Analysis of multichannel measurements of rare processes with uncertain expected background and acceptance

I. B. Smirnov

Seminar HEPD PNPI,
April 31, 2015

Introduction

Multivariate methods:

- Response variable from 0 to 1.
- Reducing expected background, possibly zero at the end, where the expected signal is maximal.
- This is the most important region.
- How to treat?

The options:

- To accept (to ignore) zeros?
- To unify the channels?
- To smooth?

The results:

- Deterioration of resolution or “look elsewhere” effect?
- Optimization of division?
- What is the meaning of the result?

No recipe in the literature.
Statistical analysis with the nuisance parameters is not developed.
It needs to start from the basics.

Introduction

Statistical data analysis:

- There is one or several measurements of an unknown parameter or its function.
- What is the true value of this parameter?

Two conceptually different approaches:

- To estimate, how probable or unprobable are given measurements, if the true value of the parameter is equal, larger, or smaller than some value.
  - **Likelihood (maximum or ratio).**
  - **Frequentism (the method of confidence intervals of Neyman).**
  - **Significance of hypothesis.**
- To estimate, how probable is assumed value of unknown parameter, given observations.
  - **The Bayesian method.**

Two types of estimates:

- Point estimates (likelihood, Bayes)
- Interval estimates (likelihood ratio (profile), frequentism, Bayes).

Figure 1: Reverend Thomas Bayes (presumably).

Reverend Thomas Bayes, 1702–1761, England. The paper about probabilities was published in 1763, after his death. “An Essay towards solving a Problem in the Doctrine of Chances”, Philosophical transactions, 1763.

Figure 2: Pierre-Simon, marquis de Laplace

Pierre-Simon, marquis de Laplace, 1749–1827, France. The “Bayes’ theorem” is formulated in “Essai philosophique sur les probabilites” (1814). There is am English translation available in Internet: “A Philosophical essay on probabilities” (1902).

The Bayes' theorem in contemporary form

Formula of the full probability:
Let H_1, H_2, ... , H_n be the full system of events (hypothesis). Let A be some event. Then

$$P(A) = \sum_{i=1}^{n} P(A|H_k)P(H_k)$$

According to conditional probability

$$P(H_k A) = P(H_k|A)P(A) = P(A|H_k)P(H_k)$$

The "Bayes' theorem":
Let A be some event, for which $P(A) \neq 0$. Then

$$P(H_k|A) = \frac{P(A|H_k)P(H_k)}{\sum_{i=1}^{n} P(A|H_k)P(H_k)}$$

The integral form:

$$p(\mu|x) = \frac{P(x|\mu)p(\mu)}{\displaystyle\int P(x|\mu)p(\mu)d\mu}.$$

$p(\mu)$ is interpreted as a prior distribution, denoted later as $\pi(\mu)$.

More about priors

If $\eta = \eta(\mu)$ is monotonically increasing or decreasing function,

$$P(\eta \in [\eta(\mu_1), \eta(\mu_2)]) = P(\mu \in [\mu_1, \mu_2]) \Rightarrow p(\eta(\mu))|\eta'(\mu)| = p(\mu)$$
$$p(\mu|x) \sim P(x|\mu)\pi(\mu)$$
$$p(\eta|x) \sim P(x|\eta)\pi(\eta) \Rightarrow p(\mu|x) \sim P(x|\mu)|\eta'(\mu)|\pi(\eta)$$

Consequense:

- The posteriors expressed in terms of μ and η are identical, if $\pi(\mu) \sim \pi(\eta)|\eta'(\mu)|$.
- If $\pi(\mu) = 1/\mu$, $\eta = \mu^q$, where q is any non-zero power then $\pi(\eta) = 1/\eta$. The prior is "invariant" for transformations $\eta = \mu^n$ for interval estimations. But such prior:
  - Shifts the maximum, differently for each q!
  - Can make the posterior infinite at zero.
  - With any other $\eta = \eta(\mu)$ the inveriance for interval estimations does not hold too.

Other variants, reference priors etc., and philosophy: [R. E. Kass, L. Wasserman J. Am. Atatist. Assoc. 91(1996) 1343-1370, R. D. Cousins, HCPSS 2009; R. D. Cousins, Am. J. Phis. 63(5) 1995; PDG; J. Heinrich et. al., CDF note 7117, 2004].

The uniform prior is the most suitable for the main parameter.

More about priors

Assume you have performed an experiment and receive Bayesian probability density, rejected the tails with area of 5% each, and receive credible interval [2,3]. *Can you suggest that the searched parameter is in the interval [2,3] with probability of 90%?*

→ Yes, but the meaning of this probability is a subject of discussions.

At the multiple repetition of the experiment will the interval obtained by this method include the true value of unknown parameter in 90% cases?

→ No, at fixed s this is not guaranteed.
But at averaging by $\pi(s)$: yes, it will include. Trivial consequense of the Bayes theorem (interpreted in terms of sets). Proven once again in terms of "the average is calculated with respect to the prior density" in [J. Heinrich et. al., CDF note 7117, 2004].

Prior, which guarantee the frequentist coverage completely: "probability matching priors", or approximately: "priors of the first order". Mathematical difficulties: [P.D. Baines, X.-L. Meng, PHYSTAT-LHC 2007 pp. 135-138].

Figure 3: Jerzy Neyman.

Jerzy Neyman, 1894–1981, Russia (Bessarabia)—Poland—USA The author of the method of confidence intervals, the main publication about the method: “Outline of a Theory of Statistical Estimation, Based on the Classical Theory of Probability.” Phil. Trans. Roy. Soc. of London, A236(1937)333–380

Figure 4: Sir Ronald Aylmer Fisher

Sir Ronald Aylmer Fisher, 1890–1962, Britain—Australia.
Another school of frequentist statistics, hypothesis tests, significance.

