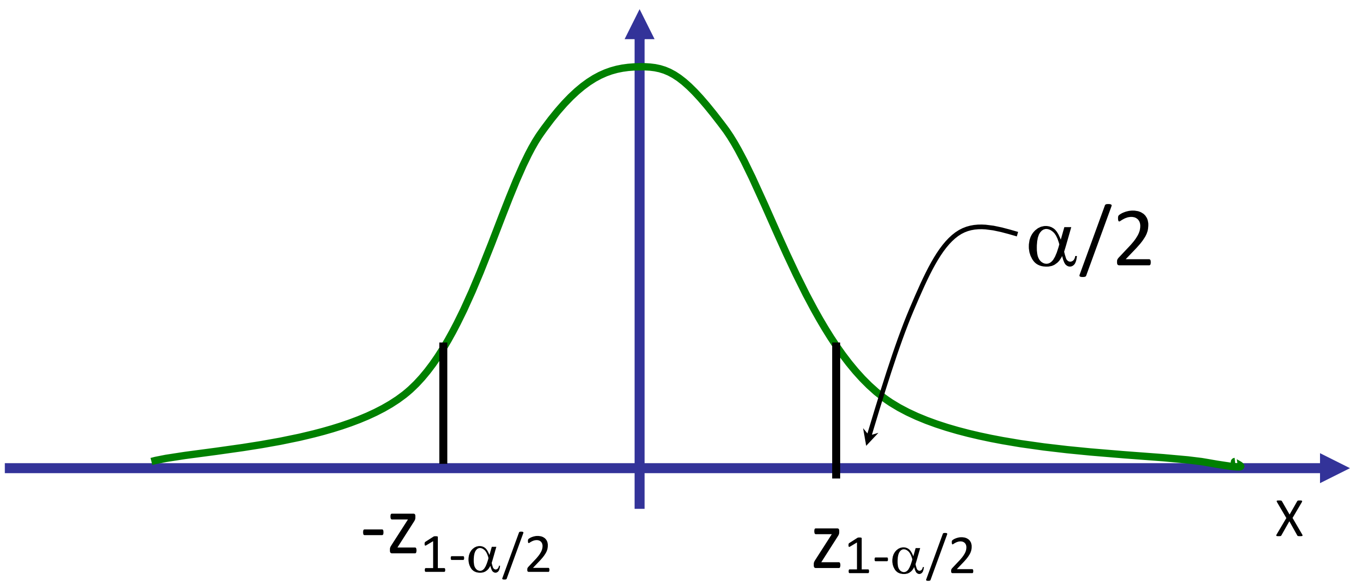
Confidence intervals

- The variance of the estimation decreases with *n*
- From the estimate distribution we can evaluate the error in the estimate itself
- Derive the variable *z* distributed like N(0,1) with the transformation

$$z = \frac{\hat{x} - m}{s / \sqrt{n}}$$

Confidence intervals

- We use the variable *z* to evaluate the error in the estimation
- The probability that *z* is in the interval $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ is $(1-\alpha)$



Confidence intervals

$$P\{-z_{1-\alpha/2} \le z \le z_{1-\alpha/2}\} = 1-\alpha$$

$$P\left\{-z_{1-\alpha/2} \le \frac{\hat{x}-m}{s/\sqrt{n}} \le z_{1-\alpha/2}\right\} = 1-\alpha$$

$$P\{-z_{1-\alpha/2}s/\sqrt{n} \le \hat{x}-m \le z_{1-\alpha/2}s/\sqrt{n}\} = 1-\alpha$$

$$\boxed{P\{\hat{x}-z_{1-\alpha/2}s/\sqrt{n} \le m \le \hat{x}+z_{1-\alpha/2}s/\sqrt{n}\} = 1-\alpha}$$

