Calculation of errors using the general theory of goodness of fit

Rajendran Raja
Fermilab
Phystat05, Oxford England

See physics/0509008 for further details

- We introduced a likelihood ratio of the theoretical likelihood to "data likelihood" derived from data as the goodness of fit measure.
- Obtain general theory of gof for both binned and unbinned likelihood fitting.
- We now use Bayes' theorem with the new theory to calculate posterior densities.
- Surprising result– No Bayesian prior needed. Frequentist formula for posterior density of fitted parameter.
- Transformation properties of posterior densities

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) signifies the conditional probability density in c, given s. i.e It defines the theoretical model-theory pdf which obeys the normalization condition.

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

Let $\vec{c}_n$ denote the dataset $c_i\ , i = 1, n$

Then

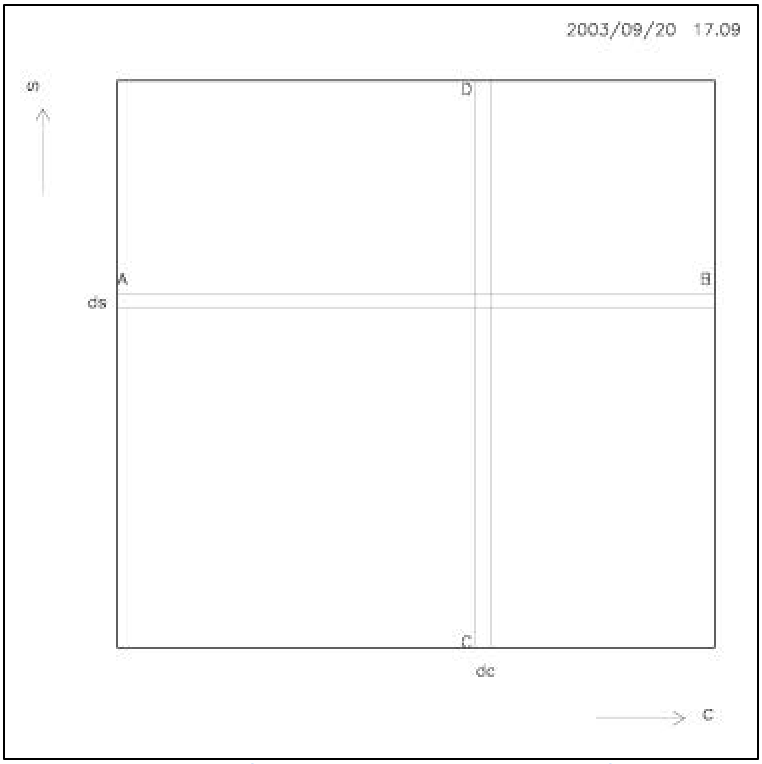
$$L \equiv P(\vec{c}_n\,|\,s) = \prod_{i=1}^{i=n} P(c_i\,|\,s)$$

is the likelihood of observing the dataset $\vec{c}_n$

Likelihood Ratio

$$L_R = \frac{P(\vec{c}_n\,|\,s)}{P^{data}(\vec{c}_n)}$$ provides gof

Bayes' Theorem-Simple derivation



Define a joint probability density P(s,c) such that

$$\iint P(s,c)\,ds\,dc = 1$$

Then define projections P(c), P(s) such that

$$\int P(s,c)\,dc = P(s); \int P(s,c)\,ds = P(c);$$

$$\int P(s)\,ds = 1; \int P(c)\,dc = 1$$

Define conditional probability P(c|s) along line AB $P(c\,|\,s) = \dfrac{P(s,c)}{\int P(s,c)\,dc} = \dfrac{P(s,c)}{P(s)}$

Define conditional probability P(s|c) along line CD $P(s\,|\,c) = \dfrac{P(s,c)}{\int P(s,c)\,ds} = \dfrac{P(s,c)}{P(c)}$

