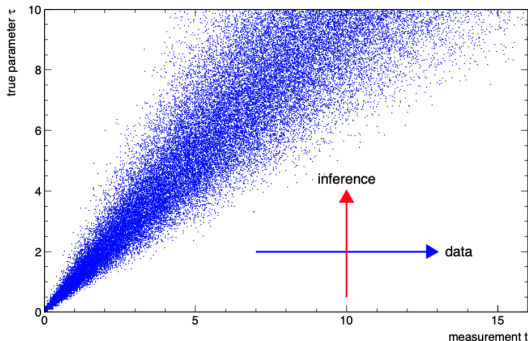


3 Information, inference, confidence level intervals

❖ (assumed) known: distribution of measurements t as a function of a parameter τ



► actual data are on the line $\tau = \tau_{\text{true}}$

from data to parameter →

The likelihood function

❖ consider measurements x which sample a PDF that depends on parameters θ

■ function of two types of - in general multidimensional - arguments

$$f(x, \theta) \quad \text{with} \quad \int dx f(x, \theta) = 1 \quad \forall \theta \quad \text{and thus} \quad \int dx \frac{\partial}{\partial \theta} f(x, \theta) = 0$$

■ inferring θ from information contained in measurements x

► when x given, then f is the **likelihood function** for the parameters θ

$$L = L(\theta) = L(\theta; x) = f(x, \theta) \quad \text{with} \quad x = \text{fixed}$$

► the likelihood function is **not normalized** when integrated over θ

► the likelihood function is **invariant** under re-parametrisation

$$L(\phi; x) = L(\theta; x) \quad \text{when} \quad \theta = \theta(\phi)$$

► **always remember:** the likelihood function is **NOT** a PDF

measures of information →

Poisson: information $\propto$ sample size of a counting experiment

❖ consider weighted events w_i , $i = 1, \dots, N$, from a Poisson process

■ total sum of weights and variance

$$W = \sum_{i=1}^N w_i \quad \text{and} \quad V = \sum_{i=1}^N w_i^2$$

■ relate significance to be different from zero to Poisson error of unweighted events

$$\frac{W}{\sqrt{V}} = \sqrt{N_{\text{eq}}}$$

■ result

$$N_{\text{eq}} = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2} \rightarrow N_{\text{eq}} = N \quad \text{if} \quad w_i = \text{const}$$

Shannon, Kullback-Leibler: information $\propto$ bits learned/event

■ **starting point:** equally probable outcomes for 2^n possibilities

▶ $n = 0 \rightarrow$ probability $p = 1 \rightarrow I = 0$: no information obtained

▶ $n = 1 \rightarrow$ probability $p = 1/2 \rightarrow I = 1$: one bit learned

▶ $n = 2 \rightarrow$ probability $p = 1/4 \rightarrow I = 2$: two bits learned

general pattern: $I = -\log_2 p$ or equivalently $S = -\ln p$

■ **generalization:** weighted sum for non-uniform probabilities p_i

$$S = -\sum_i p_i \ln p_i = \langle -\ln p \rangle \quad \text{Shannon entropy}$$

■ **further generalization:** information gain w.r.t. prior knowledge q

$$S = \sum_i p_i (\ln p_i - \ln q_i) = \sum_i p_i \ln \frac{p_i}{q_i} = \left\langle \ln \frac{p}{q} \right\rangle \quad \text{Kullback-Leibler divergence}$$

a.k.a “cross-entropy”, well defined for continuous PDFs

Fisher: information $\propto 1/\text{variance}$

❖ considerations regarding quantifying the information content of a data set

- information should scale linearly with the number of independent measurements
- information is defined for a parameter that affects the measurement
- information reflects the precision of a parameter estimate based on a measurement

