Remarks on confidence interval classes in ROOT (what should I use ?)

J. Conrad (CERN)

30.09.2005

- The confidence interval classes in ROOT
  - 1) Concepts
  - 2) Performance: Coverage
  - 3) Properties and Applicability
- How to decide what to use ?
- What is missing (IMHO) ?
- Summary

What exists ?

- TFeldmanCousins (written by A. Bevan)
- TRolke (written by J.C.)
- TLimits (written by C. Delaere)
- TMinuit (MINOS errors) (written by R. Brun (?), F. James)

Since it is not rocket science: code up your intervals using existing ROOT functionality. **RooFit** seems a very good candidate (not covered here).

1) The concepts

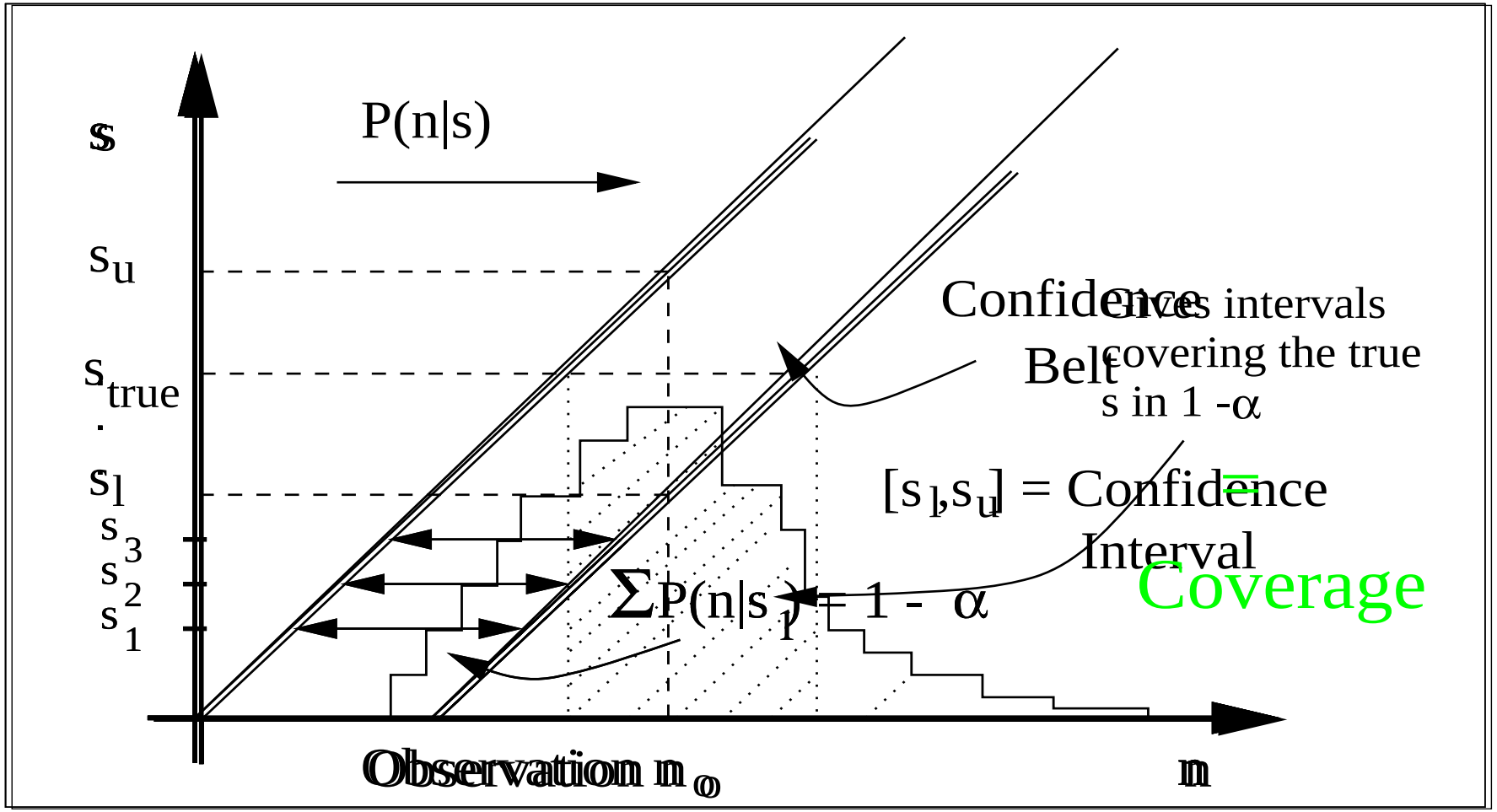
Key performance figure of frequentist confidence intervals: **coverage**

