Optimization and Data Science

Lecture 15: Statistical Estimators

Prof. Dr. Thomas Slawig

Kiel University - CAU Kiel
Dep. of Computer Science

Summer 2020

Contents

Statistic Estimators

- What is it?

  Estimation of statistic properties of the distribution of data
- Why are we studying this?

  Important tools in data analysis taking uncertainty into account
- How does it work?

  Definition of estimator functions

  Analysis of their properties (expectation, variance)

  Often: assumption of normal distribution of data
- What if we can use it?

  Analyze data

  Detect "typical" behavior and outliers in data

  Quantify uncertainties

Contents

Random samples and estimators

- We regard given sets of data as sets of realizations of random variables X_i with the same probability distribution.
- We call this a *(random) sample*.

Definition (Estimator)

- A function e of a sample $\{X_i, i = 1, \ldots, n\}$ is called a **(point) estimator** (or also a **statistic**).
- For given realizations x_i of the $X_i, i = 1, \ldots, n$, the realization or value of the estimator is called **estimate**.

- An estimator is a random variable itself.
- Since the X_i in a sample have the same distribution (they are independent and identically distributed: iid), estimators are used for parameters p of the underlying distribution, e.g., expectation, variance.

Estimators: Examples

- For a random sample $\{X_i, i = 1, \dots, n\}$, the **mean value**, defined as

$$e(n, X_i, \dots, X_n) := \bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{1}$$

  is an estimator for the expected value $\mathbb{E}(X_i)$.
- An estimator for the variance is given by

$$e(n, X_i, \dots, X_n) := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{2}$$

  with $\bar{X}$ as in (1).
- Why $\frac{1}{n-1}$ (and not $\frac{1}{n}$) in the last estimator?

$\rightsquigarrow$ We want to have special properties of estimators.

Properties of estimators

Definition (Bias)

- An estimator e is called **unbiased** if, for all lengths n of the sample, its expectation equals the estimated parameter p of the underlying distribution, i.e.,

$$\mathbb{E}(e(n, X_1, \ldots, X_n)) = p \quad \text{for all } n \geq 1.$$

- It is called **asymptotically unbiased** if the above property holds in the limit, i.e.,

$$\lim_{n\to\infty} \mathbb{E}(e(n, X_1, \ldots, X_n)) = p.$$

- The difference

$$\mathbb{E}(e(n, X_1, \ldots, X_n) - p$$

  is called **bias** of the estimator.

Estimator for the expectation

Theorem

For a random sample $\{X_i, i = 1, \ldots, n\}$, the mean value $\bar{X}$ defined in (1) is an unbiased estimator for the expectation $\mathbb{E}(X_i)$ with variance $\mathbb{V}(\bar{X}) = \frac{1}{n}\mathbb{V}(X_i)$.

Proof.

Using linearity of the expectation, we get:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n} n\,\mathbb{E}(X_i) = \mathbb{E}(X_i) \text{ for all } n \geq 1.$$

Using the rules $\mathbb{V}(\alpha X) = \alpha^2\mathbb{V}(X), \mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ for independent X, Y, we get

$$\mathbb{V}(\bar{X}) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n}\mathbb{V}(X_i). \tag{3}$$

$\square$

Unbiased estimator for the variance

Theorem

For a random sample $\{X_i, i = 1, \ldots, n\}$, the estimator (2)

$$e(n, X_1, \ldots, X_n) := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is an unbiased estimator for the variance $\mathbb{V}(X_i)$.

Proof.