Confidence intervals

- The interval

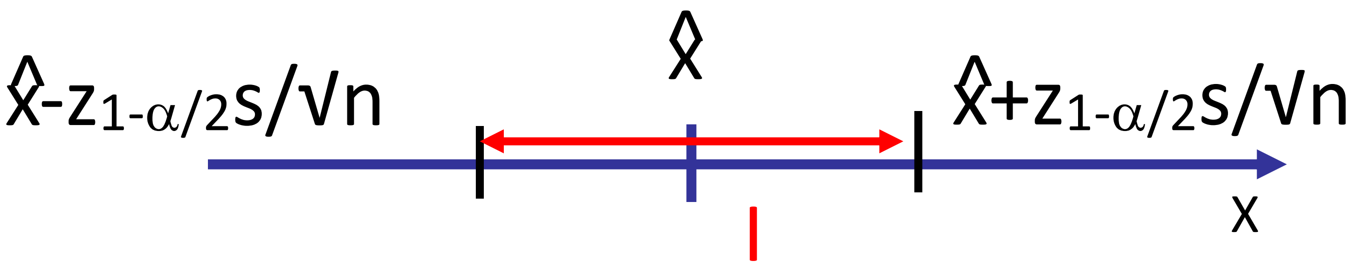
$$I = [\,\hat{x} - z_{1-\alpha/2} s/\sqrt{n}, \hat{x} + z_{1-\alpha/2} s/\sqrt{n}\,]$$

  is called *confidence interval*
- 1-α is the *confidence level*
- Commonly used values for α are:
  - Confidence 90% (α=0.1)
  - Confidence 95% (α=0.05)
  - Confidence 99% (α=0.01)

Confidence intervals

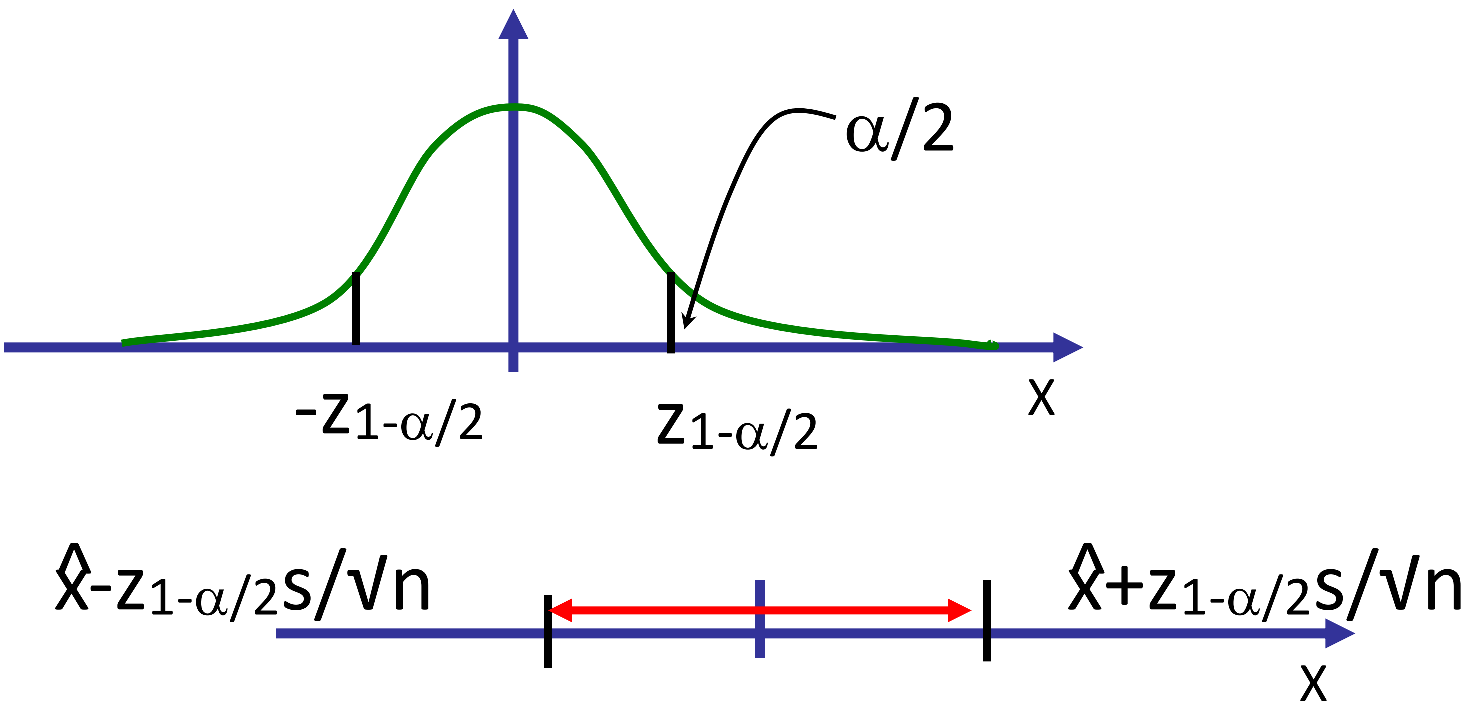
- The confidence interval I with confidence level 1-α means that

if we repeat the procedure many times, (1-α) intervals “I” will actually contain the value m we are trying to estimate