- A method is said to have **coverage** (1-α) if, in infinitely many repeated experiments the resulting confidence interval includes the true value with probability (1-α) irrespective of what the true value is

Bayesian dude: don't worry about coverage

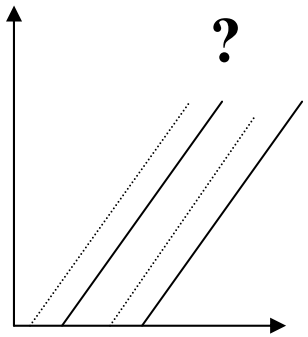
Bayesian physicist: use Bayesian methods, but check coverage !

TFeldmanCousins concept: frequentist confidence belt construction



J. Neyman, Phil. Trans. Royal Soc. London, A236 (1937)

TFeldmanCousins: where to start and stop the sum ? Likelihood ratio ordering



- Calculate likelihood ratio:

$$R = \frac{\mathcal{L}(n|b+s)}{\mathcal{L}(n|b+s_{best})}$$

- Rank n according to likelihood ratio
- Include n in descending order of the ratio until sum condition fulfilled

TFeldmanCousins: exactly known background

so called "nuisance parameter"

Feldman Cousins (1998) Phys.Rev.D 57:3873-3889 (1998)

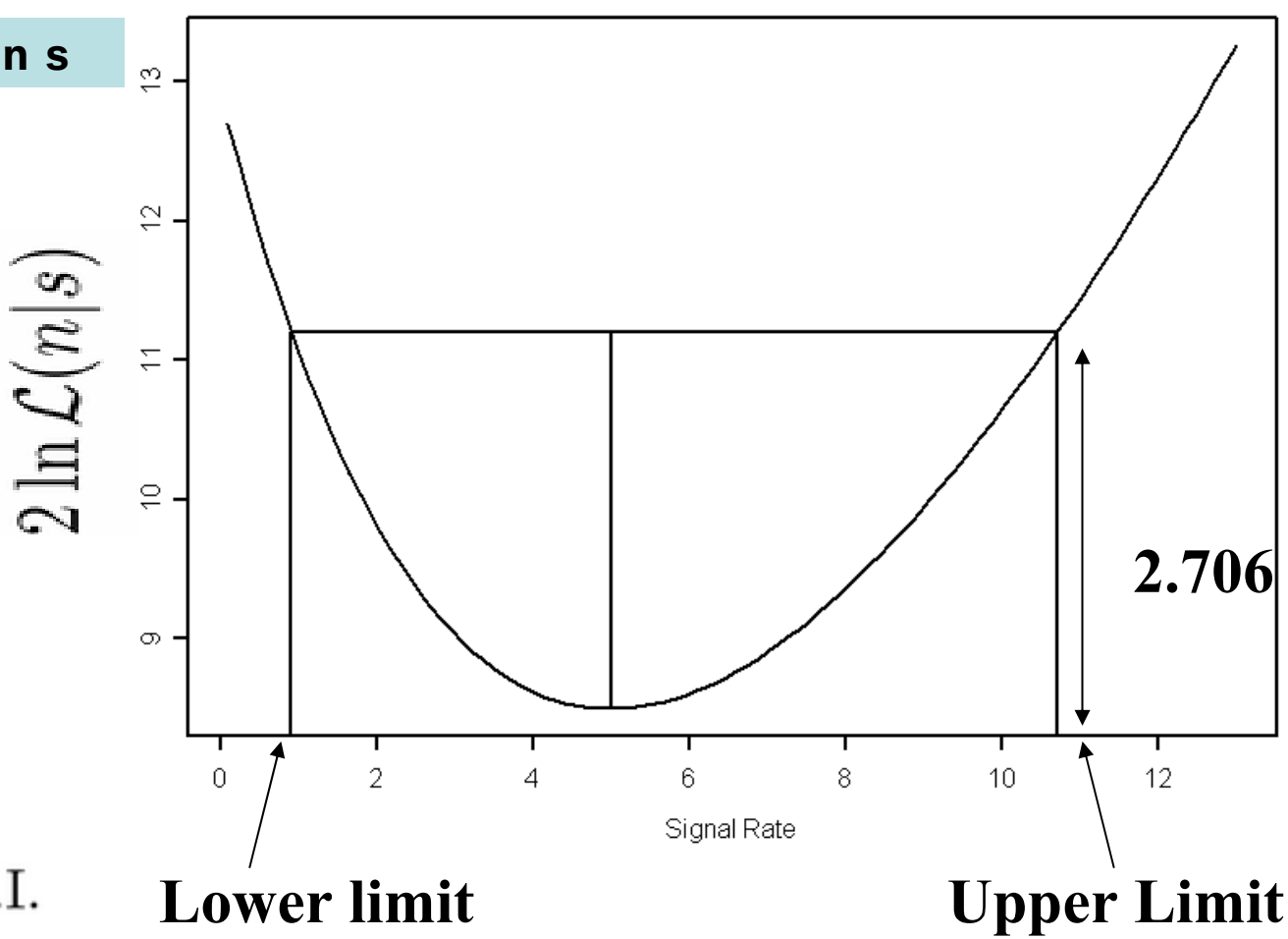
TMinuit (MINOS) and TRolke: profile likelihood