→ ansatz: information = precision = $1/\text{variance}$ of a parameter estimate

Study an unbiased estimator $\hat{\theta}(x)$ for a parameter θ from a measurement x in a system described by the PDF $f(x, \theta)$:

$$\langle \hat{\theta} - \theta \rangle = \int dx f(x, \theta) (\hat{\theta}(x) - \theta) = 0 \quad \text{and thus} \quad \frac{\partial}{\partial \theta} \langle \hat{\theta} - \theta \rangle = 0$$



source: wikipedia.org

it follows

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int dx f(x, \theta) (\hat{\theta}(x) - \theta) = \int dx \frac{\partial f(x, \theta)}{\partial \theta} (\hat{\theta}(x) - \theta) - \int dx f(x, \theta) \\ &= -1 + \int dx f(x, \theta) \frac{\partial \ln f(x, \theta)}{\partial \theta} (\hat{\theta}(x) - \theta) \\ &= -1 + \int dx \left[(\hat{\theta}(x) - \theta) \sqrt{f(x, \theta)} \right] \left[\frac{\partial \ln f(x, \theta)}{\partial \theta} \sqrt{f(x, \theta)} \right] \end{aligned}$$

and with the Cauchy-Schwarz inequality

$$\left(\int dx A(x) B(x) \right)^2 \leq \left(\int dx A^2(x) \right) \cdot \left(\int dx B^2(x) \right)$$

one finds for the square of the x -integral

$$1 \leq \underbrace{\left[\int dx (\hat{\theta}(x) - \theta)^2 f(x, \theta) \right]}_{\text{Var}(\hat{\theta})} \cdot \underbrace{\left[\int dx \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) \right]}_{I(\theta)}$$

❖ result: the “information inequality” or “Cramér-Rao bound”

$$\frac{1}{\text{Var}(\hat{\theta})} \leq I(\theta) \quad \text{with} \quad I(\theta) = \left\langle \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 \right\rangle$$

► $\text{Var}(\hat{\theta})$ cannot be smaller than the inverse of the Fisher information $I(\theta)$

❖ additional properties

■ $I(\theta)$ is linear in the number of IID measurements $x = \{x_1, \dots, x_N\}$

$$f_N(x, \theta) = \prod_{i=1}^N f(x_i, \theta) \quad \text{and} \quad I_N(\theta) = \left\langle \left(\frac{\partial \ln f_N(x, \theta)}{\partial \theta} \right)^2 \right\rangle = N I_1(\theta)$$

■ higher-dimensional parameter spaces: Fisher information matrix

$$I_{kl}(\theta) = \left\langle \frac{\partial \ln f(x, \theta)}{\partial \theta_k} \cdot \frac{\partial \ln f(x, \theta)}{\partial \theta_l} \right\rangle$$

❖ example: Fisher information for the mean of the Poisson distribution

$$f(n, \mu) = e^{-\mu} \frac{\mu^n}{n!} \rightarrow \frac{\partial \ln f}{\partial \mu} = \frac{\partial}{\partial \mu} (-\mu + n \ln \mu - \ln n!) = \frac{n}{\mu} - 1$$

the Fisher information then is given by

$$I(\mu) = \left\langle \left(\frac{\partial \ln f}{\partial \mu} \right)^2 \right\rangle = \left\langle \frac{n^2}{\mu^2} - \frac{2n}{\mu} + 1 \right\rangle$$

for the Poisson distribution one has

$$\langle n \rangle = \mu \quad \text{and} \quad \langle n^2 \rangle = \mu^2 + \mu$$

and it follows

$$I(\mu) = \frac{1}{\mu}$$

- ▶ Fisher information depends on the unknown true parameter
- ▶ more information per measurement from more narrow distributions

problems and exercises

Determine the Fisher information for the mean value of a gaussian with mean μ and standard deviation σ .

The gaussian PDF and its logarithm are given by

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad \text{and} \quad \ln f(x, \mu, \sigma) = -\frac{(x - \mu)^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma)$$

it follows

$$\frac{\partial \ln f(x, \mu, \sigma)}{\partial \mu} = \frac{x - \mu}{\sigma^2}$$

and using $\langle (x - \mu)^2 \rangle = \sigma^2$, the Fisher information is found to be:

$$I(\mu) = \left\langle \frac{(x - \mu)^2}{\sigma^4} \right\rangle = \frac{1}{\sigma^2}$$

❖ estimating the mean value of a PDF

- Fisher information about the mean of N IID gaussian measurements

$$I(\mu) = \frac{N}{\sigma^2}$$

- Cramer-Rao bound

$$\text{Var}(\hat{\mu}) = \frac{1}{I(\mu)} = \frac{\sigma^2}{N}$$

- the bound is saturated by the arithmetic mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{with error} \quad \sigma(\hat{\mu}) = \frac{\sigma}{\sqrt{N}}$$