Bayes' Theorem-Simple derivation

- Then

$$P(s,c) = P(s \mid c)P(c) = P(c \mid s)P(s)$$

$$\text{or } \mathrm{P(s \mid c)} = \frac{\mathrm{P(c \mid s)P(s)}}{P(c)} \text{ Bayes' Theorem}$$

$$\text{and } P(c) = \int P(c \mid s)P(s)ds$$

$$P(s) = \int P(s \mid c)P(c)dc$$

- Generalizing to dataset $\vec{c}_n$

$$P(s,\vec{c}_n) = P(s \mid \vec{c}_n)P(\vec{c}_n) = P(\vec{c}_n \mid s)P(s)$$

$$\text{or } \mathrm{P(s \mid \vec{c}_n)} = \frac{\mathrm{P(\vec{c}_n \mid s)P(s)}}{P(\vec{c}_n)}$$

$$\text{and } P(\vec{c}_n) = \int P(\vec{c}_n \mid s)P(s)ds$$

$$P(s) = \int P(s \mid \vec{c}_n)P(\vec{c}_n)d\vec{c}_n$$

Data Likelihood from Data is Incompatible with Bayesian Prior

Define terms:- pdf of a fixed parameter s.

We define $P_n(s)$ as the distribution of the maximum likelihood value s^* when the experiment with data set with n members is repeated N times (ensemble) and $N \to \infty$. We expect $P_n(s)$ to narrow as $n \to \infty$.

The true value of s is the maximum likelihood point of $P_n(s)$. This is an assumption of unbiasedness in the experiment. All distributions of s imply a distribution of s^* and vice versa. The true value is a number and does not have a distribution. The true value is unknown and unknowable with infinite precision. The function $P_n(s)$ is also unknowable.

To calculate errors we assume that given a single dataset $\vec{c}_n$, not only is the maximum likelihood value s^* knowable, but there is information present on the distribution of s^* as well-i.e errors on s^* are computable. We call such a function $P(s|\vec{c}_n)$.

Data Likelihood from Data is Incompatible with Bayesian Prior

- Then we can write

$$P_n(s) = \int P(s\,|\,\vec{c}_n)P^{data}(\vec{c}_n)d\vec{c}_n$$

$$\text{but } P^{data}(\vec{c}_n)d\vec{c}_n = \frac{dN}{N} \text{ the ensemble density}$$

$$\text{Leads to } P_n(s) = \lim N \to \infty \left(\frac{1}{N} \sum_{k=1}^{k=N} P(s\,|\vec{c}_n) \right)$$

- Bayesian way

$$P(s\,|\,\vec{c}_n) = \frac{P(\vec{c}_n\,|\,s)P(s)}{P^{Bayes}(\vec{c}_n)} = \frac{P(\vec{c}_n\,|\,s)P(s)}{\int P(\vec{c}_n\,|\,s)P(s)ds}$$

where $P^{Bayes}(\vec{c}_n) = \int P(\vec{c}_n\,|\,s)P(s)ds$ an uninteresting theoretical constant!

Bayesians then compute

$\int \mathrm{P}(\mathrm{s}\,|\vec{c}_n)P^{Bayes}(\vec{c}_n)d\vec{c}_n = P(s)$ an n independent Bayesian prior!

- Bayes statistics is incompatible with goodness of fit. You can use Bayes theorem to compute $P(s\,|\,\vec{c}_n)$ without the use of Bayseian priors-

Error Bootstrap

- The quantity $P_n(s)$ is the ensemble average of all the posterior densities $P(s\,|\,\vec{c}_n)$. Its maximum likelihood value is the true value s_T.
- We only have measurements from one member of the ensemble namely $\vec{c}_n$