Method of confidence intervals

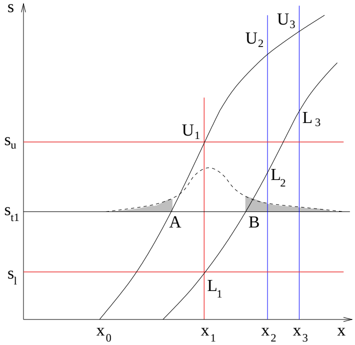


Figure 5: Usual confidence belt.

If true $s = s_{t1}$, if x_2 is measured, then L_2 is higher than s_{t1}, and the probability of this non-coverage is represented by the shaded area.
$x < x_0$: an empty or negative interval!

The sense of confidence intervals

You apply the method of confidence intervals and obtain that for the probability level 90% the searched parameter is in interval, for example, $[2, 3]$. What does that numbers mean?

Searched parameter is in interval $[1, 3]$ *with probability of 90%.*

$\rightarrow$ Wrong!

Searched parameter is with probability of 90% inside an interval, calculated by this receipt.

$\rightarrow$ Right.

What to do with numbers $[1, 3]$? Nothing!

For what reason they were obtained?

Ar the practice it is assumed that the unknown parameter "most likely" reside inside or around.

"Most likely" = probably = Bayesian approach.

Intervals and hypothesis

Another example. If the scale at the picture $s_1 \sim 1$, and the experiment has resulted in interval $[0.0001, 2]$. What is the probability that $s = 0$?

The probability $\approx 5\,\%$

$\rightarrow$ Wrong!

Correct:

If the interval is central, the probability of lower border computed by this receipt to be greater than the true parameter value is $= 5\,\%$*. Nothing more can be said. If the interval is not central, than on the base of the information given nothing can be said at all.*

Intervals and hypothesis

If the experiment has resulted in registration of $x < x_0$, than the interval is empty (no s is possible) or negative (signal is negative). Neither this nor that can be understood. If the probability of such x does not exceed 10% (5% for central intervals), than formally everything is correct. But for practical application this is yet more senseless, than the “normal” intervals like $[1, 3]$.

The only possible conclusion:

There is a downward fluctuation of background or the model is wrong.

If the experiment has resulted in registration of $x > x_0$, but $x \approx x_0$, than only upper border will be non-zero, but it will be almost zero and formal conclusion about very lower upper border will be doubtful.

The only practical conclusion is the same:

There is a downward fluctuation of background or the model is wrong.

Significance

Significance = the level of incompatibility of measurements with the background hypothesis

= a measure of probability for background to imitate the observed or larger signal.

Ambiguous, if the background is not known precisely!

Criterium of uniformity (of p-value distribution): valid only when there no nuisance parameters.

Alternative definition: a measure of incompatibility *between* the measurement of the signal+background and the measurement of the background (or its calculation) taken alone.

Implies statistic and methods of generation of its distribution such that the distribution depends only on observable values and does not depend or depends negligibly on background hypothesis.

The test problem. Conditions, notations

$$\prod_i P(n_i, n_{ai}, n_{bi}|s, a_i, b_i) = P(\vec{n}, \vec{n}_a, \vec{n}_b|s, \vec{a}, \vec{b}) =$$

$$= P(\vec{n}|t_a\vec{a}s + t_b\vec{b})P(\vec{n}_a|\vec{a})P(\vec{n}_b|\vec{b}).$$

$$\prod_i P(n_i, n_{ai}, n_{bi}|s, a_i, b_i) = P(\vec{n}, \vec{n}_a, \vec{n}_b|s, \vec{a}, \vec{b}) = P(\vec{n}|t_a\vec{a}\ s + t_b\vec{b}). \quad (1)$$

If $\vec{a}$ or $\vec{b}$ are assumed to be equal to their most probable values, which for Poisson distributions results into replacements

$$\prod_i P(n_i, n_{ai}, n_{bi}|s, a_i, b_i) \approx P(\vec{n}|t_a\vec{n}_a\ s + t_b\vec{n}_b). \quad (2)$$

Conditions of the task, parameters

The true background and signal distribution is

$$\begin{aligned} f_b(x) &= Ce^{-3x} \\ f_a(x) &= Ce^{-3(1-x)} \\ C &= (1-e^{-3})/3 \end{aligned} \tag{3}$$

The true parameters a_i and b_i are determined by equalities

$$a_i = N_a \int_{x_i}^{x_{i+1}} f_a(x)dx, \tag{4}$$

$$b_i = N_b \int_{x_i}^{x_{i+1}} f_b(x)dx, \tag{5}$$

Auxiliary experiments: Poisson with parameter N_a and N_b.
Main experiment: Poisson with mean $t_a N_a s_{\text{true}} + t_b N_b$,
where s_{true} is the true signal rate, which is then "forgotten" and has to be reconstructed.
In this work it is adopted $t_a N_a = 25$, $s_{\text{true}} = 2$, $t_a = 0.25$, $N_a = 100$.
$N_b = 250$, $t_b = 5$.

