A Measure of the Goodness of Fit in Unbinned Likelihood Fits

Rajendran Raja
Fermilab

Talk to the Run2 Advanced Analysis Group

Format of talk

- State Maximum likelihood formulas for unbinned fits.
- Quote "theorem" as to why Goodness of Fit(GoF) cannot exist for unbinned likelihood fits.
- Derive Bayes' Theorem
- Motivate Likelihood Ratios
- Use PDE's to estimate the pdf of data and introduce the concept of feeding in data into Bayes' theorem.
- Illustrative Example for 1d unbinned fits.
- Empirical Measure of GoF
- Determining the a priori likelihood distributions using data.
  - » Rigorously derive the Posterior likelihood function
- Rewriting Bayes' Equations to take into account the fact that a priori distributions depend on the number of events.
- Bootstrap arguments (incidental to general argument)
- Application to binned cases
- Conclusions
  - » GoF exists for unbinned fits
  - » A priori ``function" in Bayes is the value of the `A priori distribution at the true value. This depends on the event sample
  - » Frequentist error formulae obtain
  - » Bayes Theorem lives, but Bayesianism is no longer needed.

Notation

s denotes signal. Can be multi-dimensional.

c denotes configurations and signifies data. Can be multi-dimensional

P(s|c) signifies the conditional probability density in s, given c.

P(c|s) siginifies the conditional probability density in c, given s. pdf's obey normalization condition.

e.g.

$$\int P(c|s)dc = 1$$

Maximum Likelihood method for unbinned data

- Due to R.A. Fisher
- If there are n events in our sample, then the likelihood of observing the n events is given by

$$\mathcal{L} = \prod_{i=1}^{i=n} P(c_i|s)$$

- Fisher finds the maximum likelihood point s*, by minimizing the negative log likelihood.

$$-log_e\mathcal{L} = -\sum_{i=1}^{i=n} log_e(P(c_i|s))$$

- This yields the optimum estimate for the true value of s. The fit is unbinned, since we have evaluated the theoretical curve at each data point. Advantage, do not have to worry about ``bin systematics". However, no goodness of fit criterion exists. Likelihood at maximum likelihood point is NOT such a measure. Unsolved problem in statistics.

“Theorem” About the Non-existence of GoF in unbinned Likelihood Fits.

Find the maximum likelihood point in a variable c’ in which the theory curve is flat!

$$c^{'}(c) = \int_0^c P(c^{''}|s)dc^{''}$$

$$P(c^{'}|s) = P(c|s) \times |\frac{dc}{dc^{'}}| = 1$$

Such a transformation in multi-dimensions is known as a hyper-cube transformation. If we evaluate the likelihood in this frame it comes out to 1 ! I.e. Likelihood is not “Metric” invariant. So no goodness of fit measure can exist!

Derivation of Bayes' theorem

$$dP_{joint} = P_{joint}(s,c)ds\,dc$$

$$dP_{conditional} = P(s|c)ds$$

$$dP_{conditional} = P(c|s)dc$$

$$dP_{joint} = P_{joint}(s,c)ds\,dc = P(c|s)dc \times P(s)ds$$

Derivation of Bayes' Theorem

$$dP_{joint} = P_{joint}(s,c)dsdc = P(s|c)ds \times P(c)dc$$

$$dP_{joint} = P(c|s)dc \times P(s)ds = P(s|c)ds \times P(c)dc$$

$$P(c|s) \times P(s) = P(s|c) \times P(c)$$

$$P(s|c) = \frac{P(c|s) \times P(s)}{P(c)}$$

Derivation of Normalization formulae

$$P(c) \equiv \int P_{joint}(s,c)ds = \int P(c|s) \times P(s)ds$$

$$P(c) = \int P(c|s) \times P(s)ds$$

Derivation of Bayes' Theorem

$$P(s) \equiv \int P_{joint}(s,c)dc = \int P(c|s) \times P(s)dc$$

$$P(s) = \int P(c|s) \times P(s)dc$$

$$or \int P(c|s)dc = 1$$

- This yields the following normalization formulae

Derivation of Bayes' Theorem

$$P(c) = \int P(c|s) \times P(s) ds$$

$$P(s) = \int P(s|c) \times P(c) dc$$

$$\int P(s) ds = 1$$

$$\int P(c) dc = 1$$

$$\int P(c|s) dc = 1$$

$$\int P(s|c) ds = 1$$

- Leading to Bayes' Theorem Familiar expression

$$P(s|c) = \frac{P(c|s) \times P(s)}{\int P(c|s) \times P(s) ds}$$

Observation of Many Configurations

$$P(s|c_1) = \frac{P(c_1|s) \times P(s)}{\int P(c_1|s) \times P(s)ds}$$

- Replace P(s) by $P(s|c_1)$

$$P(s|c_1, c_2) = \frac{P(c_2|s) \times P(s|c_1)}{\int P(c_2|s) \times P(s|c_1)ds}$$

$$P(s|c_1, c_2) = \frac{P(c_2|s)P(c_1|s)P(s)/\int P(c_1|s)P(s)ds}{\int P(c_2|s)P(c_1|s)P(s)ds/\int P(c_1|s)P(s)ds}$$

$$\text{yielding } P(s|c_1, c_2) = \frac{P(c_2|s)P(c_1|s)P(s)}{\int P(c_2|s)P(c_1|s)P(s)ds}$$

$$\text{generalizing, } P(s|c_1, c_2 \ldots c_n) = \frac{P(c_n|s) \ldots P(c_2|s)P(c_1|s)P(s)}{\int P(c_n|s) \ldots P(c_2|s)P(c_1|s)P(s)ds}$$

- Think of $\mathbf{c}_n$ as another object consisting of $c_1, c_2 .. c_n$

$$P(s|\mathbf{c_n}) = \frac{P(\mathbf{c_n}|s) \times P(s)}{\int P(\mathbf{c_n}|s) \times P(s)ds}$$

$$\text{where } P(\mathbf{c_n}|s) = P(c_n|s) \ldots P(c_2|s)P(c_1|s)$$

Likelihood Ratios

Re-write Bayes' Theorem equations as a Likelihood ratio

$$\mathcal{L}_{\mathcal{R}} = \frac{P(s|\mathbf{c_n})}{P(s)} = \frac{P(\mathbf{c_n}|s)}{P(\mathbf{c_n})}$$

Notice that the likelihood ratio is invariant under change of variable c→c' and s→s'. Though

$$P(c'|s) = |\frac{dc}{dc'}|P(c|s)$$

Likelihood ratio is invariant,since the Jacobian cancels in denominator and numerator. Similarly for transformations s→s'. These are extremely important properties, and we henceforth work with Likelihood ratios and not Likelihoods.