- ▶ the arithmetic mean is the best possible unbiased estimator for a gaussian mean
- ▶ it is in fact unbiased for any PDF

check whether it is also optimal →

- find the PDF $f(x - \mu)$ with unit variance that minimizes the Fisher information

$$I_f(\mu) = \left\langle \left(\frac{\ln f(x - \mu)}{\partial \mu} \right)^2 \right\rangle = \int dx f(x - \mu) \left(\frac{\partial \ln f(x - \mu)}{\partial \mu} \right)^2 \stackrel{!}{=} \min$$

- standard problem of variational calculus: solution with minimum Fisher information

$$f(x - \mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \mu)^2\right) \quad \text{with} \quad I_f(\mu) = 1$$

- example for an alternative with more information

$$g(x - \mu) \propto \frac{1}{1 + \left(\frac{x - \mu}{a}\right)^{2m}} \quad \text{with} \quad a^2 = 1 + 2 \cos \frac{\pi}{m} \quad \text{and} \quad I_g(\mu) = \frac{2m - 1}{2a^2}$$

- $I_g(\mu) > 1$ for $m \geq 2$
- rectangular distribution for $m \rightarrow \infty$
- Fisher information becomes infinite, i.e. uncertainty $\rightarrow 0/\sqrt{N}$

problems and exercises

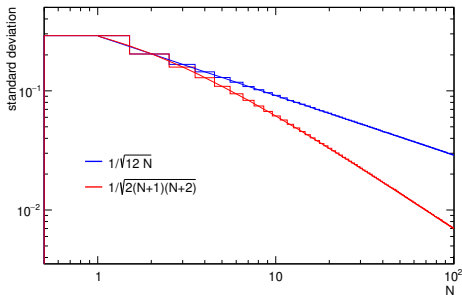
In order to find out what “infinite Fisher information” means, generate 10^5 n-tuples of uniformly distributed random numbers x_i determine numerically, as a function of the size N of the n-tuples, the standard deviation for two estimators a and b of the central point:

$$a = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad b = \frac{1}{2} \left(\min_{i=1}^N(x_i) + \max_{i=1}^N(x_i) \right)$$

hint: since both estimators are unbiased, it is OK to simply determine the RMS difference between the estimator and $1/2$, i.e. to determine $\sqrt{\langle (a - 1/2)^2 \rangle}$ and $\sqrt{\langle (b - 1/2)^2 \rangle}$.

► result

- the error of the arithmetic average scales with $1/\sqrt{N}$
- the average of the extremes has asymptotic uncertainties $\propto 1/N$



- Fisher information determines the coefficient of the $1/\sqrt{N}$ term
- discontinuities with infinite Fisher information allow for convergence with $1/N$

Bayesian confidence level intervals

❖ example: Poisson measurements without background

A counting experiment has observed $n = 2$ events, and it is known that n is a Poissonian random variable, with the probability p_n to observe n events

$$p(n, \mu) = e^{-\mu} \frac{\mu^n}{n!}$$

remember: μ fixed $\rightarrow$ probability distribution; n fixed $\rightarrow$ likelihood function

■ use the likelihood function $L(\mu) = p(n, \mu)$ to get information about μ :

▶ $L(\mu = 0.1) \approx 0.0045$

▶ $L(\mu = 1.0) \approx 0.1839$

▶ $L(\mu = 10.) \approx 0.0023$

■ in principle any value for μ is possible, but values of $O(1)$ seem more plausible

wanted: best estimate and uncertainty $\rightarrow$

❖ the Bayesian approach

- assign a probability $P(\mu)$ to every value of μ
 - ▶ encode what is known about μ even if μ has a well defined true physical value
 - ▶ can be completely ad-hoc without leading to mathematical inconsistencies
 - ▶ makes $p(n, \mu)$ a conditional probability $p(n|\mu)$
- use Bayes' theorem to improve the knowledge by the measurement:

$$p(\mu|n) = \frac{p(n|\mu) P(\mu)}{P(n)} \quad \text{or} \quad p(\mu|n) = \frac{L(\mu) P(\mu)}{\int d\mu p(\mu) P(\mu)}$$

