Interval estimation

Jan Conrad

Oskar Klein Centre, Stockholm University

INFN School of Statistics, Ischia, 2015

Acknowledgments

- Thanks to
  - **Fred James** for his inspiration, **Alex Read** for discussions on CL_s and **Knut Morå**, **Brandon Anderson** for preparing exercises and producing some of the plots in this lectures.

Interval estimation

- Confidence intervals are used to quantify the statistical accuracy of a measurement. The simplest example is the **standard error**, or standard deviation (square-root of the variance of an estimator), which provides such an estimate for the Gaussian case.

- More general: the goal of **interval estimation** is to estimate intervals that **contain the true value of a parameter with given probability.** The standard error for the Gaussian case will turn out to give the interval for which this probability is 68.3 %.

- The meaning of probability, and the operational definition of how the interval is estimated will differ between **Bayesian and Frequentist intervals**. A special frequentist case are so called **Likelihood intervals.**

- The choice of interval is not unique.

Interval estimation

- *Bayesian intervals ("credibile (-ility) intervals")*
- *Exact frequentist intervals ("confidence intervals")*
  - *Neyman construction (exact method) in particular: unified approach*
- *Likelihood intervals*
- *CL_s*
- *Nuisance parameters and their treatment including real life examples*
- *Inference with high dimensional complicated likelihood functions.*
- *Summary*

Interval estimation

CREDIBILE INTERVALS

Bayesian interval estimation

- The Bayesian interval can be constructed from the **posterior distribution**

$$p(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{\int P(X \mid \theta)P(\theta)d\theta}$$

- The interval which contains the true value with a certain **degree of belief** is then given by an integration of the posterior distribution.

$$1-\alpha = \int_{\theta^{LL}}^{\theta^{UL}} p(\theta \mid X)d\theta$$

1- α is chosen to be 0.683 (1σ) or 0.9 or 0.95. The corresponding interval is called **credible interval (cf: frequentist: confidence interval).**

Credible intervals -uniqueness

- The condition that the credible interval should have probability 1-α is not sufficient to make it unique.
- Other conditions can be imposed:

  – Accept points of **highest posterior density**

  – **Central Interval:**

$$\frac{1}{2}\alpha = \int_{\theta^{UL}}^{\infty} p(\theta \mid X)d\theta = \int_{-\infty}^{\theta^{LL}} p(\theta \mid X)d\theta$$

  – **One sided interval (upper or lower limit)**

$$\alpha = \int_{\theta^{UL}}^{\infty} p(\theta \mid X)d\theta$$

Upper limit (U.L) in words

An upper limit is of particular interest in particle physics, as it is often the result of searches for unknown physics in case of no detection

Bayesian: *the degree of belief that the signal is larger than the U.L. is small.*

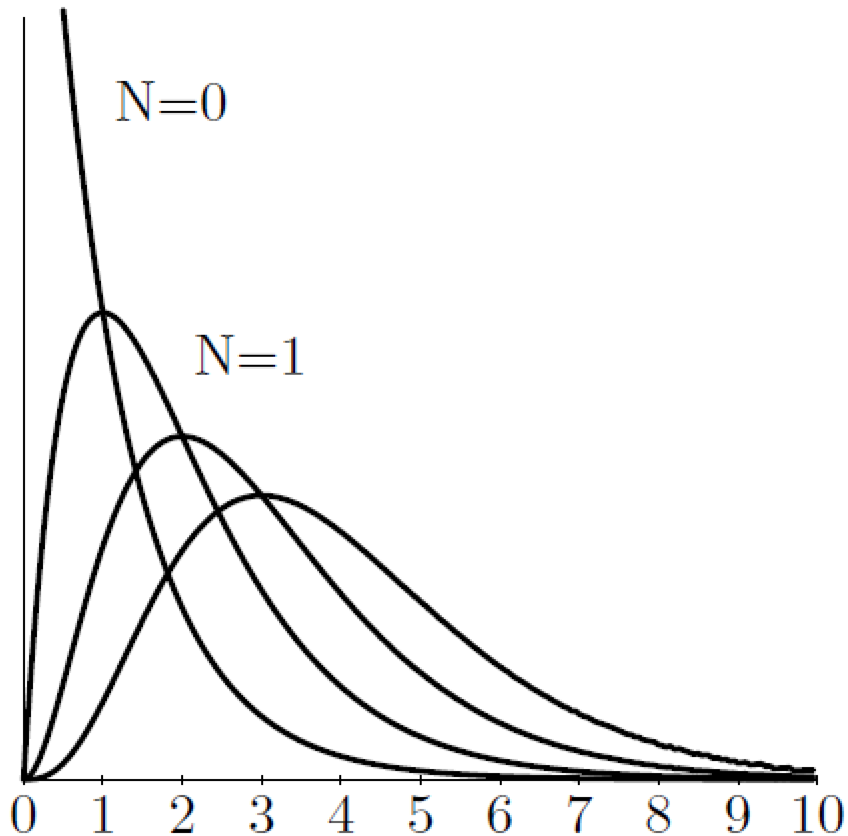
Frequentist: *if the signal is larger than U.L., the probability for the experimental outcome is small.*

Upper limit and detection limit

In particular an upper limit should not be confused with the minimal detectable signal.

Minimal detectable signal: signal that on average yields a result that is unlikely under the background only hypothesis.

Posterior for Poisson process with uniform prior

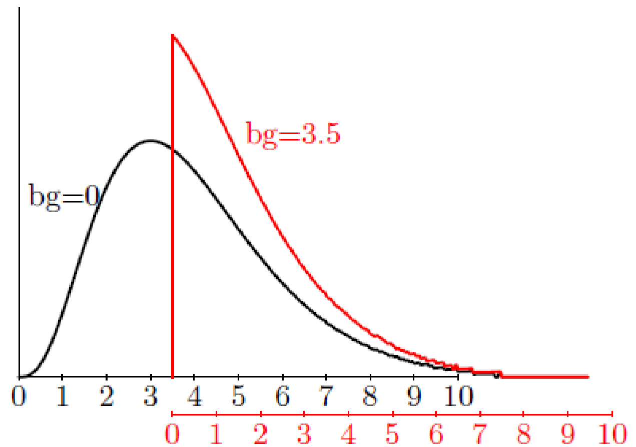


A nice property of Bayesian intervals

- Bloom: ”We are restricting the parameter to be larger than 0, aren’t we biasing the result, Jeff?”
- Scargle: ”Is the parameter supposed to be larger than 0?”
- Bloom: ”Yes, it is a cross-section”
- Scargle:”Then it should be biased in this direction”.

Credible upper limits- Poisson distribution with known background (90 % confidence level).

Observed	0	1	2	3
b = 0.0	2.3	3.89	5.32	6.68
b = 0.5	2.3	3.50	4.83	6.17
b = 1.0	2.3	3.26	4.44	5.71
b = 2.0	2.3	3.00	3.87	4.92



Some words on Jeffrey's prior.

- Jeffrey's prior is defined as the sqrt of the determinant of the Fisher Information

- Jeffrey's prior for Poisson => (1/sqrt(s)) minimizes the Fisher information in the prior and is scale invariant and a proper prior → gives unreasonable upper limits, unfortunately.

Interval estimation

EXACT FREQUENTIST INTERVALS

Frequentist intervals

- As we know the PDF for given parameter, a method to calculate confidence intervals can be reformulated as a method that finds the function $Z = f(\theta,d)$ such that the PDF in Z becomes independent of the parameter θ.

- Example: Normal theory: $Z = (X-\mu)/\sigma$. Confidence intervals can then be readily obtained from evaluating (or tabulating) the error function.

  – This is all I am going to say about Normal theory, if you want more go to the excellent books by Glen Cowan or Fred James.

- In the general case → Neyman Construction

Jerzy Neyman (1894-1981)

- born in Russia to polish parents
- Studied in Charkiv (Ukraine) and Warsaw, later also active in Berkeley.
- "Such confidence sets are easily obtained under the Bayesian assumption that the parameter is itself random with a known probability distribution, but Neyman's aim was to dispense with such an assumption, which he considered arbitrary and unwarranted."

Frequentist intervals

- Find the interval $[\theta_{ll,}\ \theta^{ul}]$ in θ – space, such that:

$$1-\alpha = p(\vec{d}\,|\,\theta_{ll} < \theta_{true} < \theta_{ul})$$

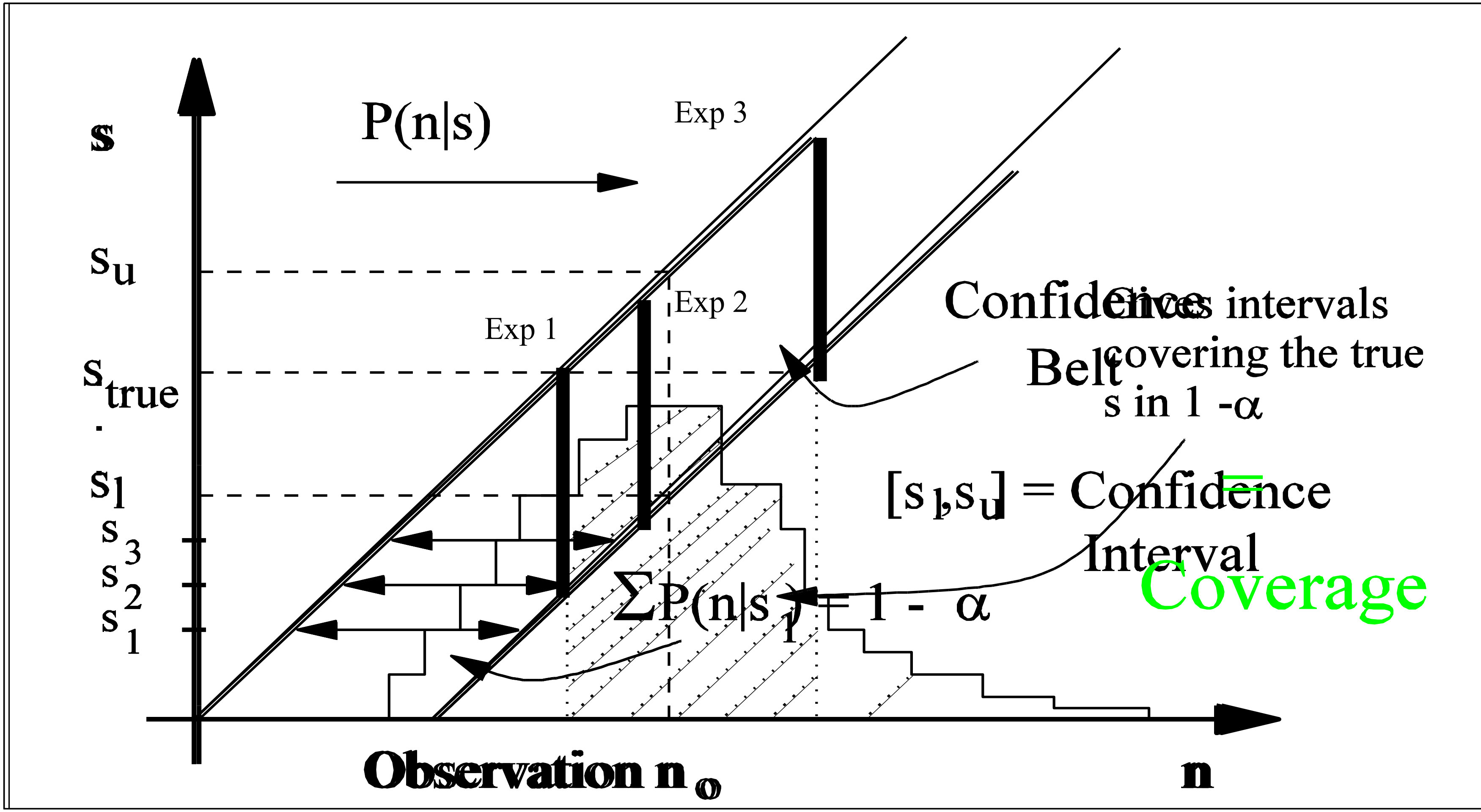
- The property needs to be fulfilled independent of the true value.

- The interval is called the **confidence interval.** The property described by above equation is called **coverage**:

In a very large number of experiments, each providing a confidence interval $[\theta_{ll,},\ \theta^{ul}]$, the fraction of intervals that contain the true value is 1- α, independent of what the true value is.

The random variable is the interval $[\theta_{ll,}\theta^{ul}]$

Exact frequentist intervals- The Neyman construction



"Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability". *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **236** (767): 333–380.

Upper limits and Central intervals

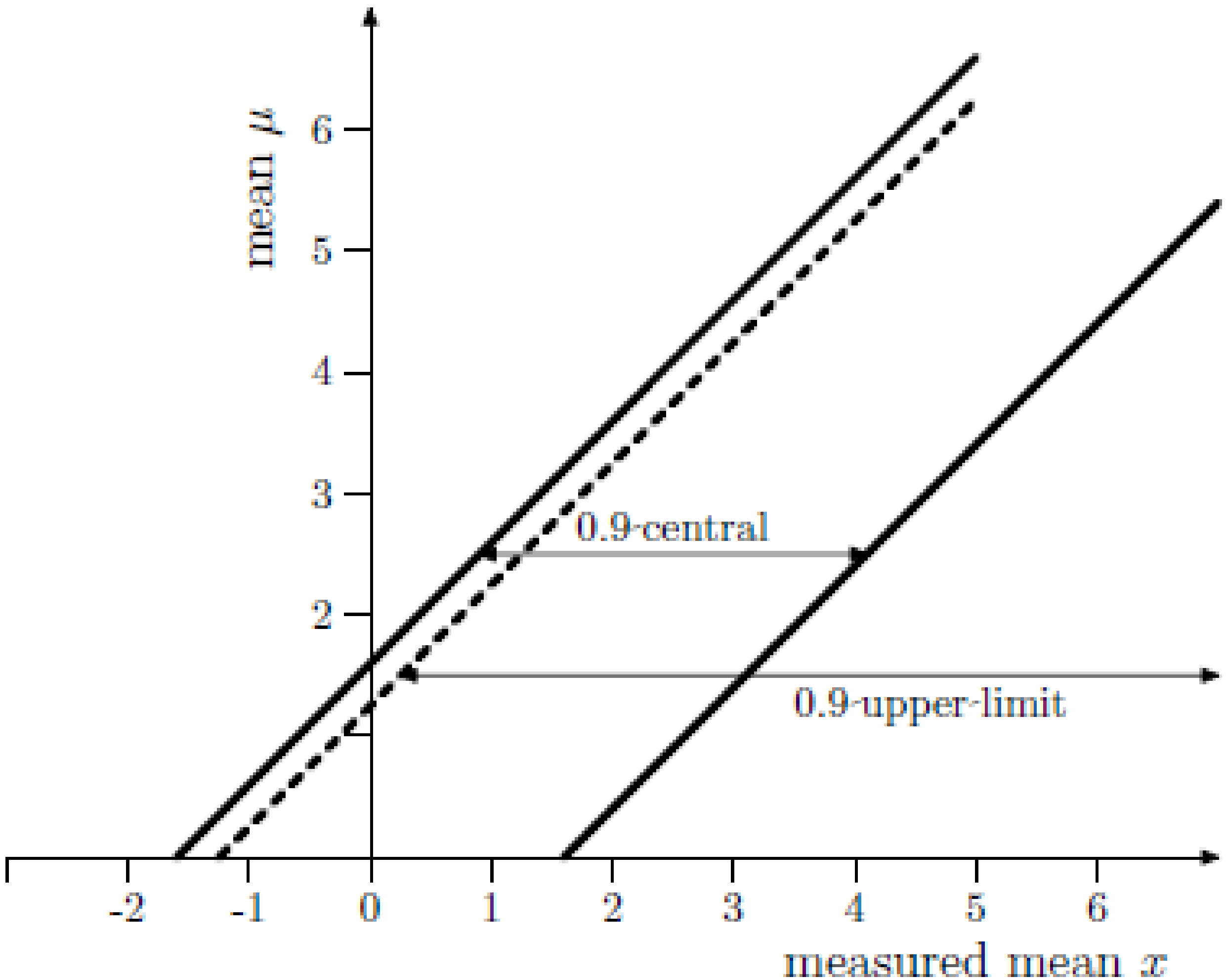
- As mentioned earlier, the requirement on the confidence belt to contain a fraction 1-α or experimental outcomes does not define it uniquely.

- There is complete freedom to choose the observational outcomes (you can come up with some yourself!)

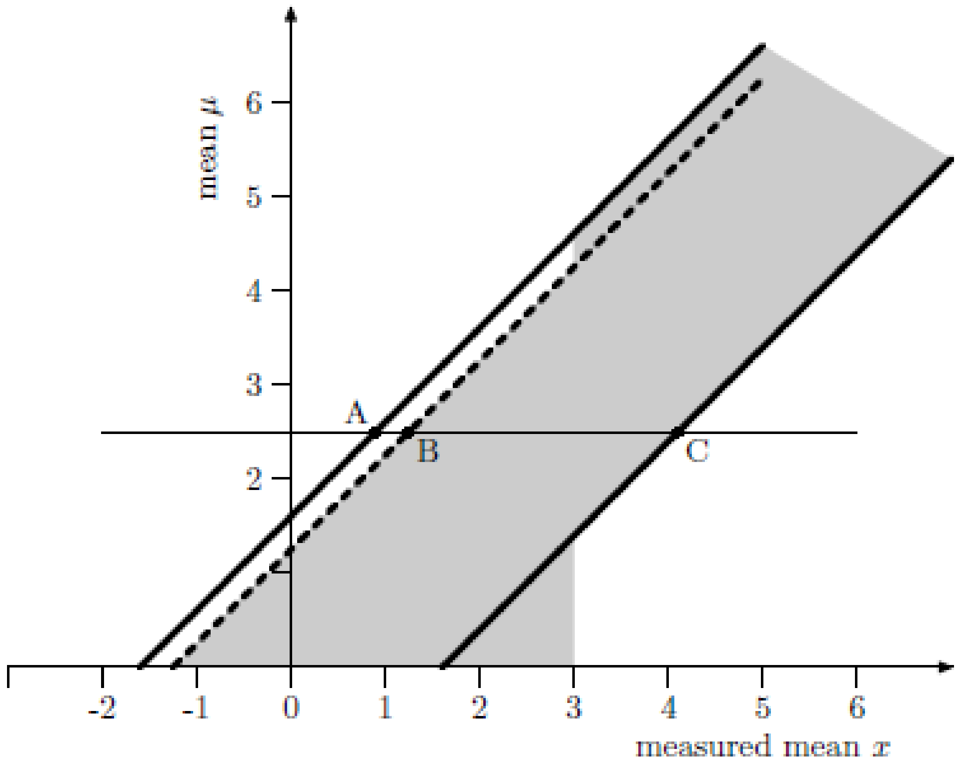
- Central interval: $\frac{1}{2}\alpha = \int_{n_{ul}}^{\infty} p(n \mid \theta) dn = \int_{-\infty}^{n_{ll}} p(n \mid \theta) dn$

- Upper limit: $\alpha = \int_{n^{ll}}^{\infty} p(n \mid \theta) dn$

Central intervals and upper limits.



Flip-flopping



Flip-flopping for a Gaussian measurement. The shaded area represents the effective confidence belt resulting from choosing to report an upper limit only when the measurement is less than 3σ above zero. This effective belt undercovers for $1.2 < \mu < 4.3$, for example at $\mu = 2.5$ where the intervals AC and $B\infty$ each contain 90% probability but BC contains only 85%.

Unified confidence intervals.

- A more clever way to choose which observations to include in the confidence interval by is computing the likelihood ratio and rank the observations accordingly

$$R = \frac{L(X\,|\,\theta_o)}{P(X\,|\,\theta_{best})}$$

- For given observation x and parameter θ_o include first the one with highest R, then next-to highest R until you reach 1-α.