Principle of Maximum Likelihood Ratios

The quantity P(c) we interpret as the a priori distribution of data. It does not know anything about the theory. So the individual likelihoods multiply.

$$\mathcal{L}_{\mathcal{R}} = \frac{P(\mathbf{c_n}|s)}{P(\mathbf{c_n})} = \frac{P(c_1|s)}{P(c_1)} \times \frac{P(c_2|s)}{P(c_2)} \cdots \times \frac{P(c_n|s)}{P(c_n)}$$

$$\mathcal{L}_{\mathcal{R}} = \frac{P(s|\mathbf{c_n})}{P(s)} = \frac{P(s|c_1)}{P(s)} \times \frac{P(s|c_2)}{P(s)} \cdots \times \frac{P(s|c_n)}{P(s)}$$

$$-\frac{\partial log_e \mathcal{L}_{\mathcal{R}}}{\partial s} = -\sum_{i=1}^{i=n} \frac{\partial log_e P(c_i|s)}{\partial s} = 0$$

This is the same as the maximum likelihood equations.

Maximizing wrt c

- We can keep s constant and differentiate wrt c

$$-\frac{\partial log_e \mathcal{L}_{\mathcal{R}}}{\partial c_i} = -\frac{\partial log_e P(c_i|s)}{\partial c_i} + \frac{\partial log_e P(c_i)}{\partial c_i} = 0$$

$$\frac{\partial log_e P(c_i|s)}{\partial c_i} = \frac{\partial log_e P(c_i)}{\partial c_i}$$

$$P(c_i|s) = P(c_i)$$

- i.e. Likelihood ratio is maximum (=1), when the theoretical density and the data pdf are equal at all points.

Evaluating the function P(c) and the GoF

- The key point to note is that just as P(s) is the a priori probability of the parameter s, P(c) is the a priori probability density function of the data. In order to evaluate the likelihood ratio L_R at the maximum likelihood point, we need to provide it with the pdf of data, given the event configurations $c_1, c_2\ c_3 \ldots c_n$.

- Well known methods exist to do this, these are collectively titled PDE's (Probability Density Estimators). They have recently found application in HEP analyses (Knutson, Holstrom, Miettinen et all).

- In previous uses of Bayes' theorem, to the author's best knowledge, P(c), the data pdf was subsumed into the integral

$$P(c) = \int P(c|s) \times P(s) ds$$

- In binned likelihood fits, one is always comparing a theory histogram (P(c|s)) with a data histogram (P(c)). Two pdf's are involved. The absence of two pdf's in Fisher maximum Likelihood method is the primary reason for absence of GoF. You need two pdf's to make an invariant Likelihood Ratio!

Probability Density Estimators

- If d is the dimension of the vector c, then

$$< c^{\alpha} >= \frac{1}{n}\sum_{i=1}^{i=n} c_i^{\alpha}$$

$$E^{\alpha,\beta} =< c^{\alpha}c^{\beta} > - < c^{\alpha} >< c^{\beta} >$$

- $H=E^{-1}$. h is a smoothing factor

$$\mathcal{G}(c) = \frac{1}{(\sqrt{2\pi}h)^d\sqrt{(det(H))}} exp(\frac{-H^{\alpha\beta}c^{\alpha}c^{\beta}}{2h^2})$$

$$P(c) \approx PDE(c) = \frac{1}{n}\sum_{i=1}^{i=n} \mathcal{G}(c - c_i)$$

$$P(c) = \int P(c)G_{\infty}(c - c_i)dc_i$$

Probability Density Estimators

$$G_\infty(c-c_i) \equiv \lim_{n\to\infty} G(c-c_i) = \delta(c-c_i)$$

- This is assured by making the smoothing factor depend on the number of events.

$$h \approx n^{-1/(d+4)}$$

- PDE's are generalizable to arbitrary dimensions.

Illustrative Example

$$P(c|s) = \frac{1}{s}\exp(-\frac{c}{s})$$

$$\mathcal{G}(c) = \frac{1}{(\sqrt{2\pi}sh)}\exp(-\frac{c^2}{2s^2h^2})$$

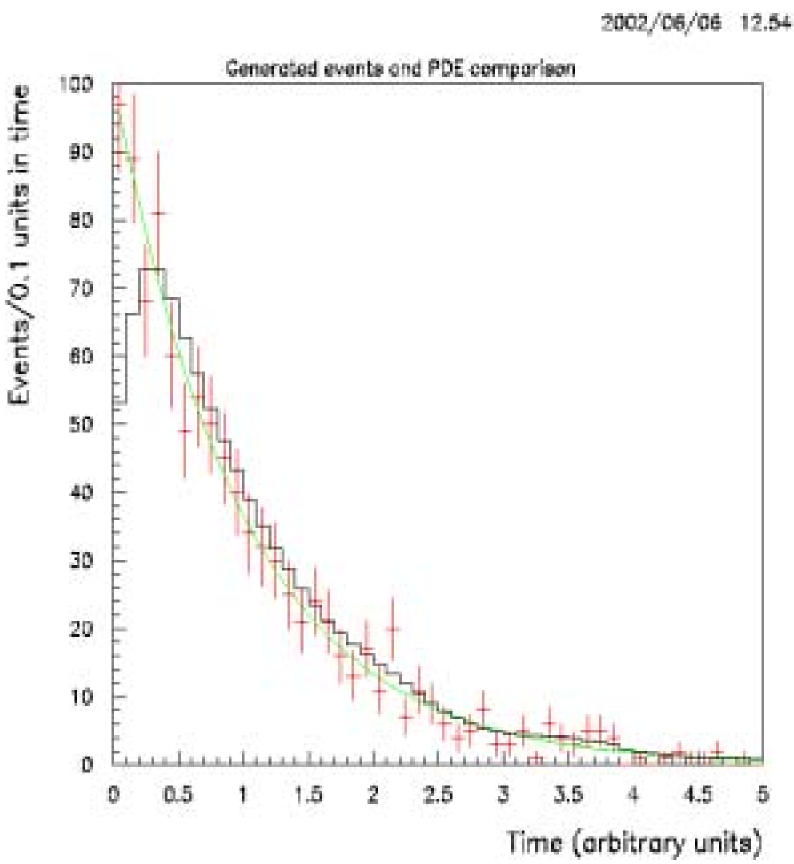


FIG. 1: Figure shows the histogram (with errors) of generated events. Superimposed is the theoretical curve $P(c|s)$ and the PDE estimator (solid) histogram with no errors.

Improve the smoothing factor

- Smoothing factor should be allowed to vary as a function of event density. Estimate event density using constant smoothing factor and then apply the formula