Distrubutions of expected signal and background

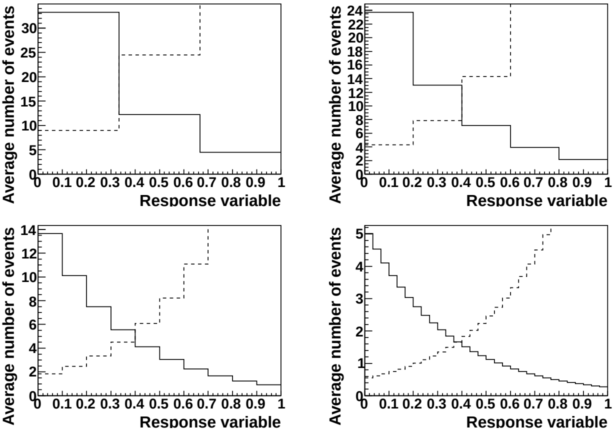


Figure 6: Distributions of a_i and b_i for 3, 5, 10, and 30 channels if $N_a = 25$ and $N_b = 50$.

Optimization of divisions

How should the data be divided into channels?
Possible criteria:

- Seaching of intervals with minimal width.
  - → Danger of the Look-elsewhere effect.
- Dividing untill a zero channel appears in the expected background distribution.
  - → Loss of precision if there is truly zero background channel.
  - → Too many-channel distributions and a danger of unexpected numerical problems for experiments with high number of events.
- Dividing untill a zero channel appears in the expected signal distribution.
  - → Too many-channel distributions and a danger unexpected numerical problems for experiments with high number of events.

The only universal solution found so far:
Seaching of division with minimal width from those which do not have zeros in the expected signal distribution.
The limits (for the purpose of optimization) are calculated by the modified Bayesian method with safe priors (see later).

The Bayesian probability density

$$p(s|\vec{n}) = \frac{P(\vec{n}|s)\pi(s)}{\displaystyle\int P(\vec{n}|s)\pi(s)ds},$$

$$P(\vec{n}|s) = \prod_i \int\int P(n_i,|s,a_i,b_i)p(a_i|n_{ai})p(b_i|n_{bi})da_i db_i.$$

$$p(a_i|n_{ai}) = \frac{P(n_{ai}|a_i)\pi(a_i)}{\displaystyle\int P(n_{ai}|a_i)\pi(a_i)da_i} \qquad p(b_i|n_{bi}) = \frac{P(n_{bi}|b_i)\pi(b_i)}{\displaystyle\int P(n_{bi}|b_i)\pi(b_i)db_i}$$

This is equivalent to:

$$p(s|\vec{n}) = \frac{N(s)}{\displaystyle\int N(s)ds},$$

where

$$N(s) = \prod_i \int\int P(n_i,|s,a_i,b_i)\pi(s)P(n_{ai}|a_i)\pi(a_i)P(n_{bi}|b_i)\pi(b_i)da_i db_i$$

Prior for s

Total absence of information $\Longrightarrow$ no preference to any specific value $\Longrightarrow$ uniform or flat distribution.

Problem: Metric-dependent. Transormations of $s \Longrightarrow$ either non-uniform prior or different probabilities. But:

- s seems to be in the same metric as n at one channel.
- No need to consider other metric.
- No need to make the prior invariant for particular sort transformations.
- No magic prior, which is invariant for all transormations.
- Uniform prior for s will be uniform after translations $s + b$ (with no respect to zero threshold).
- Probability density obtained with the uniform prior can be converted to that of any other prior.
- Prior $1/s$ and in general the prior with any negative power of s creates maximum with the infinite amplitude in zero.
- Priors are factorized, therefore combinations suppressing singularity, like $1/(s + b_i)$, are not allowed.

Conclusion: the uniform prior for s.

The priors for a_i and b_i

prior	mean	σ^2	mode
uniform	$n+1$	$n+1$	n
$1/\sqrt{\mu}$	$n+0.5$	n + 0.5	$\max(n-0.5,0)$
$1/\mu$	n	n	$\max(n-1,0)$

Table 1: Parameters of Bayesian posterior probability density distributions for Poisson distribution of observations with different priors. The mode is the most probable value.

In one case mean can be important, in other maximum.

Prior exhagarates the parameter $\Longrightarrow$ less signal is needed to describe the measurements.

Prior underastimates the parameter $\Longrightarrow$ more signal is needed to describe the measurements.

Safe priors:

- For lower limit: both the mean and the maximum should be not smaller than the measurement.
- For upper limit: both the mean and the maximum should be not larger than the measurement.

[J. Heinrich et al., CDF/MEMO/STATISTISC/PUBLIC/7117] claimed that "inverse priors are matched to this Poisson case".
However, there is strong evidence that for the lower border and for significance the uniform priors are needed.

Central intervals

The area rejected from each side is denoted by α. Then the central intervals:

$$\int_{s_{\mathrm{L}}}^{\infty} p(s|\vec{n})\,\mathrm{d}s = 1 - \alpha \tag{6}$$

and

$$\int_{0}^{s_{\mathrm{U}}} p(s|\vec{n})\,\mathrm{d}s = 1 - \alpha \tag{7}$$

Benefits:

- Fixed probability of violation of each separated border.

Drawbacks:

- Shortage of frequentist coverage.
- The low border is always through out some region around zero, even if the latter has the maximum probability density. The most probable point can also be cut off.