Meas. back

MLE of b given s

$$\lambda(s) = \frac{\mathcal{L}(n, b_m | s, \hat{\hat{b}}(s))}{\mathcal{L}(n, b_m | \hat{s}, \hat{b})}$$

MLE of b and s given observations



To extract limits:

$$-2 \ln \lambda \approx \chi^2$$

$$\chi^2 - \chi^2_{min} = 2.706 \quad \equiv 90\% \text{ C.I.}$$

F. James, e.g. Computer Physics Comm. 20 (1980) 29-35
W. Rolke, A. Lopez, J. C. Nucl.Instr.Meth A 551 2-3 (2005) 493-503

TRolke vs. TMinuit

- TRolke calculates analytically the solution for the seven (most common ?) problems in Particle Physics:

  - Signal process: Poisson plus background
  - Nuisance parameters:
    -1: Background - Poisson, Efficiency - Binomial
    -2: Background - Poisson, Efficiency - Gaussian
    -3: Background - Gaussian, Efficiency - Gaussian
    -4: Background - Poisson, Efficiency - known
    -5: Background - Gaussian, Efficiency - known
    -6: Background - known, Efficiency - Binomial
    -7: Background - known, Efficiency - Gaussian

  → requires less thinking and less code AND:

- Implements special treatment for some cases (to improve coverage)

  → means that TRolke works "better" than TMinuit for the models above

TLimit (The CL_s method)

- Class description: "Class to calculate 95 % confidence limits", violent understatement:
  "Class for Hypothesis Testing"
  → Caluclates p-values, power etc …..

Better name:

TLevel (TCLs)

- Can also be used to calculate upper limits at ANY given confidence → needs a loop over hypotheses

- Test statistics: likelihood ratio based on Neyman Pearson Lemma

$$-2\ln Q = -2\ln\frac{\mathcal{L}(n|s+b)}{\mathcal{L}(n|b)}$$

Obtained from histograms

Renormalizes on the background only case.