$$h(c) = \left(\frac{n\,PDE(c)}{(t_2 - t_1)}\right)^{-0.6}$$

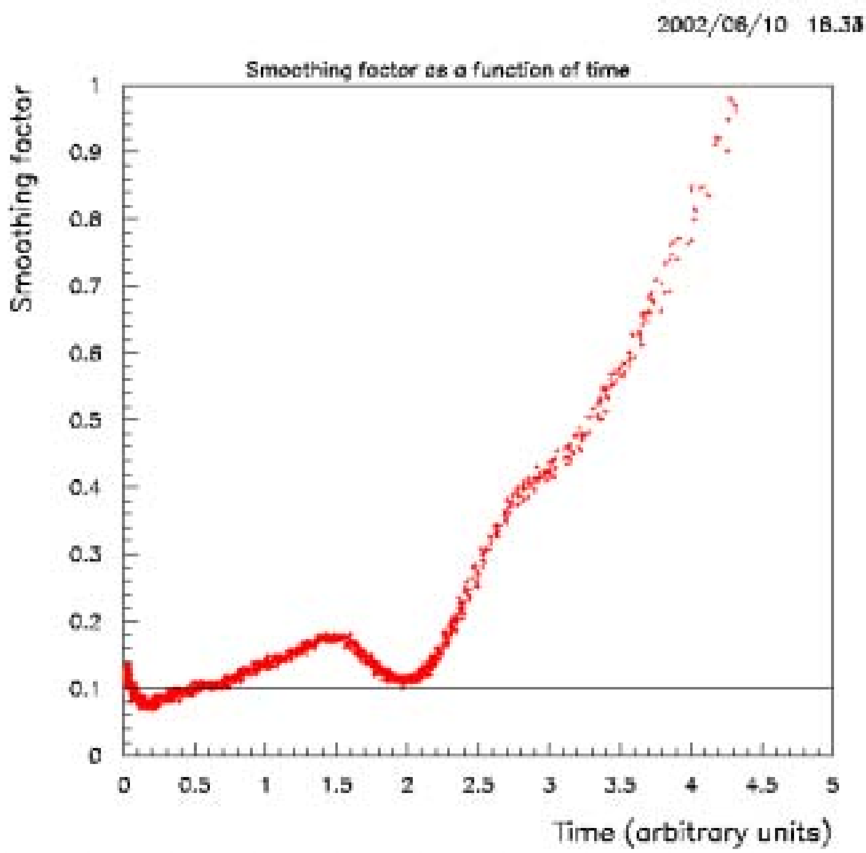
- 

FIG. 3: The variation of h as a function of c for the example shown in Figure 2. The variation of the smoothing parameter is obtained iteratively as explained in the text. The flat curve is a smoothing factor resulting from the formula $h \approx 0.5n^{-1/(d+4)}$.

PDE tracks data

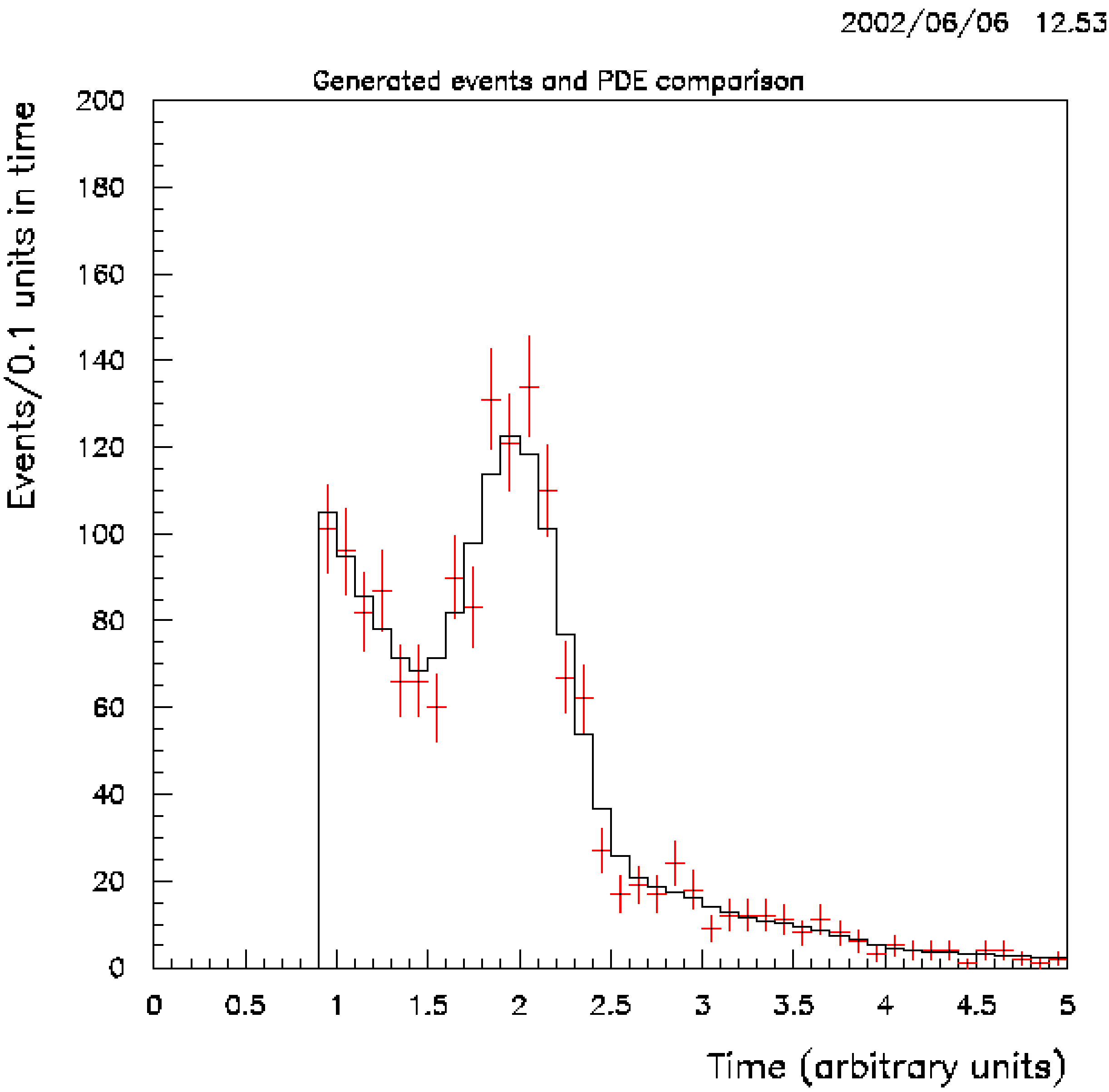


FIG. 2: Figure shows the histogram (with errors) of 1000 events in the fiducial interval $1.0 < c < 5.0$ generated as an exponential with decay constant s=1.0. with a superimposed Gaussian of 500 events centered at c=2.0 and width=0.2. The *PDE* estimator is the (solid) histogram with no errors.