- This is known as the ”Feldman&Cousins Confidence Interval”

  G. Feldman + R. Cousins, Phys.Rev. D. 57, 1998

Illustration: Poisson case confidence belt

n	$P(n\|\mu)$	μ_{best}	$P(n\|\mu_{\text{best}})$	R	rank	U.L.	central
0	0.030	0.	0.050	0.607	6		
1	0.106	0.	0.149	0.708	5	√	√
2	0.185	0.	0.224	0.826	3	√	√
3	0.216	0.	0.224	0.963	2	√	√
4	0.189	1.	0.195	0.966	1	√	√
5	0.132	2.	0.175	0.753	4	√	√
6	0.077	3.	0.161	0.480	7	√	√
7	0.039	4.	0.149	0.259		√	√
8	0.017	5.	0.140	0.121		√	
9	0.007	6.	0.132	0.050		√	
10	0.002	7.	0.125	0.018		√	
11	0.001	8.	0.119	0.006		√	

Taken from original FC paper

- Can you guess what μ and b are ?

Comparison, U.L. m unified intervals.

Observed	0	1	3
b = 0.0	**2.30**	3.89	6.68
b = 1.0	**1.30**	2.89	5.58
b = 3.0	**-0.70**	0.89	3.68

Upper limits

Observed	0	1	3
b = 0.0	**2.44**	**4.36**	**7.42**
b = 1.0	**1.61**	3.36	**6.42**
b = 3.0	**1.08**	1.88	4.42

Unified intervals

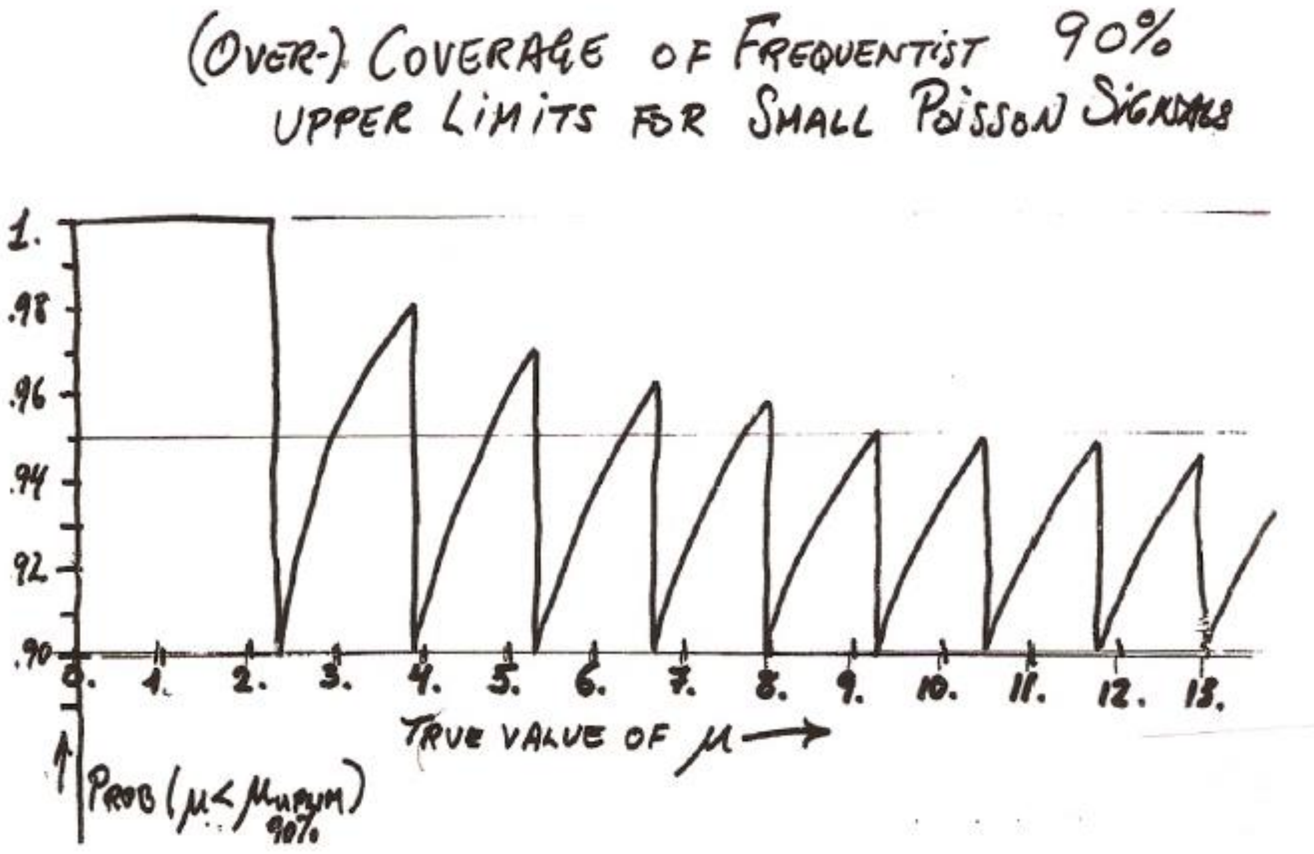
Confidence intervals for discrete data

- Confidence intervals can exactly acquire probability 1 - α only if the observable is continous.
- If the data is discrete (as is the case in the Poission distribution),we need to replace the integral with a sum and the requirement will also have to be altered as:

$$1-\alpha = \int_{x_{ll}}^{x^{ul}} p(x|\theta)dx \rightarrow 1-\alpha \geq \sum_{i=n_{ll}}^{n_{ul}} P(n\,|\,s)$$

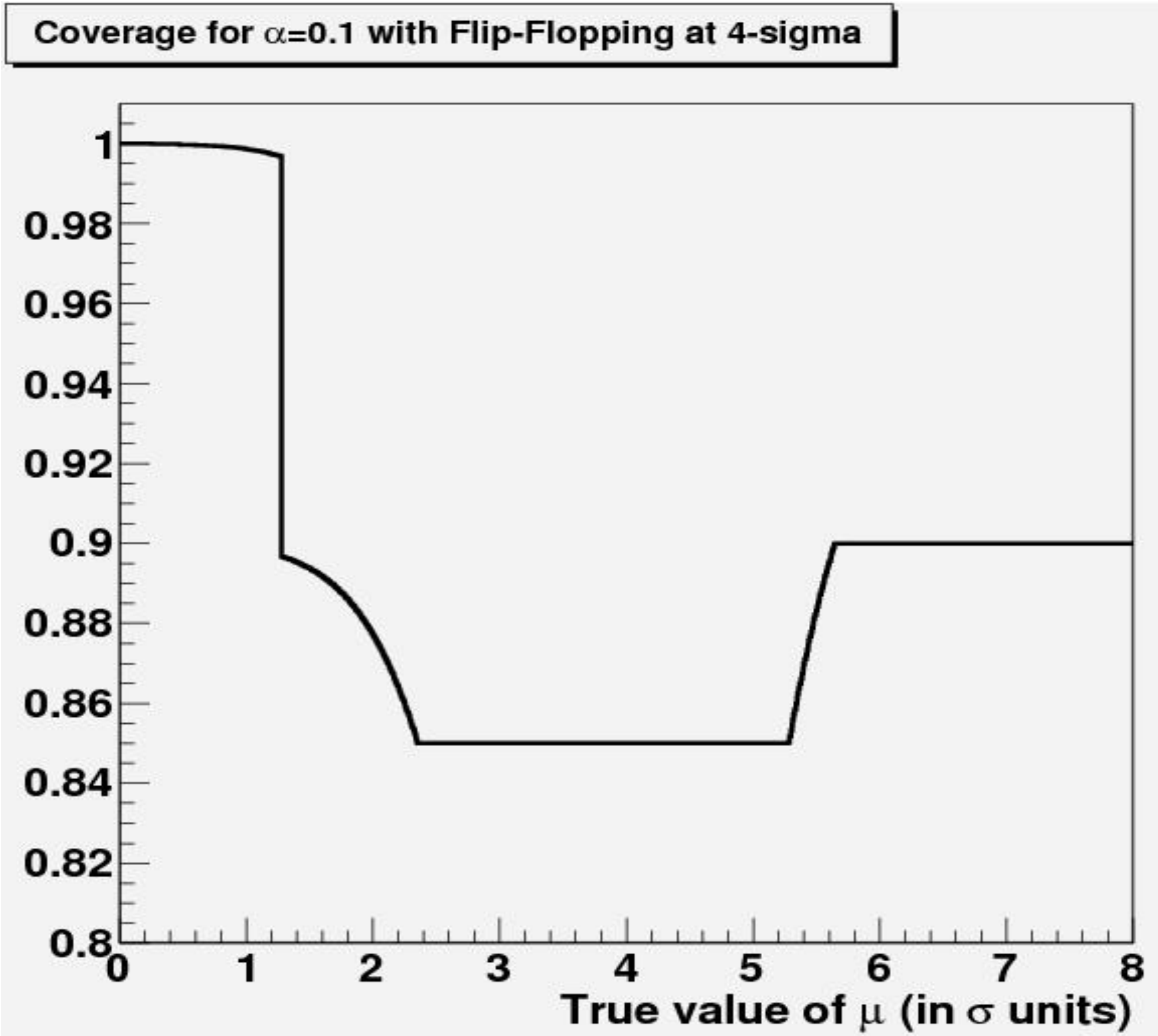
The coverage will be exactly fulfilled only for certain values of the nuisance parameter.

Upper Limits



events observed =	0	1	2	3
upper limit =	2.30	3.89	5.32	6.68

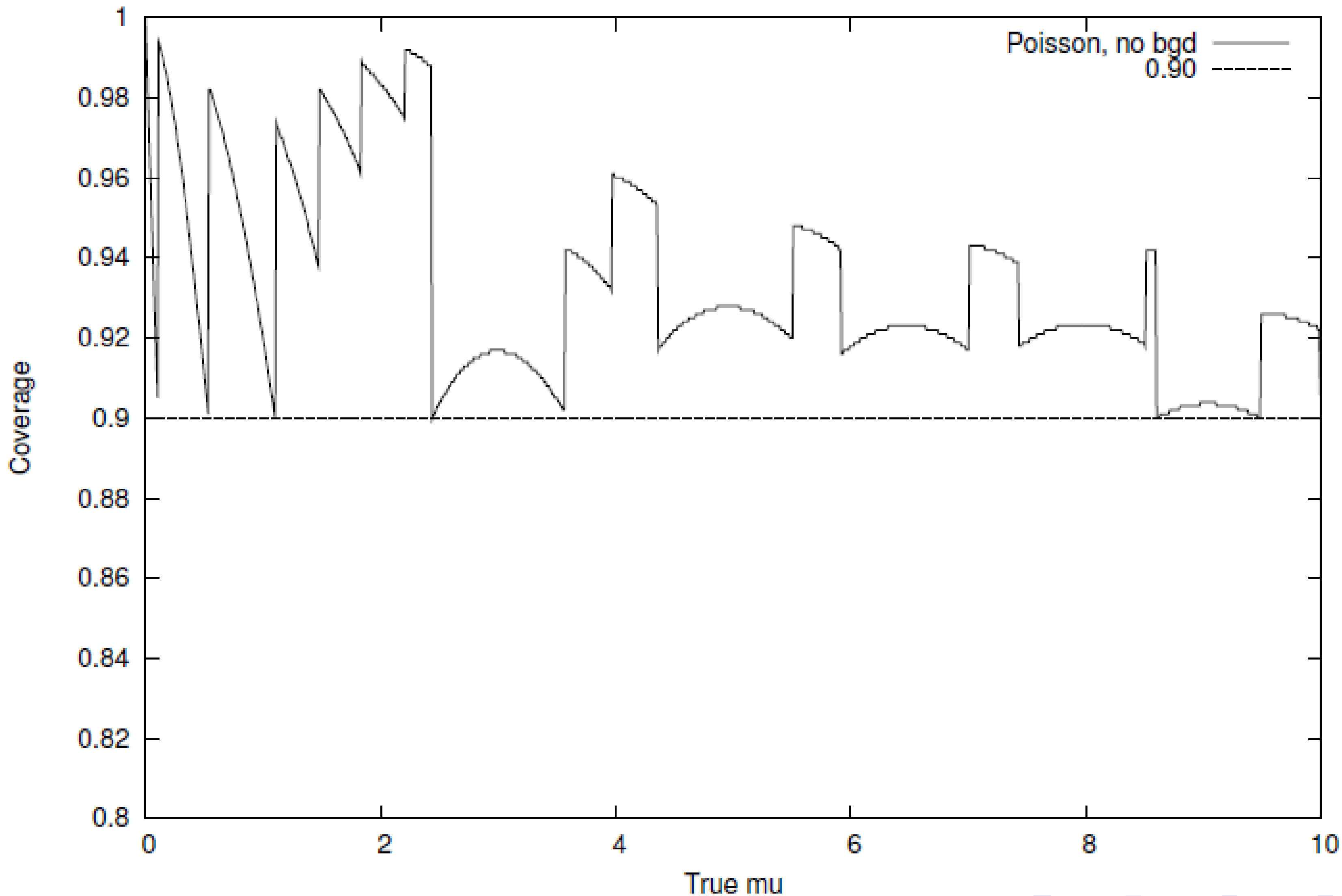
Coverage with Flip-flopping



T. Dorigo?

Two-sided intervals when measurement > 4σ away from zero (Gaussian process).

Feldman-Cousins

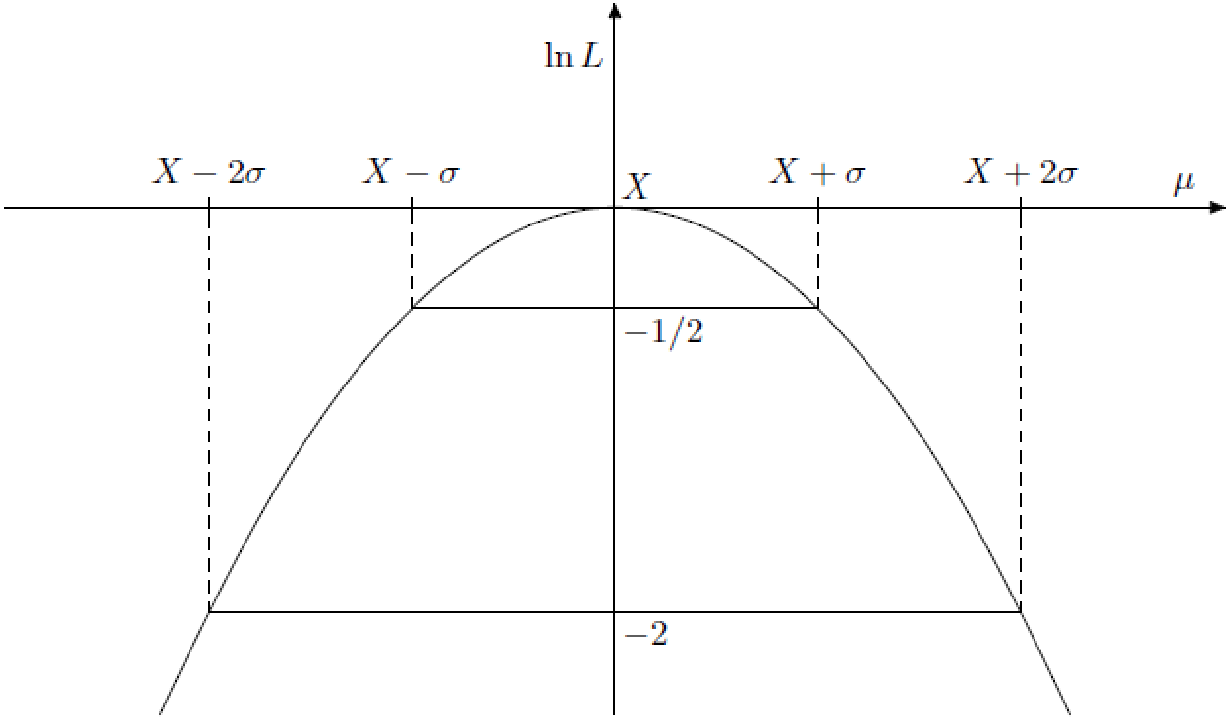


Interval estimation

LIKELIHOOD INTERVALS

Likelihood intervals

- Normal theory: pdf can be converted into a likelihood function (L) by exchanging X and μ. Then: lnL becomes a parabola:



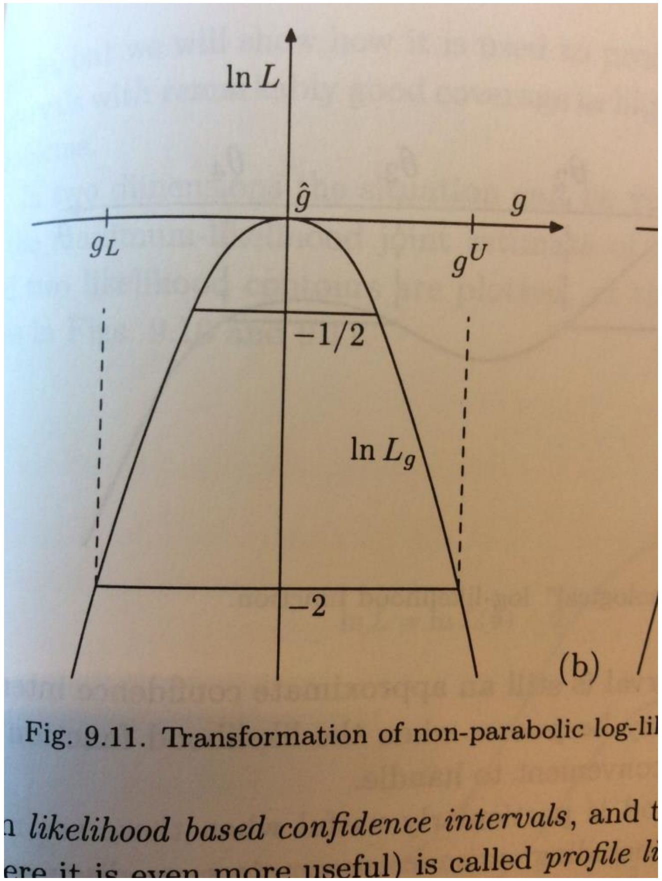
Why is this useful?

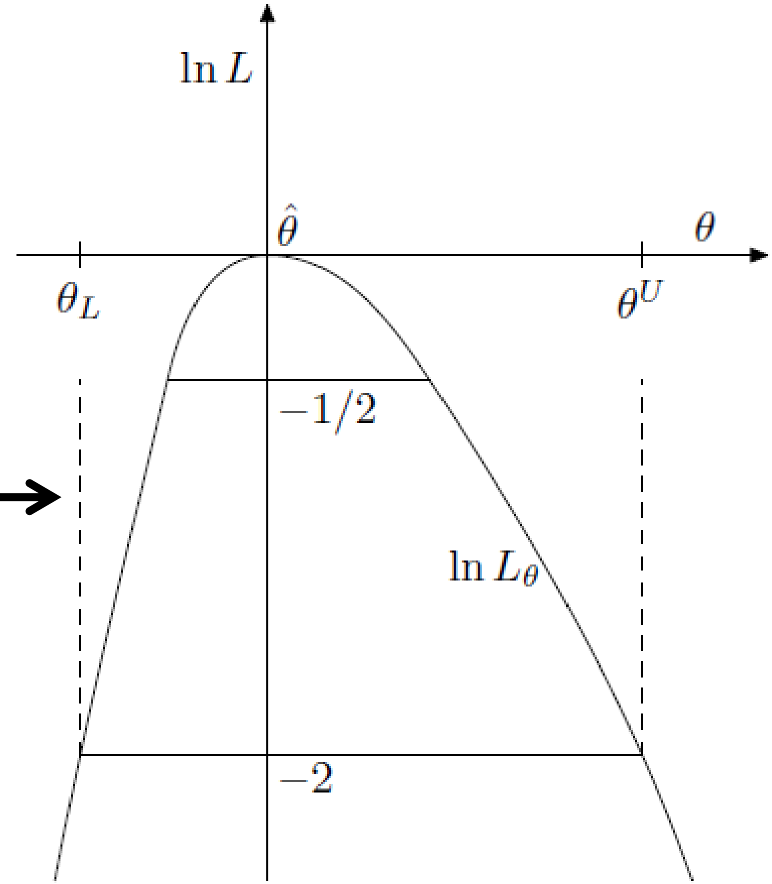
- If Normal theory is applicable, log-likelihood is parabolic
- If likelihood is parabolic, normal theory is applicable
- Assume: the likelihood is non-parabolic, but it can be transformed into a parabolic one (by a transformation of the parameter)
- But the likelihood values are invariant under this transformation, thus even in this case:

$$\ln L = \ln L_{\max} - 1/2$$

… for 1σ intervals…

Non- parabolic likelihood





F. James, Statistical Methods in Experimental Physics.

Other confidence levels and least-square fits.

$$\ln L \sim -\frac{\chi^2}{2} \rightarrow \chi^2 = \chi^2_{\min} + 1$$

Number of Parameters	Confidence level (probability contents desired inside hypercontour of $\chi^2 = \chi^2_{\min} + \mathrm{UP}$)				
	50%	70%	90%	95%	99%
1	0.46	1.07	2.70	3.84	6.63
2	1.39	2.41	4.61	5.99	9.21
3	2.37	3.67	6.25	7.82	11.36
4	3.36	4.88	7.78	9.49	13.28
5	4.35	6.06	9.24	11.07	15.09
6	5.35	7.23	10.65	12.59	16.81
7	6.35	8.38	12.02	14.07	18.49
8	7.34	9.52	13.36	15.51	20.09
9	8.34	10.66	14.68	16.92	21.67
10	9.34	11.78	15.99	18.31	23.21
11	10.34	12.88	17.29	19.68	24.71
	If `FCN` is $-\log(\text{likelihood})$ instead of χ^2, all values of `UP` should be divided by 2.				