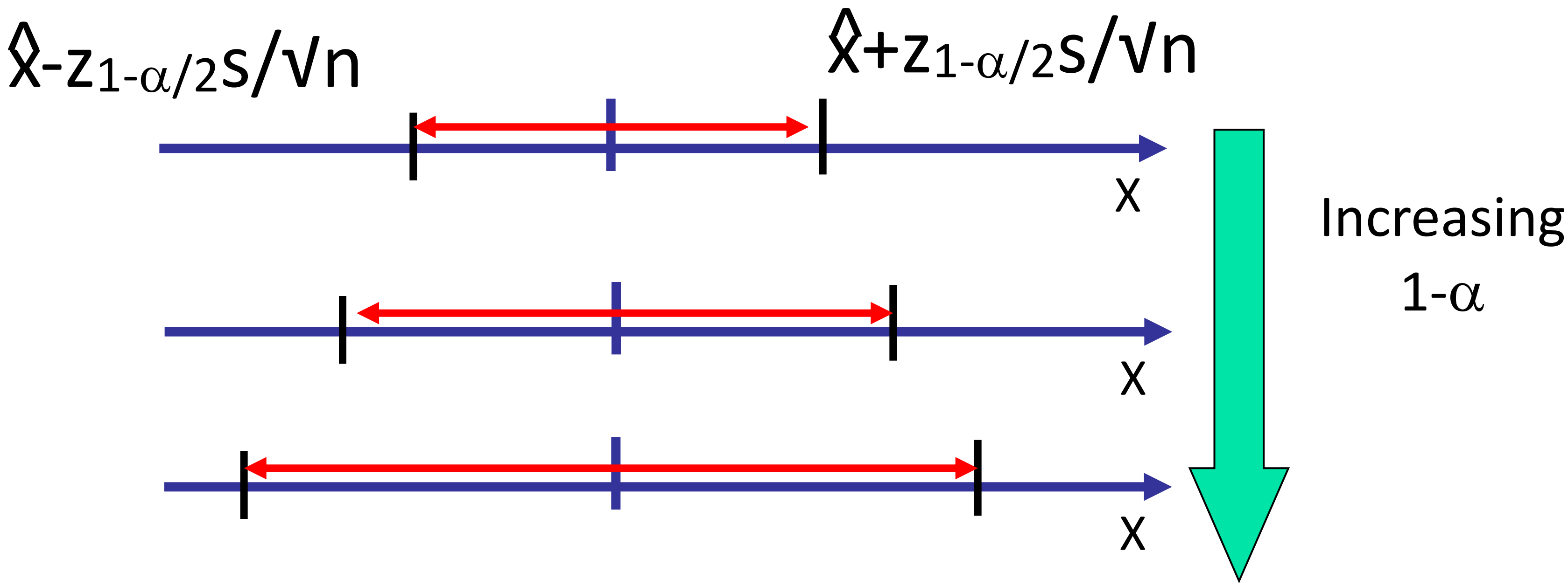
Confidence intervals

- Given an experiment, a higher level of confidence corresponds to a wider confidence interval