Goodness of fit

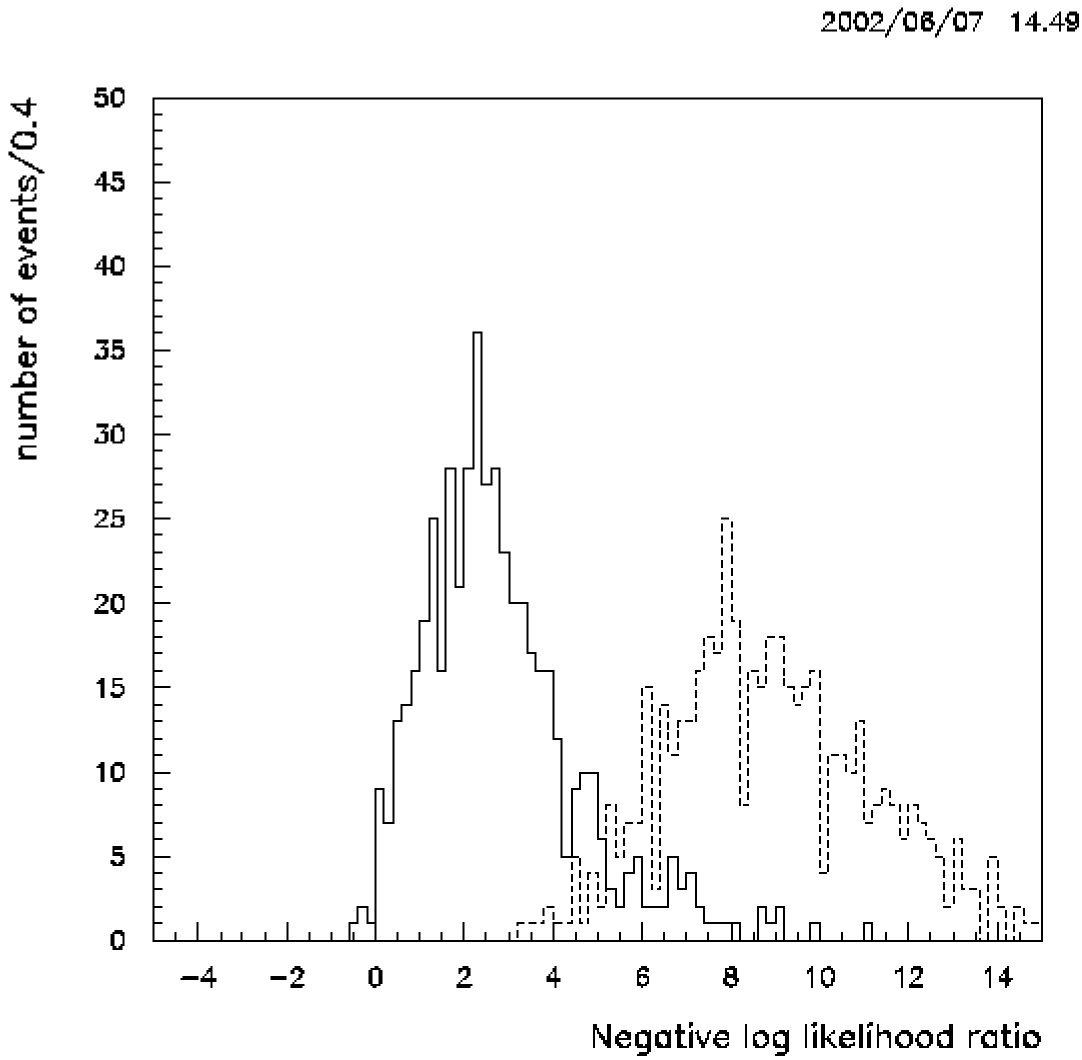
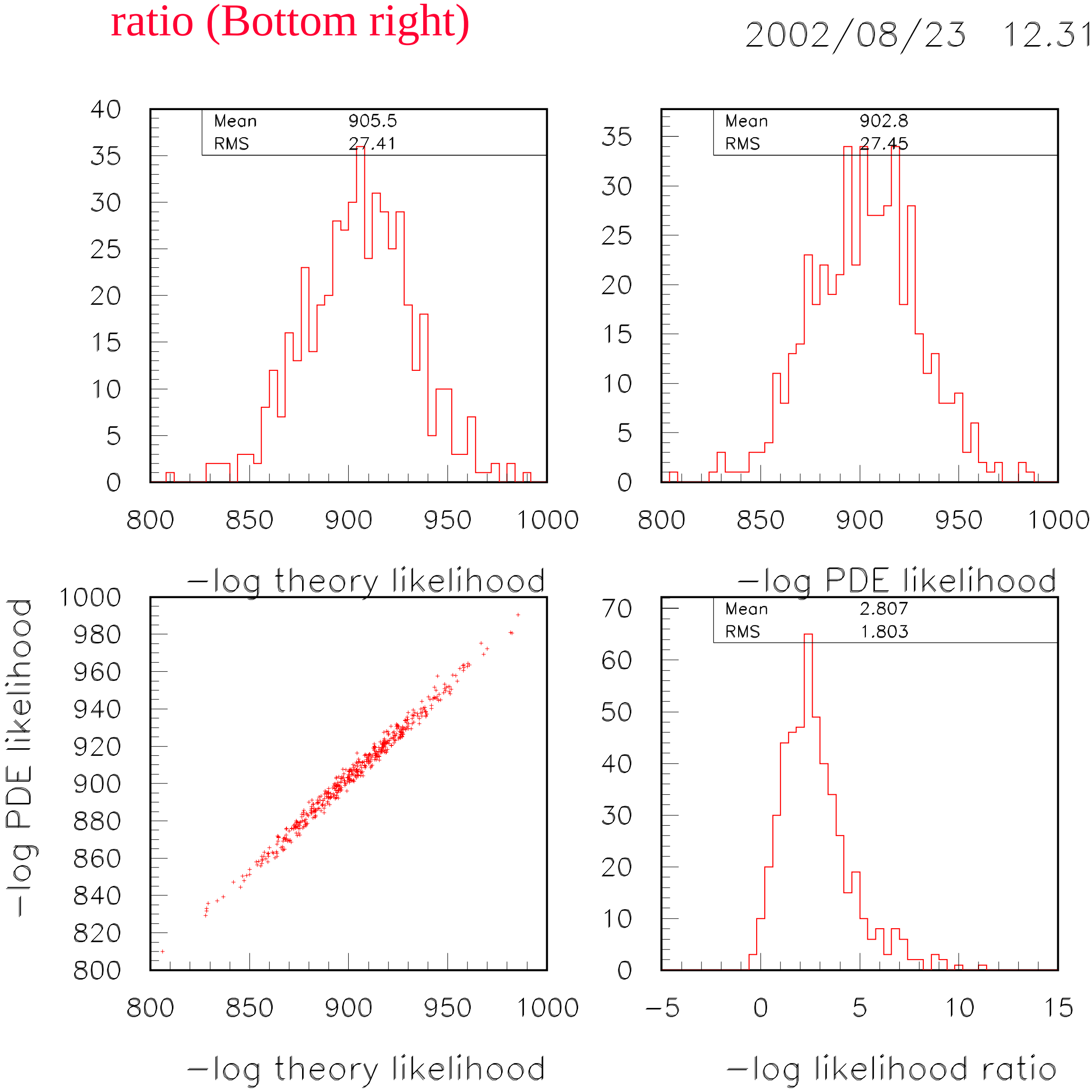


FIG. 4: The solid curve shows the distribution of the negative log likelihood ratio $\mathcal{NLLR}$ at the maximum likelihood point for 500 distributions, using the iterative smoothing function mechanism. The dashed curve shows the corresponding distribution in the case of a constant smoothing function.