- We note that for any random variable X we have

$$(X - \mathbb{E}(X))^2 = X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2,$$

  and using linearity of the expectation:

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X)^2\right) = \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

  or

$$\mathbb{E}(X^2) = \mathbb{V}(X) + \mathbb{E}(X)^2. \tag{4}$$

Unbiased estimator for the variance

- We compute

$$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}(X_i^2-2\bar{X}X_i+\bar{X}^2)=\sum_{i=1}^{n}X_i^2-2\bar{X}\overbrace{\sum_{i=1}^{n}X_i}^{=n\bar{X}}+n\bar{X}^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2.$$

- Using linearity of the expectation and (4) for $X := X_i$ and $X := \bar{X}$, we get

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i^2)-n\mathbb{E}(\bar{X}^2)\\ &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbb{V}(X_i)+\mathbb{E}(X_i)^2\right)-n\big(\overbrace{\mathbb{V}(\bar{X})}^{=\frac{1}{n}\mathbb{V}(X_i)}+\overbrace{\mathbb{E}(\bar{X})^2}^{=\mathbb{E}(X_i)^2}\big)\\ &= \frac{1}{n-1}\left(n\mathbb{V}(X_i)+n\mathbb{E}(X_i)^2-\mathbb{V}(X_i)-n\mathbb{E}(X_i)^2\right)=\mathbb{V}(X_i). \qquad \square\end{aligned}$$

- With n instead of $n-1$ in the denominator of the estimator (2), there would be a bias.

Contents

Discrete distributions

- Uniform distribution (e.g., rolling dice, $n = 6$):

$$P(X = k) = \frac{1}{n}, \quad k = 1, \dots, n.$$

- Geometric distribution (coin toss, success in k-th toss, $p = \frac{1}{2}$):

$$P(X = k) = (1 - p)^{k-1} p, \quad k \in \mathbb{N}_{\geq 1}.$$

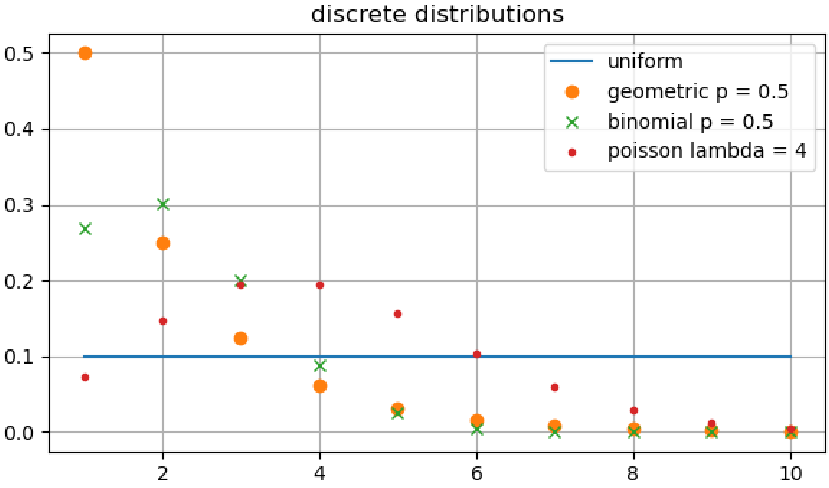
- Binomial distribution (exactly k times success in n tries, probability in every try: p):

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, \dots, n.$$

- Poisson distribution, parameter λ (number of occurences in a given time interval, distribution of fish in a lake):

$$P_\lambda(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

Discrete distributions



Continuous distributions

- Uniform distribution on $[a, b]$, see random number generator:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{else.} \end{cases}$$

- Logistic distribution:

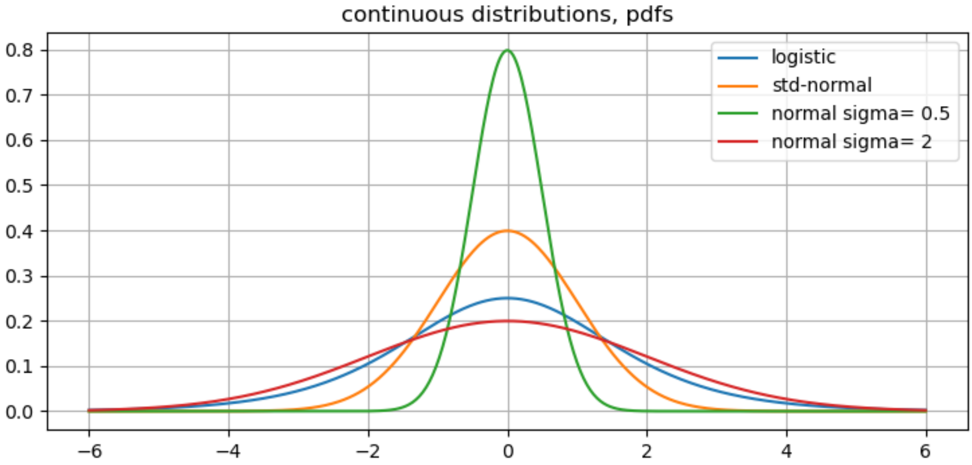
$$f_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

- Normal (Gaussian) distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mathbb{E}(X) = \mu, \mathbb{V}(X) = \sigma^2.$$

- Standard normal distribution: $\mu = 0, \sigma^2 = 1$.
- Central limit theorem: $n \to \infty \rightsquigarrow$ Gauss is good approximation for many distributions.

Continuous distributions

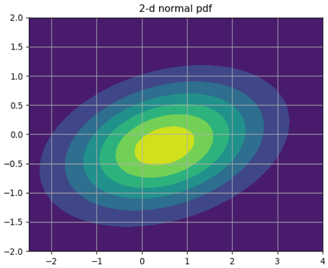


Multi-variate Gauss/normal distribution: $X \sim \mathcal{N}(\mu, \Sigma)$

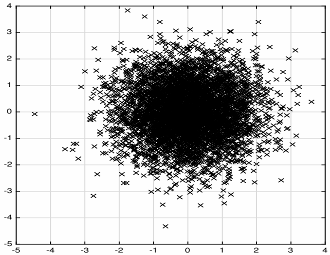
- Random vector: $X = (X_i)_{i=1}^n$. Probability density function:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^n.$$

- Expectation: $\mathbb{E}(X) = \mu \in \mathbb{R}^n$, covariance matrix $Cov(X) = \Sigma \in \mathbb{R}^{n\times n}$.



pdf, $\mu = (0.5, 0.2), \Sigma = \begin{pmatrix} 2 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}$

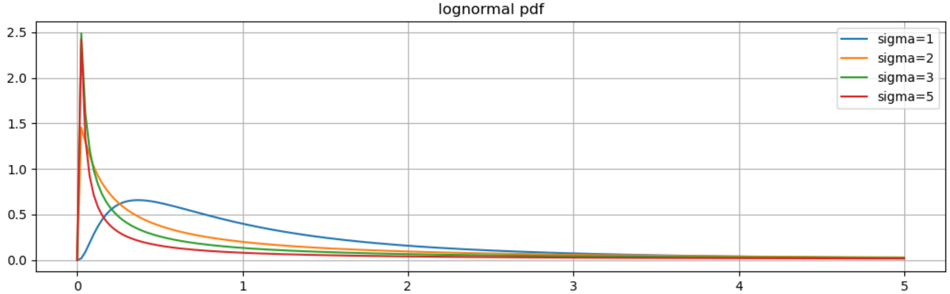


5’000 samples of 2-d standard-normal distributed random vector

Log-normal distribution: $\log X \sim \mathcal{N}(\mu, \sigma^2)$

- Normal distribution has non-zero probabilities for negative values of X.
- If $X > 0$: consider $X := \exp(Y)$ with $Y \sim \mathcal{N}(\mu, \sigma^2) \rightsquigarrow X \sim \log \mathcal{N}(\mu, \sigma^2)$.
- Probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \mathbb{E}(X) = e^{\mu + \frac{\sigma^2}{2}}, \mathbb{V}(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}.$$