- ▶ $P(\mu)$: prior PDF of μ - to be defined
- ▶ $L(\mu) = p(n, \mu)$: Likelihood function
- ▶ $P(n)$: probability for n , unknown constant, normalization of $p(\mu|n)$
- ▶ $p(\mu|n)$: posterior PDF for μ after the measurement

❖ implementation of the Bayesian approach

- choice of prior distribution

$$P(\mu) \propto \mu^k$$

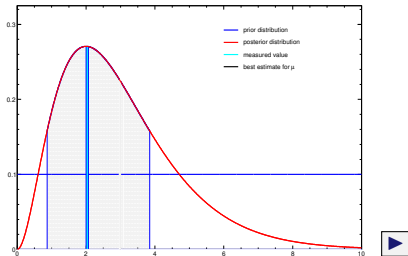
- ▶ ad hoc, but easy to implement
- ▶ variations of k probe the sensitivity to the choice of the prior

- posterior PDF

$$p(\mu|n) = e^{-\mu} \frac{\mu^{n+k}}{\Gamma(n+k+1)}$$

- many possibilities to interpret the posterior
 - ▶ preferred values
 - ▶ credible intervals or limits

example →



■ results sensitive to prior PDF

- ▶ $P(\mu) \propto 1$ uniform prior in μ
- ▶ $P(\mu) \propto \mu^{-1}$ scale invariant prior
- ▶ $P(\mu) \propto \mu^{-1/2}$ Jeffreys prior

■ different possible conclusions

- ▶ for best estimates
- ▶ intervals or limits

→ whether a prior density is uniform depends on the choice of variables e.g.:

$$\frac{dn}{d\mu} = 1 \quad \text{implies for } z = \mu^2 \quad \frac{dn}{d\mu^2} = \frac{dn}{2\mu d\mu} = \frac{1}{2\mu} \quad \text{or} \quad \frac{dn}{dz} = \frac{1}{2\sqrt{z}}$$

→ need other criteria to define the prior, e.g. scale invariance

$$\frac{dn}{d\mu} = \frac{1}{\mu} \quad \text{scale transformation } z = \alpha\mu \quad \frac{dn}{d(\alpha\mu)} = \frac{dn}{\alpha d\mu} = \frac{1}{\alpha\mu} \quad \text{or} \quad \frac{dn}{dz} = \frac{1}{z}$$

The Jeffreys Prior

❖ construction of prior PDF suggested by Sir Harold Jeffreys

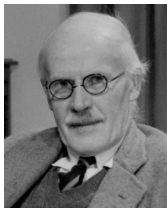
- a function of all available prior knowledge $f(x, \alpha)$
- independent of actual data, i.e. integral over x
- transform like a PDF under change of variables

$$P(\beta) = P(\alpha) \left| \frac{d\alpha}{d\beta} \right|$$

➤ Jeffreys: the square root of the Fisher information buys the list

$$P(\alpha) = \sqrt{\left\langle \left(\frac{\partial \ln f(x, \alpha)}{\partial \alpha} \right)^2 \right\rangle} = \sqrt{I(\alpha)}$$

- the transformation property holds since $|d\alpha/d\beta|$ does not depend on x
- $P(\alpha)$ is not unique, e.g. $\langle (\partial \exp(f(x, \alpha))/\partial \alpha)^n \rangle^{1/n}$ would work, too



source: npg.org.uk

(c) Gonville & Caius College, Cambridge

Bayesian combination of results

- OK: using the posterior of experiment 1 as prior for experiment 2

$$P_1(\mu|n_1) \propto p(n_1|\mu)P(\mu)$$

$$P_{21}(\mu|n_2, n_1) \propto p(n_2|\mu)p(n_1|\mu)P(\mu) \propto P(n_2|\mu)p_1(\mu|n_1)$$

- ▶ same result from joint analysis and updating approach
- ▶ consistent formalisation for gain of knowledge by successive measurements
- ▶ prior $P(\mu)$ enters once

- NOT OK: multiplicative combination of posteriors from different experiments

$$P_1(\mu|n_1) \propto L(n_1|\mu)P(\mu) \quad \text{and} \quad P_2(\mu|n_2) \propto L(n_2|\mu)P(\mu)$$

- ▶ multiplication of posteriors is proportional to $P^2(\mu)$
- ▶ consistent only for uniform priors $P(\mu) = \text{const}$