Likelihood vs Likelihood Ratio

- Negative log likelihoods, beside not being invariant are broad. (Top Left)
- Negative Log PDE likelihood are similar (Top right)
- The two correlate (Bottom left)
- The difference is the narros GoF, negative log likelihood ratio (Bottom right)



Results of unbinned and binned fitting

TABLE I:

Number of Gaussian events	Unbinned fit $\mathcal{NLLR}$	Unbinned fit $N\sigma$	Binned fit χ^2 39 d.o.f.
500	189.	103	304
250	58.6	31	125
100	11.6	4.9	48
85	8.2	3.0	42
75	6.3	1.9	38
50	2.55	-0.14	30
0	0.44	-1.33	24

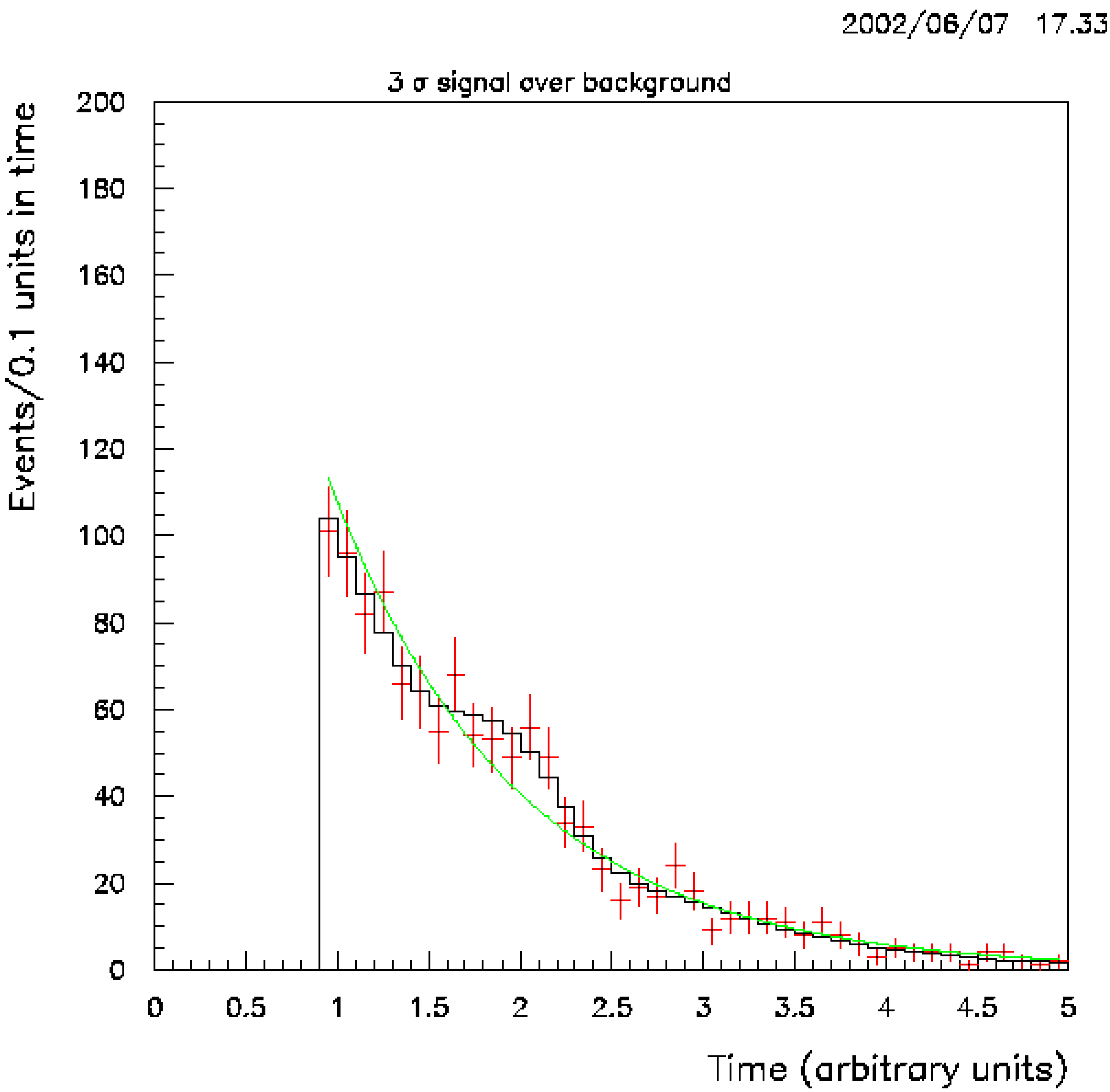


FIG. 5: Figure shows the histogram (with errors) of 1000 events in the fiducial interval $1.0 < c < 5.0$ generated as an exponential with decay constant s=1.0. with a superimposed Gaussian of 85 events centered at c=2.0 and width=0.2. The PDE estimator is the (solid) histogram with no errors. The data are fitted with a goodness of fit that is 3σ away from the average value of $\mathcal{NLLR}$. The continuous curve shows the fit to an exponential.