Types of intervals

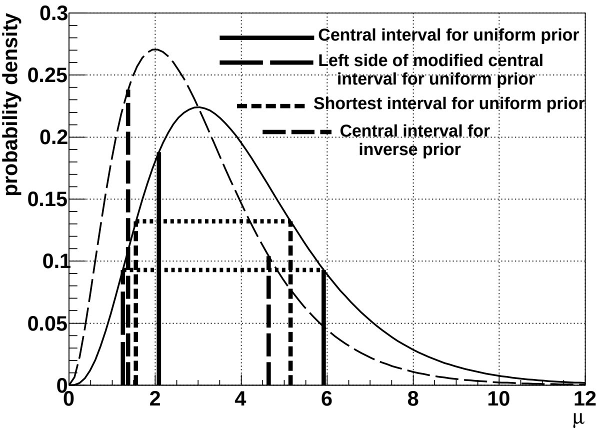


Figure 7: Probability densities for μ of the Poisson distribution (the one-channel problem without background), if 3 events are observed. The solid (dashed) smooth line depicts the probability density distribution for the uniform (inverse) prior. The intervals calculated with $\alpha = 0.158655$ are shown by solid and dashed thick lines with different length of dashes. Dotted lines show horizontal levels.

The modified central intervals

s_U is the same as for the central interval.
s_L is the minimum of the central s_L and s_L at the same level of density as it is for s_U:

$$p(s_L|\vec{n}) \quad = \quad p(s_U|\vec{n}).$$

Benefits:

- Good coverage provided correct priors are used.
- Always includes the most probable point and a region with the largest probability density, the lower limit can be zero.
- Probability of violation of upper border is fixed.

Drawbacks:

- Probability of violation of lower border is variable, but can be determined.

The modified central intervals

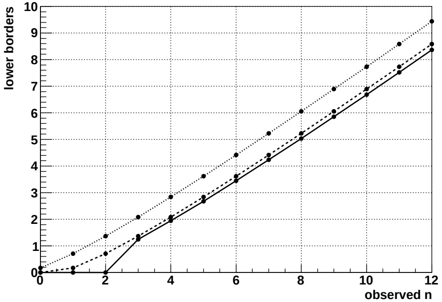


Figure 8: The left border of the central interval with uniform prior (dotted line), the left border of the central interval with inverse prior (dashed line), the left border positioned at the same probability density level as the right border of the central interval with uniform prior (solid line), as functions of observed n for the single measurement of the Poisson distribution without background.

Frequentist Treatment of Maximal Likelihood Estimate

Works

M. Mandelkern J. Schultz, J. Math. Phys. 41 (2000) 5701.

S. Ciampolillo, Nuono Cimento A 111 (1998) 1415.

No nuisance parameters.

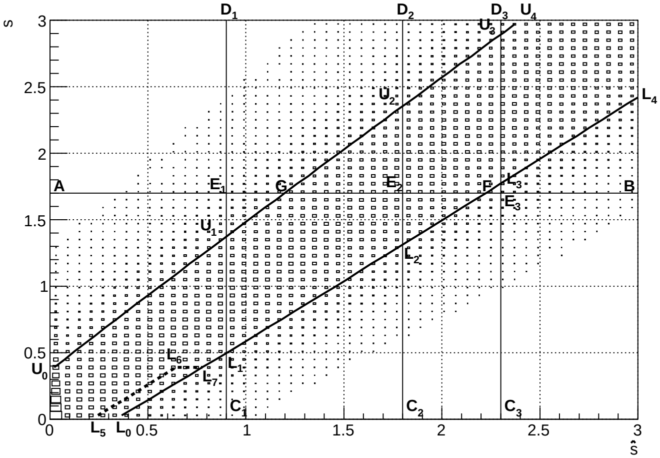


Figure 9: Distribution of observed most probable s (denoted as s_m) as function of true s for 5 channels without uncertainties of expected signal and background. See text for other notations.

Frequentist Treatment of Maximal Likelihood Estimate options for nuisance parameters

- Subgeneration:
  - Nuisance parameters:
    - Random according to Bayesian approach with $\vec{n}_a$ and $\vec{n}_b$ with safe priors
      (SSP: Subgeneration with Safe Priors).
    - Fixed most probable for $\vec{n}_a$ and $\vec{n}_b$
      (SM: Subgeneration with the most probable nuisance parameters).
    - Variable most probable for $\vec{n}$, $\vec{n}_a$ and $\vec{n}_b$ at given s
      (SA: Subgeneration with the adjusted nuisance parameters).
  - Nuisance measurements:
    - Fixed. (Results in good modeling interpretation.)
    - Random (RN: Random Nuisance). (Sorter intervals, but no modeling interpretation.)
- Analysis:
  - $\hat{s}$ with marginalization by the nuisance parameters
    (FMML: Frequency of Marginalized Maximum Likelihood)
  - $\hat{s}$ with global maximization by the nuisance parameters
    (FGML: Frequency of Global Maximum Likelihood)

Frequentist Treatment of Maximal Likelihood Estimate, studied combinations

- SSP–FMML.
- SSP–FGML.
- SMRN–FGML.
- SARN–FGML.
- SSPRN–FMML.
- SSPRN–FGML.

The Method of Likelihood Ratio (profile)

$$P(\vec{n}, \vec{n_a}, \vec{n_b} | s, \vec{a}, \vec{b}) = \prod_i P(n_i, | s, a_i, b_i) P(n_{ai} | a_i) P(n_{bi} | b_i)$$

The likelihood ratio is

$$R(s) = \frac{\sup_{s,\vec{a},\vec{b}} (P(\vec{n}, \vec{n_a}, \vec{n_b} | s, \vec{a}, \vec{b}))}{\sup_{\vec{a},\vec{b}} (P(\vec{n}, \vec{n_a}, \vec{n_b} | s, \vec{a}, \vec{b}))}$$

Writing the integral of Gaussian in the form

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} \, \mathrm{d}t$$

one obtains z for given p-value by $z = F^{-1}(1 - p)$. Finding interval borders s_{lower} and s_{upper} such that

$$2 \ln(R(s_{\mathrm{lower}})) = 2 \ln(R(s_{\mathrm{upper}})) = z^2$$

Frequentist Treatment of Likelihood Ratio, Statistics

$$Q(s) = \frac{P(\vec{n}|\hat{s})}{P(\vec{n}|s_{\rm ref})}. \tag{8}$$

Initially proposed $s_{\rm ref} = 0$.
A. L. Read, 1st Workshop on Confidence Limits, CERN, Geneva Switzerland, 2000, CERN-2000-005, pp 81-101.
Local notation: Background Ralated (BR).

