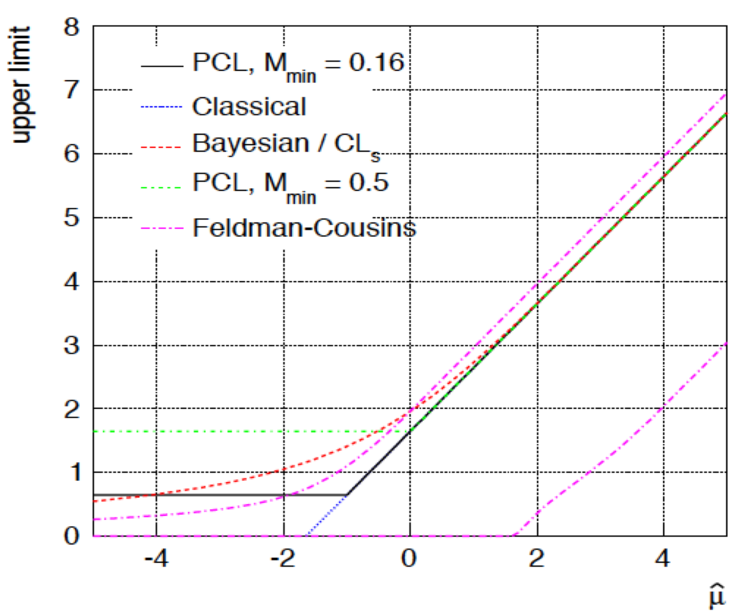
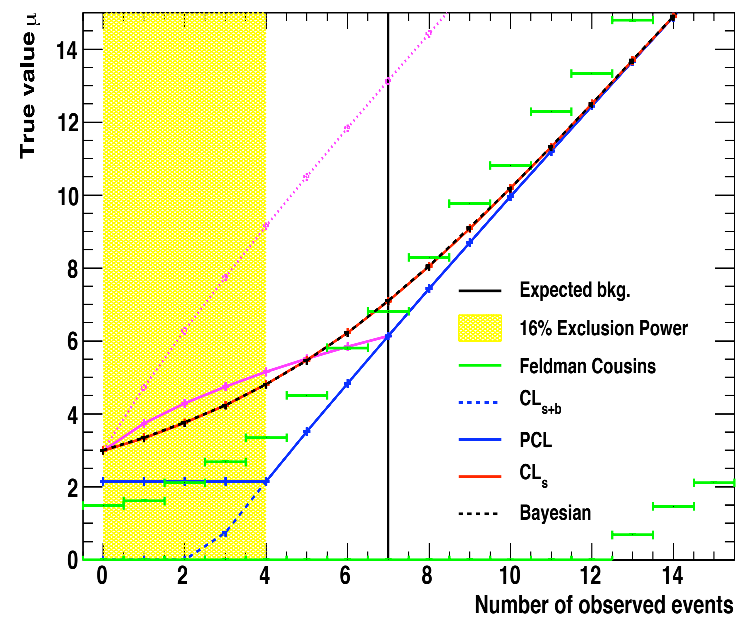
Confidence Intervals and Limits for Pedestrians

Markus Schumacher

GK Lecture, Freiburg, 26 - 28 September 2017





Goal of the lecture: understand the content and interpretation of the two figures

Outline

Lecture 1: Basics (26.9.)

- Motivation
- Frequentist and Bayesian Probability
- Parameter Estimation from Maximum Likelihood
- Frequentist Confidence Intervals a la Neyman and Coverage
- Bayesian Credibility Interval from Likelihood Principle

Lecture 2: Limits for Gaussian Probability Distribution (27.9)

- Connection of Frequentist Limit to Frequentist Hypothesis Test
- Limits close to physical boundary
- Frequentist and Bayesian Limits
- Modified Frequentist: CL_s Method and Power Constrained Limit (PCL)
- Unified Approach, Feldman- Cousins Intervals (FCL)

Lecture 3: Limits for Poisson Distribution (28.9.)

- Confidence Intervals
- Limits close to physical boundary
- Frequentist, Bayesian, PCL, CL_s, FC Limits

Outline

Lecture 1: Basics (Today)

- Motivation
- Frequentist and Bayesian Probability
- Parameter Estimation from Maximum Likelihood
- Frequentist Confidence Intervals a la Neyman and Coverage
- Bayesian Credibility Interval from Likelihood Principle

Lecture 2: Limits for Gaussian Probability Distribution (27.9)

- Connection of Frequentist Limit to Frequentist Hypothesis Test
- Limits close to physical boundary
- Frequentist and Bayesian Limits
- Modified Frequentist: CL_s Method and Power Constrained Limit (PCL)
- Unified Approach, Feldman- Cousins Intervals (FCL)

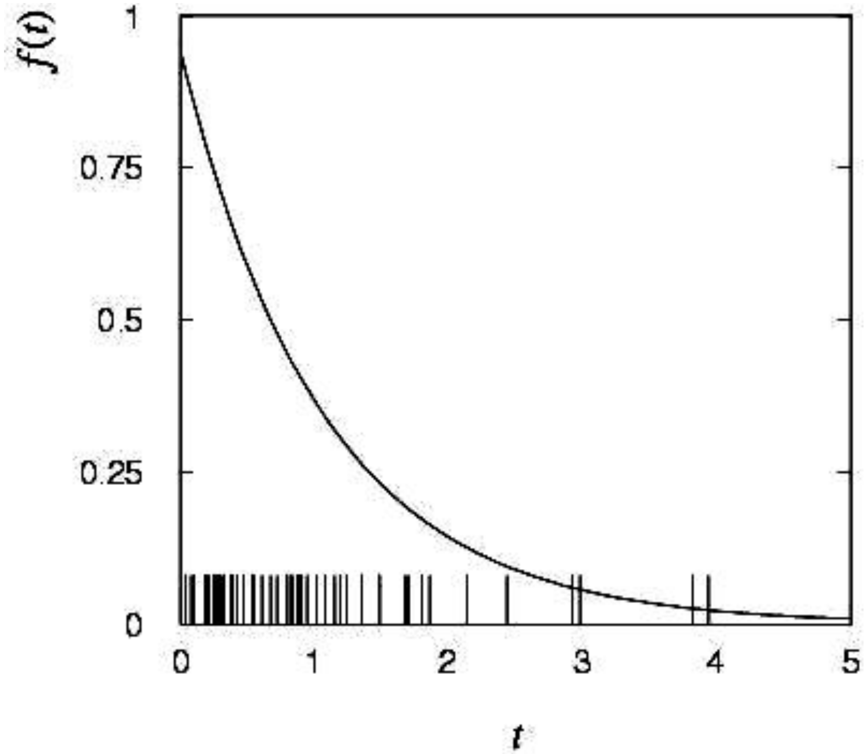
Lecture 3: Limits for Poisson Distribution (28.9.)

- Confidence Intervals
- Limits close to physical boundary
- Frequentist, Bayesian, PCL, CL_s, FC Limits

Motivation

Consider measurement of 50 decay times $_t$ of instabile particle (τ_{true}= 1)
Random variable (RV) t follows exponetial PDF.

$$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$$



Goal 1: estimate of life time

$$\hat{\tau} = 1.062$$

and estimate for dispersion of estimate in repeated identical experiments
→ variance and standard deviation

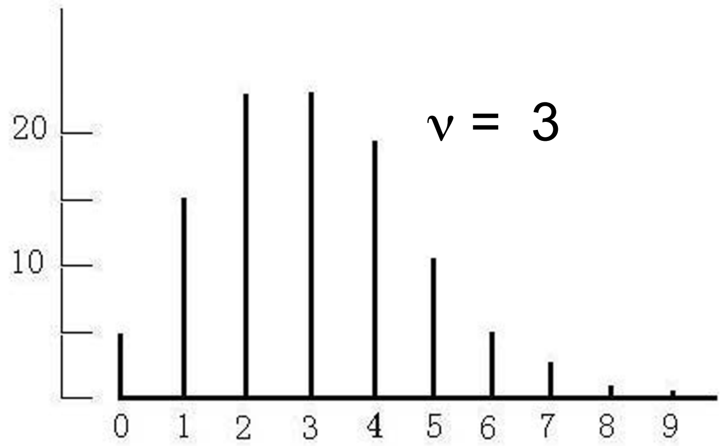
$$\widehat{V[\hat{\tau}]} = \frac{\hat{\tau}^2}{n} \qquad \hat{\sigma}_{\hat{\tau}} = 0.151$$

Goal 2: try to make a probabilistic statement connecting measured value and true value
→ confidence interval [a,b] and/or limit c_{95}

Motivation (2)

Consider measurement of a counting rate $n_{obs}=4$ ($n_{true} = \nu = 3$)
Random variable (RV) n follows Poisson PDF.

$$f(n;\nu) = \frac{\nu^n}{n!}e^{-\nu} \quad (n \geq 0) \qquad E[n] = V[n] = \nu$$



Goal 1: estimate for ν: $\hat{\nu}$

$$\hat{\nu} = n_{obs} = 4$$

estimate of variance $\hat{V}[\hat{\nu}] = \hat{\nu} = 4$
and
of standard deviation $\hat{\sigma}[\hat{\nu}] = \sqrt{\hat{\nu}} = 2$