- Confidence Level calculated from histograms of -2 ln Q

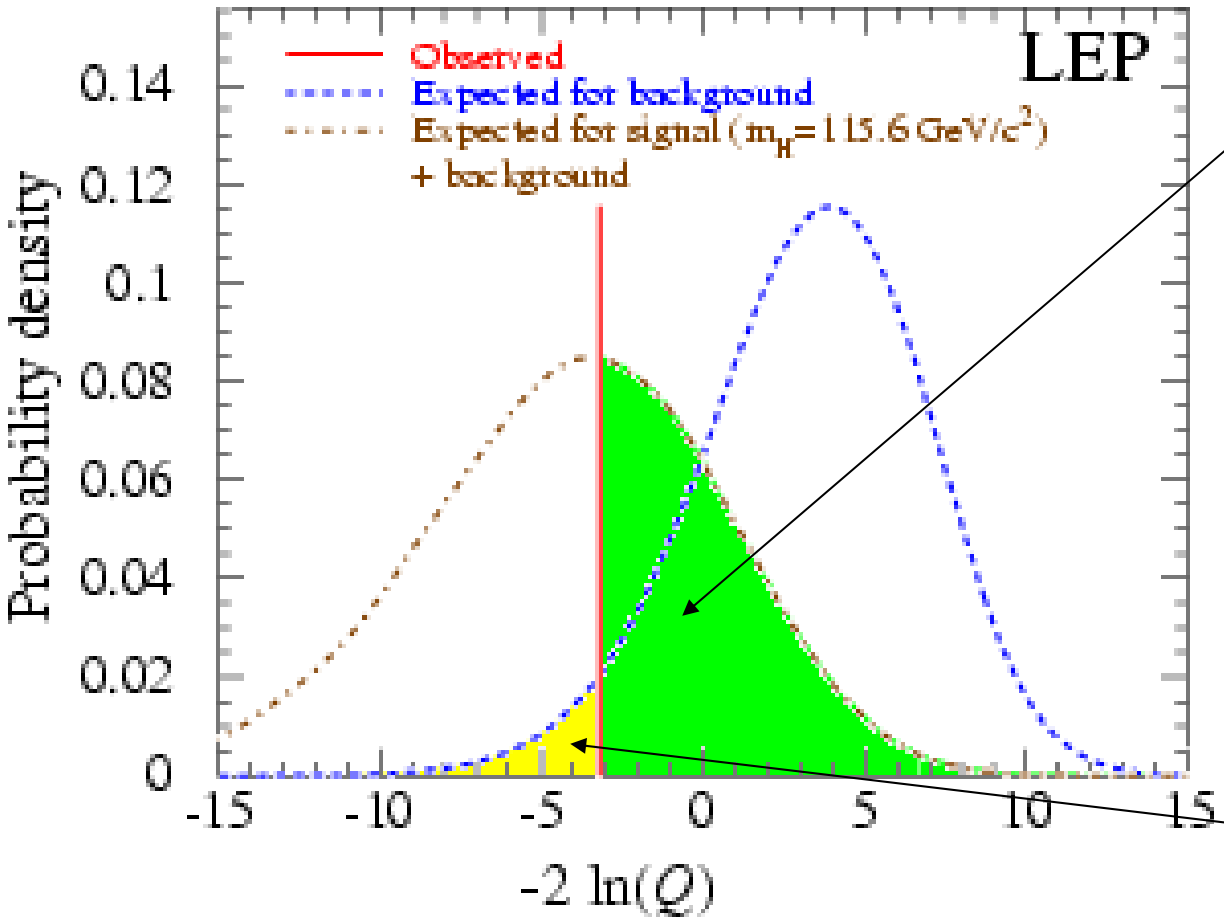
$$CL_s = \frac{P(Q \le Q_o|s+b)}{P(Q \le Q_o|b)} = \frac{CL_{s+b}}{CL_b} = \frac{\text{p-value(s+b)}}{1-(\text{p-value(b)})}$$

Note: Q monotonically increasing for more signal like obs

- Uncertainties on signal & background can be included (Gauss), using