Contents

Distribution of estimators of random variables with normal distribution

Theorem

Let a sample $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \dots, n$, *be given. Then, the mean value* (1) *taken as estimator for the expectation* $\mu = \mathbb{E}(X_i)$, *satisfies*

$$e_\mu(n; X_1, \dots, X_n) := \bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \ \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

and the estimator (2) *for the variance* $\sigma^2 = \mathbb{V}(X_i)$ *satisfies*

$$e_{\sigma^2}(n; X_1, \dots, X_n) := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}) \sim \frac{\sigma^2}{n-1} \chi^2(n-1).$$

χ^2 distribution

Definition (χ^2-distribution)

For $k \in \mathbb{N}$ the probability distribution with density function

$$f_k(x) = \frac{1}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x \in \mathbb{R}_{>0}, k \in \mathbb{N}.$$

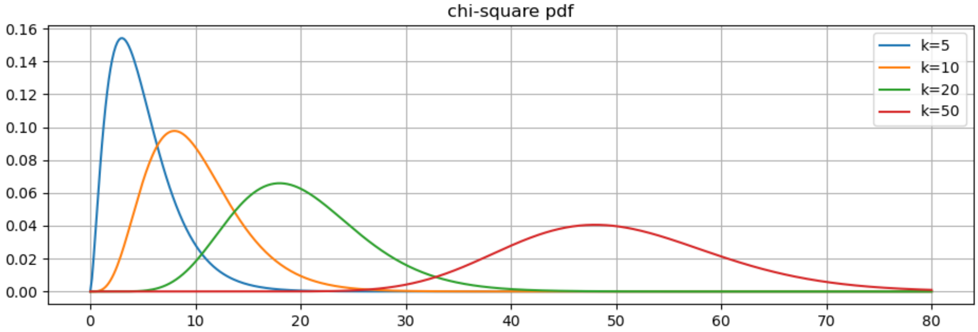
is called χ^2-**distribution of dimension** k or $\chi^2(k)$.

- Gamma function (interpolates the factorial: $x\Gamma(x) = \Gamma(x+1), x \in \mathbb{R}_{\geq 0}$):

$$\Gamma : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, \quad \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

- Values of the χ^2-distribution can be found in tables or in numerical libraries.
- Important for the computation of confidence intervals.

χ^2 distribution



Contents

Confidence intervals

A confidence interval of a random variable is an interval in which the value of the variable lies with a given probability:

Definition

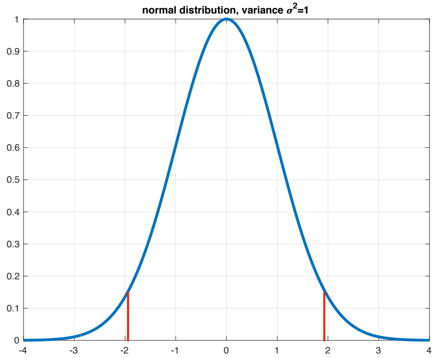
Let X be a random variable and $\gamma \in (0,1)$. Then we call the intervals

$$(-\infty, b) \text{ such that } P(X \leq b) = \gamma,$$
$$(a, \infty) \text{ such that } P(X \geq a) = \gamma,$$

the **one-sided γ-confidence intervals** and

$$(a, b) \text{ such that } P(a \leq X \leq b) = \gamma$$

a **two-sided γ-confidence interval**.



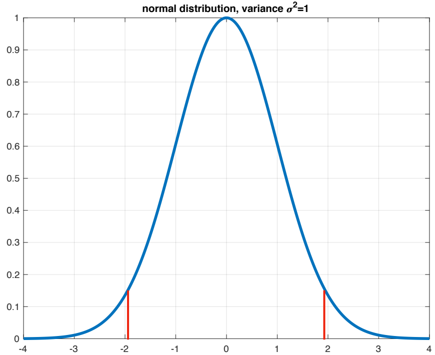
a two-sided 0.95-confidence interval

Interval estimators

- Often, we consider symmetric both-sided confidence intervals centered at the expectation, i.e., we are looking for $c > 0$ such that

$$P(\mathbb{E}(X) - c \leq X \leq \mathbb{E}(X) + c) = \gamma.$$

- For a given sample, a confidence interval can only be estimated.
- These estimators are called **interval estimators**.
- To estimate the interval bounds, we need estimates for expectation and variance ...
- ... and we need to know their distributions.