Naive estimate of confidence interval to CL = 68%

naive CI = $[n_{obs}-\sigma, n_{obs}+\sigma]$ = [2,0;,6,0] length = 4,0
correct frequentist CI = [2,1; 7,2] length = 5,1

→ estimate ± 1-sigma only correct if estimate follows Gauss PDF

Motivation (3)

First step in interpretation:
Estimate of parameter and its variance (often with ML method)

Second step: estimate of
- a two-sided confidence interval [a,b] at 68% confidence level CL
- or single-sided confidence interval = limit c_{95} at 95% CL

which make a statistical statement between outcome of experiment and the true value of a parameter

Two statistical schools: Frequentist and Bayesian statistics
- different method for construction of confidence interval
- numerical identical for sample size $n_{sp} \rightarrow \infty$ and estimated value not close to physical boundary (e.g. estimates $m_\nu^2 = -5 \pm 2$ eV^2 $s = n-b = 0 - 3 = -3$)
- interpretation always different

Modified (pseudo)-frequentist methods:
Power Constrained Limit (PCL), CL_S Limit, Feldman-Cousins Limit (FCL)

Axiomatic Definition of Probability

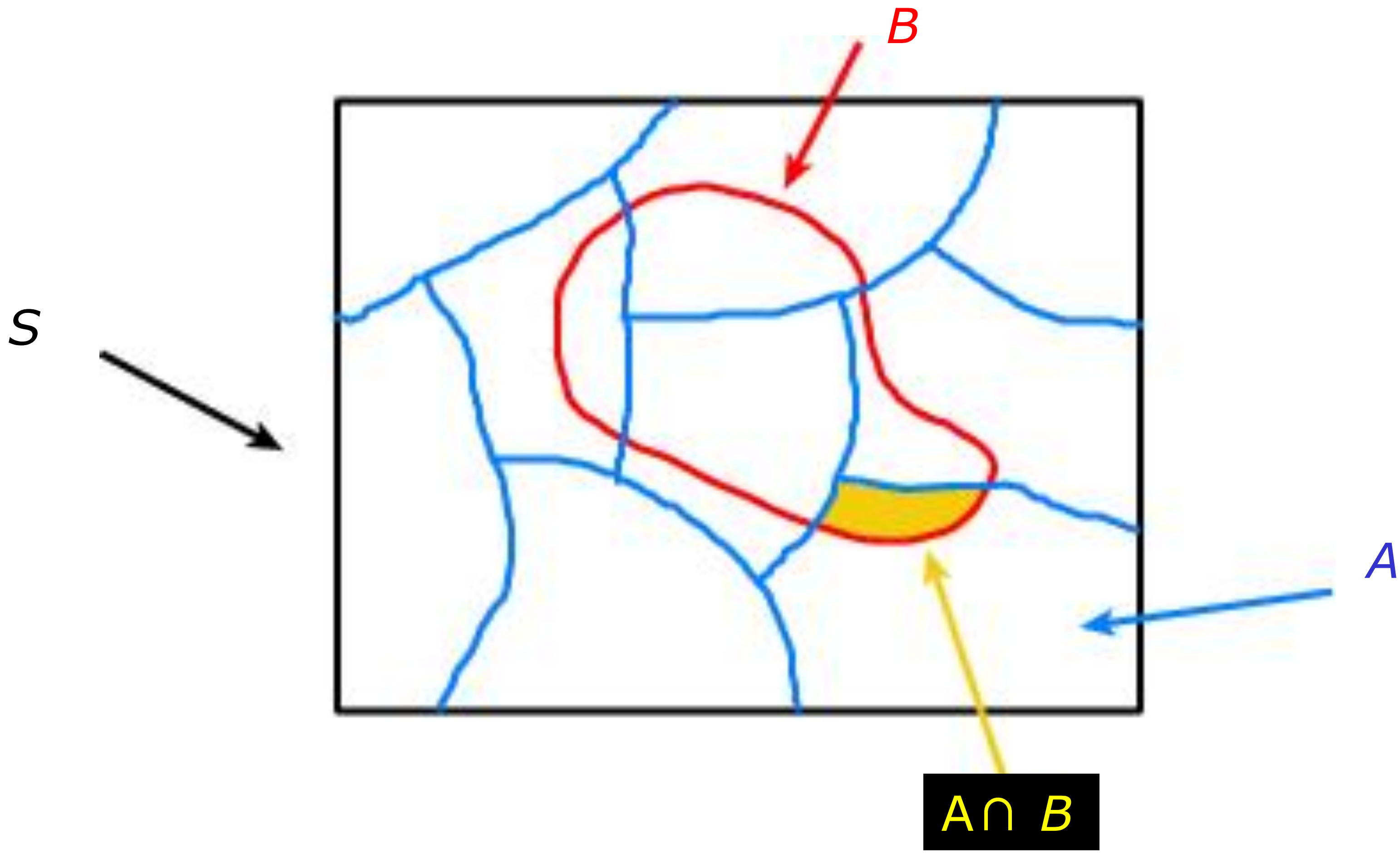
For all $A \subset S, P(A) \geq 0$

$P(S) = 1$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$

Kolmogorov Axioms (1933)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



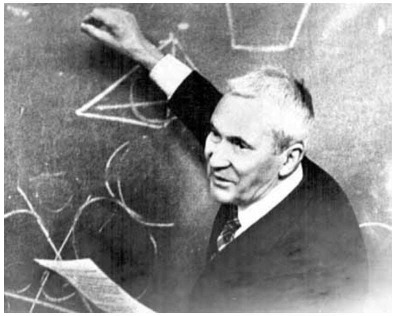
Axiomatic Definition of Probability

Consider set S with subsets A, B, ...
Assign to each set a number between 0 and 1 with

$$\text{For all } A \subset S, P(A) \geq 0$$

$$P(S) = 1$$

$$\text{If } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$$



Kolmogorov
Axioms (1933)

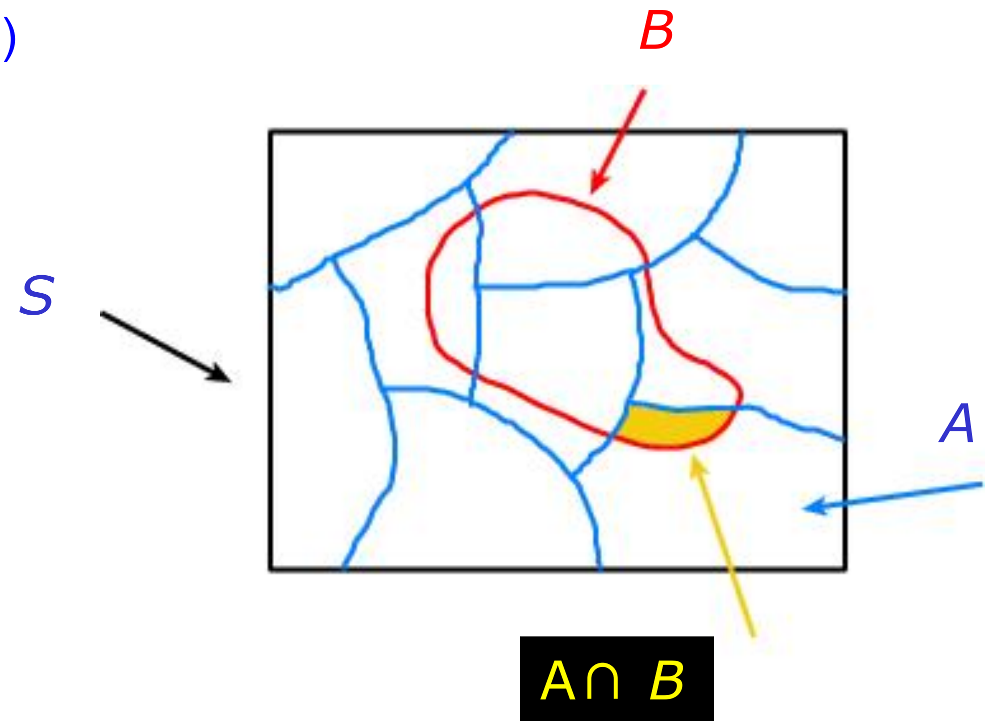
Conditional probability (for $P(B) \neq 0$))

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

If subsets A,B independent:

$$P(A \cap B) = P(A)P(B)$$

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$



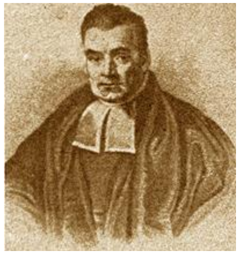
Axiomatic Definition of Probability (2)

From the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \cap B) = P(B \cap A) \qquad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