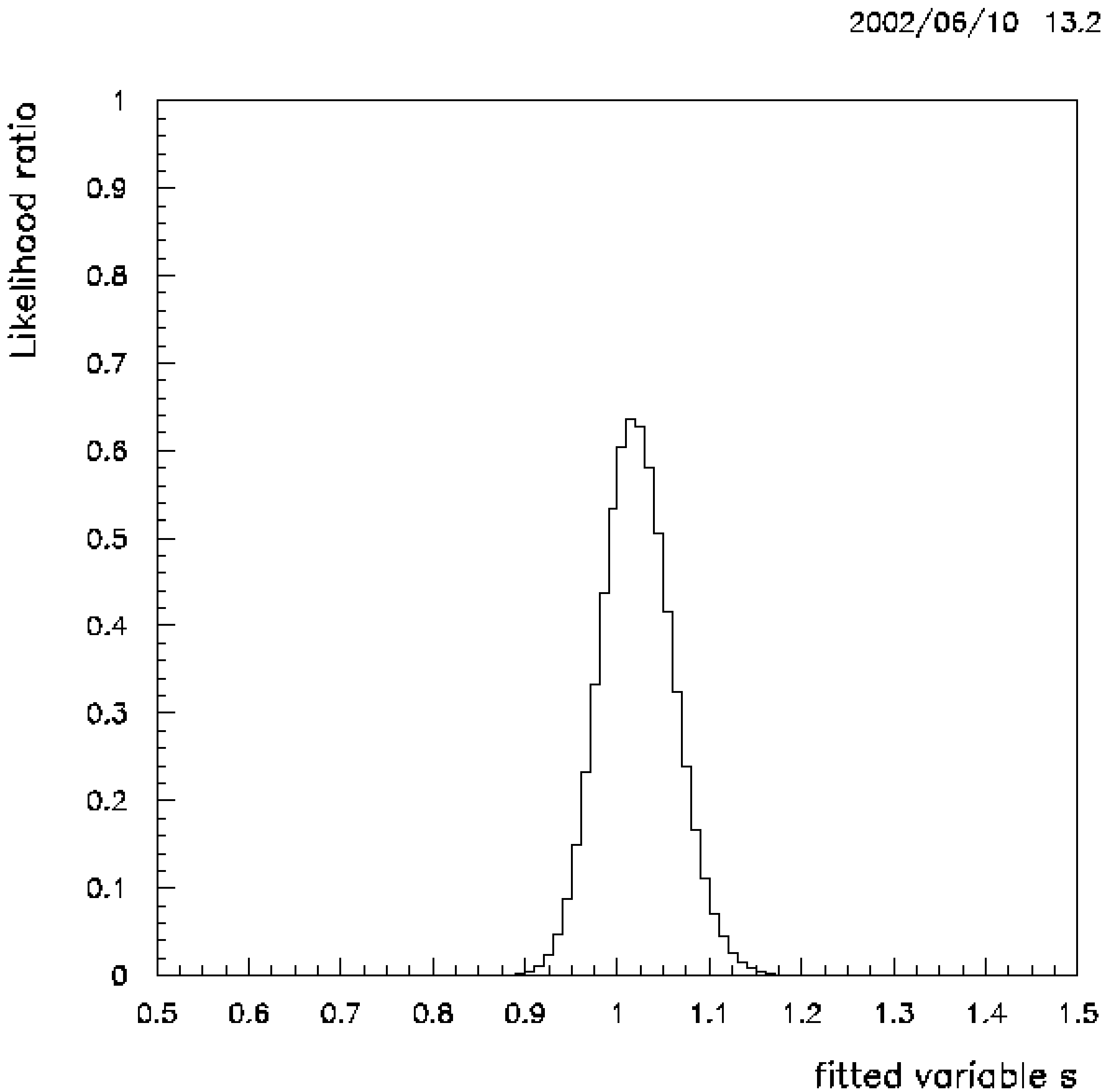


FIG. 6: Figure shows the likelihood ratio $\mathcal{L}_{\mathcal{R}} = \frac{P(c|s)}{P(c)}$ as a function of the fitted parameter s. The maximum likelihood point is at $s = 1.019$.

Determination of the a priori likelihood P(s)

- We want to elucidate the nature of P(s). Bayes' theorem applies for two problems. One with fixed s=s*, and the other where the data contain a mixture of proper lifetimes distributed according to P(s). This case also results in the same Bayes' equations. For the fixed parameter case, which is where Bayesianism comes in, the equation below should be treated as an OR of the possible values and P(s) is the value of P(s) at the true value of s(=s*).

- For the case where data is a mixture of proper lifetimes, the equation below is an AND over all the possible values of s.

$$P(c) = \int P(c|s) \times P(s) ds$$

- Since P(c) is given, P(s) can be inferred.

Determine of the a priori likelihood P(s)

- One can write the following equation

$$P(s|\mathbf{c_n}) = P(s) \times \mathcal{L}_{\mathcal{R}} = P(s) \times \frac{P(\mathbf{c_n}|s)}{P(\mathbf{c_n})}$$

- As $n \rightarrow \infty$, $P(s|\mathbf{c}_n)$ will tend to a delta function. However in this limit, the ratio $P(c_n|s)/P(c_n)$ will tend to unity at s=s*, if the fit is good, since the data pdf and the theory pdf will be identical to each other. The only way out of this is to allow P(s) to depend on n and let $P(s) \rightarrow \delta(s-s^*)$ as $n \rightarrow \infty$. We can see this further in the equation

$$P(c) = \int P(c|s) \times P(s)ds$$

- As $n \rightarrow \infty$, P(c) will have the form P(c|s*), if it fits theory and so, $P(s) \rightarrow \delta(s-s^*)$. So we should write P(s) as $P_n(s)$. We are using posterior information to deduce priors. This leads to a re-write of Bayes' theorem as follows.

Rewriting Bayes' theorem equations

$$\mathcal{L}_{\mathcal{R},n} = \frac{P(s|\mathbf{c_n})}{P_n(s)} = \frac{P(\mathbf{c_n}|s)}{P(\mathbf{c_n})}$$

$$\mathcal{L}_{\mathcal{R},n} = \frac{P(s|\mathbf{c_n})}{P_n(s)} = \frac{P(s|c_1)}{P_1(s)} \times \frac{P(s|c_2)}{P_1(s)} \cdots \times \frac{P(s|c_n)}{P_1(s)}$$

The recursive chain rule now becomes

$$\mathcal{L}_{\mathcal{R},k} = \frac{P(s|\mathbf{c_k})}{P_k(s)} = \prod_{i=1}^{i=k} \frac{P(c_i|s)}{P(c_i)}$$

$$\mathcal{L}_{\mathcal{R},l} = \frac{P(s|\mathbf{c_l})}{P_l(s)} = \prod_{i=1}^{i=l} \frac{P(c_i|s)}{P(c_i)}$$

$$\mathcal{L}_{\mathcal{R},k+l} = \mathcal{L}_{\mathcal{R},k} \times \mathcal{L}_{\mathcal{R},l} = \frac{P(s|\mathbf{c_k})}{P_k(s)} \times \frac{P(s|\mathbf{c_l})}{P_l(s)} = \frac{P(s|\mathbf{c_{k+l}})}{P_{k+l}(s)} = \prod_{i=1}^{i=k+l} \frac{P(c_i|s)}{P(c_i)}$$

i.e Likelihood ratios multiply.

An ansatz for $P_n(s)$

- The expression $P(c) = \int P(c|s) \times P(s)ds$ thought of as the theoretical prediction for P(c) and the PDE can be thought of as the experimental measurement of P©. Then one can write

$$\frac{P^{pred}(\mathbf{c_n})}{P^{exp}(\mathbf{c_n})} = \int \frac{P(\mathbf{c_n}|s)}{P^{PDE}(\mathbf{c_n})} \times P_n(s)ds$$

$$= \int \mathcal{L}_{\mathcal{R},n}(s) \times P_n(s)ds = \int P(s|\mathbf{c_n})ds = 1$$

- $L_{R,n}(s)$ cannot be unity in general, due to statistical fluctuations. If one takes the view that $P_n(s)$ is the value of the distribution of $P_n(s)$ at s=s* , s* being the true value and unknown, then $P_n(s)$ is constant and the solution is

$$P_n(s) = \frac{1}{\int \mathcal{L}_{\mathcal{R},n}(s)ds} \equiv \frac{1}{2\lambda}$$

Determination of a priori likelihoods

- This is the value of the $P_n(s)$ at s=s*. This is all that is needed for use in Bayes' theorem. Then one can substitute this into Bayes' theorem equations and get

$$P(s|\mathbf{c_n}) \times P(\mathbf{c_n}) = P(\mathbf{c_n}|s) \times P_n(s) = P(\mathbf{c_n}|s) \times \frac{1}{2\lambda}$$

$$P(s|\mathbf{c_n}) = \frac{P(\mathbf{c_n}|s)}{P(\mathbf{c_n})} \times \frac{1}{2\lambda} = \frac{\mathcal{L}_{\mathcal{R}}(s)}{\int \mathcal{L}_{\mathcal{R}}(s)ds} = \frac{P(\mathbf{c_n}|s)}{\int P(\mathbf{c_n}|s)ds}$$

- $P(s|\mathbf{c}_n)$ is the a posteriori likelihood for s, and is independent of P(c) data!. We get the ``Frequentisit expression".
- Notice it is possible to derive the expression only

$$P_n(s) = \frac{1}{\int \mathcal{L}_{\mathcal{R},n}(s)ds} \equiv \frac{1}{2\lambda}$$

because of our use of the PDE. It would be dimensionally impossible to get the expression otherwise.