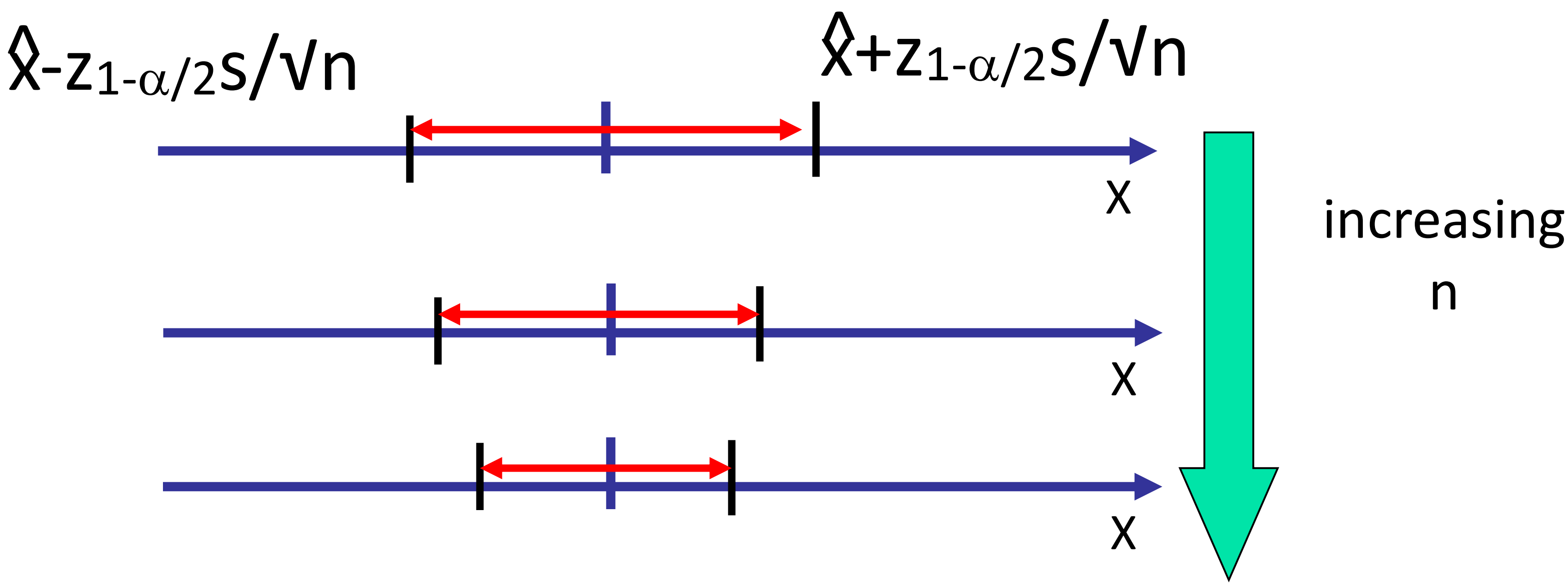
Confidence intervals

- Given an experiment, a higher level of confidence corresponds to a wider confidence interval



Confidence intervals

- Fixing the level of confidence $1-\alpha$, to shrink the confidence interval we must add new samples, incrementing *n*



Confidence intervals

- The central limit theorem is valid only for large *n* (larger than 30)
- When *n* <30, we must substitute the normal distribution with the Student' s *t* distribution with *n-1* degrees of freedom
- We define the variable *t*

$$t = \frac{\hat{x} - m}{s / \sqrt{n}}$$

Student' s *t* distribution

- If W is a r.v. with normal distribution N(0,1) and V is chi-square distributed with *k* degrees of freedom, the ratio

$$\frac{W}{\sqrt{V/k}}$$

  is *t* distributed with *k* degrees of freedom
- The *t* distribution is also called Student distribution or Student' s *t*