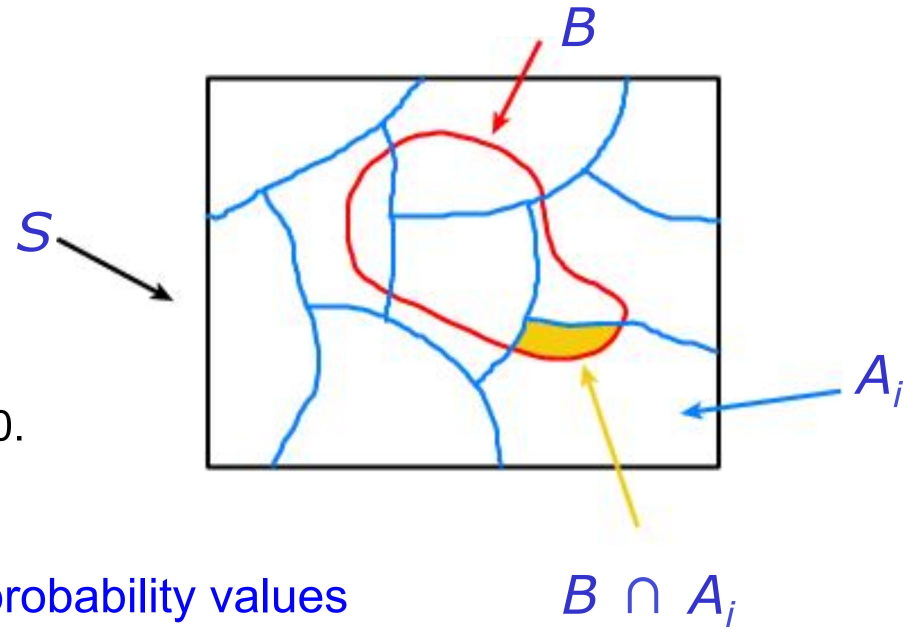
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(B) = \sum_i P(B|A_i)P(A_i)$$



Thomas Bayes (1702−1761)

An essay towards solving a problem in the doctrine of chances, Philos. Trans. R. Soc. **53** (1763) 370.



Axiomatic definition not helpful in real life.
Need: definition of subsets, rule to assign probability values

2 Schools: Frequentists and Bayesians

Bayes Theorems holds and is accepted in both schools

Controversy about: what are the subsets, to which probability values can be assigned

Frequentist and Bayesian

Subsets:

Outcome of (repeatable) experiment

Any hypothesis

Assignment of probabilities:

Relative frequency in limit nr of trials → inf.

$$P(A) = \lim_{n\to\infty} \frac{\text{times outcome is in } A}{n}$$

Degree of belief in hypothesis

$$P(A) = \text{degree of belief that } A \text{ is true}$$

P (SUSY exists)

P (9.81 m/s^2 < g < 9.82m/s^2)

P (rain in Freiburg on 27.9.2017)

Not defined. Either 0 or 1.

No problem. This is the goal.

Bayesian definition: More general (includes Frequentist definition)

Applicable to singular events, "true" values, …

Does not care about repeatability of experiment

Needs a-priori probability in application of Bayes theorem

Bayesian Statistics: General Philosophy

How to use Bayes theorem to update "degree of belief" in light of data

Probability to observe data assuming a hypothesis H (true value of a parameter)
Likelihood function (also used by Frequentists)

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

A-priori probability, i.e. before data taking (not defined in Frequentist school)

Posterior probabilty, i.e. after analysis of the data (not defined in Frequentist school)

Normalisation includes sum/integral over all possible hypothesis/par. values

No general rule for choice of a-priori probability → "subjective"

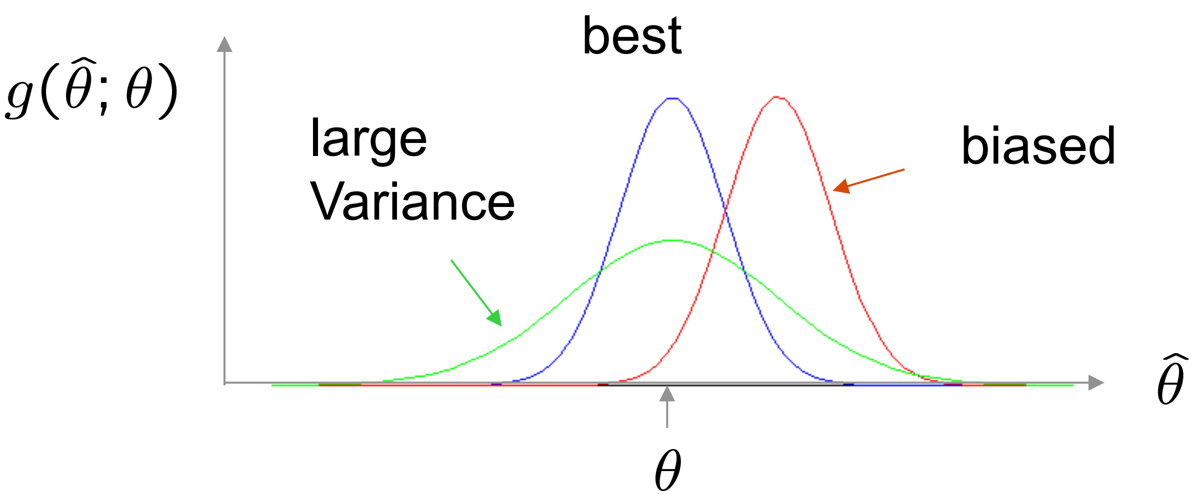
"Objective" prior = uniform?
- not well defined probability for infinite parameter space
- uniform in θ, θ^2 sqrt(θ), ln θ, … ?

→ Jeffrey Prior   p(θ) = sqrt(Information (θ))
uniform for mean μ of Gauss pdf
1/sqrt(μ) for Poisson   1/τ for exp(-t/ τ)

Properties of estimators

Estimator is a function of the sample to determine an unknown parameter

Estimator is a randomvariable and hence has a PDF $g(\hat{\theta};\theta)$



We want small (or vanishing) Bias (systematic error) $b = E[\hat{\theta}] - \theta$

$\rightarrow$ expectation value from repeated measurment should be = true value

We want small variance (statistical uncertainty): $V[\hat{\theta}]$

$\rightarrow$ small bias and small variance are are in general competing criteria

Minimum Variance Bound

In informaton theory one can show, that there is a lower limit for the variance for the estimator of an parameter
(if the sample range is independent on the true paramater value)

Minimmum Variance Bound (MVB) from Rao-Cramer-Frechet-Ineuqality

$$V\left[\hat{\theta}\right] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]}.$$

$$V\left[\hat{\theta}\right] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right]}.$$

Information according to R.A Fisher: $I(\theta) \equiv E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right] = E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]$

$\rightarrow$ the large the information, the smaller the statistical uncertainty

Likelihood and Desired Properties of Estimators

Given a sample of measurements ($x_1..x_n$) for a RV x folllowing PDF f(x;θ) the common PDF for the sample is given by:

$$f(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Considerering the samples fixed, this is called the likelihood:

$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i;\vec{\theta})$$

Consistency

$$\widehat{\theta}^{(n)} \xrightarrow{n\rightarrow\infty} \theta.$$

Bias

$$b^{(n)} = E\left[\widehat{\theta}^{(n)}\right] - \theta$$

bias should be small / „0“
b=0 estimator unbiased

consistent estimators with finite variance are asymptotiically (n→∞) unbiased

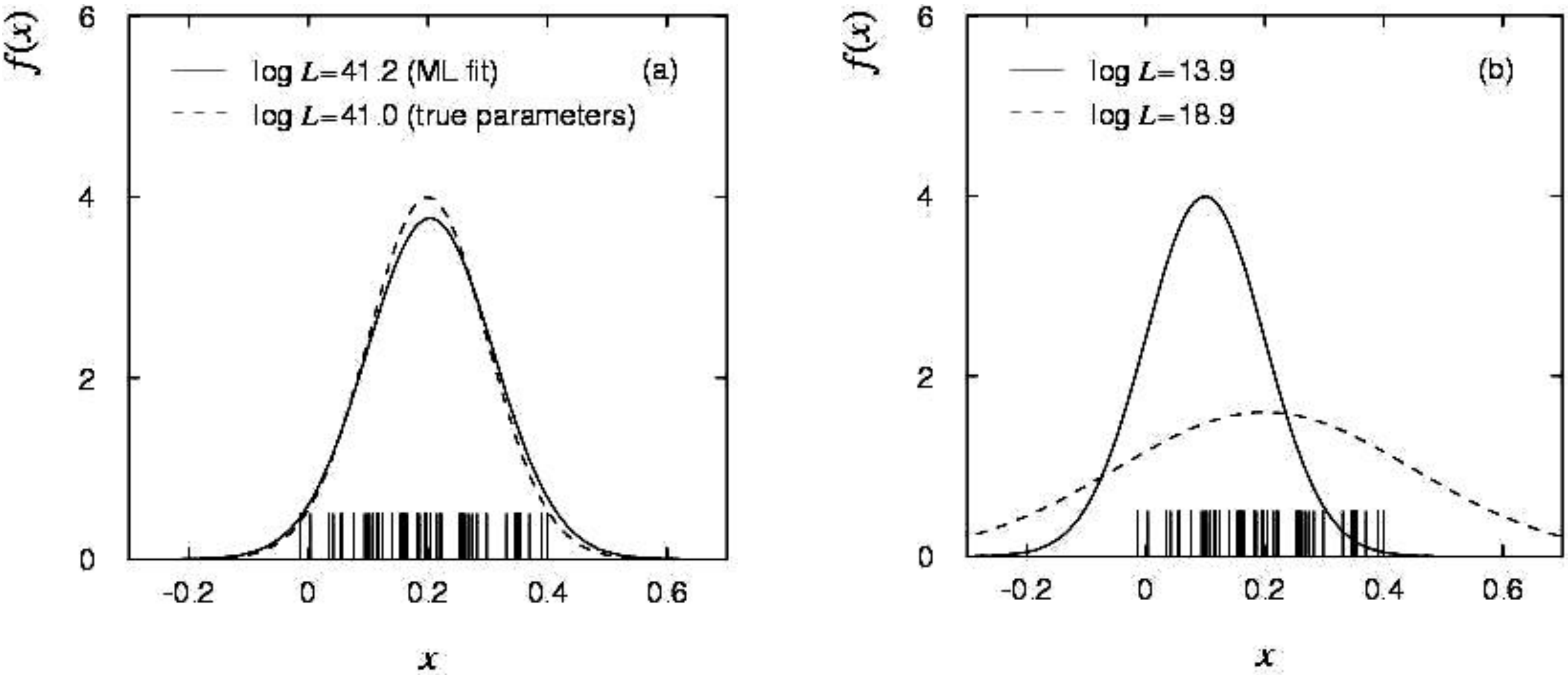
Efficency

$$\text{Effizienz}\left[\widehat{\theta}^{(n)}\right] = \frac{SMV}{V\left[\widehat{\theta}^{(n)}\right]}$$

Efficiency should be close to “1”

Maximum-Likelihood: Basic Idea

If hypothetically value θ close to true value θ_{true},
then probability to observe actual measured sample to is large



Hence define "Maximum Likelihood (ML)" estimator as parameter value, wich maximises likelihood den Parameterwert

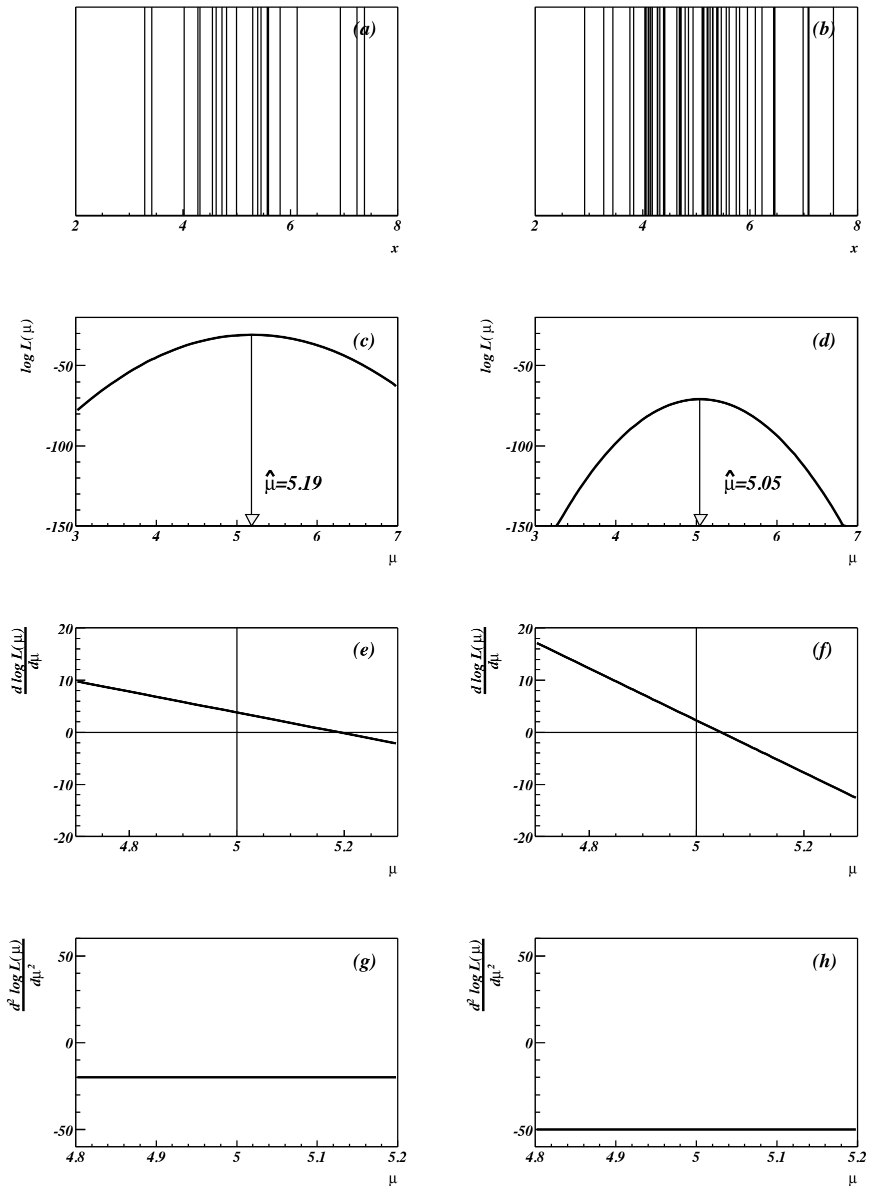
$$\mathcal{L}(\hat{\theta}) = \prod_{i=1}^{n} f(x_i;\hat{\theta}) = \text{Maximum.} \qquad \left.\frac{\partial \log \mathcal{L}}{\partial \theta}\right|_{\hat{\theta}} = 0$$

Estimator for Mean Value of Gauss-PDF

Gauss-PDF
μ=5, σ=1
sample size N= 20

$$\hat{\mu} \approx 5.19$$

$$I(\mu) = 20$$



Gauss-PDF
μ=5, σ=1
sample size N= 50

$$\hat{\mu} \approx 5.05$$

$$I(\mu) = 50$$

Estimator for Mean of Exponential PDF

Consider Exponential PDF: $$f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$$

and sample of n indepdent measurments $$t_1, \ldots, t_n$$

The likelihood is given by $$L(\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$$

The value of τ which maximises $L(\tau)$t, also yields the maximum of its logarthm (lthe log/ln-lIkelihood function):

$$\ln L(\tau) = \sum_{i=1}^{n} \ln f(t_i;\tau) = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

Estimator for Mean of Exponential PDF (2)

$$\log \mathcal{L}(\tau) = \sum_{i=1}^{n} \log f(t_i;\tau) = \sum_{i=1}^{n} \left(\log \frac{1}{\tau} - \frac{t_i}{\tau} \right) = n \log \frac{1}{\tau} - \frac{1}{\tau} \sum_{i=1}^{n} t_i$$

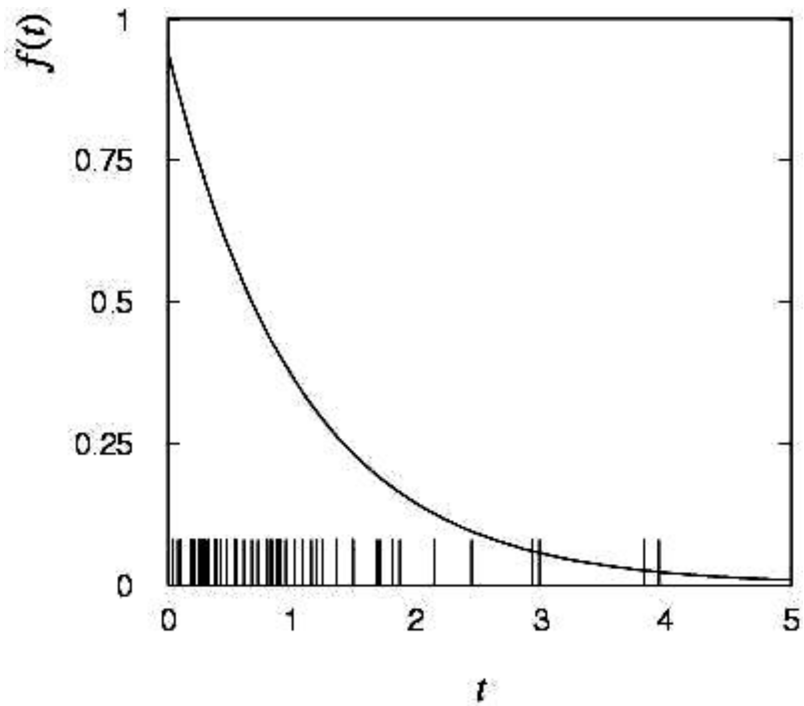
Determination of maximum

$$0 = \left. \frac{\partial \log \mathcal{L}(\tau)}{\partial \tau} \right|_{\tau=\hat{\tau}} = -n \frac{1}{\hat{\tau}} + \frac{1}{\hat{\tau}^2} \sum_{i=1}^{n} t_i$$

Yields the ML estimator:

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

which is the arithmetic mean of the sample and hence consistent and unbiased



Monte Carlo Test:
Generate 50 Measurments for $\tau = 1$.