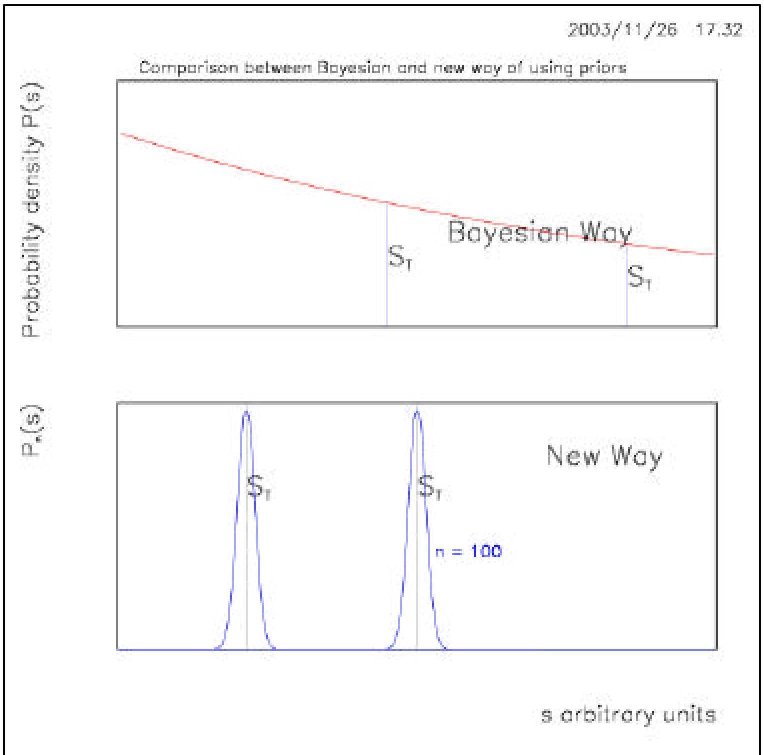
.We want to describe to the system our lack of knowledge of the true value. I.e. We want to say that it is at s_1 OR it is at s_2 OR it is at s_3. At each value of s, we hypothesize that that is the true value.

- The likelihood ratio $L_R(s)$ gives the goodness of fit at that value.
- At the true value $s=s_T$, the joint probablilty P(s,c) is given by

$$P(s,c) = P(s\,|\,\vec{c}_n)P(\vec{c}_n) = P(\vec{c}_n\,|\,s)P_n(s_T)$$

- As you change the value of s, the whole distribution $P_n(s)$ has to move so that the true value is at the new value of s. I.e. $P_n(s_T)$ in Bayes' equation is a constant.

Error Bootstrap



$$\int P(s \mid \vec{c}_n)ds = 1 = \frac{P_n(s_T)}{P(\vec{c}_n)} \int P(\vec{c}_n \mid s)ds$$

$$\frac{P_n(s_T)}{P(\vec{c}_n)} = \frac{1}{\int P(\vec{c}_n \mid s)ds}$$

$$P(s \mid \vec{c}_n) = \frac{P(\vec{c}_n \mid s)}{\int P(\vec{c}_n \mid s)ds}$$

This is the same formula as frequentists use! No Bayesian prior.

Illustrative example

- Measure a mass whose true value s is unknown with an apparatus whose standard error σ is known to be 5 gms. A single data set consists of n=100 measurements c_i,i=1,100. Then