Table 7.1: Table of `UP` for multi-parameter confidence regions

Likelihood intervals – Poisson with known background (90%/95% confidence level.)

Observed	0	1	3
b = 0.0	**1.36**	**3.65**	**6.82**
b = 1.0	**0.36**	2.65	5.81
b = 3.0	**-1.64**	0.65	3.82

likelihood

Observed	0	1	3
b = 0.0	**1.98**	**3.65**	**6.81**
b = 1.0	**0.99**	2.65	5.81
b = 3.0	**0.64**	0.65	3.81

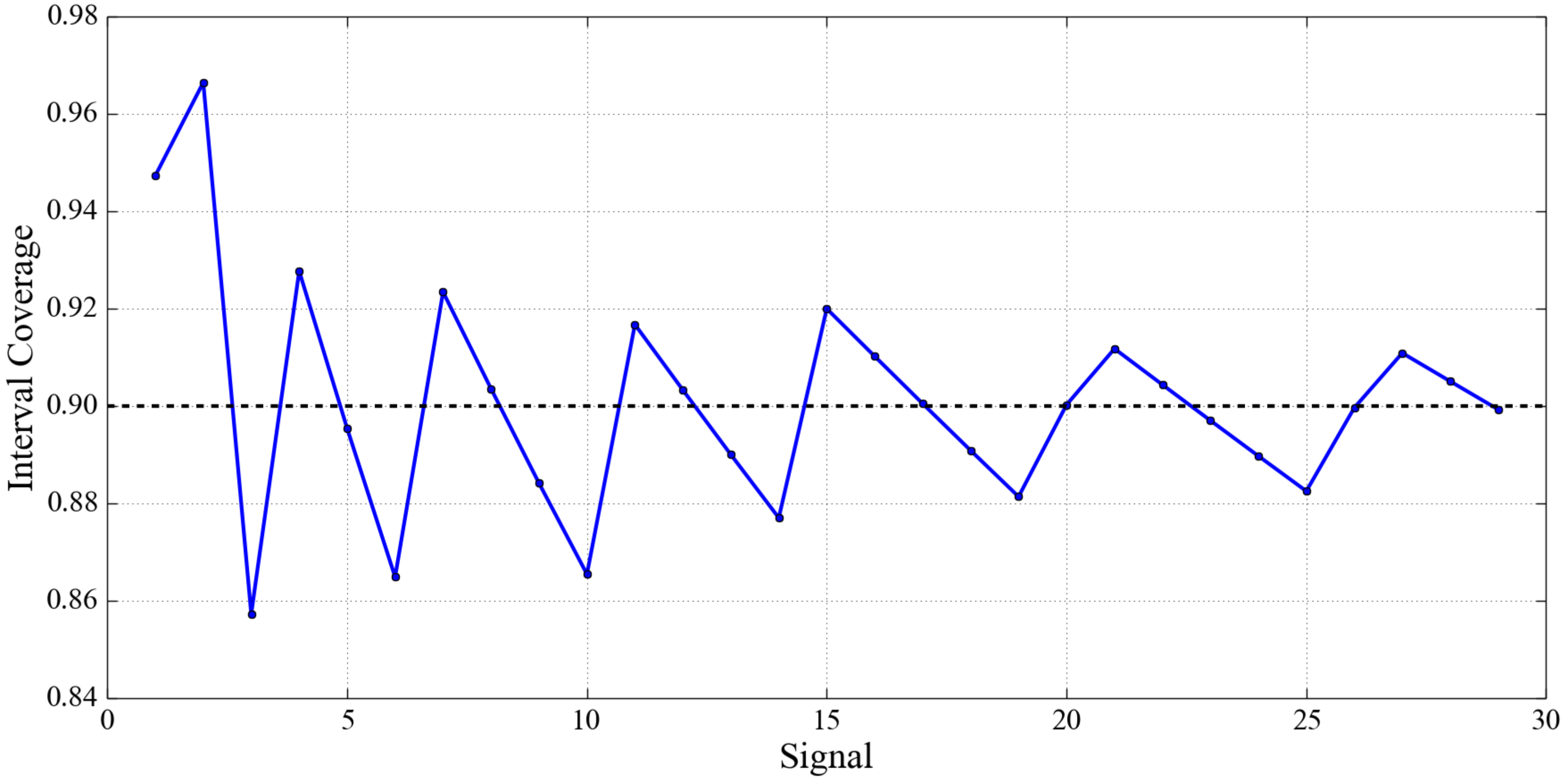
TRolke

Rolke, Lopez,Conrad, **Nucl.Instrum.Meth. A551 (2005) 493-503**

Neyman construction provide coverage, what about likelihood intervals?

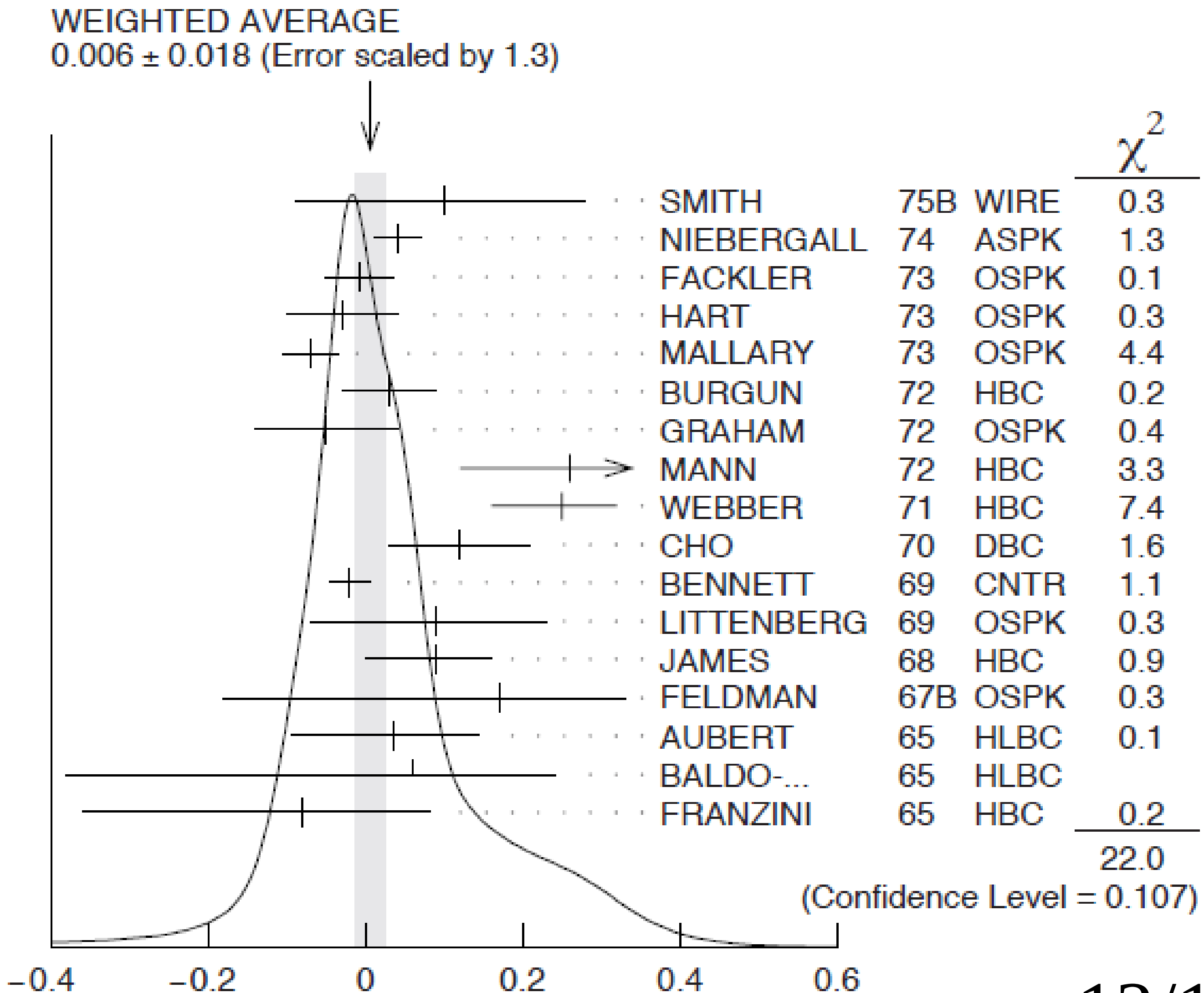
Yes, asymptotically, but not necessarily for small samples.

Coverage of likelihood intervals (Poisson 90% two sided)



Coverage in practice

•



$12/17 = 0.705$

Interval estimation

CL_S - MODIFIED FREQUENTIST

CL_s -- motivation

- The CLs method is introduced to avoid the case where a generic method (say Feldman&Cousins or likelihood) would command exclusion of signal hypotheses to which the experiment has no or little sensitivity.

- This would happen if you observe a downward fluctuation of your expected background, which might be more likely to point towards a problem in your background modelling.

- And a purely frequentist method you can produce better limits by adding background regions…..

- Can we come up with a upper limit that will allow robust statements about the signal parameter even in this case?

- A L Read 2002 *J. Phys. G: Nucl. Part. Phys.* **28** 2693

KARMEN anomaly

Liquid scintillation calorimeter and PMTs, beam-stop neutrinos

L~ 20 m

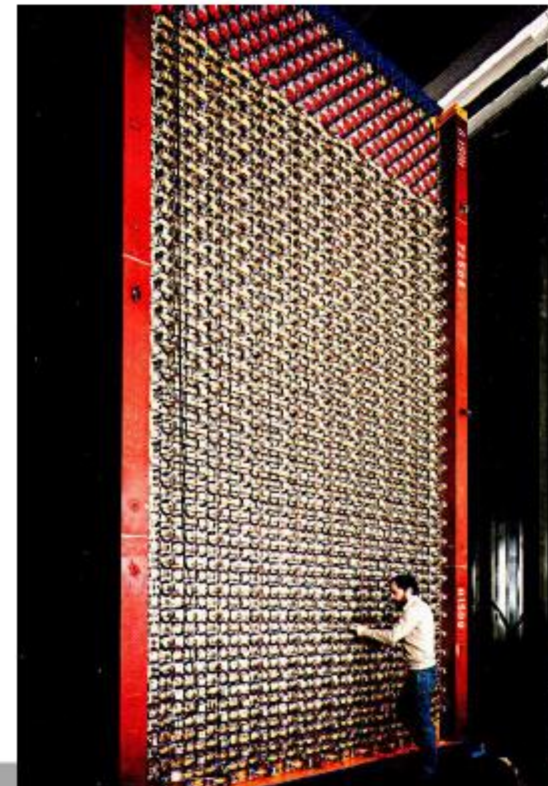
$$\nu_e + {}^{12}\mathrm{C} \rightarrow {}^{12}\mathrm{N}_{g.s.} + \boxed{e^-}$$
$$\downarrow$$
$$ {}^{12}\mathrm{C}_{g.s.} + \boxed{e^+} + \nu_e$$

<u>Neutrino experiment which</u>

→ sees no events
→ expects b = **2.9**
→ FC gives upper limit of **1.1**

<u>If experiment</u>

→ sees no events
→ expects b = **0**
→ FC gives upper limit of **2.4**



Claimed to refute LSND oscillation signal

Slide from seminar talk 2005…

Solutions to the KARMEN anomaly: **none** generally accepted !

- Roe-Woodroofe
  - Neyman Construction, FC ordering with renormalization (conditioning)

$$
q^{n_o}_{s+b}(n) = \begin{cases} \dfrac{p(n)_{s+b}}{\sum_{n'=0}^{n_o} p(n')_b} & \text{if} \quad n \leq n_o, \\[2ex] \dfrac{\sum_{n'=0}^{n_o} p(n')_b\, p(n-n')_s}{\sum_{n'=0}^{n_o} p(n')_b} & \text{if} \quad n > n_o. \end{cases}
$$

- Ciampolillo,Mandelkern & Schultz
  - based on MLE, including constraint (biased)
- Strong confidence intervals
  - consider only subset of intervals of observational space

B. Roe & M. Woodroofe, Phys.Rev.D60:053009,1999

Under-covers

B = 0	B =4
2.4	2.4

S. Ciampolillo, Nuovo Cim.A111:1415-1430, 1998
M. Mandelkern &. J. Schultz J.Math.Phys.41:5701-5709,2000

Over-covers, seriously

B = 0	B =3.0
2.6	4.7

G. Punzi, Proceedings , Durham 2002

90 % sCL

B = 0	B = 4
2.5	2.3

Generally recommended: present "sensitivity" (mean limit one would obtain in case of no signal

Now there is: CL_s

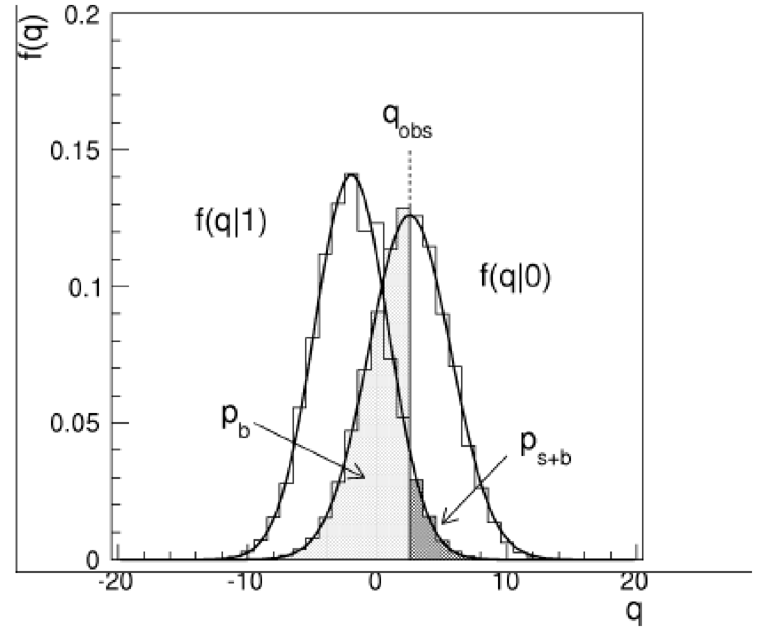
$$CL_{s+b} = P_{s+b}(q \geq q_{obs})$$

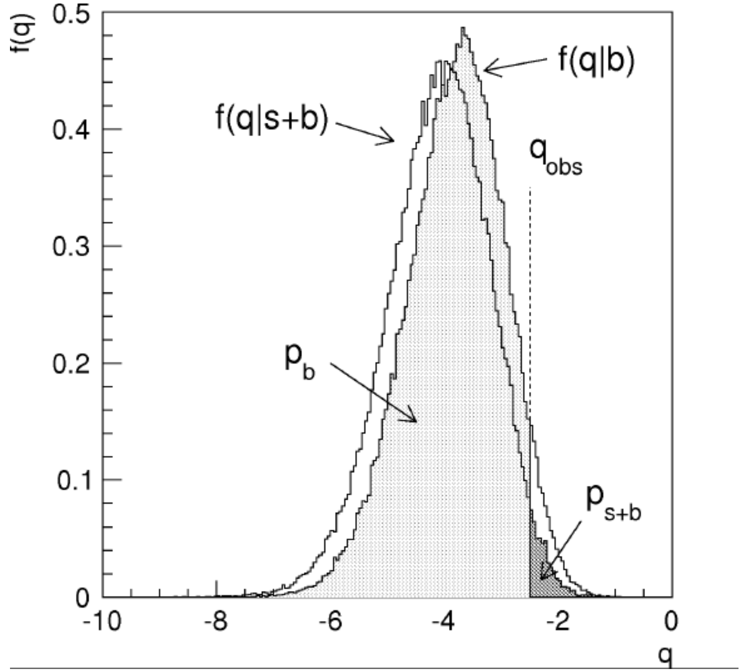
$$CL_b = 1 - p_b = P_b(q \geq q_{obs})$$

$$CL_s = \frac{CL_{s+b}}{CL_b}$$

Modified p-value:

$$p_s' = \frac{p_{s+b}}{1 - p_b}$$





Figures taken from ATLAS note for conference speakers, Glenn

Untervals CLs (Poisson with known background 90% conf. level.)

Observed CLs	0	1	3
b = 0.0	2.3	3.89	6.68
b = 1.0	2.3	3.27	5.71
b = 3.0	2.3	2.84	4.36

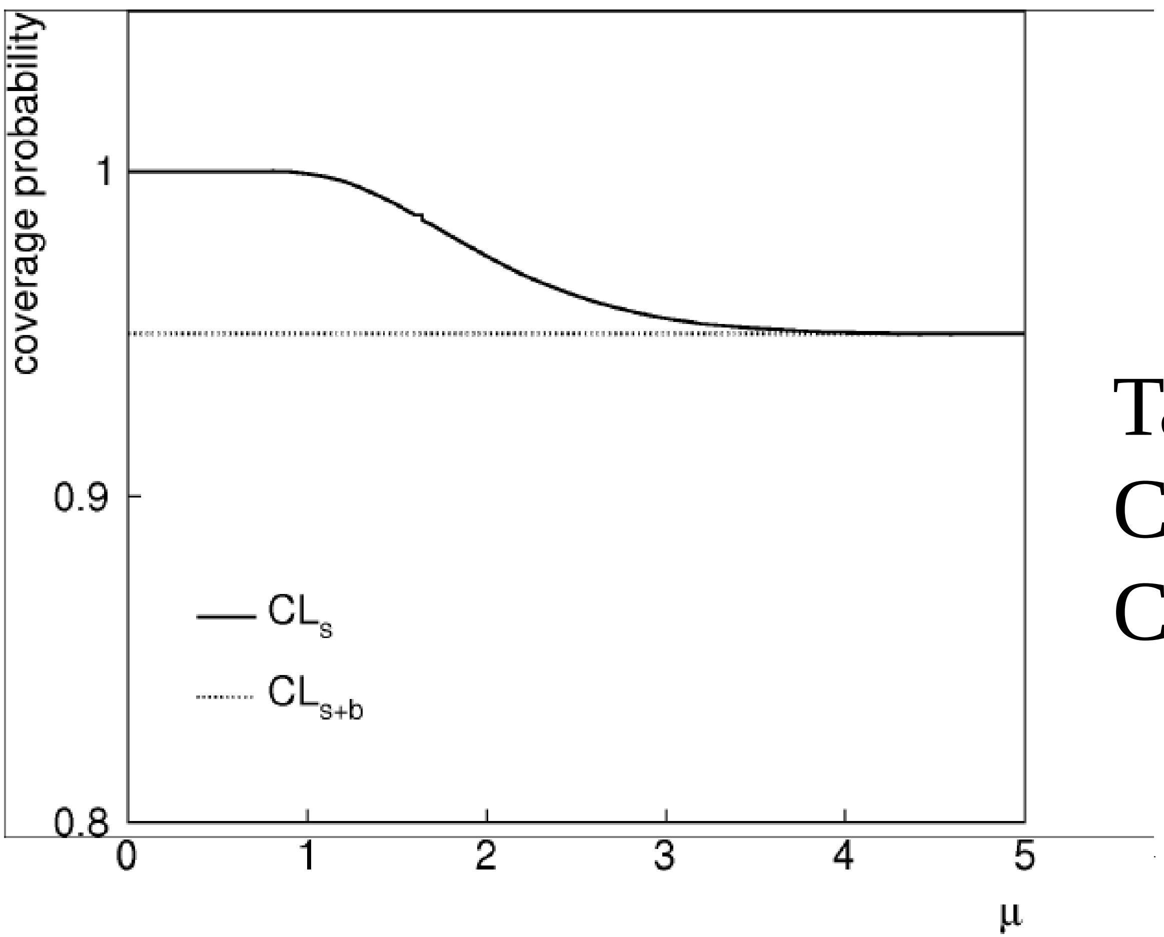
CLs

Observed Neyman	0	1	3
b = 0.0	2.3	3.89	6.68
b = 1.0	1.3	2.89	5.68
b = 3.0	-0.69	0.89	3.68

Neyman U.L.