The ML estimator yields:

$$\hat{\tau} = 1.062$$

Estimator for Mean of Exponential PDF (3)

Variance of sample mean is given by: $$V[\hat{\tau}] = \frac{1}{n}V[t] = \frac{1}{n}\tau^2$$

Comparison with Minimum Variance Bound (MVB) :

$$\frac{\partial^2 \log \mathcal{L}}{\partial \tau^2} = \frac{n}{\tau^2}\left(1 - \frac{2}{n\tau}\sum_{i=1}^{n} t_i\right) = \frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)$$

$$V[\hat{\tau}] \geq \frac{-1}{E\left[\frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)\right]} = \frac{-1}{\frac{n}{\tau^2}\left(1 - \frac{2E[\hat{\tau}]}{\tau}\right)} = \frac{\tau^2}{n}$$

Hence ML Estimator is efficient for this problem

Estimator for Variance

$$\widehat{V[\hat{\tau}]} = \frac{\hat{\tau}^2}{n}$$

Maximum-Likelihood: Estimate of Variance

Graphically $\Delta \ln L = 0{,}5$

$$[1.02 - 0.12, 1.02 + 0.16]$$

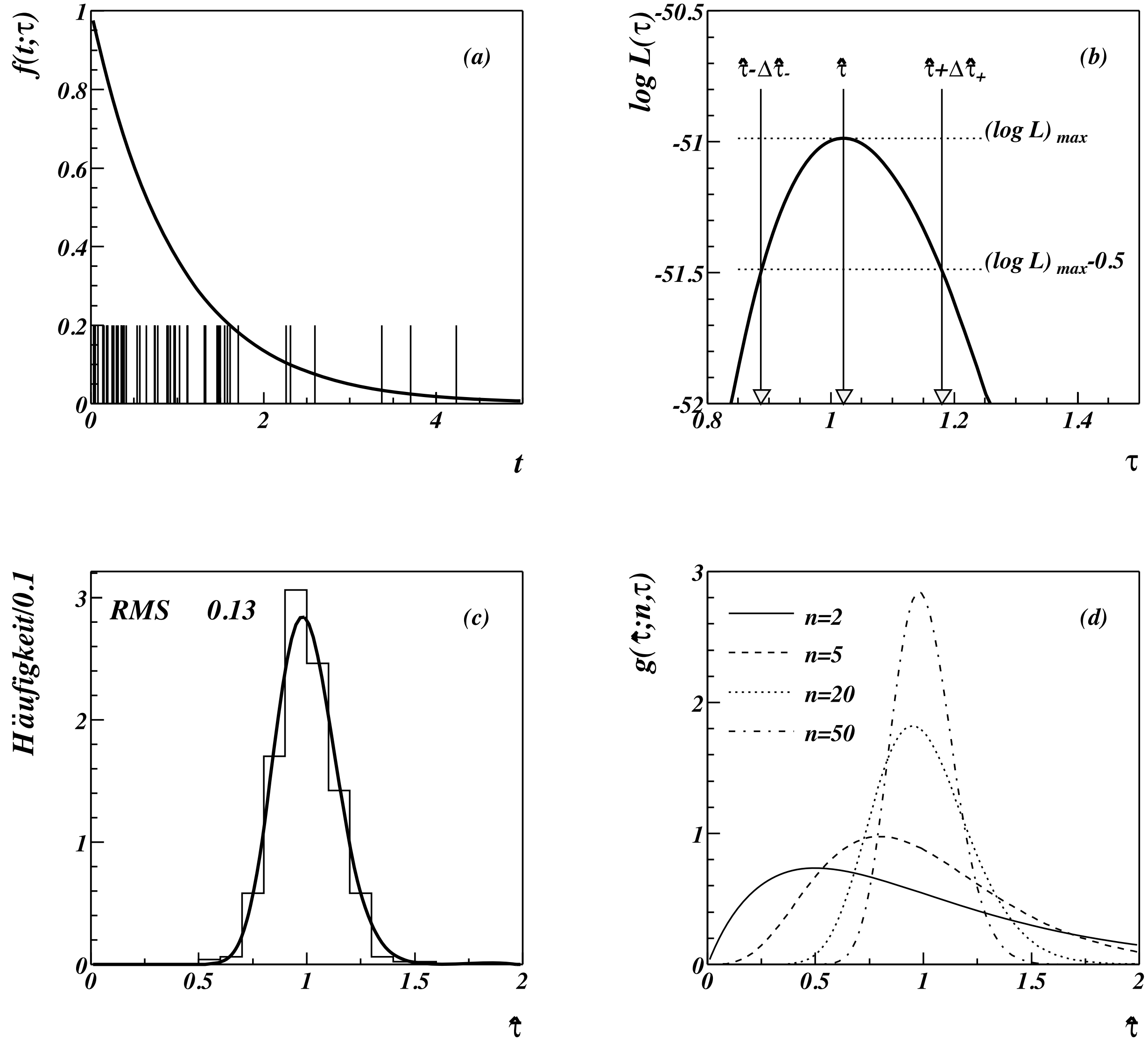
Analytically or curvature in minimum

$$\widehat{V[\hat{\tau}]} = \frac{\hat{\tau}^2}{n} \approx 0.021$$

$$\hat{\sigma} \approx 0.14$$

MC-Method:

$$0.13,$$



Estimator for Mean and Variance of Gauss PDF

Consider n independent measurements $x_1, \ldots, x_n$, from Gauss PDF $f(x;\mu,\sigma^2)$

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

The Log-Likelihood Function is given by:

$$\begin{aligned} \ln L(\mu,\sigma^2) &= \sum_{i=1}^{n} \ln f(x_i;\mu,\sigma^2) \\ &= \sum_{i=1}^{n} \left(\ln \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \ln \frac{1}{\sigma^2} - \frac{(x_i-\mu)^2}{2\sigma^2} \right) . \end{aligned}$$

Set derivatives w.r.t. μ, σ^2 to zero, and solve equations

$$0 = \left. \frac{\partial \log \mathcal{L}(\mu,\sigma^2)}{\partial \mu} \right|_{\mu=\hat{\mu}} = \sum_{i=1}^{n} \frac{(x_i-\hat{\mu})}{\sigma^2} \qquad 0 = \left. \frac{\partial \log \mathcal{L}(\mu,\sigma^2)}{\partial \sigma^2} \right|_{\sigma^2=\widehat{\sigma^2}}$$

Estimator for Mean and Variance of Gauss PDF (2)

Yields the maximum likelihood estimators:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \,, \qquad \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \,.$$

Arithmetic mean is estimator for μ, hence consistent and unbiased

But estimator for variance σ^2 is biased

$$E[\widehat{\sigma^2}] = \frac{n-1}{n} \sigma^2 \,,$$

only asymphthotically unbiased: $b \to 0$ für $n \to \infty$.

Reminder:
is unbiased estimator for variance

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

Properties of ML Estimators

Consistency: if exectation value and variance of estimator finite and sample space independent of parameter

Bias: No general statement possible. Investigate with MC method
Asymptotically ($n \to \infty$) unbiased if consistent.

Efficiency: If an efficient estimator exists it is given by ML method (if sample range independent of parameter)
Efficiency → 1 for $n \to \infty$.

Asymptocially ($n \to \infty$) it holds. WDF für ML-Schätzer:

PDF for estimator converges Gauss PDF.

Likelihood → Gauss and log-Likelihood → parabola

$$\mathcal{L}(\theta) = \mathcal{L}(\hat{\theta}) \exp\left(-\frac{(\hat{\theta}-\theta)^2}{2V[\hat{\theta}]}\right) \qquad \log \mathcal{L}(\theta) = \log \mathcal{L}(\hat{\theta}) - \frac{(\hat{\theta}-\theta)^2}{2V[\hat{\theta}]}$$

Bayesian Parameter Estimator

"Likelihood principle": the results of the measurments is summarised by the likelihood function $L(\theta) = L(\vec{x}|\theta) = f_{\text{joint}}(\vec{x}|\theta)$

Knowledge about parameter updated via Bayes theorem:

$$p(\theta|\vec{x}) = \frac{L(\vec{x}|\theta)\pi(\theta)}{\int L(\vec{x}|\theta')\pi(\theta')\,d\theta'}$$

Bayesian parameter estimate, by maximising the posterior probability $\widehat{\theta}_{\text{Bayes}}$

How to choose prior $p(q)$? Often $\pi(\theta)$ = constant considered.