the symmetric centered two-sided 0.95-confidence interval

Confidence intervals for normal-distributed random variables

- We assume that a given sample $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1 \ldots, n$, is normal-distributed.
- By the Theorem on page 19, the mean $\bar{X}$ as estimator for the expectation $\mathbb{E}(X_i)$ satisfies

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{which gives} \quad \bar{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right).$$

$\leadsto$ Deviation of the mean $\bar{X}$ from the true, but unknown expectation μ, is normal-distributed.

- By the rule for the variance, $\mathbb{V}(\alpha X) = \alpha^2 \mathbb{V}(X), \alpha \in \mathbb{R}$, we get

$$\bar{X} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) \iff (\bar{X} - \mu)\frac{\sqrt{n}}{\sigma} \sim \mathcal{N}(0, 1). \tag{5}$$

$\leadsto$ The deviation (scaled with the factor $\sqrt{n}/\sigma$) of the mean $\bar{X}$ from the true expectation μ, is standard-normal-distributed. Here, σ^2 is the (as well unknown) variance.

- We would like to have the distribution of this deviation, but using the estimate e_{σ^2} from (2) instead of σ^2 itself.

Confidence intervals for normal-distributed random variables

- We want the distribution of the deviation (scaled with the factor $\sqrt{n/e_{\sigma^2}}$) of the mean $\bar{X}$ from the true expectation μ, using the estimate e_{σ^2} from (2).
- We see that

$$(\bar{X}-\mu)\sqrt{\frac{n}{e_{\sigma^2}}} = \underbrace{(\bar{X}-\mu)\frac{\sqrt{n}}{\sigma}}_{\sim\mathcal{N}(0,1)}\sqrt{\frac{n-1}{\frac{n-1}{\sigma^2}e_{\sigma^2}}}.$$

- Again by the Theorem on page 19, we have for the estimator of the variance:

$$e_{\sigma^2} \sim \frac{\sigma^2}{n-1}\chi^2(n-1) \quad \text{or} \quad \frac{n-1}{\sigma^2}e_{\sigma^2} \sim \chi^2(n-1).$$

- Thus, the desired distribution is a combination of two distributions, which is known ...:

$$(\bar{X}-\mu)\sqrt{\frac{n}{e_{\sigma^2}}} = \underbrace{(\bar{X}-\mu)\frac{\sqrt{n}}{\sigma}}_{\sim\mathcal{N}(0,1)}\sqrt{\frac{n-1}{\underbrace{\frac{n-1}{\sigma^2}e_{\sigma^2}}_{\sim\chi^2(n-1)}}}.$$

Confidence intervals for normal-distributed random variables

Theorem

For $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(k), k \in \mathbb{N}$, the random variable $X\sqrt{\frac{k}{Y}}$ is student-$t(k)$-distributed, or short:

$$X\sqrt{\frac{k}{Y}} \sim t(k).$$

Definition (Student-t-distribution)

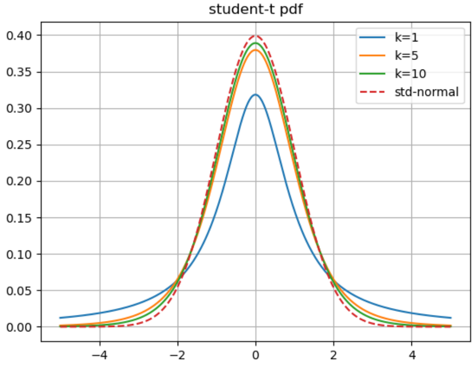
For $k \in \mathbb{N}$ the probability distribution with density function