Bayesian PDF integration (ala Highland & Cousins)

T. Junk: Nucl.Instr.Meth.A434:435-443,1999
A. Read: e.g. J.Phys.G28:2693-2704,2002

CL_s: example



CL_s : generic method

TLimit:

designed for multi-channel measurements with sufficient statistics

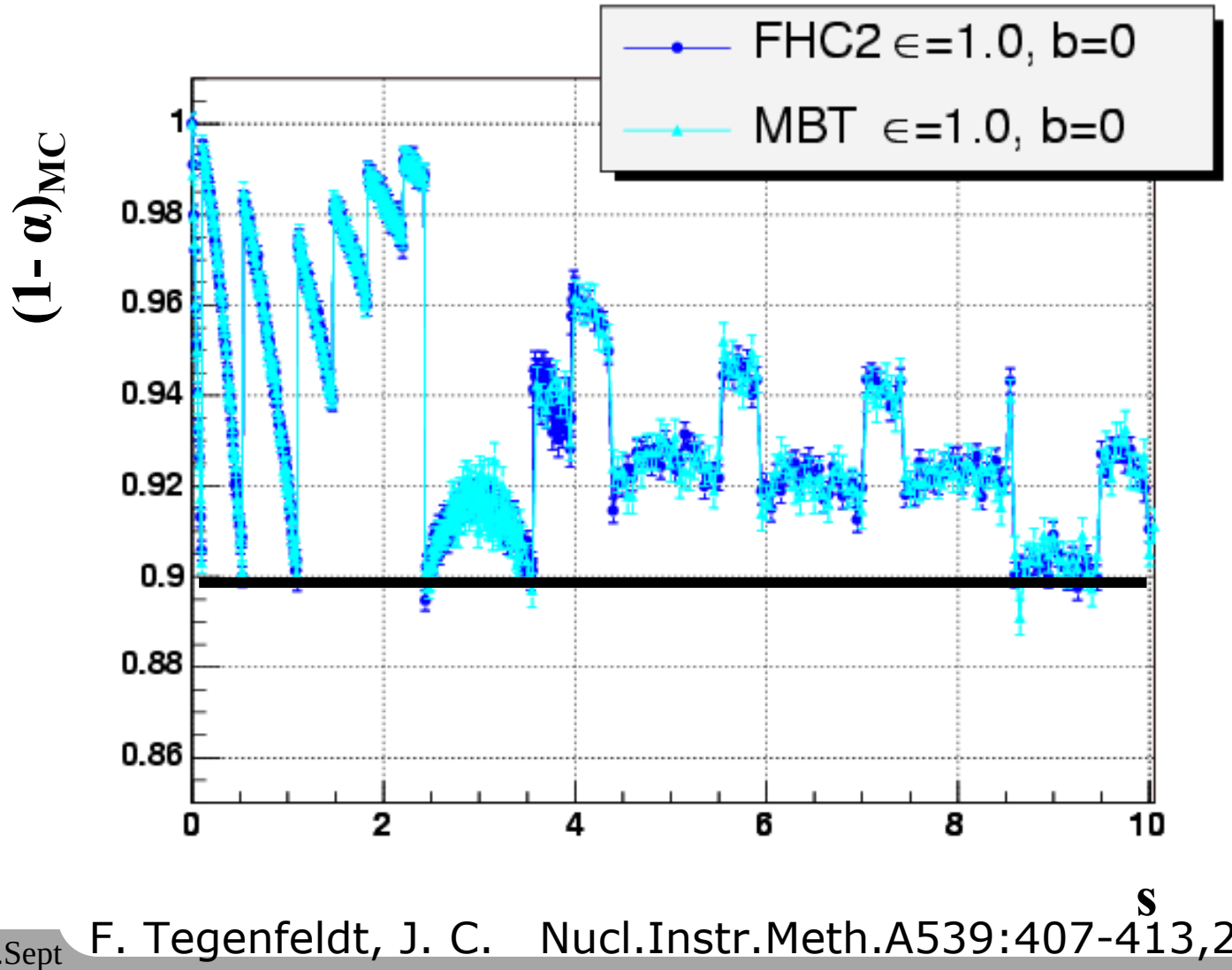
(you don't provide the PDF but the observed signal)