Then maximum likelihood and Bayesian estimators identical:

$$\widehat{\theta}_{\text{Bayes}} = \widehat{\theta}_{\text{ML}}$$

Example for Parameter Estimation

Frequentist: maximise likelihood

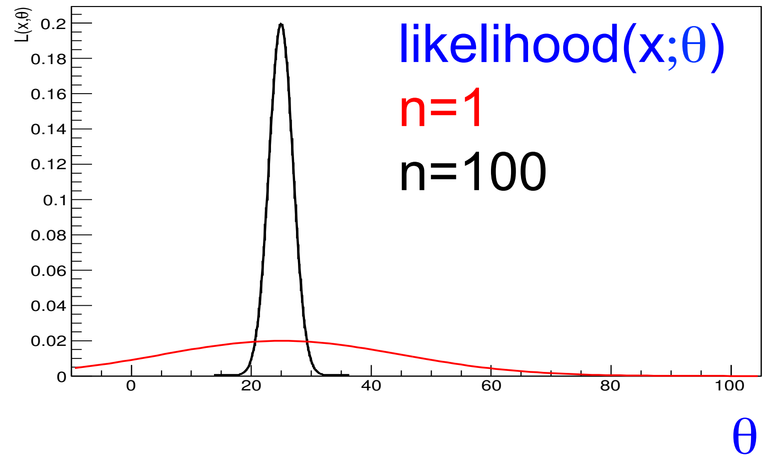
Bayesian: maximise posterior probability

Estimation of mean value θ of Gaussian PDF

Resolution $\sigma = 20$. Sample mean yields: $x = 25$

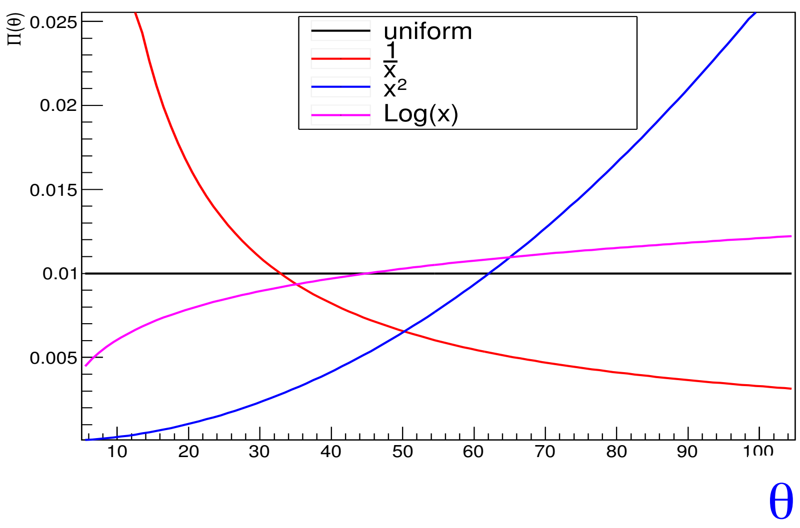
Consider two sample sizes: n= 1 (100)

→ Likelihood functions are Gaussians with $\sigma/\sqrt{n} = 20$ (2)



Four different a-priori probabilities for Bayesian estimate normalised in range 5 to 105

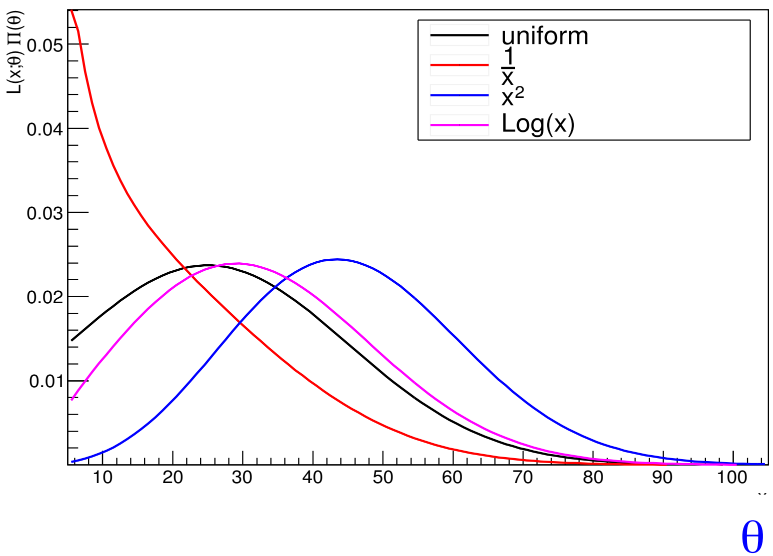
uniform, $1/x$, x^2, $\ln(x)$

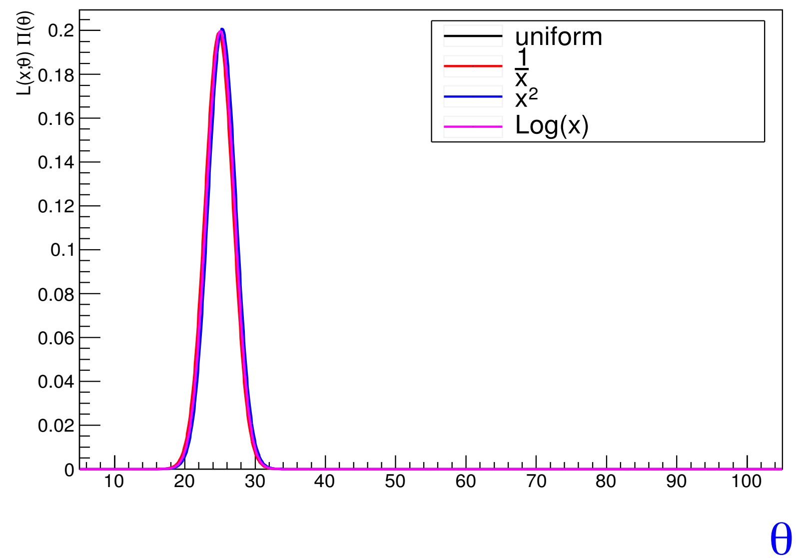


Parameter Estimation - Posterior Probabilities

Sample size n = 1
Large spread in posterior prob.

Large samples size n = 100
Small spread in posterior prob.





Significant dependence of mode on a-priori probability

Small dependence of mode on prior probability

For sample size n → infinity Bayesian and Frequentist results identical
Bayesian with uniform a-priority prob. and Frequentist numerical identical

Exception: in special situations e.g. close to a physical boundary

But interpretation is always different in both schools

Interpretation of CI: Frequentist and Bayesian

CI: Attempt for a probability statement connecting measurement with true value

Frequentist:
- objects to / can not make probability assignment to true values
- construct a confidence interval CI [a,b] at xy% CL from data in such a way that in a sequence of repeated identical measurements the fraction xy% of such intervals contains the true value
- ”the coverage probability of the interval is XY %“
- no problems with “empty” intervals: $m_\nu^2 < -1$ eV^2, $s < -0.3$ @95% CL

Bayesian:
- wants to make statement about probability of true value from single measurement
- credibility interval / Bayesian confidence interval [a,b] at xy% CL
- probability / degree of belief that true values lies in [a,b] is xy%
- coverage and outcome of not observed experiments not interesting
- all information is in observed likelihood function → likelihood principle
- „empty“ intervals are meaningless in Bayesian interpretation but are avoided by an apropiate prior probability

Classical Frequentist Intervals

Neyman construction for equal tailed CI at CL = 1 – α –β = 1- γ   α=β=γ/2

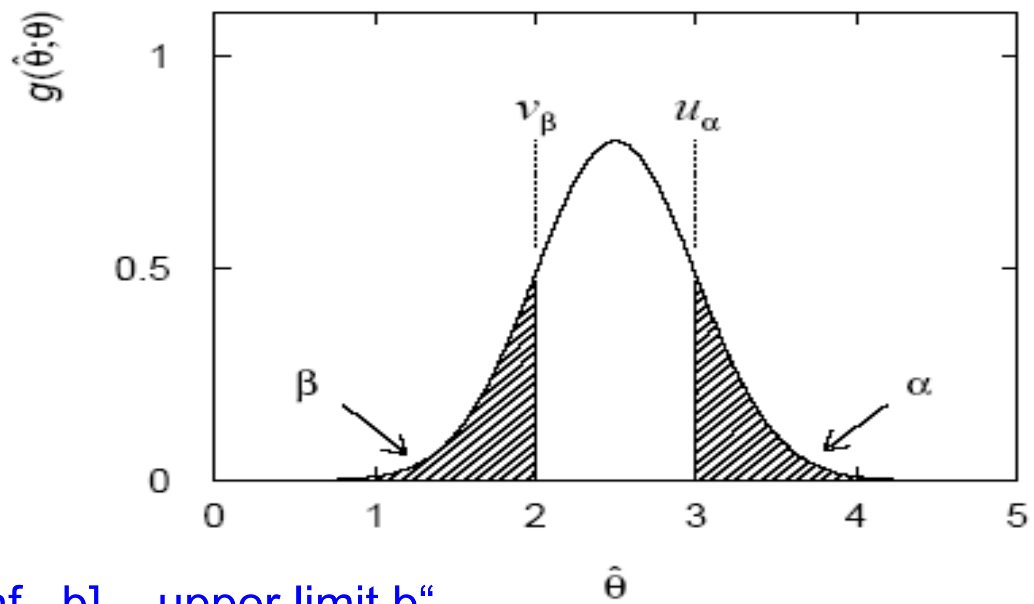
Consider: estimate $\hat{\theta}$ for parameter θ and measured value $\hat{\theta}_{\mathrm{obs}}$.

Need PDF for estimate for all possible true values θ $g(\hat{\theta};\theta)$.