Remarks on CL_s

- There is a prize to pay → over-coverage

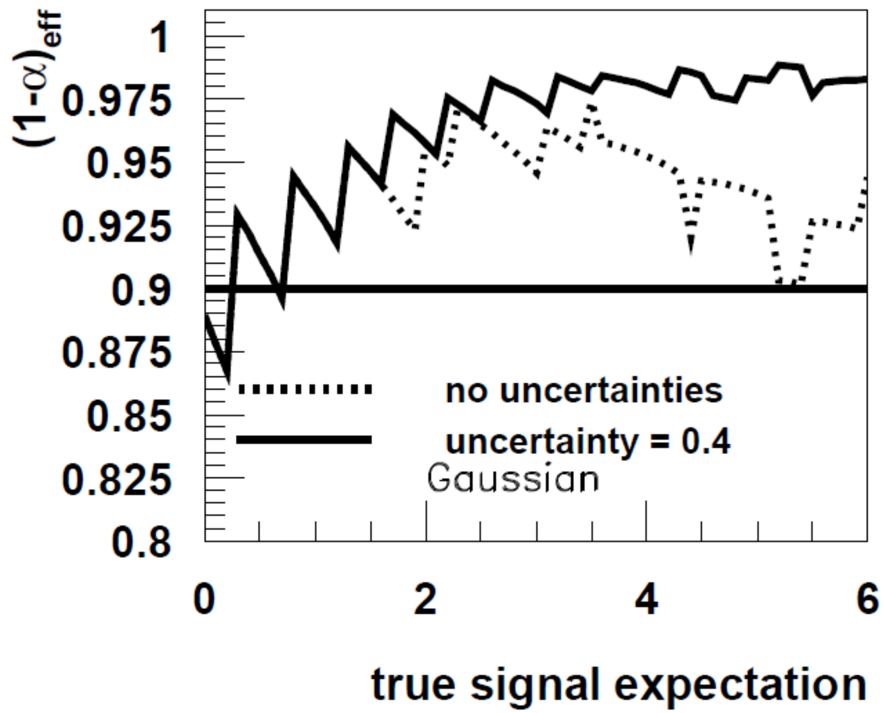


Taken from G. Cowans note on CL_s

Conditioning on the background: Roe&Woodroofe **Phys.Rev. D60 (1999) 053009**

- Unified ordering with background conditioning (conditional probability to see n events given at most n_0 background events)

$$q_{s+b}^{n_o}(n) = \begin{cases} \dfrac{p(n)_{s+b}}{\underbrace{\sum_{n'=0}^{n_o} p(n')_b}_{1-p_b}} & \text{if } n \le n_o \\[2ex] \dfrac{\sum_{n'=0}^{n_o} p(n')_b p(n-n')_s}{\sum_{n'=0}^{n_o} p(n')_b} & \text{if } n > n_o \end{cases}$$



JC+, Pre-Phyststat conference arxiv:0206034, Durham 2002

More remarks on CL_s

- CL_s applicable with any test statistic
- Same limits from Bayesian for Poisson and Gauss estimates of mean for uniform prior.
- CL_s is by now standard in ATLAS, it seems …. it is also used – guilt by association – in astroparticle physics (Xenon/LUX(?)).
- Feldman & Cousins also realized the problem of exclusion beyond sensitivity.
- They proposed to always present the actual upper limit together with the sensitivity (mean upper limit in case of background only).
- The solution of FC seems more purist in the frequentist sense.

Summary

- Parameter intervals are used to quantify the statistical accuracy of a measurement. The simplest example is the **standard error**, which is the standard deviation (**square-root of the variance of an estimate),**

- More general, the goal of **interval estimation** is to estimate intervals that **contain the true value of a parameter with given probability.** The standard error for the Gaussian case will turn out to give the interval for which this probability is 68.3 %. This will usually be a two sided interval, which we choose to report for a point measurement.

Summary

- Upper limits are a special case of confidence intervals.

Bayesian: *the degree of belief that the signal is larger than the U.L. is small.*

Frequentist: *if the signal is larger than U.L., the probability for the experimental outcome is small.*

Summary

- Four methods have been discussed so far:
  - *Credible intervals (Bayesian)*
  - *Exact frequentist intervals (Neyman construction)*
  - *Likelihood intervals*
  - CL_S

Interval estimates

NUISANCE PARAMETERS AND THEIR TREATMENT

Nuisance parameters

- Nuisance parameters are parameters in the problem which affect the result but which are not of prime interest.

- Two examples:

  - Measure the x-sec for dark matter annihilation and estimate an interval on it. Mass of dark matter particle is then a nuisance parameter.

  - Measure the rate of a process and estimate an interval on it. Background expectation is a nuisance parameter.

Nuisance parameter and systematic uncertainties

- Example 1: both parameters are of interest, a confidence interval (ellipse) in both parameters would be relevant.

- Example 2: background is an experimental uncertainty. A confidence interval in both the signal strength and background strength is not very interesting

→want to report confidence interval only in signal strength, however, taking into account the uncertainty in background → "project" on parameter of interest → how?

Nuisance parameters and systematic uncertainties

- Systematic uncertainties: uncertainties that do not become smaller with increasing size of data sample.

- I will be using the more general definition: uncertainties in parameters that are determined in ancillary experiments.

There are two general methods

- Profile likelihood (frequentist):

$$\ln L(\vec{\theta}) \to \lambda(\theta_k) = \max_{\theta_i, i \neq k} \ln L(\vec{\theta})$$

- Marginalisation (Bayesian)

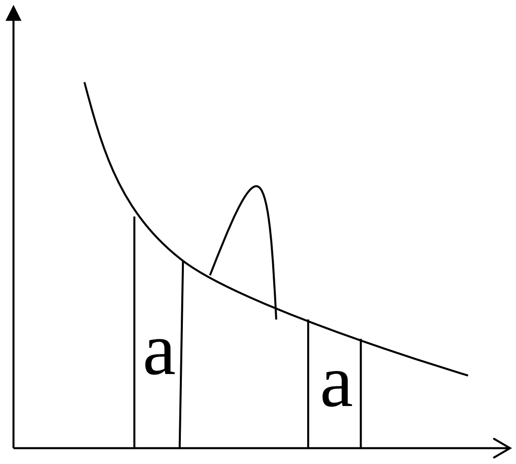
$$L(\vec{\theta}) \to L_{eff}(\theta_k) = \int d\theta_{1\ldots.i\neq k} L(\theta_{1\ldots.i\neq k})$$

Frequentist treatment:
hybrid (freq./bayesian)

Calculate posterior
from it: Bayesian

Let us consider a concrete example.

- Search for excess over background.
- Background determined by side-band/control region measurements.



Let us write down the likelihood

- Sideband measurement of background:

$$f(x, y|\mu, b) = \frac{(\mu + b)^x}{x!} e^{-(\mu+b)} \cdot \frac{(\tau b)^y}{y!} e^{-\tau b}$$

- And another common case, normal uncertainties on background and detector efficiency.

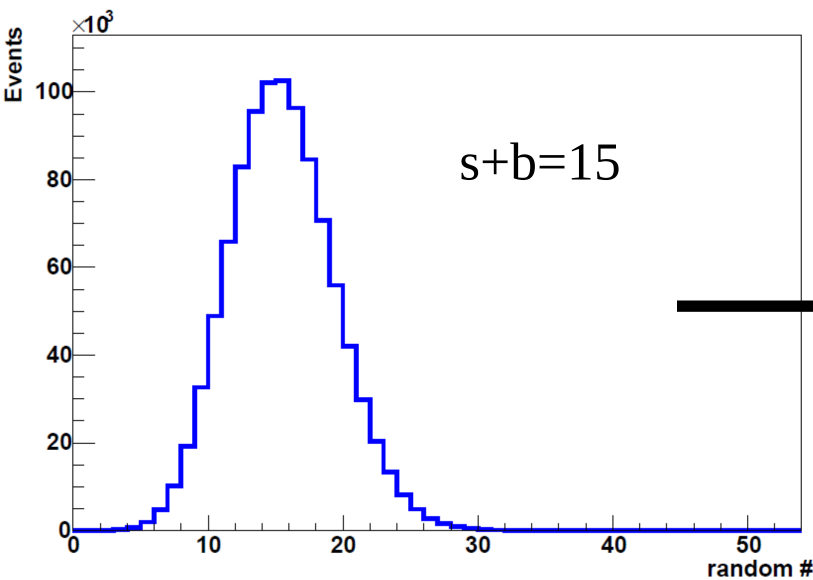
$$X \sim Pois(e\mu + b), \qquad Y \sim N(b, \sigma_b), \qquad Z \sim N(e, \sigma_e)$$

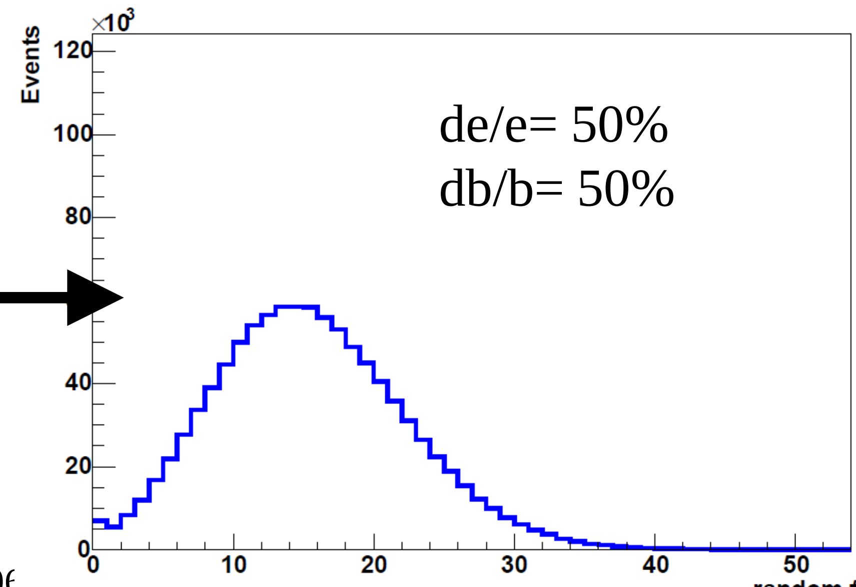
Credible Intervals –marginalisation

- In the Bayesian approach you can find the posterior for the nuisance parameter and integrate over them, eg. with Gaussian uncertainties on efficiency and background this could look like:

$$1-\alpha(s_1,s_2) \propto \iiint P(\varepsilon_{true}s+b\,|\,n)G(\varepsilon_{true}\,|\,\varepsilon_{meas})G(b\,|\,b_{meas})d\varepsilon_{true}dbds$$

- i.e. a Poisson convolved with two Gaussians.



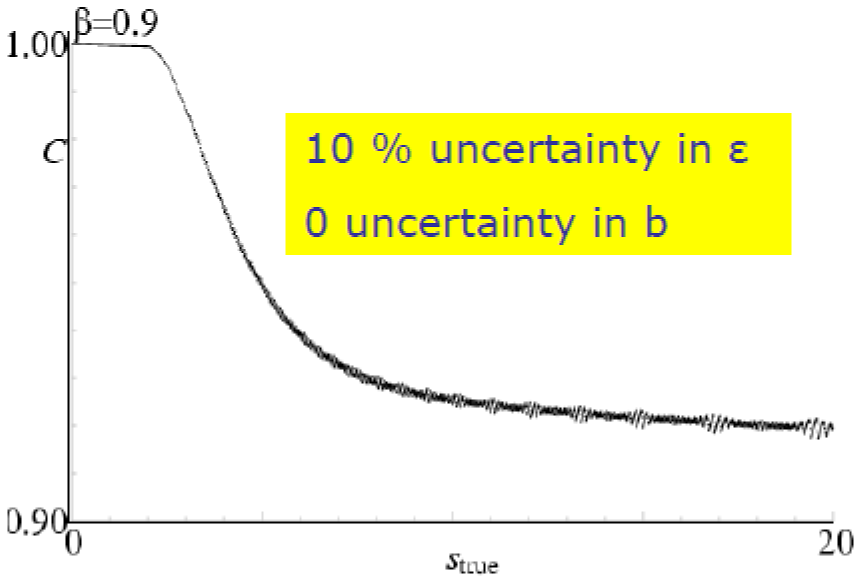


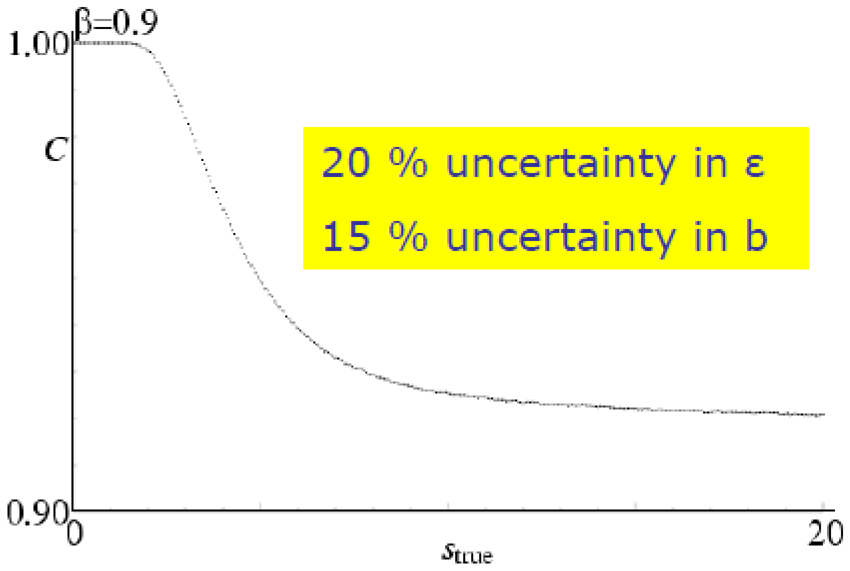
Coverage of Bayesian intervals

- A general approach for physicists (and also for statisticians) is to use Bayesian methods and to study their frequentist properties.
  - →Bayesian method acceptable if reasonable frequentist properties
  - →Frequentist properties maybe a good diagnostic for problems in the Bayesian approach.
  - → Let us give an example …..

Generically overcovering

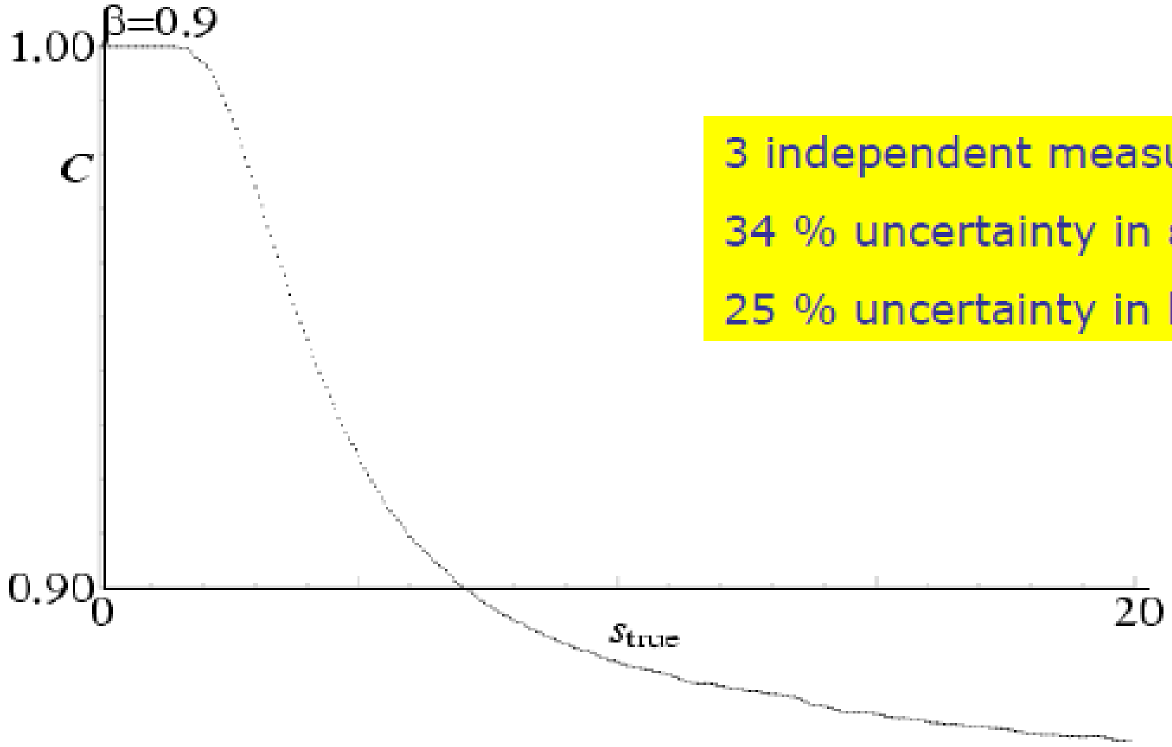
- Poisson process with uncertainty in background and efficiency





J. Heinrich, PHYSSTAT 2005

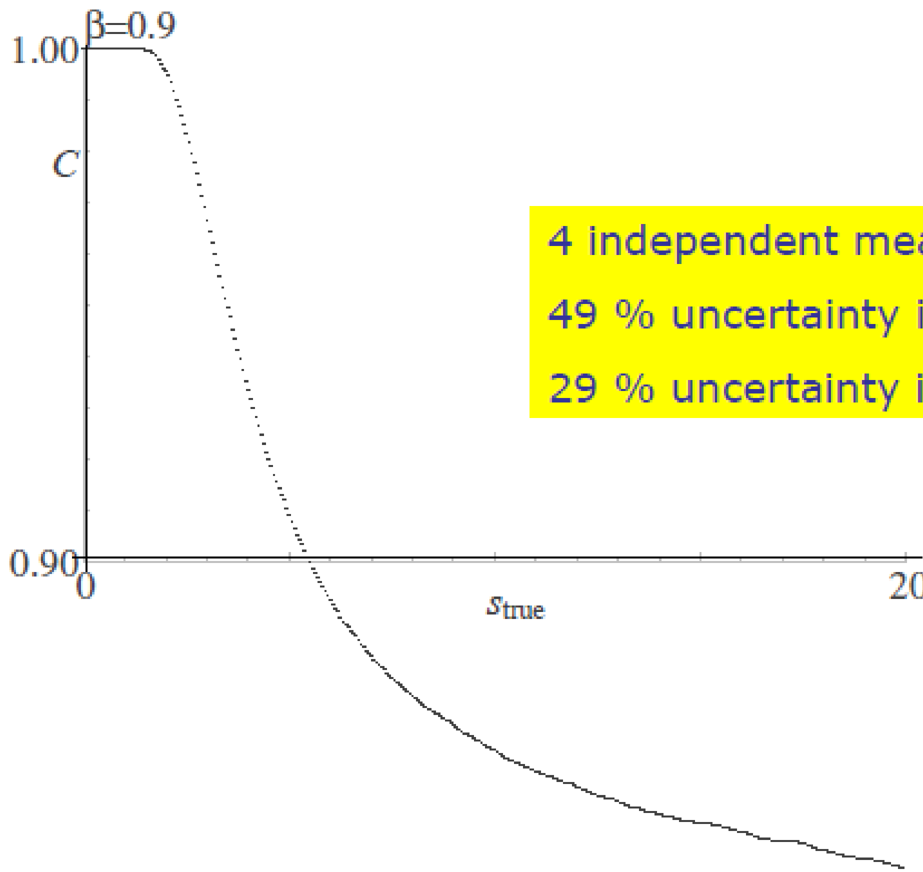
Combining several measurments



3 independent measurements

34 % uncertainty in ε (per channel)