$s_{\rm ref}$ maximizes $P(\vec{n}|s_{\rm ref})$, but constrained in the $[0, s]$ interval. This can be used for upper interval border.
The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011.* August 2011, ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005.
Significance, $s = 0$, hence $s_{\rm ref}$ simply maximizes $P(\vec{n}|s_{\rm ref})$ in the interval $[0, +\infty]$.
By analogy (my extension), for the lower border $s_{\rm ref}$ can be chosen such that it maximizes $P(\vec{n}|s_{\rm ref})$, and constrained by $s_{\rm ref} \geq s$.
Local notation: the LHC-style method, constrained-maximum-related method (CMR).
Unified approach: Local notation: unconstrained-maximum-related (UMR).

Frequentist Treatment of Likelihood Ratio, Setting Limits

General equation, the probability of signal plus background to coincide or to exceed the observed value:

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}) = \sum_{\substack{\vec{n}_\gamma: \\ Q(\vec{n}_\gamma,\vec{a},\vec{b},s) \leq Q(\vec{n},\vec{a},\vec{b},s)}} P(\vec{n}_\gamma | s, \vec{a}, \vec{b}) = \alpha \, . \tag{9}$$

Local notation: Frequency of Likelihood Ratio (FLR).
The significance, the probability of background to coincide or to exceed the observed value (at $s = 0$):

$$CL_b = P_b(Q \leq Q_{obs}) = \sum_{\substack{\vec{n}_\gamma: \\ Q(\vec{n}_\gamma,\vec{a},\vec{b},s) \leq Q(\vec{n},\vec{a},\vec{b},s)}} P(\vec{n}_\gamma | s_u = 0, \vec{a}, \vec{b}) \, . \tag{10}$$

Problems of LHC-style CL_{s+b}

- Exclusion of signal with true s with probability equal to α even for experiments microscopically dependant on signal. [Read et all, 2000]
- Ridiculously low upper limit at backward fluctuation of signal.

Solved by dividing CL_{s+b} by CL_b and substituting in Eq. (9). The researcher takes into account how well the experiment is described by the background. Local notation: normalization (NFLR).

Frequentist Treatment of Likelihood Ratio, One channel, dependence on n.

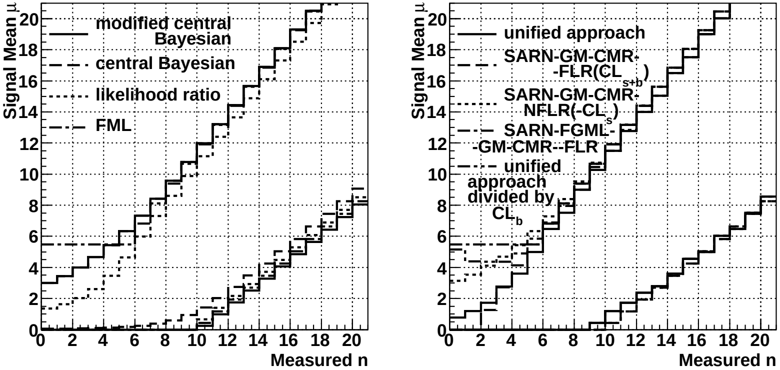


Figure 10: The confidence intervals for the one-channel problem with known auxiliary parameters $a = 1$ and $b = 5$, $t_a = t_b = 1$, for different observed n by different methods.

Frequentist Treatment of Likelihood Ratio, Maximum Related Method, Handling Nuisance Parameters

- Subgeneration:
  - Nuisance parameters:
    - Random according to Bayesian approach with $\vec{n}_a$ and $\vec{n}_b$ with safe priors
      (SSP: Subgeneration with Safe Priors).
    - Fixed most probable for $\vec{n}_a$ and $\vec{n}_b$
      (SM: Subgeneration with the most probable nuisance parameters).
    - Variable most probable for $\vec{n}$, $\vec{n}_a$ and $\vec{n}_b$ at given s
      (SA: Subgeneration with the adjusted nuisance parameters).
  - Nuisance measurements:
    - Fixed. (Results in good modeling interpretation.)
    - Random (RN: Random Nuisance). (Shorter but no interpretation.)
- Analysis:
  - Separate marginalization by the nuisance parameters
    (MM: Marginalized Maximum)
  - Separate maximization by the nuisance parameters (GM: Global Maximum)

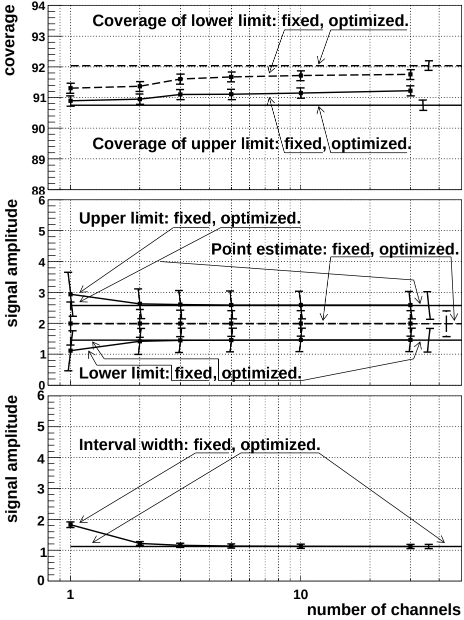
Frequentist Treatment of Likelihood Ratio, Combinations