Specify tail probabilities e.g. $\alpha = \beta$ = 0.025 (0.16) and determine functions $u_\alpha(\theta)$ und $v_\beta(\theta)$ with:

$$\alpha = P(\hat{\theta} \geq u_\alpha(\theta)) = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta};\theta)\, d\hat{\theta}$$

$$\beta = P(\hat{\theta} \leq v_\beta(\theta)) = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta};\theta)\, d\hat{\theta}$$



for α = 0 , $u_\alpha(\theta)$ → inf →]- inf., b] „upper limit b“

for β = 0 , $v_\beta(\theta)$ → - inf → [a, + inf] „lower limit a“

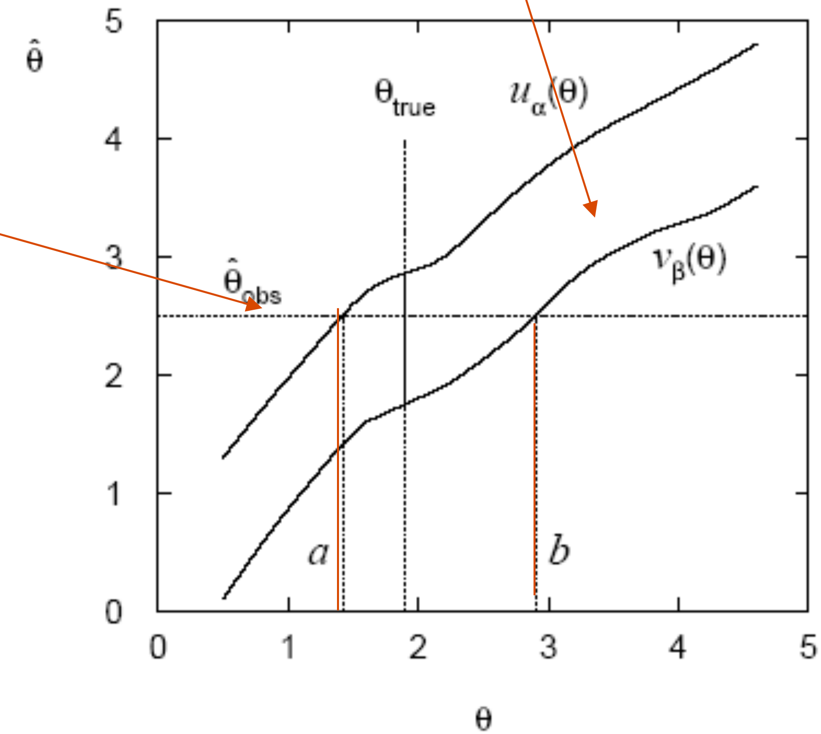
Classical Frequentist Intervals

Region btw. $u_\alpha(\theta)$ and $v_\beta(\theta)$ is the confidence belt

$$P(l_\beta(\theta) \le \hat{\theta} \le u_\alpha(\theta)) = 1 - \alpha - \beta$$

Boundaries of confidence interval given by intersect of observed value with confidence belt → [a,b]

$$a(\hat{\theta}) \equiv u_\alpha^{-1}(\hat{\theta})$$
$$b(\hat{\theta}) \equiv l_\beta^{-1}(\hat{\theta}).$$



For all possible true values θ holds:

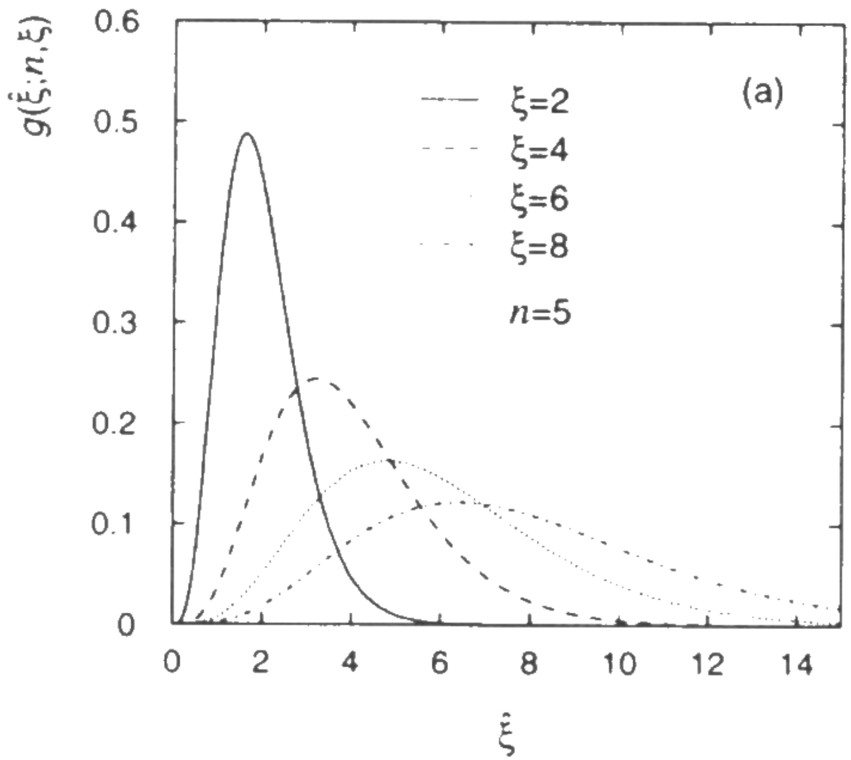
$$\hat{\theta} \ge u_\alpha(\theta) \Leftrightarrow a(\hat{\theta}) \ge \theta \qquad P(a(\hat{\theta}) \ge \theta) = \alpha$$
$$\hat{\theta} \le l_\beta(\theta) \Leftrightarrow b(\hat{\theta}) \le \theta, \qquad P(b(\hat{\theta}) \le \theta) = \beta$$

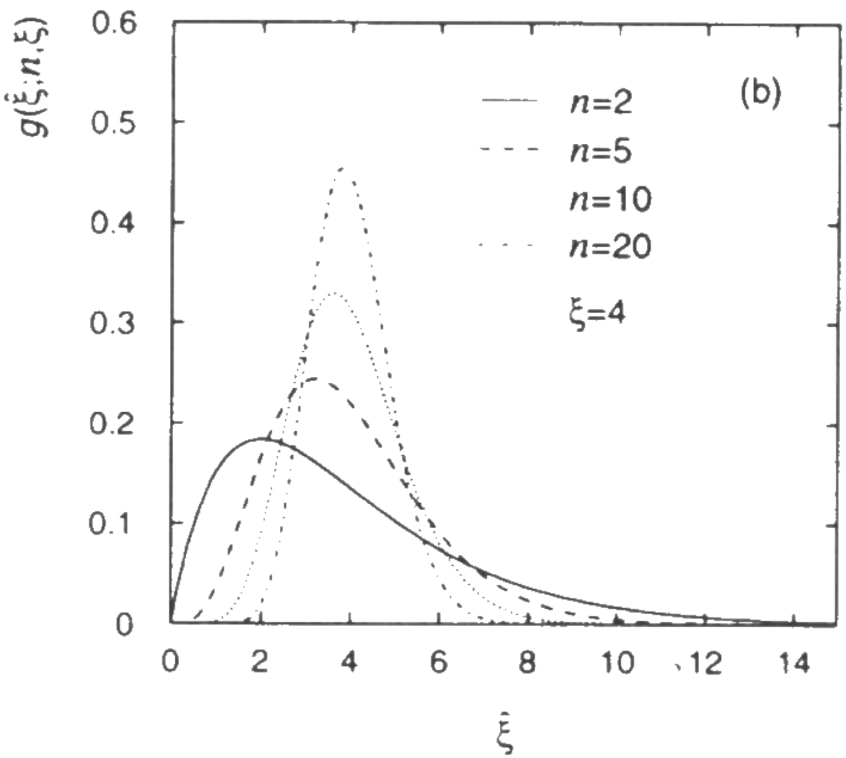
$$P(a(\hat{\theta}) \le \theta \le b(\hat{\theta})) = 1 - \alpha - \beta.$$

Correct coverage by construction

Construction of CI for Exponential PDF

ML-Schätzer Estimator = arithmetic mean of lifetimes
PDF for ML estimator is special case of gamma function