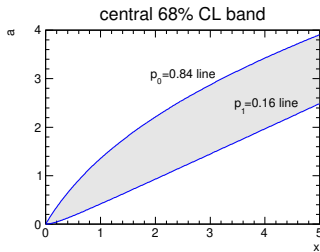
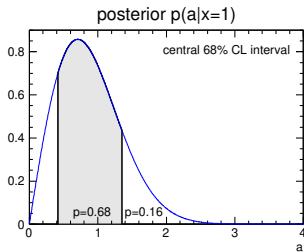
→ while likelihood functions can be multiplied, Bayesian posteriors cannot

Parameter intervals, probabilities and coverage

- from a measurement x_0 construct the **posterior density** $p(a|x_0)$ for parameter a
- determine an **interval** $[a_l(x_0), a_h(x_0)]$ where $p(a|x_0)$ has probability content p
 - ▶ the interval is a random object tied to the measurement
- you are offered to bet a sum S on the fact that a_{true} is inside $[a_l(x_0), a_h(x_0)]$
- if a is inside you are paid $R = S/p$ otherwise the money is lost
 - ▶ since the interval is a random object the outcome is also random
- the bet is fair if, for any a_{true} , the interval contains a_{true} in the fraction p of the cases
 - ▶ the expected return $\langle R \rangle = p \cdot (S/p) + (1 - p) \cdot 0$ is equal to the investment S
- frequentist language: the bet is fair if the intervals have **coverage**
 - ▶ “holy grail” in frequentist statistics, ususally ignored by bayesians

analyse what it takes to have coverage in Bayesian statistics →

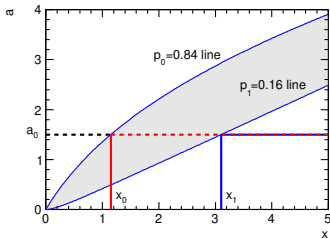
❖ schematically: consider a 1-dim PDF in $f(x, a)$ with $x, a \in [0, \infty]$



■ construction of confidence level belts

- ▶ define upper and lower probability contents p_0 and p_1
- ▶ for each value x consider the posterior $p(a|x)$
- ▶ calculate the corresponding intervals and construct a confidence level belt

❖ determine the range of x -values where a parameter a_0 is in the CL interval



■ implicit definition of x_0 and x_1

$$\int_0^{a_0} da p(a|x_0) = p_0 \quad \text{and} \quad \int_0^{a_0} da p(a|x_1) = p_1$$

■ probability p for $x \in [x_0, x_1]$

$$p = q_0 - q_1 = \int_{x_0}^{\infty} dx f(x, a_0) - \int_{x_1}^{\infty} dx f(x, a_0)$$

■ wanted: $p = p_0 - p_1 = q_0 - q_1$, which is given for $p_0 = q_0$ and $p_1 = q_1$

$$\int_0^{a_0} da p(a|x_0) = \int_{x_0}^{\infty} dx f(x, a_0) \quad \text{or} \quad p(a_0|x_0) = \frac{\partial}{\partial a_0} \int_{x_0}^{\infty} dx f(x, a_0)$$

► explicit prescription how to construct a posterior PDF that ensures coverage

→ note: this is independent of the bayesian framework

❖ examples for posterior distributions with coverage

- PDF with a location parameter (e.g. gaussian mean)

$$f(x|a) = f(x - a) \rightarrow p(a|x_0) = f(x_0 - a)$$

- PDF with a scale parameter (e.g. exponential density)

$$f(x|a) = \frac{1}{a} f\left(\frac{x}{a}\right) \rightarrow p(a|x_0) = \frac{x_0}{a^2} f\left(\frac{x_0}{a}\right)$$

- connection to Bayesian framework

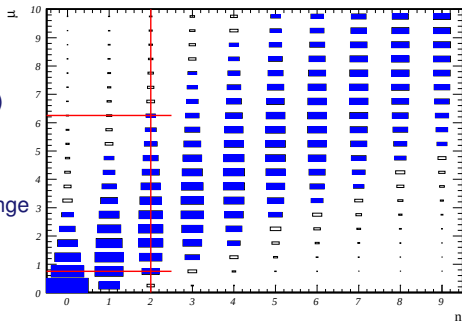
- ▶ location parameter: same result with Jeffreys prior $P(a) = 1$
- ▶ scale parameter: same result with Jeffreys prior $P(a) = 1/a$
- ▶ in general a prior that leads a posterior with coverage does not exist
(follows when trying to construct the posterior with Bayes' theorem)