1- CL_b

A. Read: e.g. J.Phys.G28:2693-2704,2002

2) Coverage

TFeldmanCousins: coverage



F. Tegenfeldt, J. C. Nucl.Instr.Meth.A539:407-413,2005

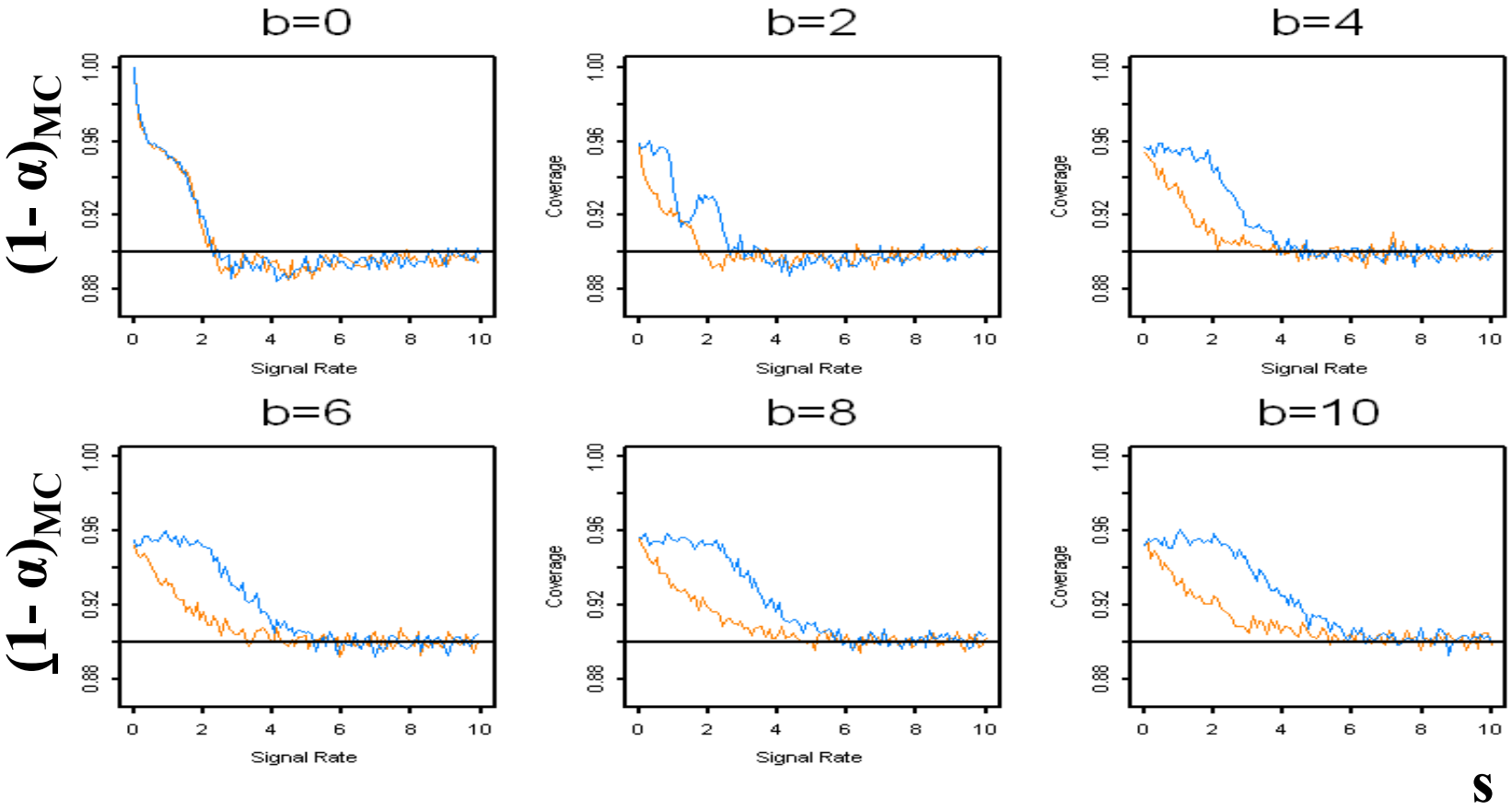
TRolke/TMinuit: coverage

Background: Poisson (unc ~ 20 % -- 40 %)

Efficiency: binomial (unc ~ 12%)