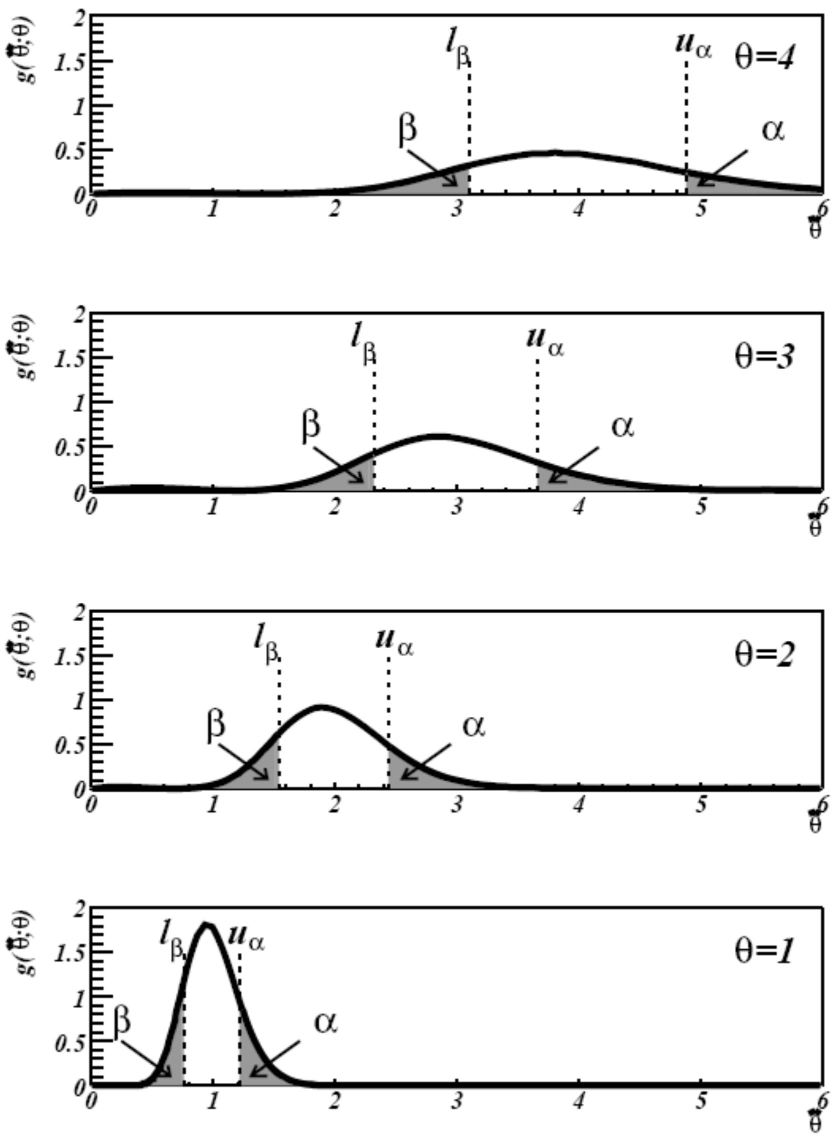
PDF for N_{SP}=5 various true lifetimes

for true lifetime = 4,various N_{SP}

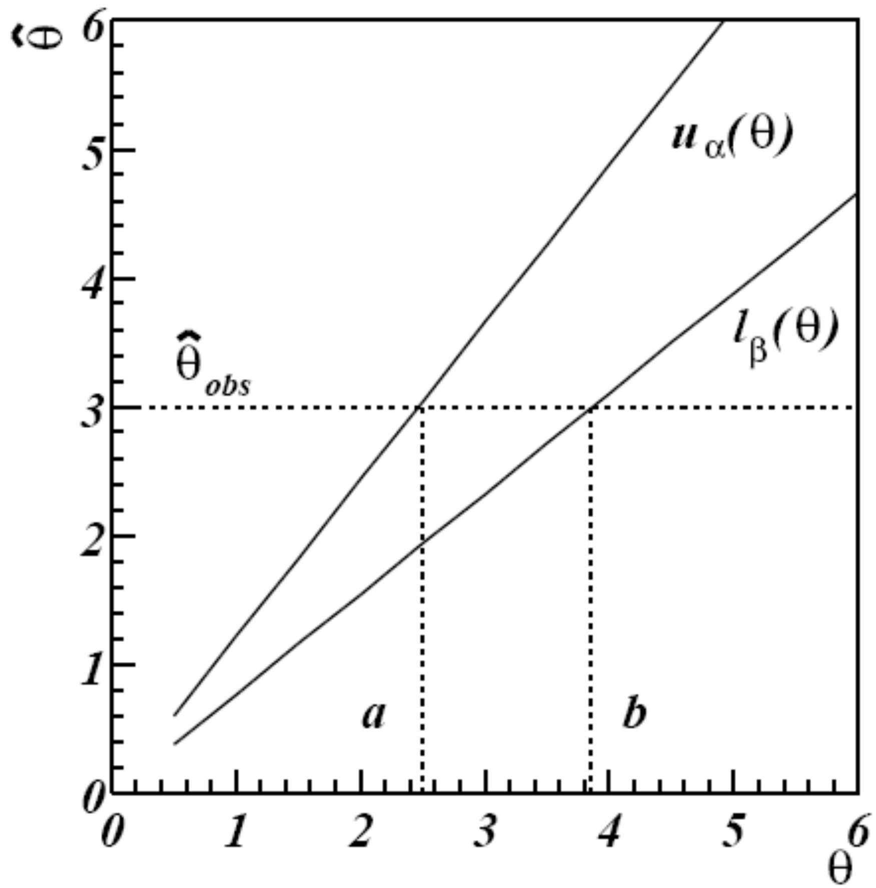
for $N_{SP} \rightarrow \infty$ PDF converegs to Gausian PDF due to Central Limit Theorem

Construction of CI for Exponential PDF (2)

PDF for ML-estimator for various true life times for sample size N_{SP} = 20



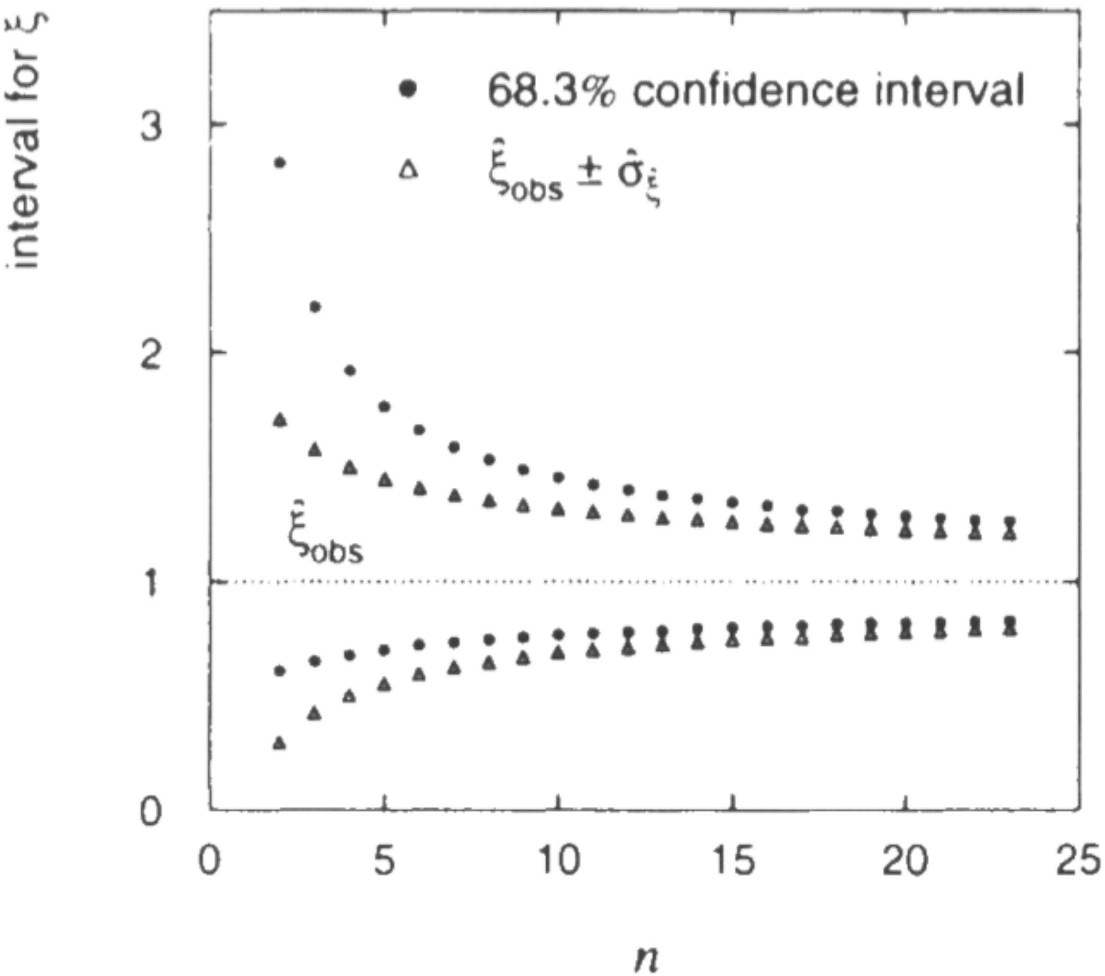
Confidence belt: u_α and l_β



Confidence interval [a,;b]

Construction of CI for Exponential PDF (3)

Comparison of CI from estimator ± 1 standard deviation (triangles)
and from correct Nyman Construction (points)



for $n = N_{SP} \rightarrow \infty$ both CI get identical as PDF for estimator → Gauss-PDF f_{Gauss}

for finite/small $n = N_{SP}$

- correct Neyman CI longer
- coverage of naive CI smaller than claimed CLI

Bayesian Credibility Interval

Result from experiment is posterior PDF for true parameter value θ

$$P(\theta; x_{SP}) = const.\ \mathcal{L}(x_{SP}; \theta)\pi(\theta)$$

Integrate posterior PDF to get CI [a,b] at credibility CL = 1 - $\alpha - \beta$

$$\alpha = \int_{-\infty}^{a} P(\theta; x_{SP}) d\theta$$

lower limit [a, ∞[

$$\beta = \int_{b}^{\infty} P(\theta; x_{SP}) d\theta$$

upper limit]-∞,b]

$$1 - \alpha - \beta = \int_{a}^{b} P(\theta; x_{SP}) d\theta$$

two sided CI [a,b]

Implement physical boundary via $\pi(\theta)$: $\pi(\theta) = 0$ in unphysical region

Repeatability of experiment and coverage is not of (main) interest for Bayesian