25 % uncertainty in b (per channel)



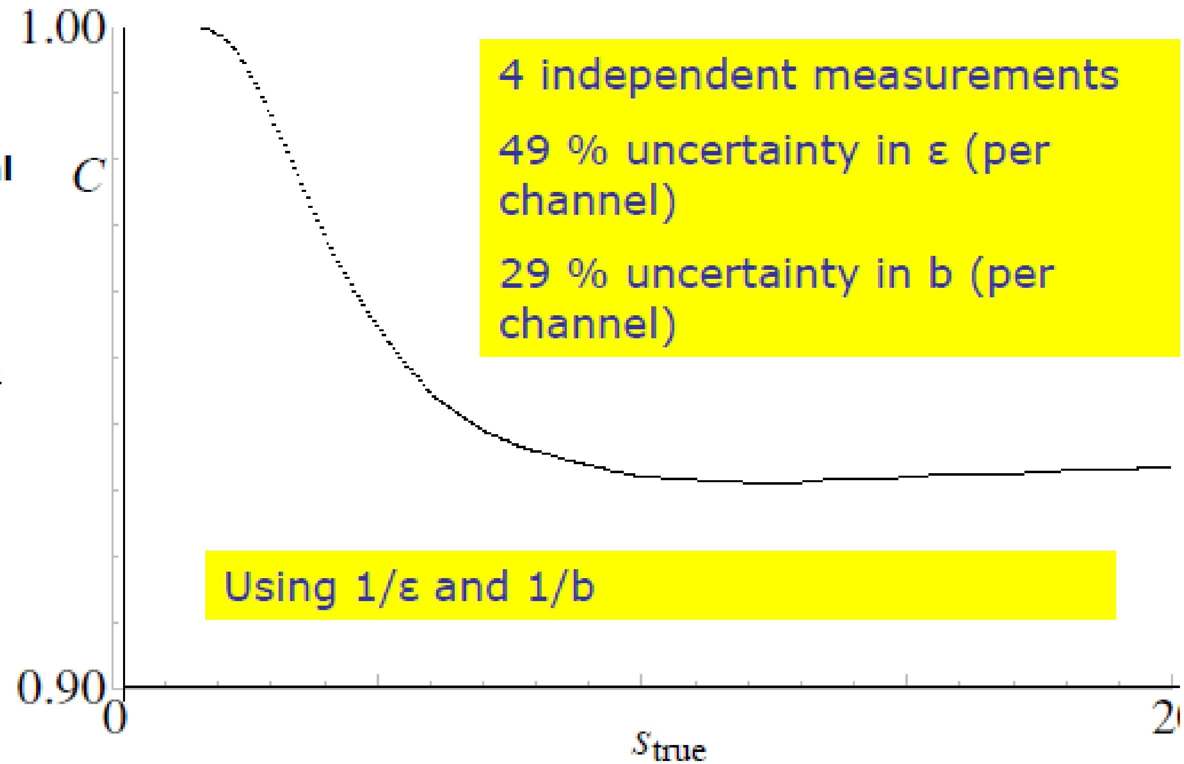
4 independent measurements

49 % uncertainty in ε (per channel)

29 % uncertainty in b (per channel)

Can be fixed with choice of prior.

- In hindisight, trivial:
  - A flat prior for each of N channels leads to a
    - **ε^{N-1} prior for total acceptance**
- So: different prior: in this case simple
  - **$1/\varepsilon$ in each channel**



Profile likelihood

See Cowan et al for an extensive discussion of asymptotic properties and

useful modifications

- A very convenient way to reduce the dimensionality to the parameter of interest is the **profile likelihood**, i.e. for given parameter of interest $x_{k,}$ the likelihood is maximized with respect to all other parameters:

$$\lambda(\theta_k) = \max_{\theta_i, i \neq k} \ln L(\vec{\theta})$$

- Interval inference is then performed on λ (i.e.: for 1σ uncertainties:

$$\lambda(\theta_k) = \lambda(\theta_k) - 1/2$$

Example: background nuisance parameter

$$f(x,y|\mu,b) = \frac{(\mu+b)^x}{x!} e^{-(\mu+b)} \cdot \frac{(\tau b)^y}{y!} e^{-\tau b}$$

$$\hat{b}(\mu) = \frac{x+y-(1+\tau)\mu + \sqrt{(x+y-(1+\tau)\mu)^2 + 4(1+\tau)y\mu}}{2(1+\tau)}$$

$$\lambda(\mu|x,y) = \frac{L(\mu,\hat{b}(\mu)|x,y)}{L(\widehat{\mu},\widehat{b}|x,y)}$$

Rolke, Lopez,Conrad, **Nucl.Instrum.Meth. A551 (2005) 493-503**

Example 2: uncertainty in detection efficiency

$$X \sim Pois(e\mu + b), \qquad Y \sim N(b, \sigma_b), \qquad Z \sim N(e, \sigma_e)$$

$$\frac{\partial}{\partial b} \log l(\mu, b, e | x, y, z) = \frac{x}{e\mu + b} - 1 + \frac{(y - b)}{\sigma_b} \doteq 0$$

$$\frac{\partial}{\partial e} \log l(\mu, b, e | x, y, z) = \frac{x}{e\mu + b} - \mu + \frac{(z - e)}{\sigma_e} \doteq 0$$

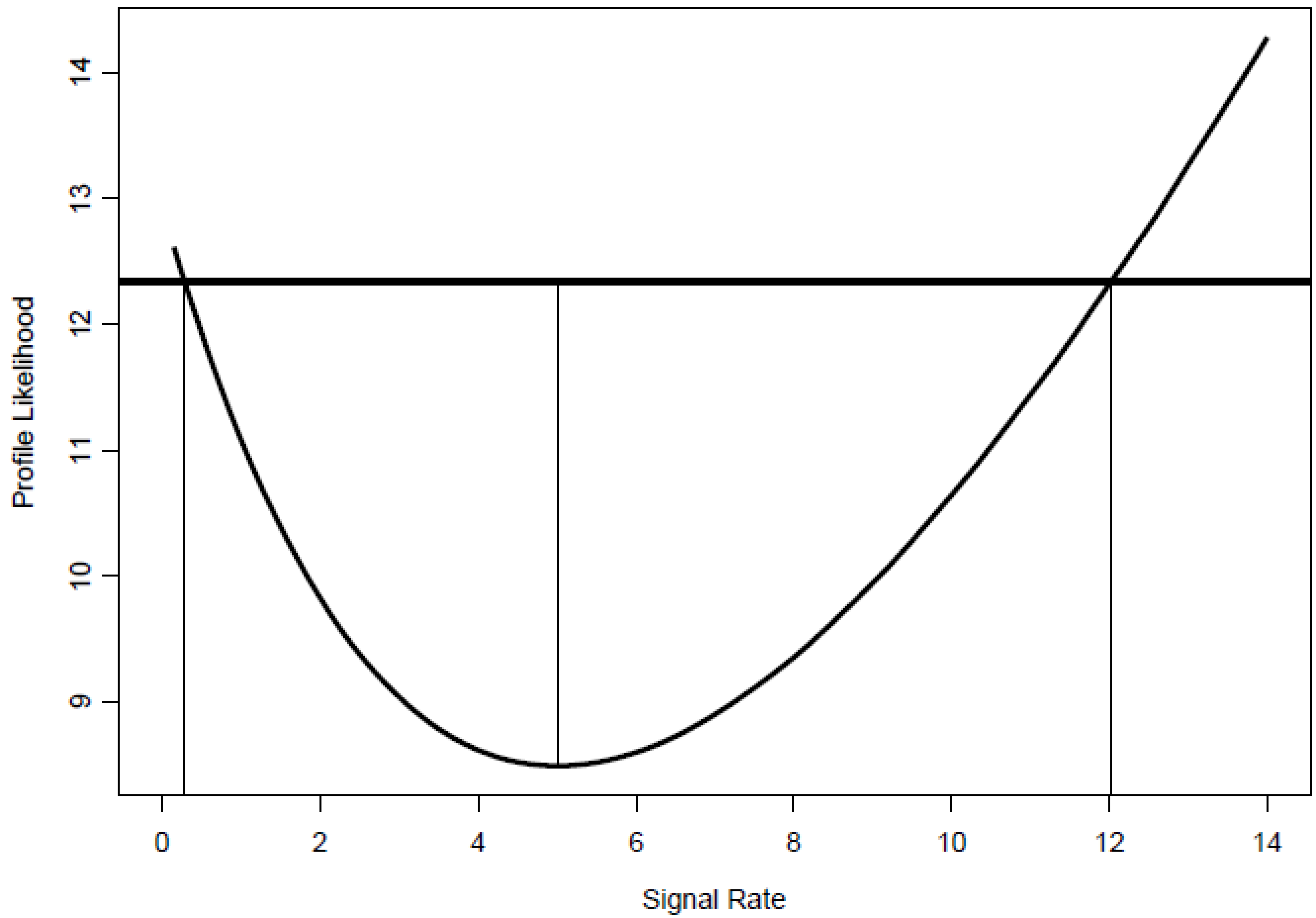
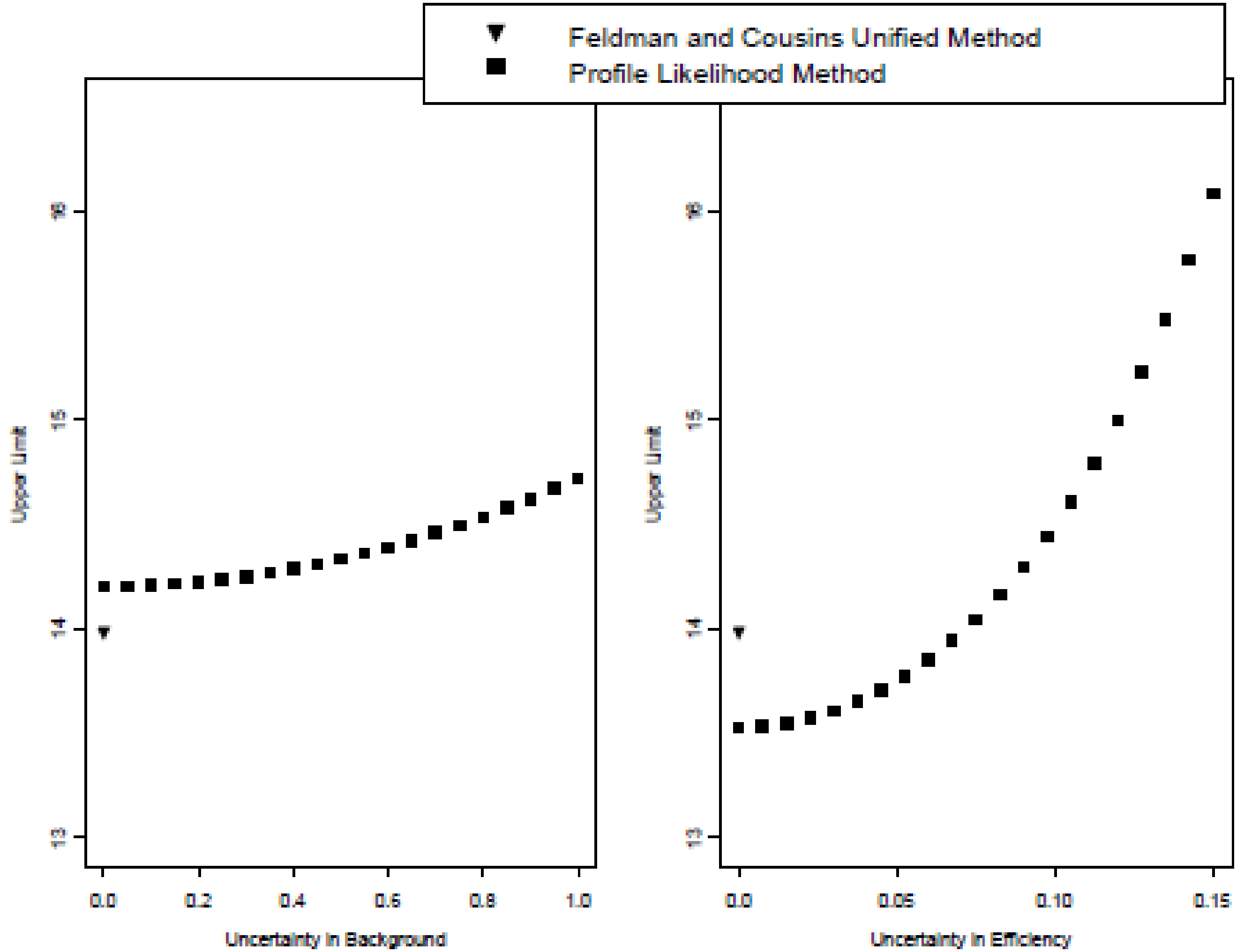
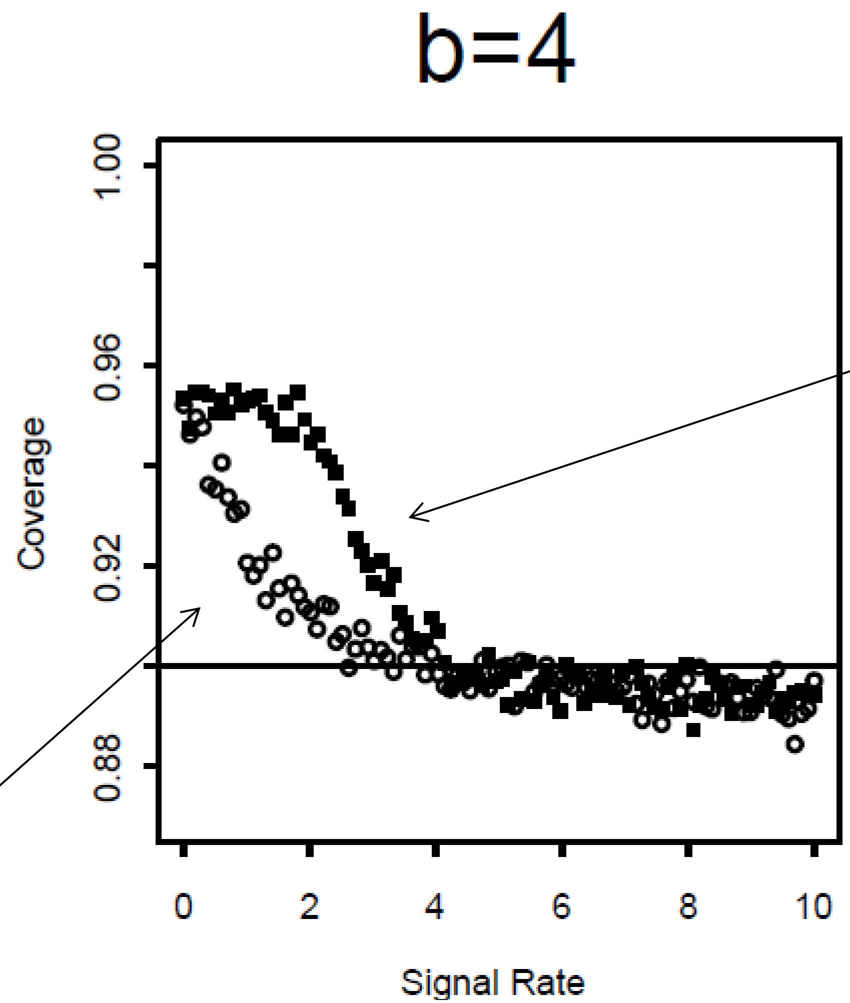


Figure 1: Example of the $-2\log\lambda$ curve. This is the case $x = 8$, $y = 15$ and $\tau = 5.0$. We find the 95% confidence interval to be $(0.28, 12.02)$.


Feldman and Cousins Unified Method
Profile Likelihood Method
Upper Limit
16
15
14
13
0.0
0.2
0.4
0.6
0.8
1.0
Uncertainty in Background
0.0
0.05
0.10
0.15
Uncertainty in Efficiency

Likelihood intervals: coverage, background as nuisance parameter



MLE set to 0 if < 0

Lower limit set to 0 if MLE<0.

Gaussians with $\sigma_b = 0.5$, $e = 0.85$ and $\sigma_e = 0.075$.

TRolke

class TRolke: public TObject

```
TRolke

This class computes confidence intervals for the rate of a Poisson
process in the presence of uncertain background and/or efficiency.

The treatment and the resulting limits are fully frequentist. The
limit calculations make use of the profile likelihood method.

    Author: Jan Conrad (CERN) 2004
    Updated: Johan Lundberg (CERN) 2009

    Copyright CERN 2004,2009           Jan.Conrad@cern.ch,
                                   Johan.Lundberg@cern.ch

For a full list of methods and their syntax, and build instructions,
consult the header file TRolke.h.


Examples/tutorials are found in the separate file Rolke.C
```

- Useful if you have Poisson process with uncertainties on efficiency or background

Historical slide: profile likelihood for marked Poisson

Higgs search

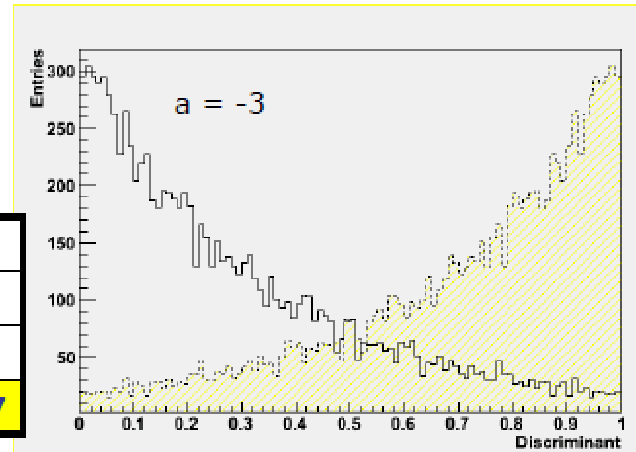
- Not straw-man, but still TOY (e.g. qqH→ qqττ → qqll(missEt):

$$\mathcal{L}(n, n_b, \vec{x}|s, b) = P(n|s+b)P(n_b|\tau b)\prod_{i=1}^{n_{obs}} \frac{s f_s(x|s) + b f_b(x|b)}{s+b}$$

$$f_s(x|s) = \frac{1}{N}e^{-a(x-1)} \quad f_b(x|b) = \frac{1}{N}e^{-ax}$$

ominal cov = 5σ

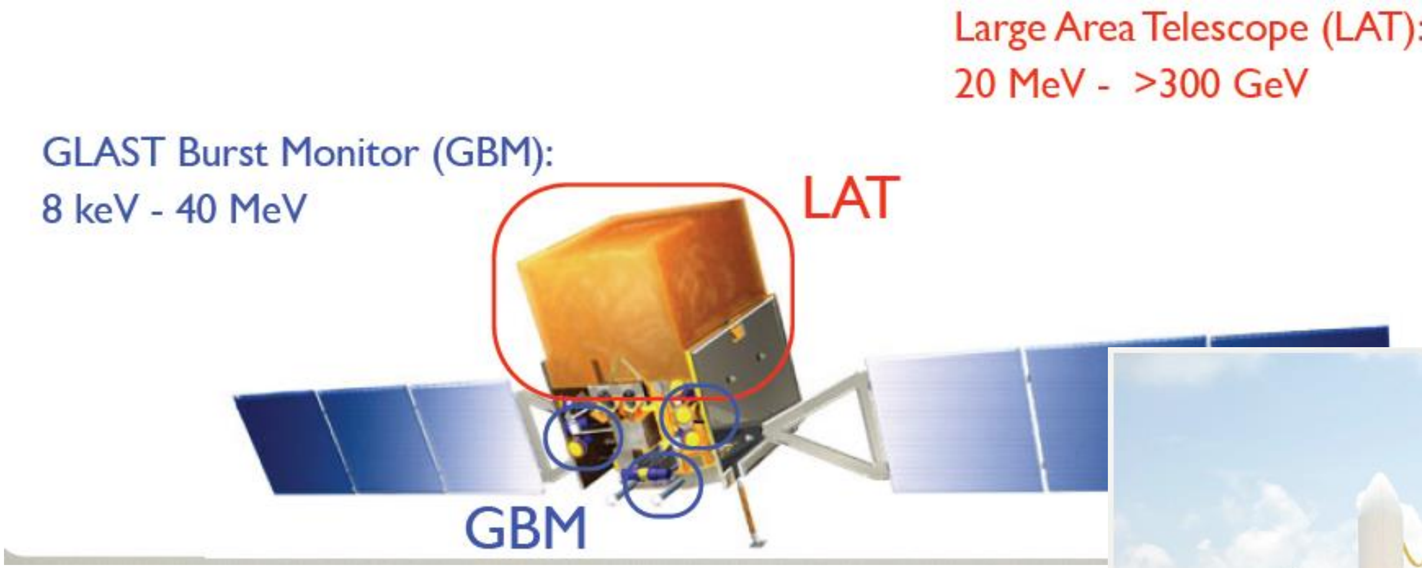
	b	τ	a	σ_{calc}	Pow
	20	1	-3	**5.2**	-
	20	1	-1	**5.1**	-
0	50	1	-3	**5.1**	**0.9997**