Student's *t* distribution

- The pdf for the *t* distribution is

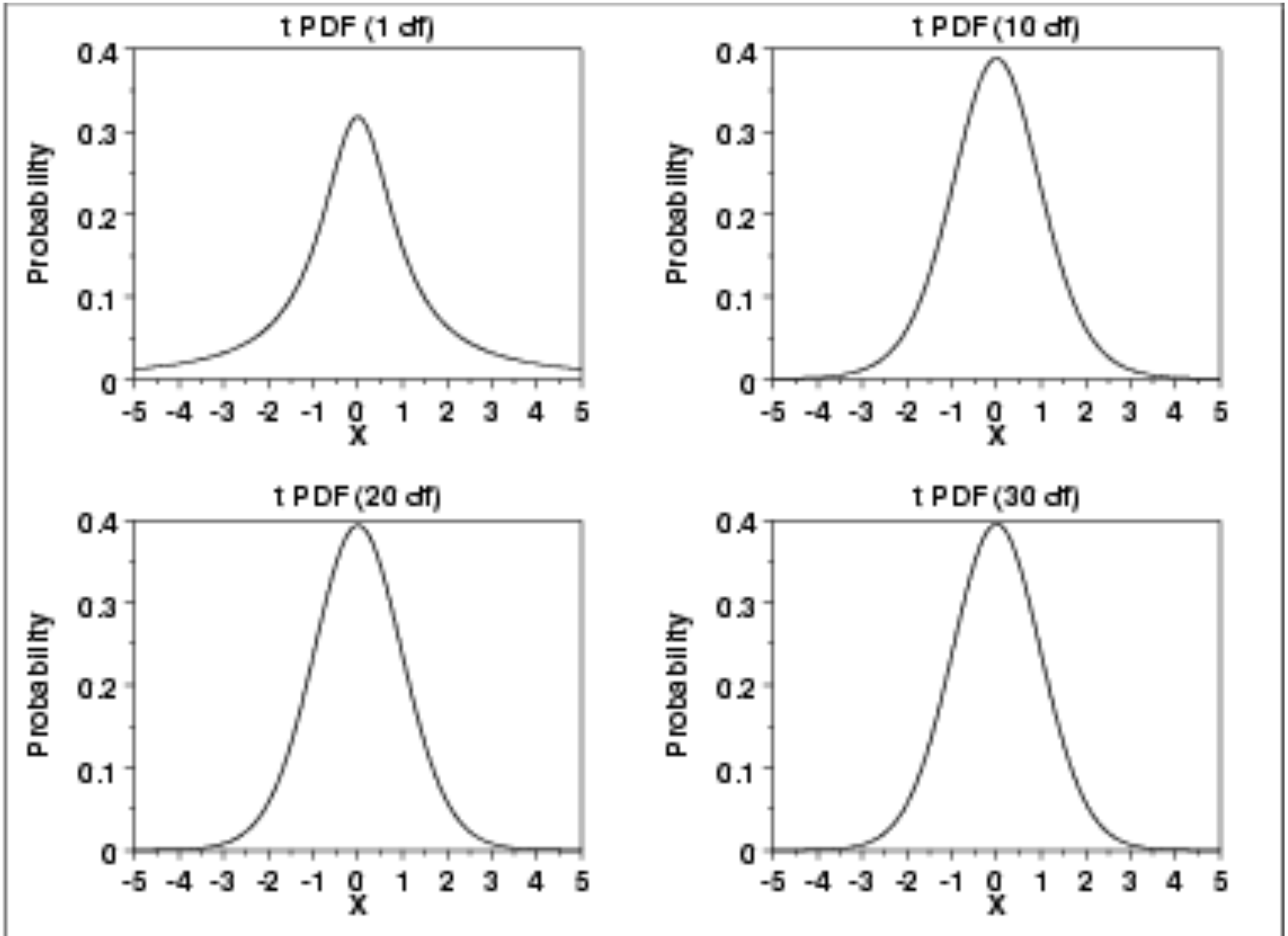
$$f(x) = \frac{1}{B(1/2, k/2)\sqrt{k}} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}$$