Frequentist confidence level intervals

❖ Neyman construction based on the data model $f(x, a)$

- ignore posterior densities and focus on CL-intervals with well defined coverage
- example: 90% CL interval for a Poisson average

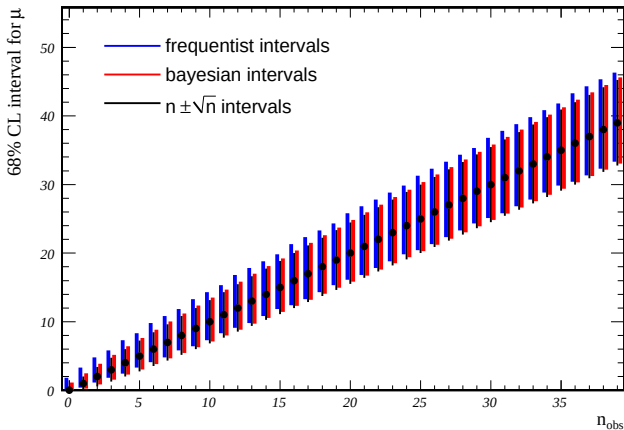
- ▶ start from table of probabilities for any observation and any value μ
- ▶ for each μ determine the (shortest) horizontal range with $\geq 90\%$ probability content
- ▶ for a measured n the vertical μ -range inside the $\geq 90\%$ area is the 90% CL interval



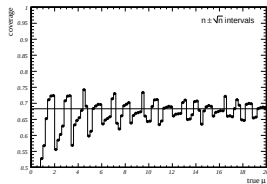
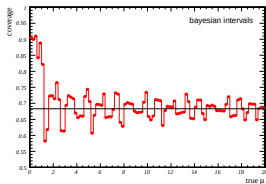
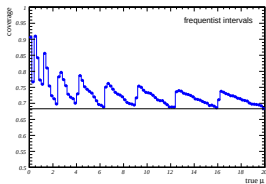
❖ discussion

- unlike in the Bayesian case, no prior distribution is needed
- the CL interval covers all true values with the measurement in the $p \geq 90\%$ range
 - ▶ it has **coverage**: from an ensemble of intervals $\geq 90\%$ contain the true value
 - whether a specific interval contains the true value is not known
 - a CL-interval is a random variable derived from observations
 - a CL-interval contains **no** information about **preferred values**
- another commonly used interval for poisson averages is $n \pm \sqrt{n}$
 - ▶ estimate for the standard deviation of the measurement
 - ▶ often taken also as approximate 68.3% confidence level interval
- for the Poisson example check which fraction of CL-intervals contains the true value
 - ▶ do the check as a function of the true μ and calculate coverage also for Bayesian intervals with uniform prior and minimum-size intervals

❖ comparison of “68.3%” confidence level intervals



- Monte Carlo simulation to check the coverage
 - ▶ step through true μ values
 - ▶ generate observations and determine the 68.3% CL interval
 - ▶ determine the fraction of the 68.3% CL intervals that contain the true μ



- exact coverage is possible only for continuous PDFs
 - ▶ for discrete distributions require coverage or over-coverage
 - ▶ largest differences observed for small μ

problems and exercises

Determine numerically the coverage probability of the interval $[n - \sqrt{n}, n + \sqrt{n}]$, where n is a Poisson distributed random variable, for Poisson averages $\mu = 10$ and $\mu = 100$.

Determine numerically the coverage probability of the interval $[x - 1, x + 1]$, where x is a gaussian distributed random variable with unit variance, for the mean values $\mu = 0$, $\mu = 1$ and $\mu = 2$.