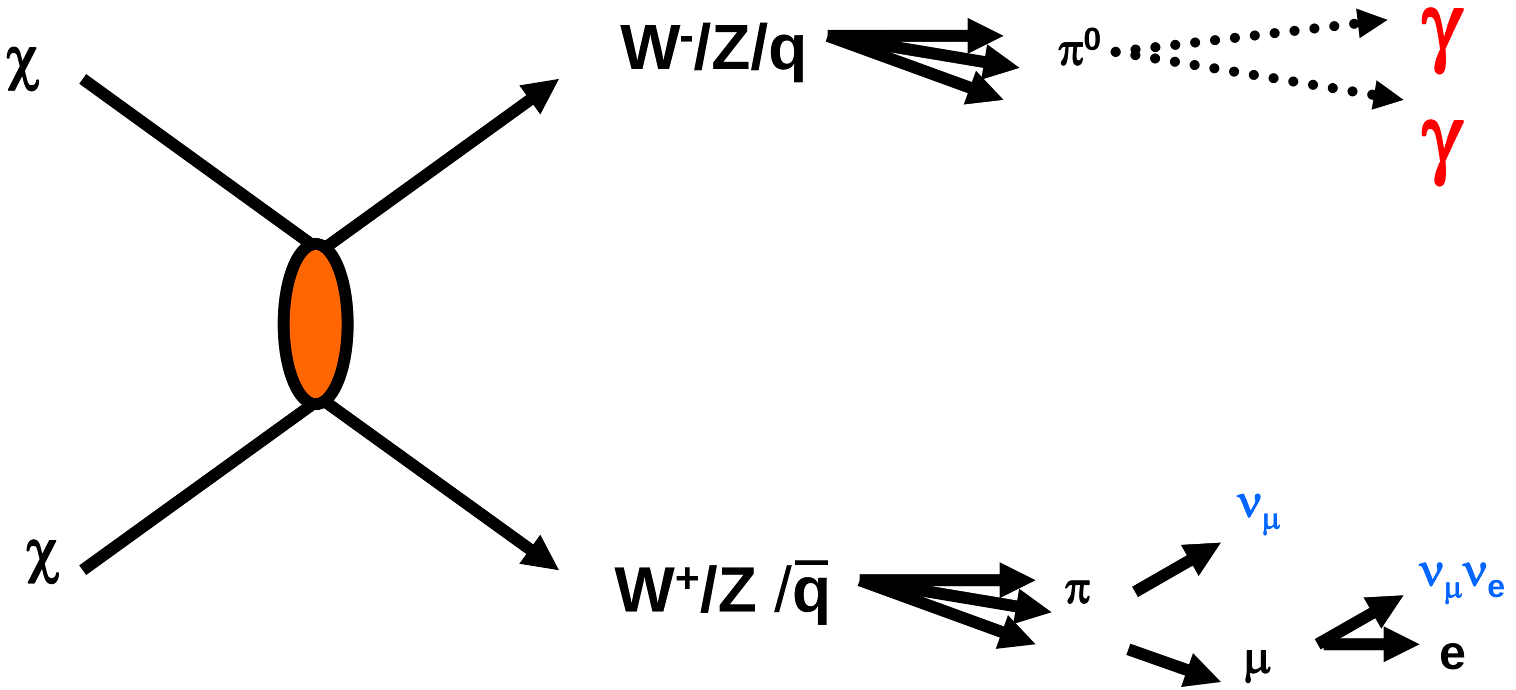
VERY PRELIMINARY

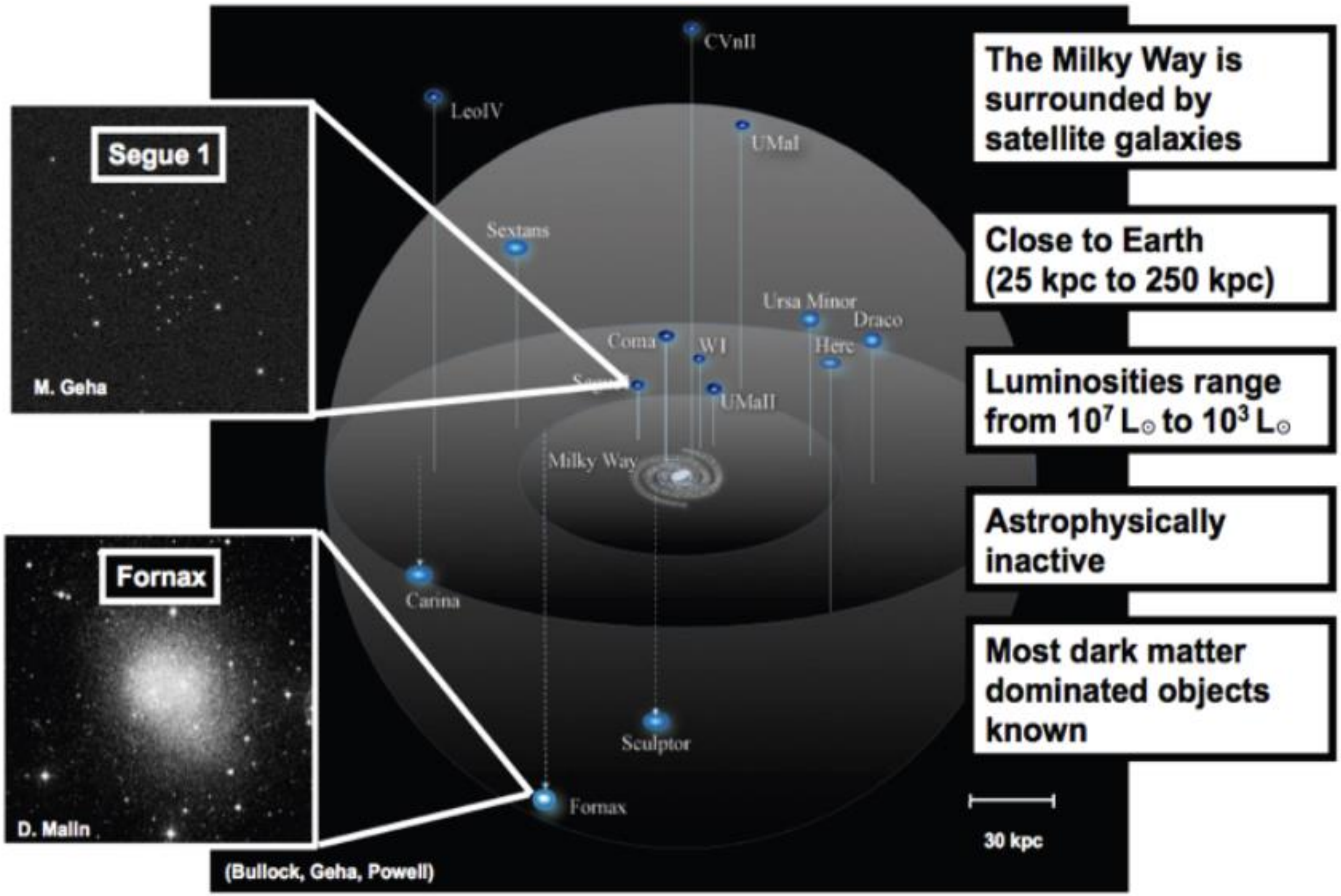
K. S. Cranmer, J. C. in preparation

Example from my own research.



Indirect detection of dark matter




CVnII
LeoIV
UMaI
Segue 1
M. Geha
Sextans
Ursa Minor
Draco
Coma
WI
Herc
UMaII
Milky Way
Fornax
D. Malin
Carina
Sculptor
Fornax
30 kpc
(Bullock, Geha, Powell)
The Milky Way is surrounded by satellite galaxies
Close to Earth (25 kpc to 250 kpc)
Luminosities range from $10^7 L_\odot$ to $10^3 L_\odot$
Astrophysically inactive
Most dark matter dominated objects known

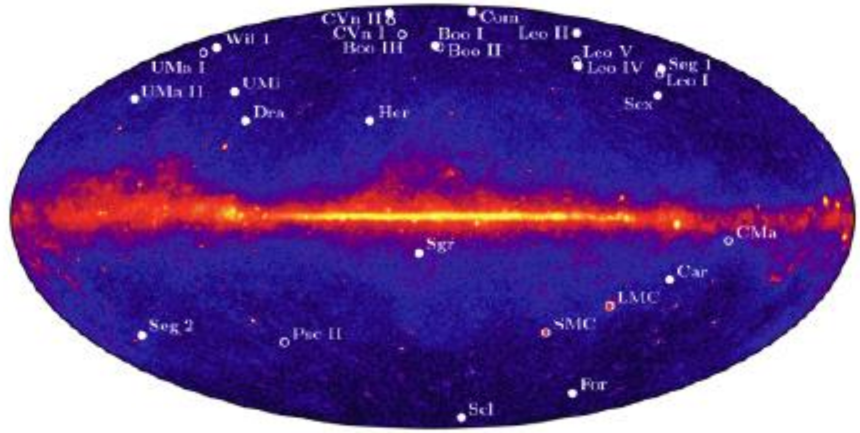
← ~25 known

Data Set & Technique

FOURTH GENERATION

arXiv	irf	time	targets	joint?
1001.4531	P6	11 mo.	10	no
1108.3546	P6	24 mo.	10	yes
1310.0828	P7	48 mo.	15	yes
	P8	60 mo.	15	yes x2!

{1} The Astrophysical Journal, Volume 712, Issue 1, pp. 147-158 (2010)
{2} Physical Review Letters, vol. 107, Issue 24, id. 241302
{3} Phys.Rev. D89 (2014) 4, 042001



EFFECTIVE LIKELIHOOD

(term accounts for uncertainty in J-factor)

$$L_2(\mathcal{D}|\boldsymbol{\mu}, \boldsymbol{\theta_t}) = L_t^{\mathrm{LAT}}(\mathcal{D}_t|\boldsymbol{\mu}, \boldsymbol{\theta_t}) \times \frac{1}{\ln(10)\mathrm{J_t}\sqrt{2\pi}\sigma_\mathrm{t}} e^{-(\log_{10}(\mathrm{J_t}) - \overline{\log_{10}(\mathrm{J_t})})^2/2\sigma_\mathrm{t}^2}$$

$$L_3(\mathcal{D}|\boldsymbol{\mu}, \{\boldsymbol{\theta_t}\}) = \prod_{targets} L_2(\mathcal{D}|\boldsymbol{\mu}, \boldsymbol{\theta_t})$$

← *(combine information from all targets)

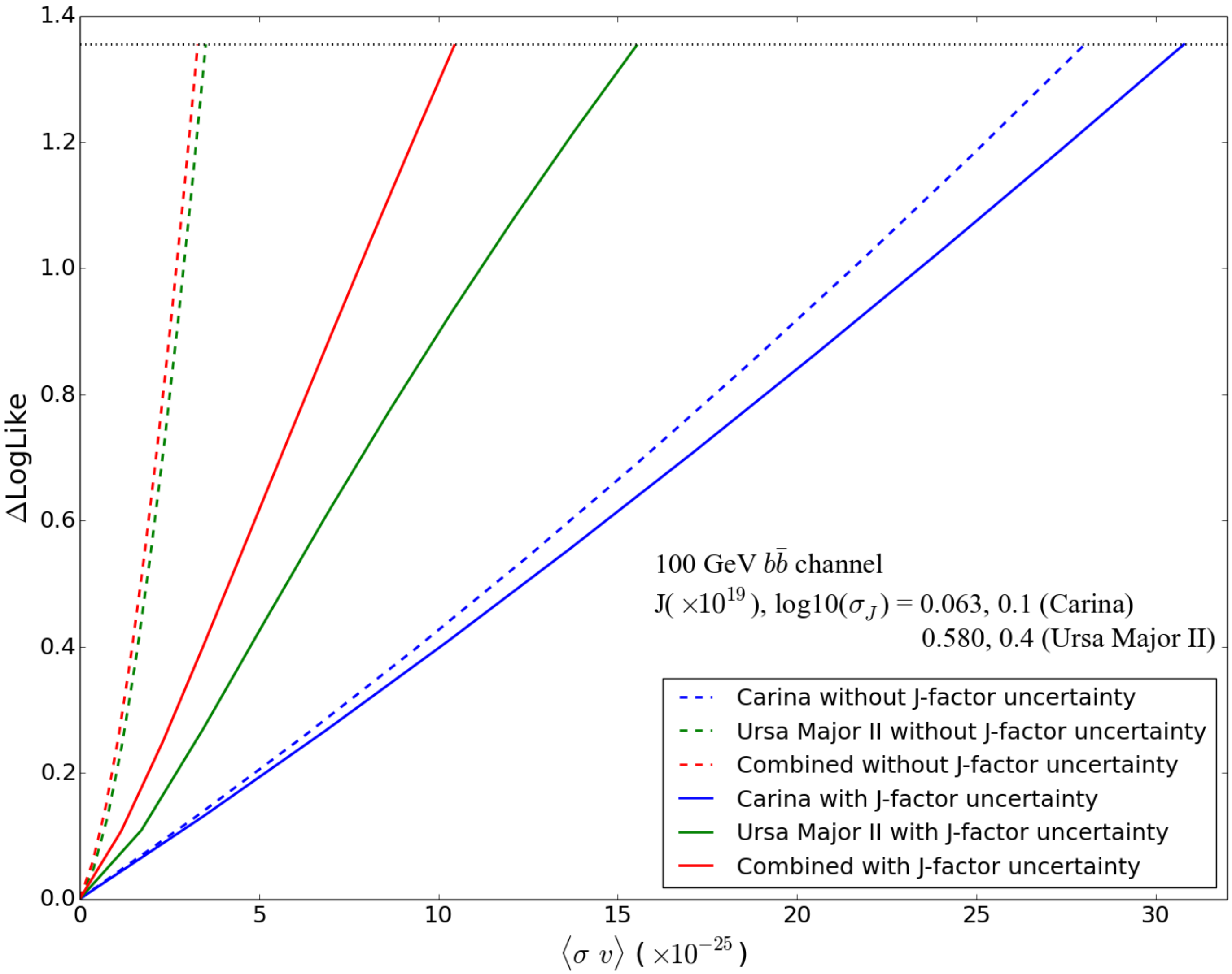
$$L_4(\mathcal{D}|\boldsymbol{\mu}, \{\boldsymbol{\theta_t}\}) = \prod_{classes} L_3(\mathcal{D}_c|\boldsymbol{\mu}, \{\boldsymbol{\theta_t}\})$$

← (combine information from all psf classes)

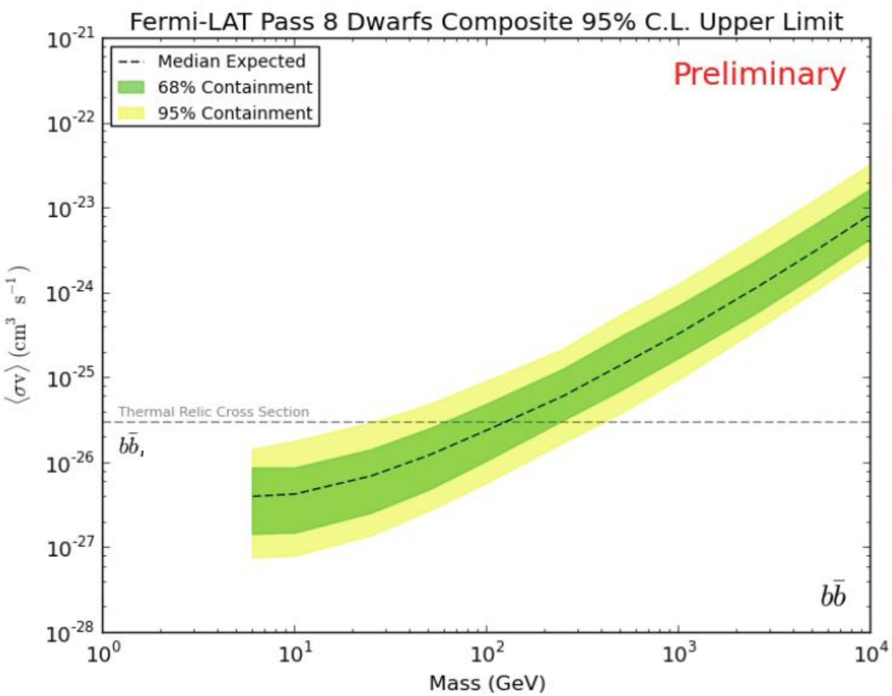
*see talk from Alex on Wednesday, also poster 2.01

Slide from Brandon Anderson (Stockholm)

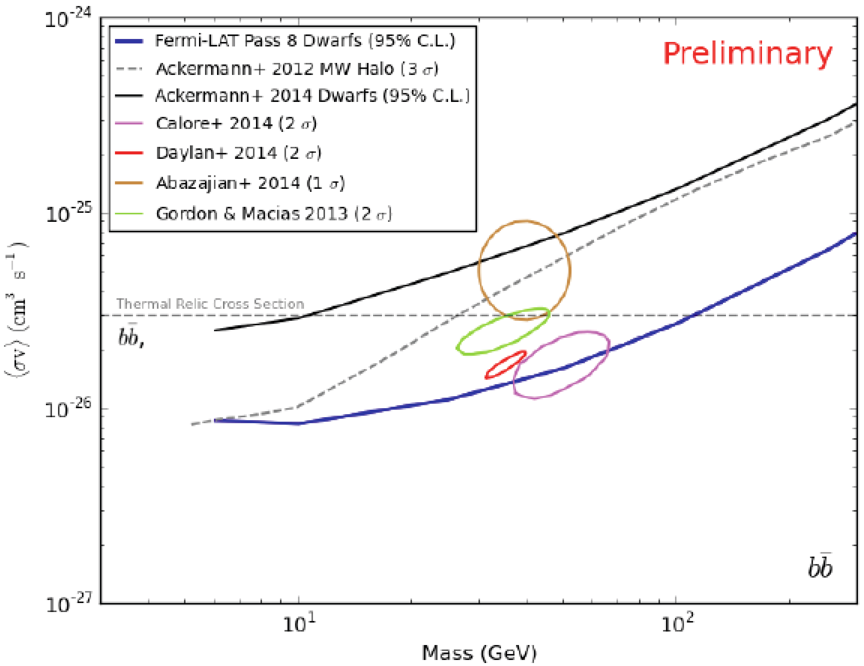
Effect on likelihood



Exciting results.