- Since $\hat{s}$ is calculated anyway for denominator, it can be compared separately with the observed one with logical AND.
  This is the merge of FML-methods and FLR-methods.
  Local notation: adding FMML or FGML.
- Lower limit by CMR–NFLR is reasonable to combine with UMR–FLR in order to reduce undercoverage at large number of zeros in the expected background.
- RN FLR methods usually have non-trivial minima and maxima at floating nuisance parameters.
  It is extremely difficult to find them.
  If they exist and can be found, they constitute very good methods.
  Local notation: "Min/Max".

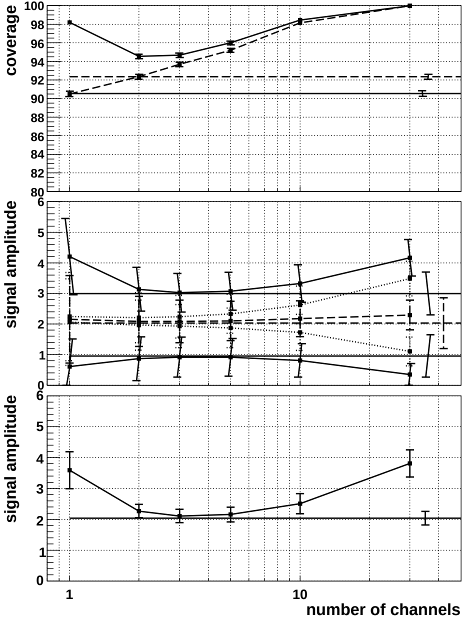
Frequentist Treatment of Likelihood Ratio, Studied Combinations

- SSP–MM–CMR–NFLR.
- SSP–GM–CMR–NFLR.
- SSP–FMML–MM–CMR–NFLR.
- SSP–FGML–GM–CMR–NFLR.
- SSPRN–MM–CMR–NFLR.
- SSPRN–GM–CMR–NFLR.
- SMRN–MM–CMR–NFLR.
- SMRN–GM–CMR–NFLR.
- SARN–MM–CMR–NFLR.
- SARN–GM–CMR–NFLR (LHC-style CL_s).
- SARN–MM–UMR–FLR.
- SARN–GM–UMR–FLR (Unified approach with globalization).
- Asymptotic SARN–MM–UMR–FLR.
- Asymptotic SARN–GM–UMR–FLR.
- Min/Max–RN–MM–CMR–NFLR.
- Min/Max–RN–GM–CMR–NFLR.
- SARN–FMML–MM–CMR–NFLR.
- SARN–FGML–GM–CMR–NFLR.
- SARN–MM–CMR–NFLR–UMR–FLR.
- SARN–GM–CMR–NFLR–UMR–FLR.

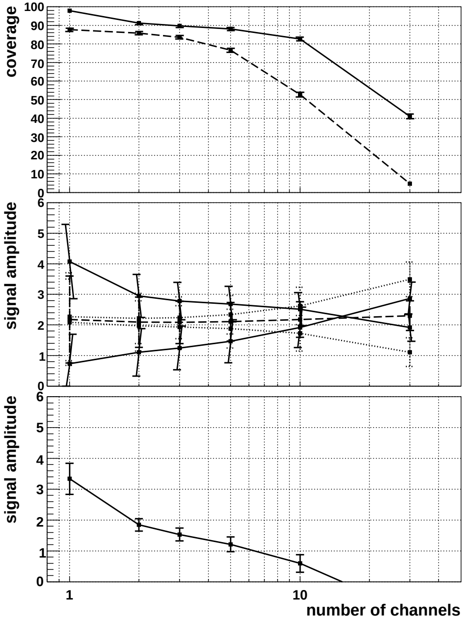
Bayes, no uncertainties



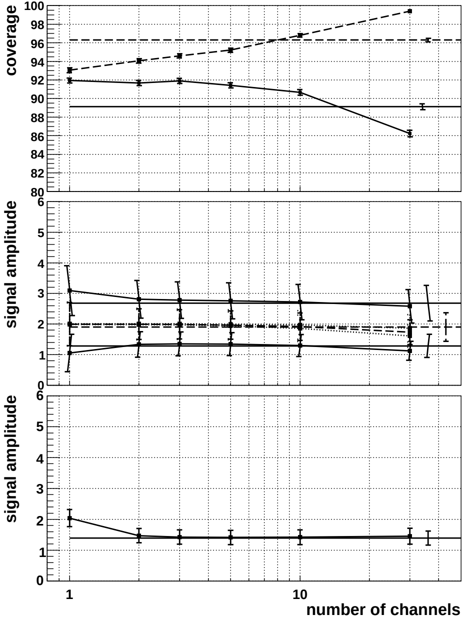
Bayes, uncertainty of the expected background



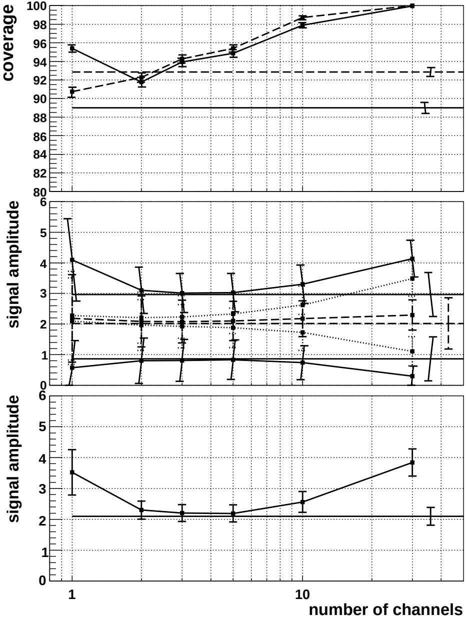
Bayes, uncertainty of the expected background echanged priors.