► solutions

■ poisson: coverage($\mu=10$) = 0.644, coverage($\mu=100$) = 0.681

■ gauss: coverage($\mu=0$) = 0.683, coverage($\mu=1$) = 0.683, coverage($\mu=2$) = 0.683

Analytical coverage calculation: Poisson intervals $n \pm \sqrt{n}$

■ check for e.g. $\mu = 3$ which measurements n have intervals that contain μ

$n = 1$	interval	$1 \pm \sqrt{1}$	
$n = 2$	interval	$2 \pm \sqrt{2}$	contains $\mu = 3$
$n = 3$	interval	$3 \pm \sqrt{3}$	contains $\mu = 3$
$n = 4$	interval	$4 \pm \sqrt{4}$	contains $\mu = 3$
$n = 5$	interval	$5 \pm \sqrt{5}$	contains $\mu = 3$
$n = 6$	interval	$6 \pm \sqrt{6}$	

► coverage probability

$$C(\mu = 3) = P_2(\mu) + P_3(\mu) + P_4(\mu) + P_5(\mu) \quad \text{with} \quad P_n(\mu) = e^{-\mu} \frac{\mu^n}{n!}$$

■ general result

$$C(\mu) = \sum_{n-\sqrt{n} \leq \mu} P_n(\mu) = \sum_{n=\mu+1/2-\sqrt{\mu+1/4}}^{\mu+1/2+\sqrt{\mu+1/4}} P_n(\mu)$$

Analytical coverage calculation: normal distribution with $x \pm 1$

■ normal distribution

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

■ probability that for a given μ and a measurement x the interval $x \pm 1$ contains μ

$$C(\mu) = \int_{x+1 \geq \mu}^{x-1 \leq \mu} dx P(x, \mu) = \int_{\mu-1}^{\mu+1} dx P(x, \mu) = \int_{-1}^1 dx \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = 0.683 \dots$$

- ▶ substitution of variables makes the integral independent of μ
- ▶ the coverage is the same for any value of μ

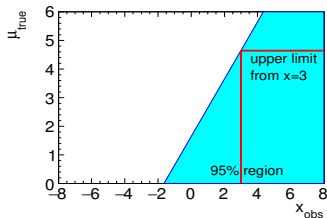
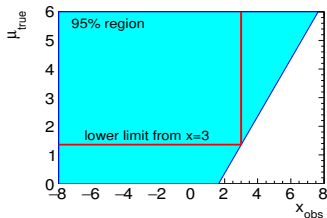
Neyman-construction for limits

❖ requires 1-sided lower or upper ranges with given probability content

► example: 95% CL limits for Gaussian PDFs with unit variance

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x dt e^{-(t-\mu)^2/2} = 0.95 \rightarrow \text{lower limit } \mu > x - 1.64485 \dots$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x dt e^{-(t-\mu)^2/2} = 0.05 \rightarrow \text{upper limit } \mu < x + 1.64485 \dots$$



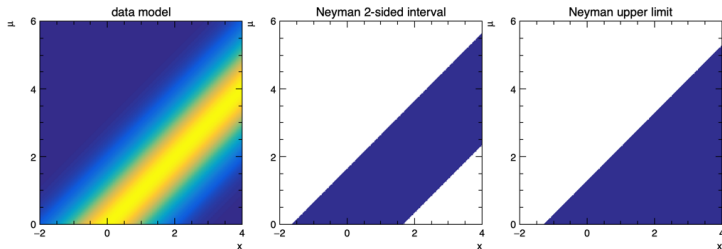
❖ observations

- ❑ if μ must be positive, it is possible that no limit with the required coverage exists
- ❑ different conclusions can be drawn from the same data
 - ▶ lower limit, upper limit, two-sided interval
- ❑ infinitely many possibilities exist to construct two-sided intervals
 - ▶ shortest intervals
 - ▶ central intervals with same quantiles above and below
 - ▶ intervals with different ratios of upper and lower tails
- ❑ **important:** coverage requires to decide the interpretation of the data beforehand
 - ▶ a decision to quote an upper limit if a measurement is different from zero with less than 3 standard deviations (“flip-flopping”) will give intervals that do not have the expected coverage

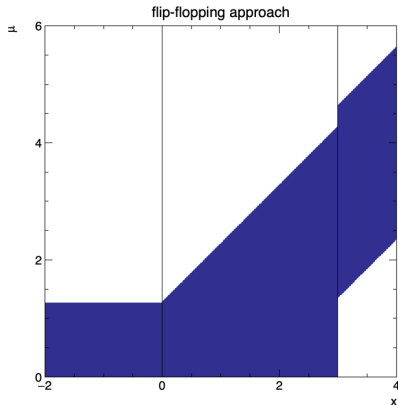
The Feldman-Cousins unified approach

❖ confidence level intervals in the presence of physical boundaries

- take Gaussian measurements x with $\sigma = 1$ and known-to-be-positive mean μ
- consider the Neyman construction for confidence level intervals
 - ▶ double sided intervals have 90% coverage for $x > 1.64$
 - ▶ upper limits have 90% coverage for $x > -1.28$
 - ▶ measurements can be in a region where no 90% CL-interval can be assigned