  where B(α,β) is the Beta function

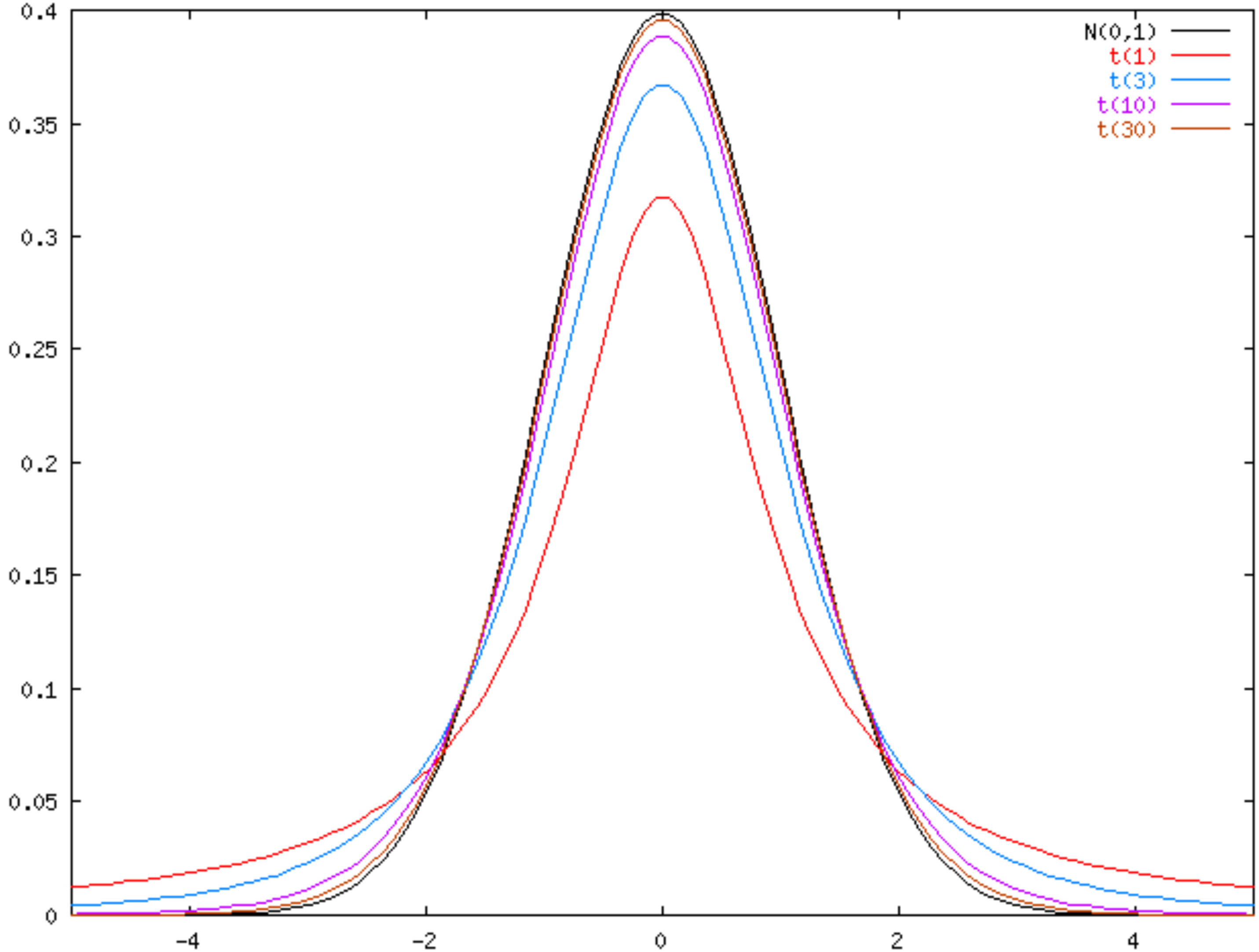
$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

- The Student's *t* distribution has mean *0* and variance *k/(k-2)*

Student's *t* distribution



Student's *t* distribution



Confidence intervals

- Besides the estimate of the mean, to apply the variable transform providing *z* or *t* , we need the estimate of the standard deviation *s*

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\hat{x}^2\right)$$

Example

- A simulation experiment runs 9 times to measure the quantity X and provides the following results:

$$\hat{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = 65 \quad \text{and} \quad \sum_{i=1}^{9}\left(x_i - \hat{x}\right)^2 = 3560$$

- Estimate the variance of X
- Compute the 90% and 99% confidence intervals

Example

- Estimator of the variance:

$$s^2 = \frac{1}{9-1}\sum_{i=1}^{9}\left(x_i - \hat{x}\right)^2 = \frac{3560}{8} = 445$$

- 90% confidence interval:
  - Use the Student's *t* with 8 degrees of freedom
  - 1-α=0.9, then α=0.1, α/2=0.05
  - Use the value $t_{8,0.05}$ =1.86

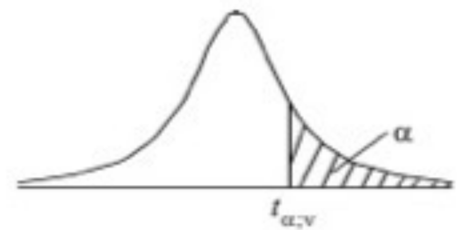
$$I = [\,\hat{x} - t_{n-1,\alpha/2}\, s/\sqrt{n}, \hat{x} + t_{n-1,\alpha/2}\, s/\sqrt{n}\,]$$