TRolke

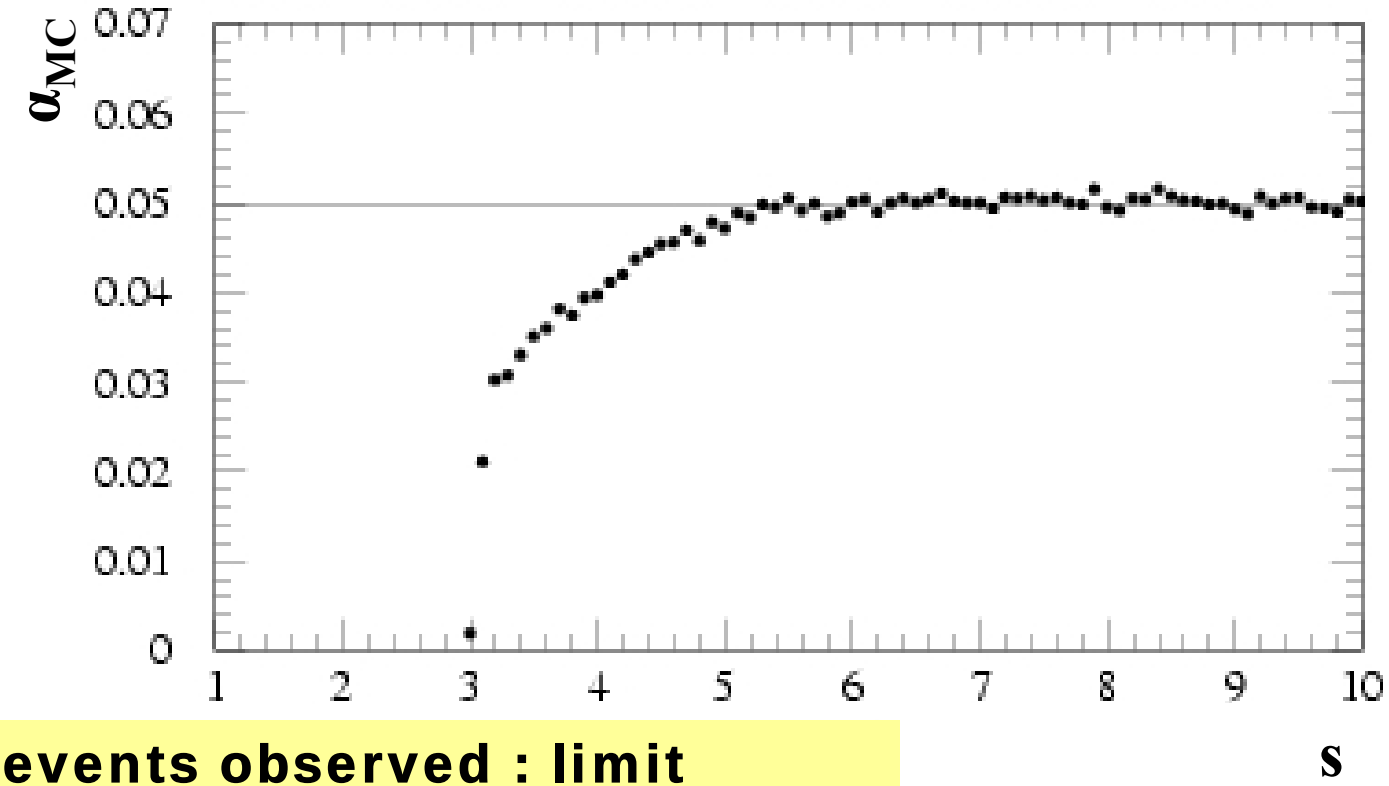
TMinuit



s

W. Rolke, A. Lopez, J. C. Nucl.Instr.Meth A 551 2-3 (2005) 493-503

TLimit (CL_s): coverage



Higgs search

(distributions in rec. higgs mass)

with b = 4 and a variable mass resolution

(therefore only little structure)

For zero events observed : limit independent of background for CL_s (not so for CL_{s+b})

T. Junk: Nucl.Instr.Meth.A434:435-443,1999

3) Properties and Applicability

Some examples: Resulting Intervals

n.b: lower limits always 0

Problem	TFeldmanC	TRolke	TMinuit	TLimit
N_{obs} = 3 b = 2 No uncertainties	**5.42**	**4.81**	**4.81**	CL_s: **4.9** CL_{s+b}: **4.7**
N_{obs} = 3 b = 2 σ_{eff} = 30 %	Not possible	**5.56**	**5.56**	CL_s: **5.7** CL_{s+b}: **5.4**

Wall clock time (being a not very clever user)

TLimit = 1 !

Problem	TFeldmanC	TRolke	TMinuit	TLimit
N_{obs} = 3 b = 2 No uncertainties	**1/100**	**1/250**	**1/250**	**1**
N_{obs} = 3 b = 2 σ_{eff} = 30 %	**Not possible**	**1/10000**	**1/10000**	**1**

Properties & Applicability

Method	Concept	Signal PDF	Uncertainties in Nuisance parameters	Coverage	Unified
TFC	Frequentist	Poisson (s+b)	No	Yes, by construction	Yes

Properties & Applicability

Method	Concept	Signal PDF	Uncertainties in Nuisance parameters	Coverage	Unified
TFC	Frequentist	Poisson (s+b)	No	Yes, by construction	Yes
TRolke	Frequentist	Poisson (s+b)	Efficiency, background, most common PDFs	Yes , proved by MC (even for “5 σ”)	Yes