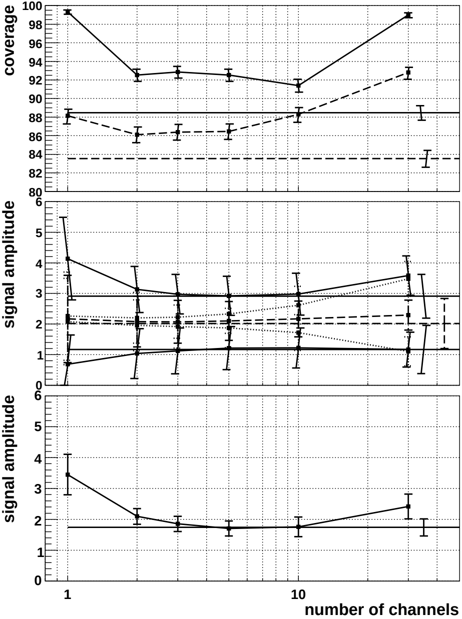
Bayes, uncertainty of the expected signal.



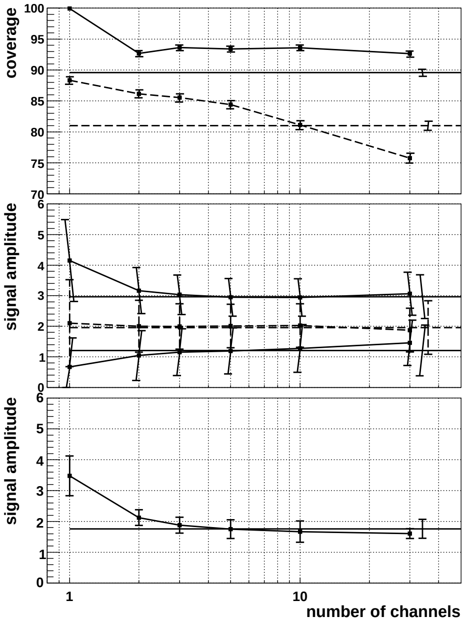
SSP–FMML, uncertainty of the expected background



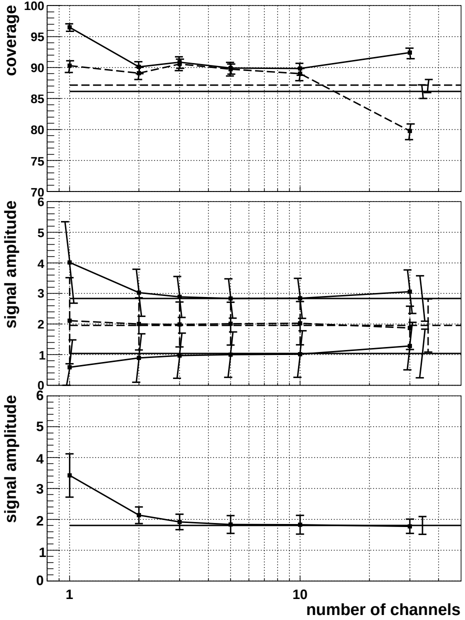
SSPRN–FMML, uncertainty of the expected background



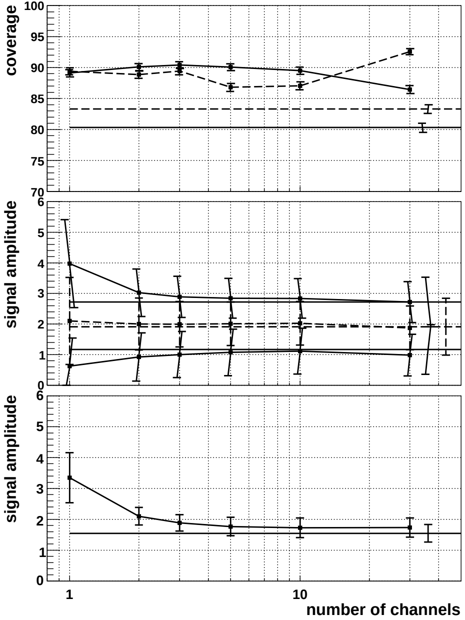
SMRN–FGML, uncertainty of the expected background



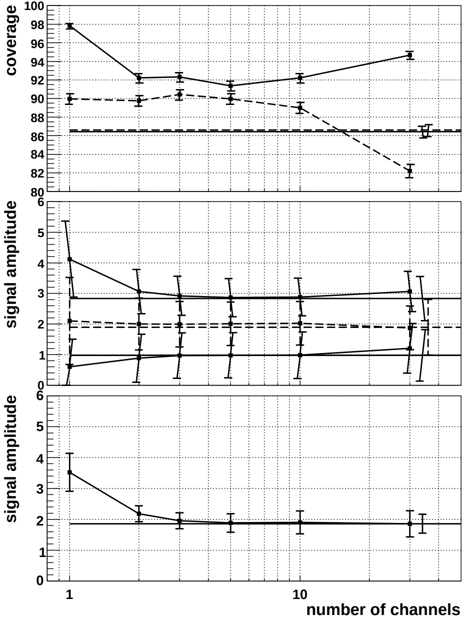
SARN–FGML, uncertainty of the expected background