Ansatz for the a priori likelihood distribution

- What follows is not rigorously needed for Bayes' equations. We include it to get more physical insight into what is going on.
- We can then ask, what can one say at this stage about the distribution $P_n(s)$? All we know further is that it integrates to unity. Then the simplest function one can write for $P_n(s)$ is a step function $\theta(s|\mu)$, where μ is the mean value of the step function and is totally unknown. s* has to be in the

$$\text{with } s_1 = \mu - \lambda$$

$$\text{and } s_2 = \mu + \lambda$$

$$P_n(s) \equiv \theta(s|\mu) = 0 \text{ if } s < s_1 \text{ or } s > s_2$$

$$P_n(s) \equiv \theta(s|\mu) = \frac{1}{2\lambda} \text{ if } s_1 \leq s \leq s_2$$

- This step function is narrower than the a posteriori likelihood distribution $P(s|\mathbf{c}_n)$! But we do not know μ, so no problem. (Explain)

Width of a posterior likelihood vs step function.

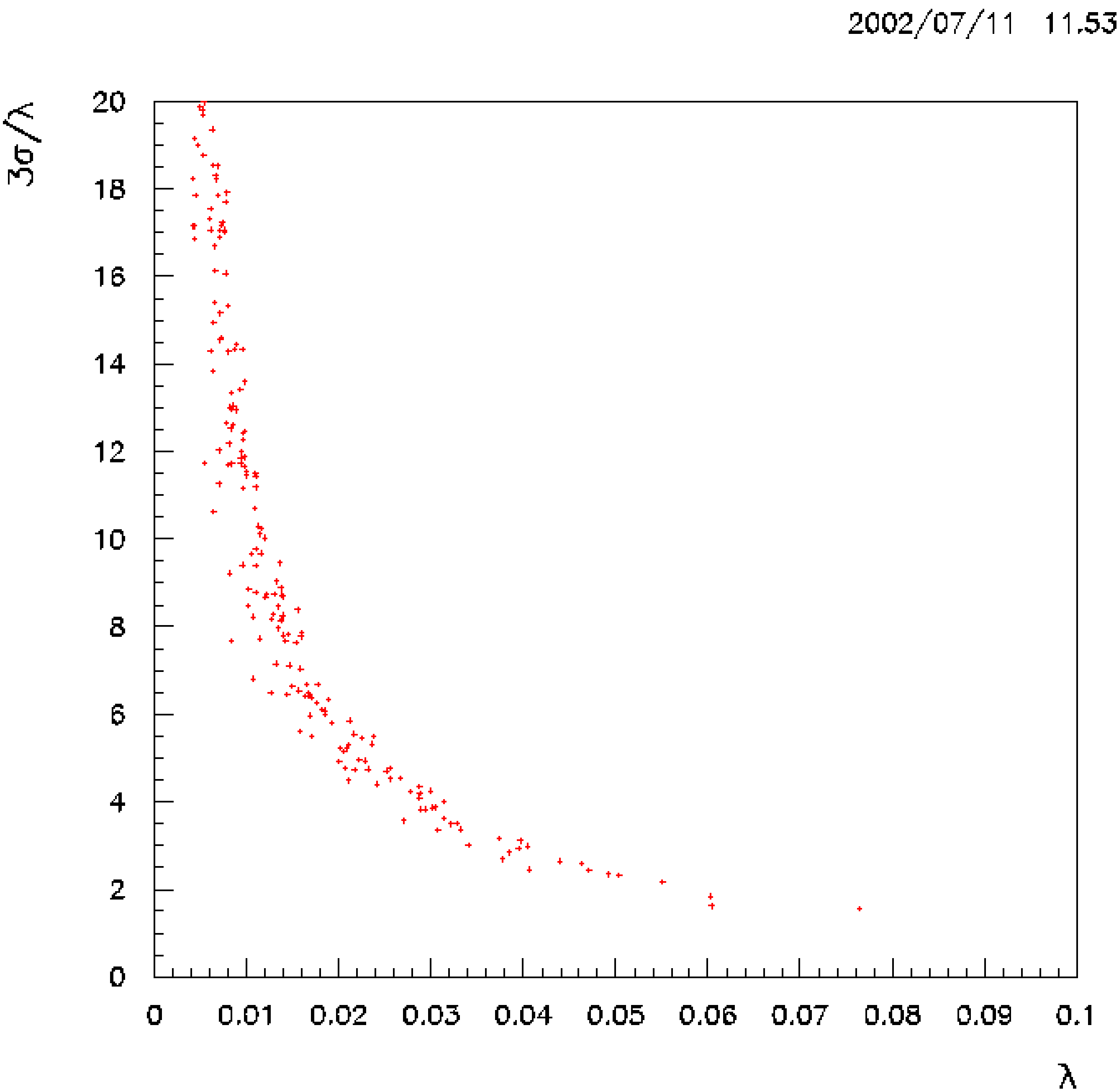


FIG. 7: Figure shows a scatter plot of λ, half the integral under the likelihood curve vs. $3\sigma/\lambda$, where σ is the width of the likelihood distribution for 500 configurations.