Example

- t-student table

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587

Example

$$I = [\, \hat{x} - t_{n-1,\alpha/2} s/\sqrt{n}, \hat{x} + t_{\alpha/2} s/\sqrt{n} \,]$$

$$I_{0.9} = [\, \hat{x} - t_{8,0.05} * s/\sqrt{9}, \hat{x} + t_{8,0.05} * s/\sqrt{9} \,]$$

$$I_{0.9} = [\, \hat{x} - t_{8,0.05} * \sqrt{445}/3, \hat{x} + t_{8,0.05} * \sqrt{445}/3 \,]$$

$$I_{0.9} = [\, 51.92, 78.08 \,]$$

Example

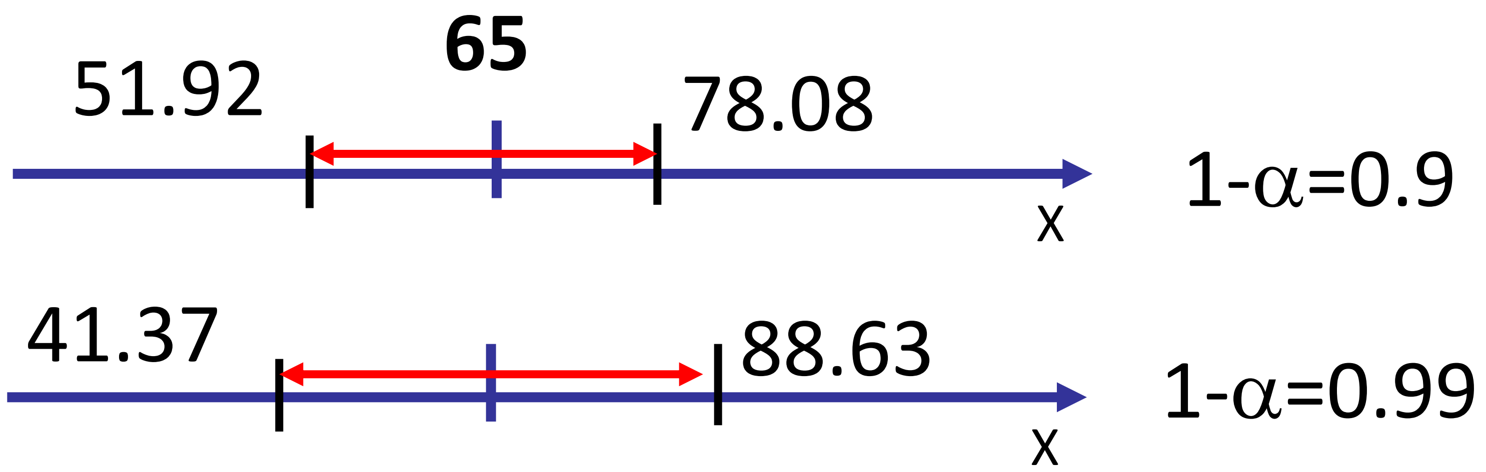
- 99% confidence interval:
  - Use the Student's *t* with 8 degrees of freedom
  - α=0.01, $\alpha/2$=0.005
  - Use the value $t_{8,0.005}$=3.36

$$I_{0.99} = [\,65 - 3.36 * \sqrt{445}/3\,,\,65 + 3.36 * \sqrt{445}/3\,]$$

$$I_{0.99} = [\,41.37, 88.63\,]$$

Example

- Increasing the level of confidence 1-α, the interval becomes wider



Example

```
from scipy.stats import t
import math

t.interval(0.99,
           8,
           65,
           math.sqrt(445/9))
```

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html

Probability of rare events

- In this case:

$$\hat{p} = \frac{\sum 1_{A \cap B}}{\sum 1_B}$$

- Typically $\sum 1_B$=n

- Note that $\hat{p}$ is a $\frac{1}{n}$Bin(n,p), which for n large tends to a Gaussian.
- The standard deviation of $\frac{1}{n}$Bin(n,p), is: s=$\sqrt{\frac{p(1-p)}{n}}$, therefore we can approximate it with: $\hat{s}$= $\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$ and obtain :

$$I = [\widehat{p} - z_{\alpha/2}\hat{s}, \widehat{p} + z_{\alpha/2}\hat{s}]$$