Likelihood Ratio, uncertainty of the expected background



LHC CL_s, uncertainty of the expected background



Simple examples

Table 2: Comparison of signifance obtained by different methods for the simple one-channel and two-channel examples.

n	67	60	88		37, 51		7
t_a	1	1	1		1		1
a / n_a	1 / –	1 / –	1 / –		1, 1 / –, –		– / 3
t_b	2	10	10		10		1
b / n_b	– / 15	– / 0	– / 3		–, – / 3, 0		2 / –
	z	z	z	limits	z	limits	z
Bayesian central	—	—	—	23.6–81.9	—	33.0–75.3	—
Bayesian central modified	—	—	—	23.6–81.9	—	33.1–75.3	—
Likelihood Ratio	—	—	—	28.0–79.9	—	49.1–74.0	—
SSP–...	2.86	2.70	1.88	19.7–81.3	2.45	32.4–74.9	2.64

SSP–... means SSP–FMML, SSP–FGML, SSP–MM–CMR–NFLR, SSP–GM–CMR–NFLR, SSP–FMML–BM–CMR–NFLR, SSP–FGML–BM–CMR–NFLR, which are approximately the same fro these examples.

Continuation on the next page.

	z	z	z	limits	z	limits	z
Bayesian central	—	—	—	23.6–81.9	—	33.0–75.3	—
Bayesian central modified	—	—	—	23.6–81.9	—	33.1–75.3	—
Likelihood Ratio	—	—	—	28.0–79.9	—	49.1–74.0	—
SSP–...	2.86	2.70	1.88	19.7–81.3	2.45	32.4–74.9	2.64
SEP–...	3.06	∞	2.33	33.0–74.3	4.75	52.2–67.3	2.64
SDSP–FMML/FGML	3.66	4.12	2.74	32.4–77.7	3.80	43.7–74.6	2.65
SSPRN–FMML/FGML	3.68	4.06	2.84	33.1–84.0	3.59	44.0–73.9	2.10
SSPRN–GM(MM)–CMR–NFLR	3.01	4.32	$\approx$ 2.1	24.1–80.1	3.86	47.9–74.0	2.69
SMRN–FGML	4.07	∞	5.17	34.2–85.1	$\gtrsim$5.2	51.9–74.6	1.43
SARN–FGML	2.94	2.89	2.08	24.7–79.5	2.74	44.0–75.0	1.43
SMRN–GM–CMR–NFLR	3.03	∞	1.76	24.1–79.4	$\gtrsim$5.2	51.0–74.5	2.68
SARN–GM–CMR–NFLR	3.01	3.07	2.09	24.4–79.6	2.85	45.0–74.5	2.68
Asymptotic GM–UMR–FLR	3.04	3.38	2.19	—	3.12	—	2.75
Used nuisance parameters	27.3	5.45	8.27		3.6, 4.6		
Minimal RM–GM–CMR–NFLR	2.83	3.04	1.74	13.7–82.6	2.71	41.7–75.1	2.68
Best nuisance parameters	5.0		2.8		$\gtrsim$20, 5.4		
SARN–FGML–GM–CMR–FLR	2.92	2.93	2.04	23.4–80.2	2.68	43.1–74.6	1.71
SARN–GM–CMR–NFLR– –UMR–FLR	3.01	3.07	2.09	24.1–79.4	2.85	45.0–74.5	2.68
SARN–FGML–GM–CMR– –FLR–UMR–FLR	2.92	2.93	2.04	23.4–80.2	2.68	43.1–74.6	1.71
SMRN–MM–CMR–NFLR	3.01	∞	2.08	26.7–79.2	$\gtrsim$4.75	51.7–74.3	2.77
SARN–MM–CMR–NFLR	3.01	3.08	2.10	24.4–79.6	2.86	44.0–74.2	2.77
Asymptotic MM–UMR–FLR	3.01	3.11	2.16	—	2.85	—	2.62
Minimal RM–MM–CMR–NFLR	3.02	3.04	2.00	24.4–84.2	2.73	37.5–74.5	2.65
Best nuisance parameters	$\gtrsim$ 15	5.5	3.4		$\gtrsim$20, 5.7		
SARN–FMML–MM–CMR–FLR	2.92	2.84	2.05	23.4–80.2	2.60	42.5–74.5	2.03
SARN–MM–CMR–NFLR– –UMR–FLR	3.01	3.07	2.10	24.4–79.6	2.86	39.9–74.2	2.77
SARN–FMML–GM–CMR– –FLR–UMR–FLR	2.92	2.84	2.05	23.4–80.2	2.60	39.9–74.5	2.03

Conclusions

Bayesian approach:

- To use safe priors.
- To use modified central intrervals.
- Simple calculations.
- Clear interpretation, including similar to the frequentist one.
- No classical significance.

Likelihood ratio (profile).

- Technically simple.
- No statistical interpretation.
- Slightly insuffisient coverage of intervals.
- No classical significance.

Frequentist approach:

- Calculates not only intervals but also significance.
- Many methods with different results, no strict rule for selection.
- Technical difficulties, computing difficulties.
- Sometimes no clear interpretation.
- The most promising: SSP–FMML, SSP–MM–CMR–NFLR, SSP–FMML–MM–CMR–FLR, Min/Max–RN–MM–CMR–FLR, Asymptotic RN–MM–UMR–FLR.