Bootstrap of the a posteriori distribution

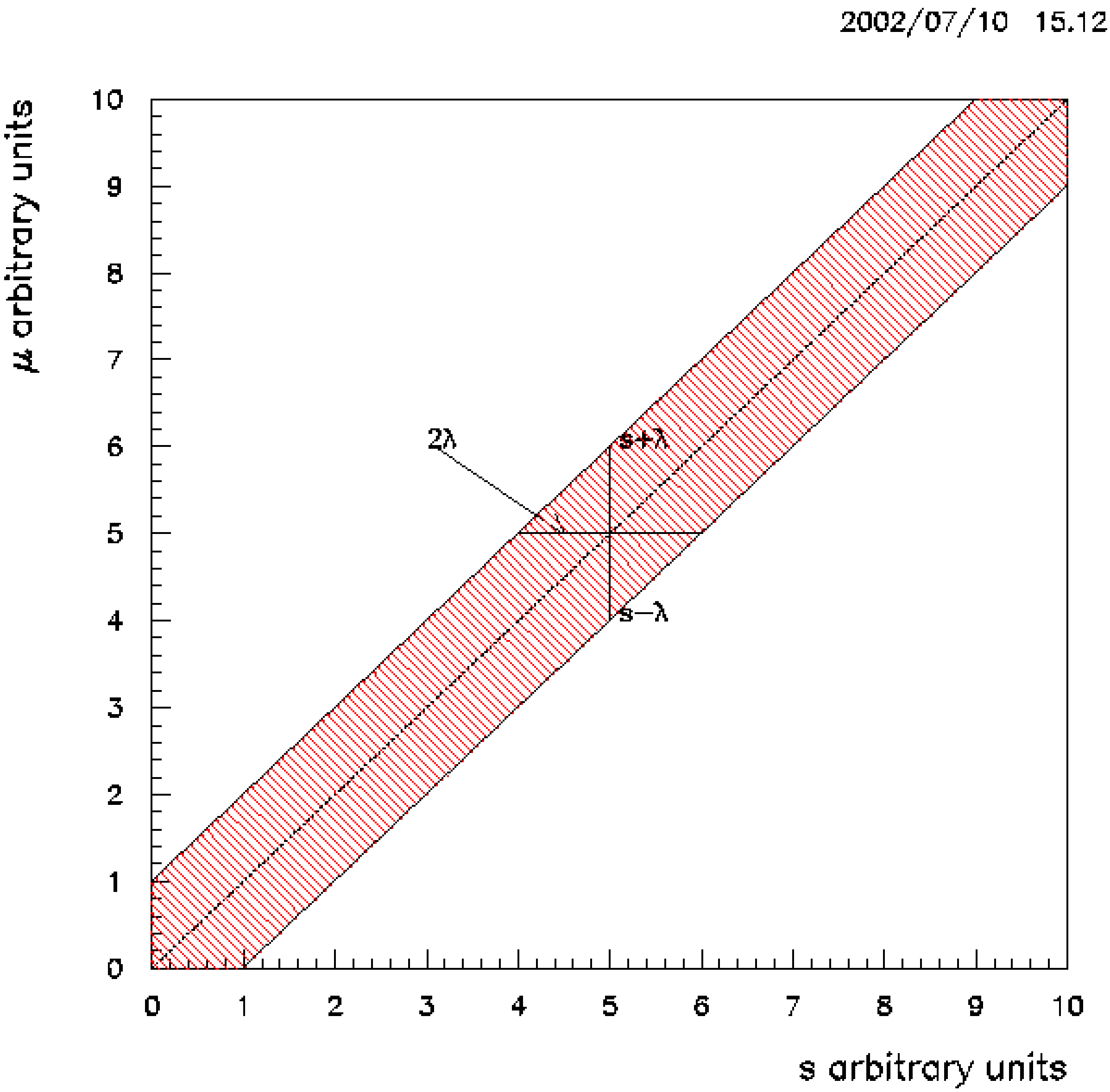


FIG. 8: The abscissa shows the variable s and the ordinate the variable μ, the mean value of the θ function distribution. The hatched region shows the area over which the probability distribution for s is non zero as a function of μ.

Bootstrap of the a posteriori distribution

$$P(\mu) \times \theta(s|\mu) d\mu ds = \frac{P(\mu)}{2\lambda} d\mu ds$$

$$\int_{\mu} P(\mu) \times d\mu \int_{s} \theta(s|\mu) ds = 1$$

$$\int_{s} \frac{1}{2\lambda} ds \int_{s-\lambda}^{s+\lambda} P(\mu) d\mu = 1$$

Bootstrap of the a posteriori distribution

$$\int_s \frac{1}{2\lambda} g(s) ds = 1$$

$$g(s) = \int_{s-\lambda}^{s+\lambda} P(\mu) d\mu$$

$$g(s) \equiv \mathcal{L}_{\mathcal{R}}(s) = \int_{s-\lambda}^{s+\lambda} P(s|\mathbf{c_n}) ds$$

$$\mathcal{L}_{\mathcal{R}}(s) \approx 2\lambda \times P(s|\mathbf{c_n})$$

Bootstrap of the a posteriori distribution

$$P(s|\mathbf{c_n}) \approx \frac{\mathcal{L}_{\mathcal{R}}(s)}{2\lambda} = \frac{\mathcal{L}_{\mathcal{R}}(s)}{\int \mathcal{L}_{\mathcal{R}}(s)ds}$$

- Also we can identify the projection along the s axis of the distribution as $P(s|\mathbf{c}_n) \equiv g(s)/2\lambda$, which completes the bootstrap. The final result– Frequentist formulas are correct. The a priori likelihood and Bayes theorem have ben used to rigorously derive the a posteriori pdf's . Having served their function, the a priori likelihood is no longer needed! (Grin on the face of the Cheshire cat).

$$P(s|\mathbf{c_n}) = \frac{\mathcal{L}_{\mathcal{R},n}(s)}{\int \mathcal{L}_{\mathcal{R},n}(s)ds} = \frac{P(\mathbf{c_n}|s)}{\int P(\mathbf{c_n}|s)ds}$$

Towards an analytic theory of the goodness of fit measure

- Am working on it. Go into Hypercube . PDE's become easier.

$$c^{'}(c) = \int_0^c P(c^{''}|s)dc^{''}$$

$$P(c^{'}|s) = P(c|s) \times |\frac{dc}{dc^{'}}| = 1$$

Conclusions

- A goodness of fit measure for unbinned likelihood fits now exists. It results from demanding invariances of the measure under variable transformations and feeding in data into Bayes' theorem.
- A priori likelihoods can be calculated. They are shown to depend on the amount of data. The a prior likelihood is the value of a distribution at the true value of the parameter. As statistics become large, this tends to infinity and the a priori distribution as well as the a posterior distribtuion tend towards delta functions.
- Implications for Bayesianism
  - » Bayes theorem lives.
  - » The practice of guessing a priori likelihoods, degree of belief etc collectively entitled Bayesianism is no longer needed.
  - » Frequentist formulae are correct. If you accept the conclusions of this paper, the word frequentist (invented to indicate "Not Bayesian" can also be made obsolete.
- We have worked in one dimension only, but the method is valid for multi-dimensions

Conclusions

- The formula for the dimensions of the step function will have to be modified thus

$$P_n(s) = \frac{1}{\int \mathcal{L}_{\mathcal{R}}, n(s) ds} \equiv \frac{1}{(2\lambda)^{\alpha}}$$

- As a by-product, we have introduced a method of iteratively determining the local smoothing factor, that is fast.

Applications to Binned likelihoods

- Consider the case where we have one bin with a large number of events c. The expected value of the number of events is s. Then Gaussian probability yields

$$P(c\,|\,s) = \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{(s-c)^2}{2s}\right) = \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{\chi^2}{2}\right)$$

- With $\chi^2 = \dfrac{(s-c)^2}{s}$

- How many think χ^2 is the negative-log of a Gaussian Likelihood? What about the log of the denominator?
- Work with likelihood ratios and the PDE for the data in that bin!

$$L_R = \frac{\dfrac{1}{\sqrt{2\pi s}} \exp\left(-\dfrac{\chi^2}{2}\right)}{\text{PDE of data} = \dfrac{1}{\sqrt{2\pi c}}} = \sqrt{\frac{c}{s}} \exp\left(-\frac{\chi^2}{2}\right)$$

Applications to Binned likelihoods

- Over many independent bins, this Likelihood ratio becomes the product over individual bins.

$$L_R = \prod_{i=1}^{nbins} \sqrt{\frac{c_i}{s_i}} \exp\left(-\frac{\chi_i^2}{2}\right)$$

- For a good fit, one can write $c_i = s_i + \varepsilon_i$

- Where ε_i are small. Then

$$\prod \frac{c_i}{s_i} \approx 1 + \sum \frac{\varepsilon_i}{s_i} \approx 1$$

- Leading to

$$L_R = \prod_{i=1}^{nbins} \exp\left(-\frac{\chi_i^2}{2}\right)$$