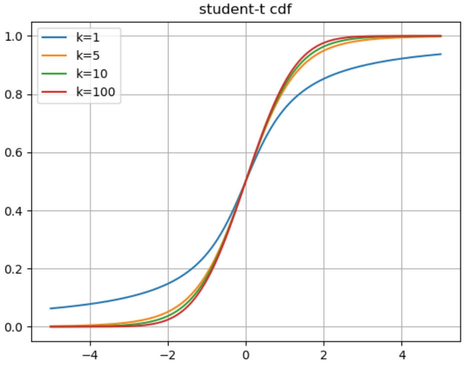
$$f_k(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}, x \in \mathbb{R}, \tag{6}$$

is called **student-** or t-**distribution of dimension** k.

- Values of the cdf and pdf of the t-distribution can be found in tables or library functions.

Student-t-distribution





Confidence intervals for normal-distributed random variables

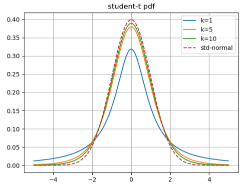
Corollary

Let a sample $X_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n$, be given. Let $\bar{X}$ from (1) be the estimator for the expectation and e_{σ^2} from (2) the estimator for the variance. Then we have

$$(\bar{X} - \mu)\sqrt{\frac{n}{e_{\sigma^2}}} \sim t(n-1).$$

$\rightsquigarrow$ The deviation (scaled with the factor $\sqrt{n/e_{\sigma^2}}$) of the mean $\bar{X}$ from the true expectation μ, is student-$t(n-1)$-distributed:

$$P\left(-c \leq (\bar{X} - \mu)\sqrt{\frac{n}{e_{\sigma^2}}} \leq c\right) = \int_{-c}^{c} f_{n-1}(x)dx = 2\int_{0}^{c} f_{n-1}(x)dx,$$



... where we used that the student-t-pdf is symmetric w.r.t. $x = 0$.

Confidence intervals for normal-distributed random variables

⇝ γ-confidence intervals for the deviation of $\bar{X}$ from the true, unkown expectation μ:

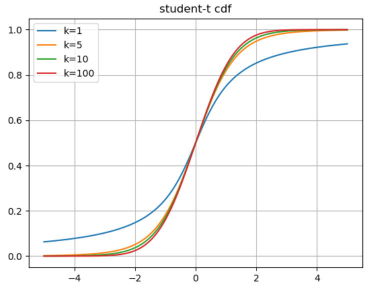
$$P\left(-c\sqrt{\frac{e_{\sigma^2}}{n}} \le \bar{X} - \mu \le c\sqrt{\frac{e_{\sigma^2}}{n}}\right) = P\left(-c \le (\bar{X} - \mu)\sqrt{\frac{n}{e_{\sigma^2}}} \le c\right)$$
$$= 2\int_0^c f_{n-1}(x)dx = 2(F_{n-1}(c) - F_{n-1}(0)) = \gamma,$$

- F_{n-1} is the student-t-cumulative distribution function.
- For given γ, we have to find $c > 0$ with

$$F_{n-1}(c) = \frac{1}{2}(\gamma + F_{n-1}(0)).$$

⇝ We need the inverse cumulative distribution function (quantile function).

- Then, $\pm c\frac{e_{\sigma^2}}{\sqrt{n}}$ are the bounds of the two-sided centered γ-confidence interval for the deviation of $\bar{X}$ from μ.



What is important

- Estimators are used to estimate parameters like expectation and variance for given samples.
- Estimators are also random variables.
- We can compute the expectation and variance of estimators.
- An estimator is (asymptotically) unbiased, if its expectation equals the estimated parameter (in the limit).
- Most important probability distributions are the uniform, the normal and log-normal distributions.
- We often assume that data and samples are normal-distributed.
- For those, we can compute confidence intervals.
- For this purpose, we need χ^2- and student-t-distributions, whose values can be found in tables or by using software libraries.