(bands derived from 300 random blank-sky sets)



Typical LHC profile likelihood (H->$\gamma\gamma$)

$$\mathcal{L}_c(\mu, \boldsymbol{\theta}_c) = e^{-N_c} \prod^{N_c} \mathcal{L}_{c,n}(m_{\gamma\gamma}(n); \mu, \boldsymbol{\theta}_c)$$

$$\mathcal{L}_{c,n}(m_{\gamma\gamma}(n); \mu, \boldsymbol{\theta}_c) = \quad N_{s,c}(\mu, \boldsymbol{\theta}_c^{norm}) f_{s,c}(m_{\gamma\gamma}; \boldsymbol{\theta}_c^{shape})$$

Mass distribution $+ N_{bkg,c} f_{bkg,c}(m_{\gamma\gamma}; \boldsymbol{\theta}_c^{bkg})$,

$$\begin{aligned} N_{s,c}(\mu, \boldsymbol{\theta}_c^{norm}) \; = \; & \mu\,[N_c^{ggH,SM}(\boldsymbol{\theta}_c^{ggH}) + N_c^{VBF,SM}(\boldsymbol{\theta}_c^{VBF}) \\ & + N_c^{WH,SM}(\boldsymbol{\theta}_c^{WH}) + N_c^{ZH,SM}(\boldsymbol{\theta}_c^{ZH}) + N_c^{ttH,SM}(\boldsymbol{\theta}_c^{ttH})] \\ & \cdot K_{BR}(\theta_{BR})\, K_{lumi}(\theta_{lumi})\, K_{eff}(\theta_{eff})\, K_{isol}(\theta_{isol}) \\ & \; K_{pile-up}(\theta_{pile-up}) K_{EScale}(\theta_{EScale}) \\ & \; K_{pile-up,c}(\theta_{pile-up,c}) K_{mat,c}(\theta_{mat}) \\ & + \sigma_{spurious,c}\, \theta_{spurious,c}\ . \end{aligned} \qquad (8.12)$$

Signal normalization

The test statistics of GCVG for intervals.

Cowan Cranmer, Gross, Vitells Eur.Phys.J. C73 (2013) 2501

- Provides common framework for upper limits and discovery, provides asymptotic properties.

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu\,, \\ 0 & \hat{\mu} > \mu\,, \end{cases}$$

- Profile likelihood ratio upper limit

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu) = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} \ge 0\,. \end{cases}$$

- Profile likelihood ratio,TS becomes constant if negative MLE, two-sided.

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & 0 \le \hat{\mu} \le \mu\,, \\ 0 & \hat{\mu} > \mu\,. \end{cases}$$

- Profile likelihood ratio upper limit, provides some CLs type protection

Neyman construction and nuisance parameters.

- There is no standard solution to the task of including nuisance parameters into the Neyman construction
- One way to deal with nuisance parameters is to use a modified PDF, e.g in presence of a Gaussian background uncertainty:

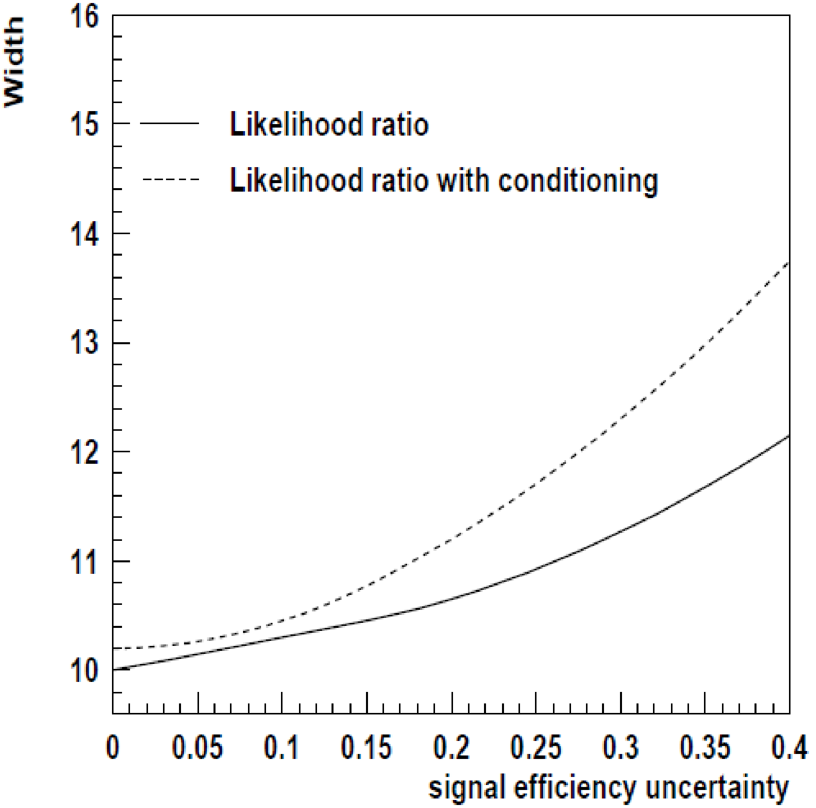
$$P(n|s) = \int_{-\infty}^{\infty} P(n|s,b)G(b|b_{est})db$$

- Then the construction proceeds as usual

see JC Phys.Rev. D67 (2003) 012002, Cousins& Highland (1992)

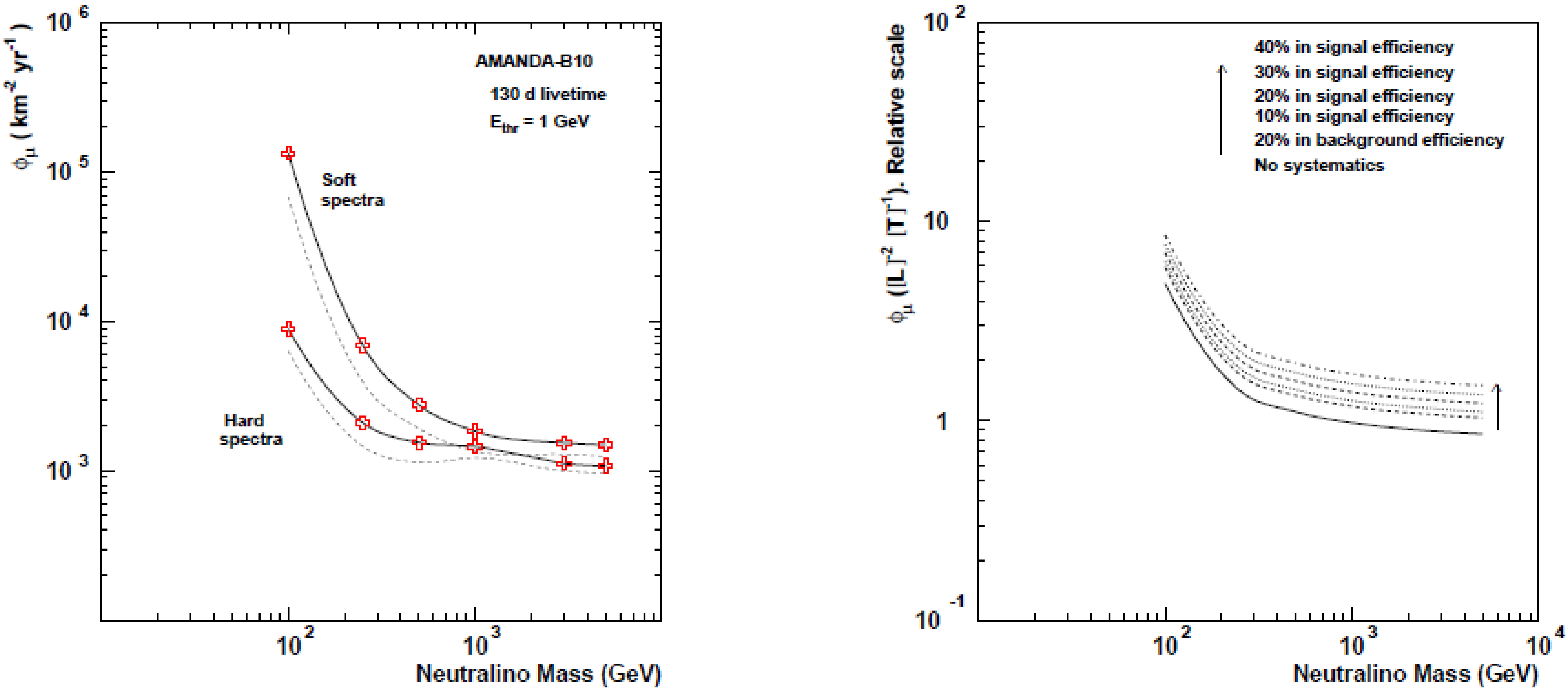
Example intervals for Feldman Cousins ordering and nusiance parameters.

n_0	b	background uncertainty	Likelihood ratio interval	Likelihood ratio interval with conditioning
2	2	0	0: 3.90	0: 4.00
		0.2	0: 3.95	0: 4.10
		0.3	0: 3.95	0: 4.25
		0.4	0: 3.95	0: 4.35



JC + Phys.Rev. D67 (2003) 012002

What would it do to an upper limit on dark matter – indirect detection with neutrinos



JC + Phys.Rev. D67 (2003) 012002

pole++
https://code.google.com/p/polepp/

- Extension of FORTRAN program pole which includes Bayesian treatment in FC ordering Neyman construction
  - treats $P(n|\varepsilon s + b)$

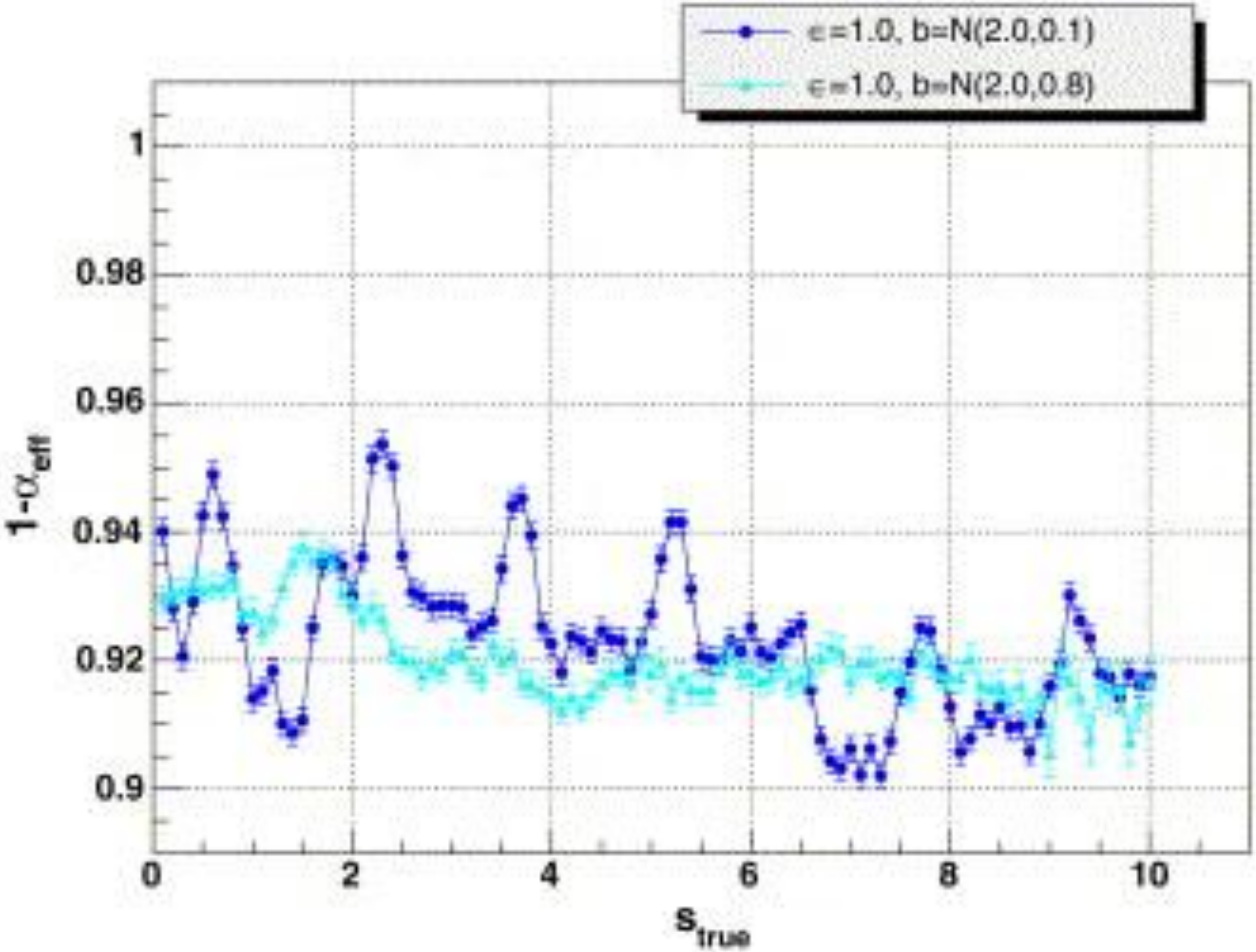
- Consists of C++ classes:
  - ***Pole*** calculate likits
  - ***Coverage*** coverage studies
  - ***Combine*** combine experiments

- Nuisance parameters
  - supports flat, log-normal and Gaussian uncertainties in efficiency and background

Neyman construction provides coverage, any modifaction needs to provide the same property.

Would the new intervals still exhibit coverage?

Feldman Cousins with Bayesian treatment of background uncertainties.



Tegenfeldt+JC, Nucl.Instrum.Meth. A539 (2005) 407-413

Feldman Cousins Profile Likelihood Ratio

- Remember likelihood ratio:

$$R = \frac{L(X \mid \theta_o)}{P(X \mid \theta_{best})}$$

- Knowing about the profile likelihood an obvious ansatz is:

$$R = \frac{L(X \mid \theta_o, \eta_{best|\theta_0})}{P(X \mid \theta_{best}, \eta_{best})}$$

Example (taken from G. Feldman talk)

Let x be a Poisson measurement of $\mu + \beta$ and
b be a Poisson measurement of β/r in an ancillary experiment (i.e., r = signal region/control region),

r / n, rb	0, 3	3, 3	6, 3	9, 3
0.0	0.00- 1.08	0.00- 4.42	0.15- 8.47	1.88-12.30
0.5	0.00- 1.11	0.00- 4.42	0.00- 8.47	1.75-12.30
1.0	0.00- 1.49	0.00- 4.73	0.00- 8.70	1.32-12.55
3.0	0.00- 1.57	0.00- 4.85	0.00- 9.36	0.00-13.03

Unified approach and nuisance parameters

- I am not aware of any practial application of this method.
- I assume that the method is largely superseeded by the profile likelihood (Rolke+, Cowan+)
- Should be useful for example in low statistics experiments (e.g. double beta decay)

Nuisance parameters and pseudoexperiments

- Assume Poisson distribution with parameter ***s*** and expected background ***b***. You choose to treat the nuisance parameter ***b*** in Bayesian way:

$$P(n|s) = \int_{-\infty}^{\infty} P(n|s,b)G(b|b_{est})db$$

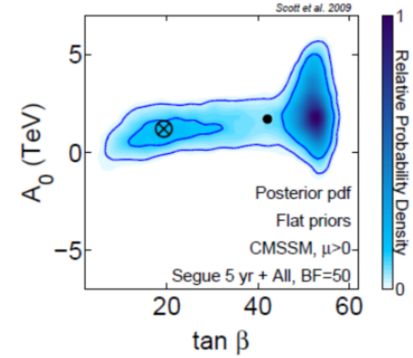
- To test this with pseudoexperiments you have to fix **s** and **b** and draw n and b_{est} i. e. in your analysis the Gaussian will be centered on a different value for each pseudoexperiment.

- Always keep track of what is measured and what is true.

Confidence intervals

COMPLICATED LIKELIHOOD SPACES

Inference on beyond the standard model physics



Parameters of the BSM theory

→ Θ

Nuisance parameters

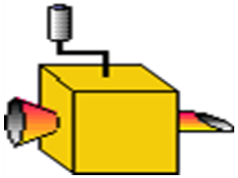
-SM

-Astrophysics

-Instrumental

→ Φ

Calculation of observables



L (Φ,Θ;d)

Scanning and inference

Data and Experimental likelihoods

→ d

Fermi,H.E.S.S.

Pamela/AMS

IceCube

Direct detection

LHC

So … what is the big deal?

Stockholms universitet

- **High dimensional parameter space (~200 parameters easily conceivable)**
- **Parameters are not coupled to observables in a linear way (RGE,astrophysical uncertainties), requires numerical calcuations**
- **Multi-modal likelihood space**
- **Non-trivial experimental likelihoods**

Most likely we are facing a formidable task in parameter estimation in the near future.

Parameters of the theory

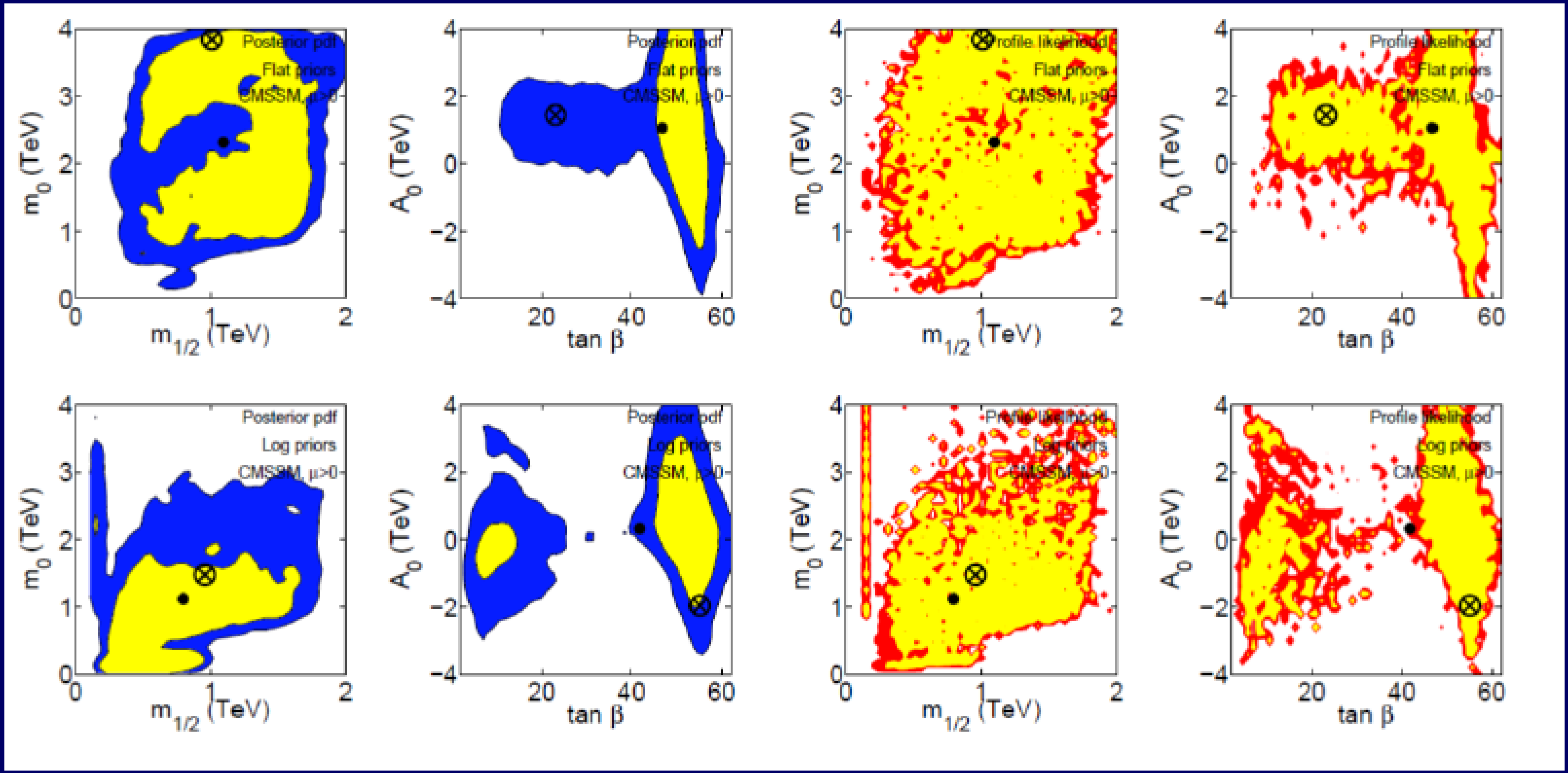
- **Example: Constrained MSSM**
  - Unification at GUT scale, gravity mediated SU-symmetry breaking and electroweak symmetry breaking
  - Gaugino masses: $m_{1/2}$ (btw: $m_{1/2} \sim 2\, m_{\chi}$)
  - Scalar masses: m_0
  - Trilinear couplings: A_0
  - Higgs vacuum expectation ratio: $\tan \beta$
  - Higgs mixing parameter: μ^2

5 free parameters

- **MSSM-7, MSSM-13, pMSSM (29 parameters) …**

Challenges even in simplest Supersymmetric (4 parameters. CMSSM) theory

- **Prior dependence**
  - Flat vs. Log priors give signifcantly different results.
  - Remedied when including more data (LHC for CMSSM, but what happens if we have to go to 100 parameters?)