Properties & Applicability

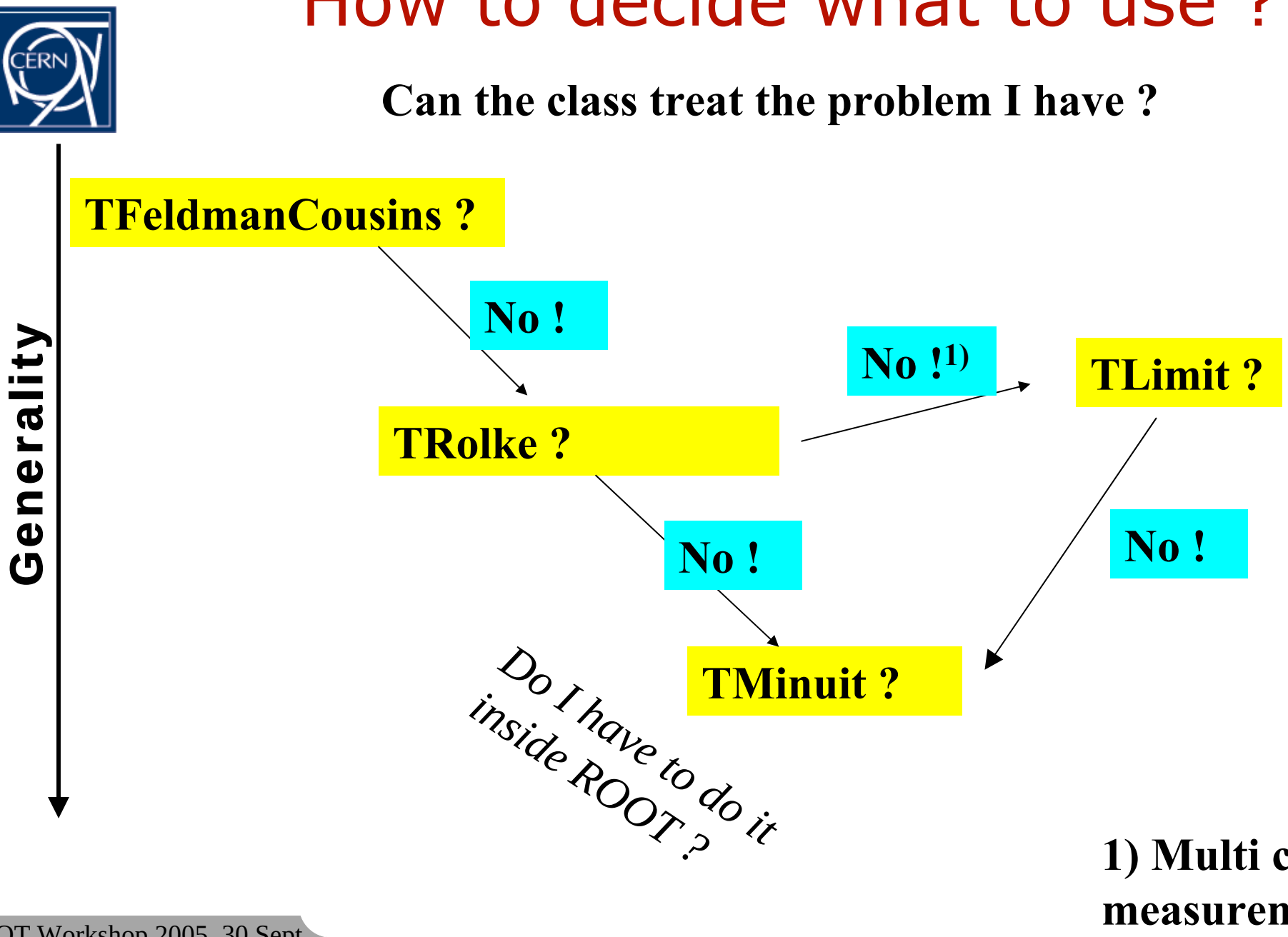
Method	Concept	Signal PDF	Uncertainties in Nuisance parameters	Coverage	Unified
TFC	Frequentist	Poisson (s+b)	No	Yes, by construction	Yes
TRolke	Frequentist	Poisson (s+b)	Efficiency, background, most common PDFs	Yes , proved by MC (even for “5 σ”)	Yes
TMinuit	Frequentist	Anything, need only likelihood function	Anything, need only likelihood function	Yes, for TRolke models, other models.	Yes

Properties & Applicability

Method	Concept	Signal PDF	Uncertainties in Nuisance parameters	Coverage	Unified
TFC	Frequentist	Poisson (s+b)	No	Yes, by construction	Yes
TRolke	Frequentist	Poisson (s+b)	Efficiency, background, most common PDFs	Yes , proved by MC (even for "5 σ")	Yes
TMinuit	Frequentist	Anything, need only likelihood function	Anything, need only likelihood function	Yes, for TRolke models, other models.	Yes
TLimit	Frequentist/ Bayesian	Anything (if several channels and suff. stats.) Otherwise: Poisson (s+b)	Efficiency, background Gauss PDF, full correlation	Yes, for CLs+b (without unc.) approx for CLs "5σ" coverage?	Only good for upper limits

All methods are in the PDG, all methods have published references

How to decide what to use ?



1) Multi channel measurement

What is missing (IMHO) ?

Mainly generalizing of Feldman & Cousins

- Feldman Cousins for other than Poisson ?
- Feldman Cousins for multiple experiments ?
- Feldman Cousins with Bayesian treatment of systematics
- Profile Likelihood for full Feldman & Cousins construction

pole++, J.C. & F. Tegenfeldt

PHYSTAT05, NIM A

→ Reorganize TFeldmanCousins:

-- ordering is the same, PDFs change

-- dimensionality might change

Summary

- TFeldmanCousins & TRolke
  - statistically well established methods with clear interpretation and good properties (even for high significance (say 5σ))
  - fast
  - should be used for the problems they can treat
- TMinuit (MINOS errors)
  - statistically well established with clear interpretation
  - properties good for TRolke models, otherwise needs testing
  - reasonably fast
  - most flexible (you need to know the likelihood function)
- TLimit
  - CL_{s+b} is well established with clear interpretation, CL_s not so much ….→ you have to be clear about what question you ask and what answer you want
  - Good for upper limits, useless for two sided intervals (mainly due to the likelihood ratio used)
  - TLimit: is designed for multiple measurements with sufficient statistics (typical Higgs search etc.), more complicated (or impossible) to use for generic problem (except Poisson distribution)
  - quite slow
  - could have problems for high significance (say 5σ) case (technically and in terms of systematics treatment), Kranmer PHYSTAT 2005

- Next extensions in ROOT: Generalizations of TFeldmanCousins ?