❖ flip-flopping if μ is known to be non-negative



■ $x > 3$: quote a 2-sided interval

$$\mu = x \pm 1.64$$

■ $x < 3$: quote an upper limit

$$\mu < x + 1.28$$

■ $x < 0$: quote the upper limit of $x=0$

$$\mu < 1.28$$

check coverage →

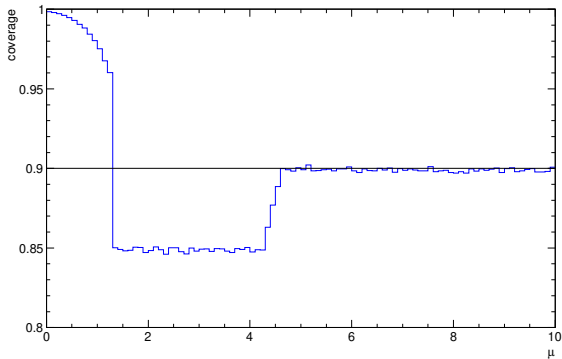
problems and exercises

In the determination of a parameter μ individual measurements are gaussian distributed around μ with unit variance. Determine numerically, for some values in the range $\mu \in [0, 10]$, the coverage probabilities of the flip-flopping 90% confidence level regions.

some numerical values:

- $\mu = 0$ coverage = 0.999, $\mu = 1$ coverage = 0.978, $\mu = 2$ coverage = 0.849
- $\mu = 3$ coverage = 0.847, $\mu = 4$ coverage = 0.847, $\mu = 5$ coverage = 0.898
- $\mu = 6$ coverage = 0.899, $\mu = 7$ coverage = 0.901, $\mu = 8$ coverage = 0.899
- $\mu = 9$ coverage = 0.898, $\mu = 10$ coverage = 0.899

► result



- approximately correct coverage for $\mu > 5$ - otherwise significant deviations

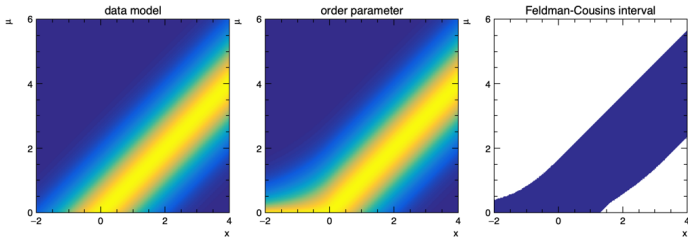
❖ Feldman-Cousins: no flip-flopping but limits or 2-sided intervals with coverage

- exploit the freedom of the Neyman construction, which is only constrained by

$$\int_{x_l}^{x_h} dx f(x; \mu) = \alpha \quad \text{for each } \mu$$

- accumulate probabilities $f(x; \mu) dx$ in decreasing order of the order parameter $R(x; \mu)$

$$R(x; \mu) = \frac{f(x; \mu)}{f(x; \mu_{\text{best}})} \quad \text{with} \quad \mu_{\text{best}} = \begin{cases} x & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}$$



Concluding remarks about frequentist confidence level intervals

- frequentist confidence level intervals are random quantities determined from data
- these intervals are constructed to have coverage
 - ▶ “90% of the 90% CL intervals contain the true value”
 - ▶ for continuous PDFs exact coverage can be obtained
 - ▶ for discrete probabilities usually intervals with over-coverage are given
- for the construction the correct data model $f(x; \mu)$ must be known
- confidence level intervals are NOT standard deviations
- confidence level intervals and limits. . .
 - ▶ (can) have exact coverage even when the true value is unknown
 - ▶ are intuitive and transform exactly under re-parametrisations
 - ▶ don't contain any information about a preferred value
- there exists no way to combine confidence level intervals from different experiments
 - ▶ always publish also the covariance matrix of a measurement
 - ▶ for estimated standard deviations give also the dependence on the true value