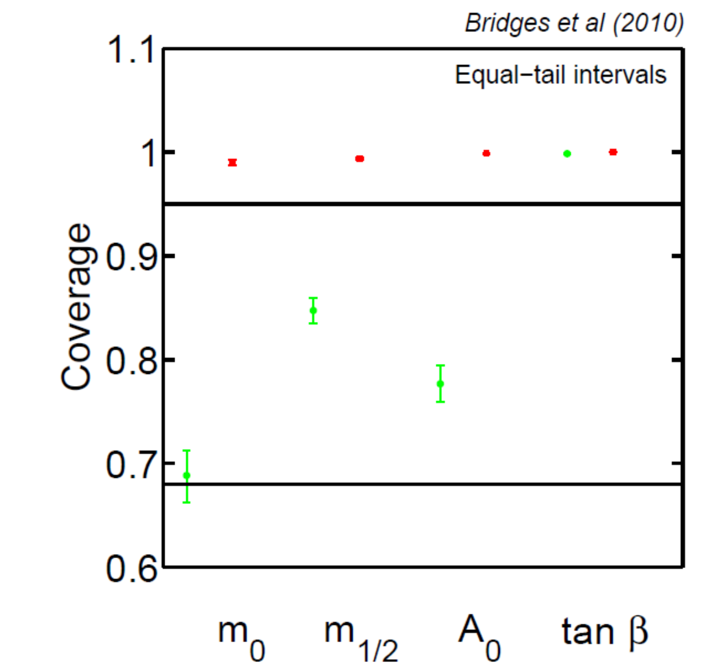
Challenges even in simplest Supersymmetric (4 parameters. CMSSM) theory

- **Frequentist properties**
  - Both over and undercoverage

Bridges+, JHEP 1103(2011) 012, LHC
Akrami+, JCAP 1107 (2011) 002

  - Bad sampling of the likelihood, boundaries on the parameters, flat prior in many dimensions (my guess) ….

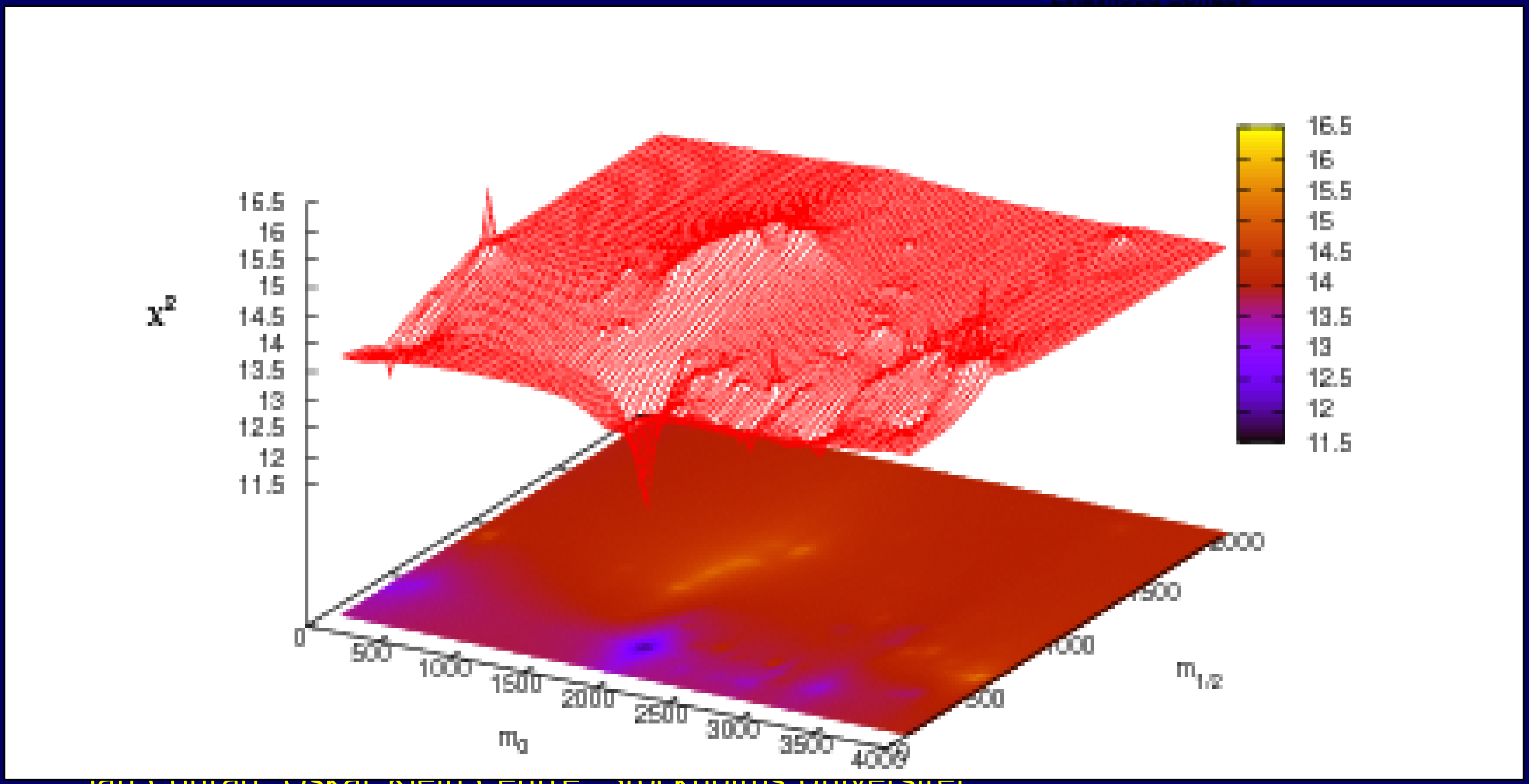
		Benchmark 1				Benchmark 2			
		Conf. int.		Cred. int.		Conf. int.		Cred. int.	
		1σ	2σ	1σ	2σ	1σ	2σ	1σ	2σ
Flat priors	m_0	96	100	98	100	62	93	0	1
	$m_{1/2}$	78	97	45	97	39	89	0	0
	A_0	100	100	91	100	96	100	0	95
	$\tan\beta$	82	100	19	100	99	100	75	100
	$m_{\tilde{\chi}_1^0}$	75	97	18	97	45	93	0	0
	σ_p^{SI}	76	98	53	96	51	87	0	22
Log priors	m_0	96	100	15	94	17	47	0	0
	$m_{1/2}$	67	92	2	30	1	17	0	0
	A_0	99	100	43	91	91	100	0	24
	$\tan\beta$	93	100	16	91	99	100	38	99
	$m_{\tilde{\chi}_1^0}$	57	88	17	37	2	15	23	23
	σ_p^{SI}	71	97	22	65	15	59	0	1



Challenges even in simplest Supersymmetric (4 parameters. CMSSM) theory

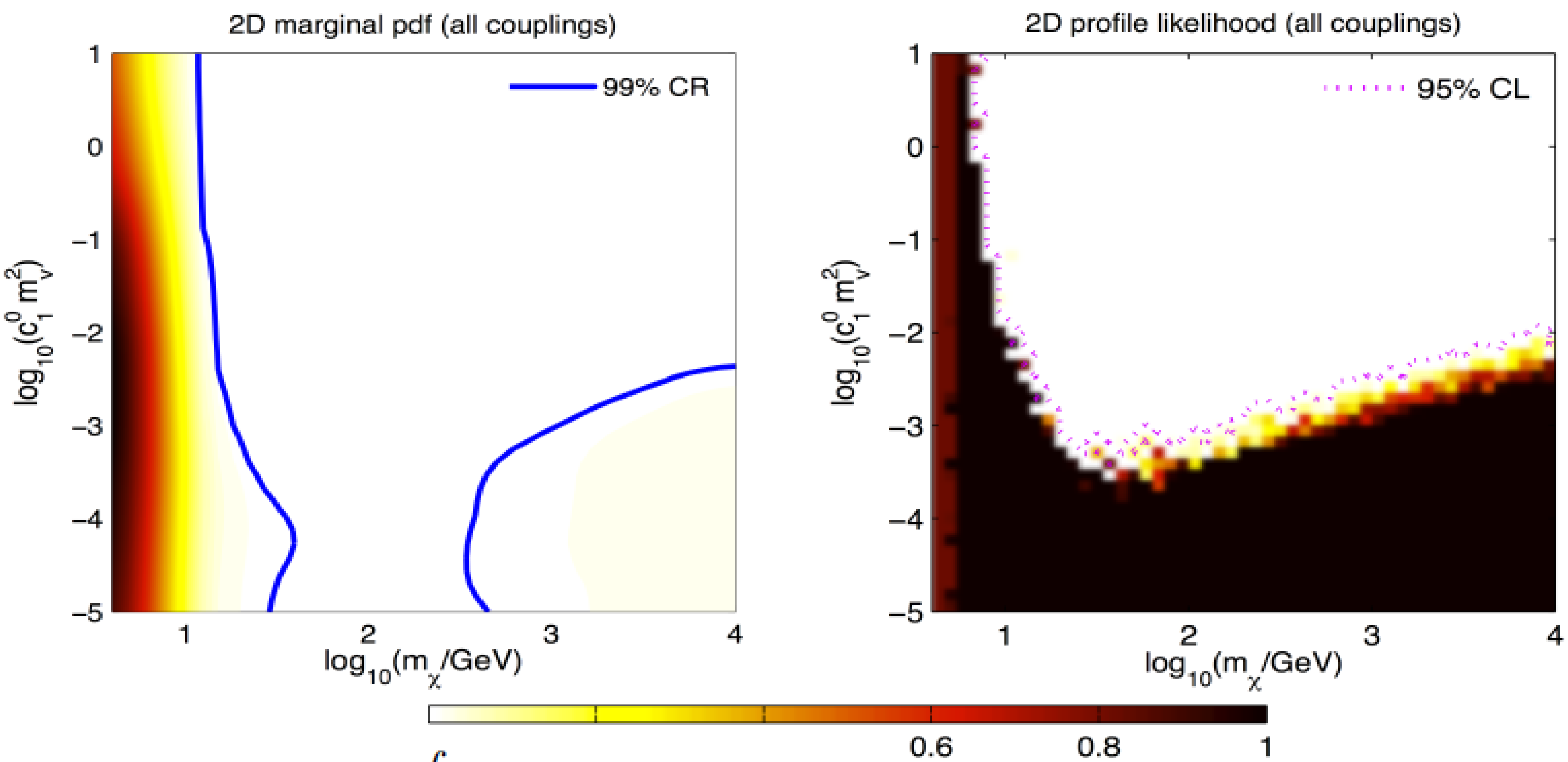
- **Sensitivity to fine-tuning (especially for profile likelihood)**
  - PL picks "false" or "true" likelihood peaks
  - PL much more sensitive to adequate sampling of the likelihood
    - Can machine learning help ??
    - likelihood free inference.

e.g. Feroz+, JHEP 1106:042,2011



Volume effects (flat priors in many dimensions)

- Example: effective field theory approach to direct detection of dark matter (11 couplings, 6 nuisance parameters), flat priors.



$$\mathcal{P}_{\mathrm{marg}}(\theta_1, \theta_2|\mathbf{d}) \propto \int d\theta_3 \ldots d\theta_m \, \mathcal{P}(\Theta|\mathbf{d}) \, .$$

Catena and Gondolo, JCAP 1409 (2014) 09, 045

Summary

- Parameter intervals are used to quantify the statistical accuracy of a measurement. The simplest example is the **standard error**, which is the standard deviation (**square-root of the variance of an estimate),**

- More general, the goal of **interval estimation** is to estimate intervals that **contain the true value of a parameter with given probability.** The standard error for the Gaussian case will turn out to give the interval for which this probability is 68.3 %. This will usually be a two sided interval, which we choose to report for a point measurement.

Summary

- Upper limits are a special case of confidence intervals.

Bayesian: *the degree of belief that the signal is larger than the U.L. is small.*

Frequentist: *if the signal is larger than U.L., the probability for the experimental outcome is small.*

Credible intervals.

- Credible intervals: intervals are obtained by integration of the posterior distribution.

$$p(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{\int P(X \mid \theta)P(\theta)d\theta}$$

- Non-unique: additional condition: Upper limits, central limits or Highest Posterior Density

- Most useful if scientist wants to avoid unintuitive results

- In many dimensions: beware volume effects

Frequentist intervals

- Frequentist intervals, coverage:

  In a very large number of experiments, each providing a confidence interval [θ_{ll}, θ^{ul}], the fraction of intervals that contain the true value is 1- α, independent of what the true value is.

  The random variable is the interval [θ_{ll}, θ^{ul}],

- As we know the PDF for given parameter, a method to calculate confidence intervals can be reformulated as a method that finds the function Z = f(θ,d) such that the PDF in Z becomes independent of the parameter θ.

- Example: Normal theory: $Z = (X-\mu)/\sigma$. Confidence intervals can then be readily obtained from evaluating (or tabulating) the error function.

Frequentist intervals.

- Otherwise: Neyman Construction.

- Flip-flopping: observer decides after the observation which confidence interval to report → effectively destroying coverage.

- Unified confidence intervals based on the likelihood ratio (Feldman & Cousins 1998). No choice necessary, will provide upper limits or ”measurements” depending on the outcome with coverage correct by construction.

Likelihood intervals.

- 68% confidence intervals:

$$\ln L = \ln L_{\max} - 1/2$$

- Steps can be chosen for required confidence level
- Works for non-Gaussian likelihoods (as long as transformable into a Gaussian)
- Good coverage properties even for relatively small statistics.

CL_s

- Modified frequentist method to obtain upper limits.

- CL_s has been developed to avoid rejection of a signal hypothesis (too good upper limit) caused by a chance fluctuation of the background hypothesis.

- Re-defined p-value for confidence interval:

$$p_s' = \frac{p_{s+b}}{1 - p_b}$$

- causes over-coverage in general.

Nuisance parameters:

- Nuisance parameters: parameters that affect the result but are not of prime interest:
- Generically two ways to ”project” on to the subspace of parameters of interest:

$$\lambda(\theta_k) = \max_{\theta_i, i \neq k} \ln L(\vec{\theta})$$

Frequentist

$$L(\vec{\theta}) \rightarrow L_{eff}(\theta_k) = \int d\theta_{1 \ldots i \neq k} L(\theta_{1 \ldots i \neq k})$$

Bayesian

Nuisance parameters

- Profile likelihood is the method adopted by LHC, laid out and unified in Cowan+, and now predominantly used in particle and astroparticle physics

- has in general excellent coverage properties (see e.g. Rolke+)

Marginalization

- Used both in posterior inference (fully Bayesian)
- Hybrid Bayesian (Highland&Cousins, Conrad+)
- In hybrid bayesian → over-coverage in general
- In fully Bayesian → don't care (except for diagnostics potentially).

Complex likelihoods (e.g. Supersymmetry)

- Bayesian methods have been used as MCMC provides sampling of posterior

- Generically, strong prior dependence, bad coverage properties → potentially overcome by better data, but not clear to my mind.

What method to use?

- Most commonly, you will use the maxmimum likelihood estimator in your problem. The simplest (and has been shown recently, also very well performing) method is then to derive the uncertainties from the likelihood function. However, in principle, the properties (coverage) would have to be proven on a case-by case basis,

- The Neyman construction gives coverage **per construction**. However might be computationally much more cumbersome, especially in the case of many parameters. If you are in a low statistics regime, or have other reasons to believe the

- Bayesian intervals might be useful if the experimental outcome of an experiment contradicts intuition. These intervals are simple to calculate. Problem here: prior dependence. This might also be a way around having to use asymptotic properties.

Interval estimation

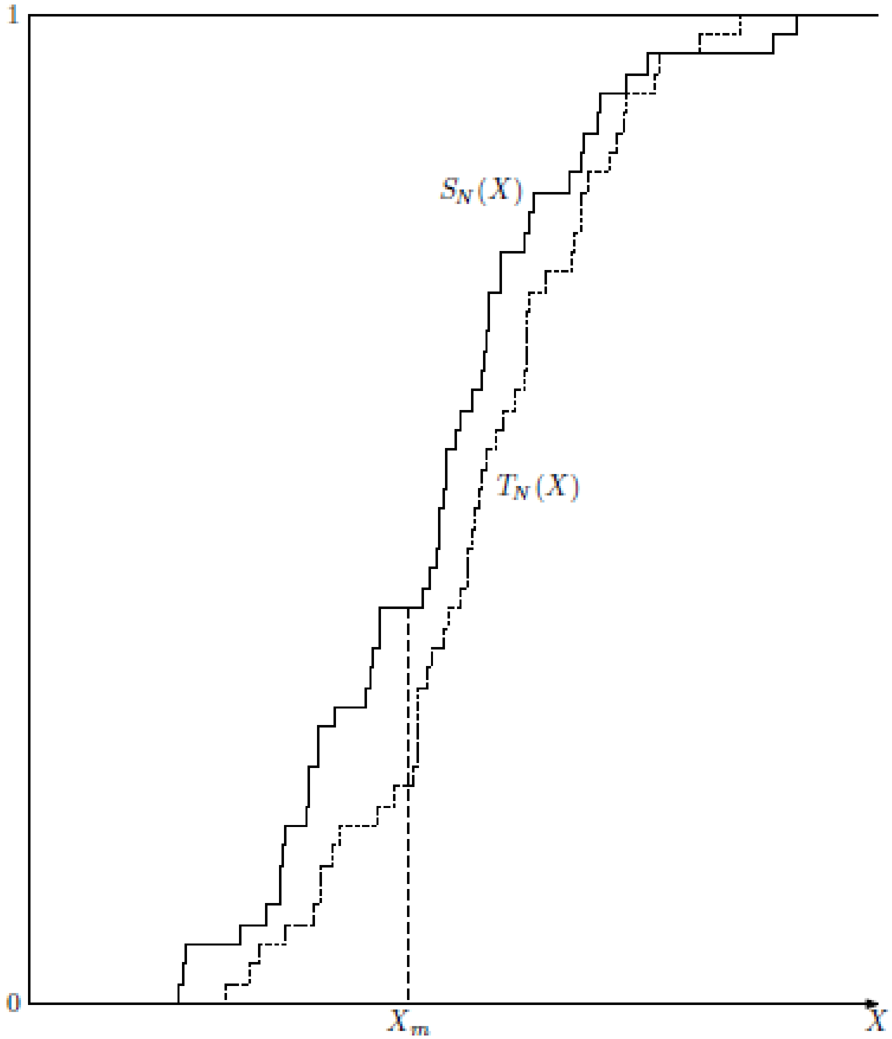
EXERCISE APPENDIX: KOLMOGOROV SMIRNOV

Goodness of fit for unbinned data

- Binning data always leads to loss of information, so in general tests on unbinned data should be superior.
- The most commonly used tests for unbinned data (that are distribution-free) are based on the **order statistics**.
- Given N independent data points $x_1,\ldots\ldots,x_N$ of the random variable X, consider the ordered sample $x_{(1)} \leq x_{(2)} \ldots \leq x_{(N)}$. This is called the order statistics, with distribution function (empirical distribution function):

$$S_N(X) = \begin{cases} 0 & X < X_{(1)} \\ i/N \quad \text{for} & X_{(i)} \leq X < X_{(i+1)}, \quad i = 1, \ldots, \quad N-1 \,. \\ 1 & X_{(N)} \leq X \end{cases}$$

Example



Difference between two EDFs, used with different norms (for different tests) is now used as a test statistics

Kolmogorov-Smirnov test

- Maximum deviation of the EDF from F(X) (expected distribution under H_0).

$$D_N = \max\ |S_N(X) - F(X)| \qquad \text{for all } X$$

$$D_N^{\pm} = \max\ \{\pm[S_N(X) - F(X)]\} \quad \text{for all } X$$

- For this test-statistics a null distribution can be f

$$\lim_{N\to\infty} P(\sqrt{N} D_N > z) = 2\sum_{r=1}^{\infty}(-1)^{r-1}\exp(-2r^2z^2)$$

$$\lim_{N\to\infty} P(\sqrt{N} D_N^{\pm} > z) = \exp(-2z^2)\,.$$