$$P(c\,|\,s) = \frac{1}{\sqrt{2p\,s}}\,e^{-\frac{(c-s)^2}{2s^{\,2}}}$$

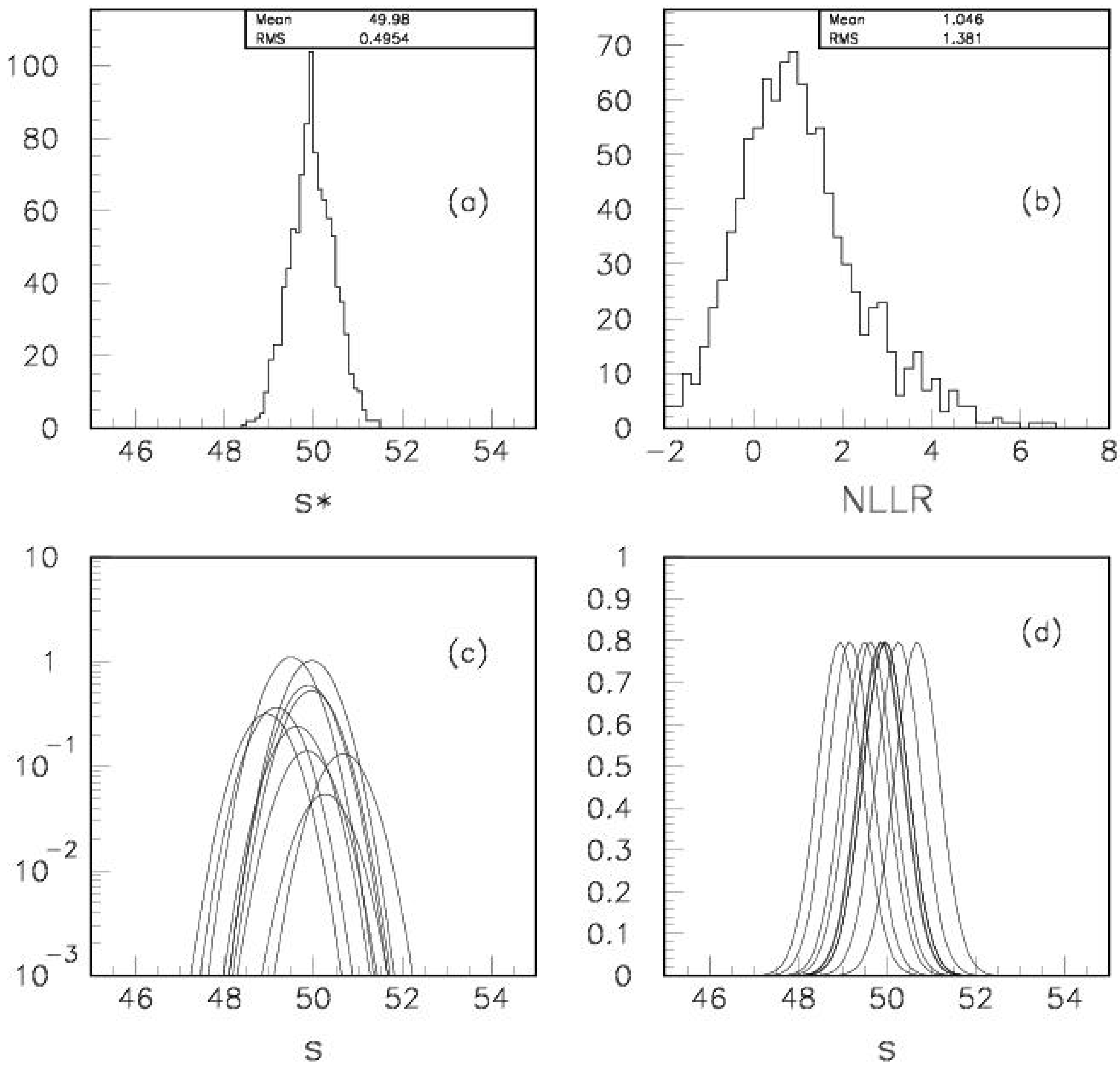
$$P(\vec{c}_n\,|\,s) = \prod_{i=1}^{i=n} P(c_i\,|\,s)$$

- Do goodness of fit using the method of unbinned likelihood fits. Obtain NLLR, likelihoods for each individual fit.
- Determine $P(s\,|\,\vec{c}_n)$ for each fit. Average over ensemble N=1000 fits to obtain a better value of $P_n(s)$. We reuse Bayes' theorem to re-evaluate posteriors, since we know $P_n(s)$ from the ensemble better than from individual measurements.
- One more iteration

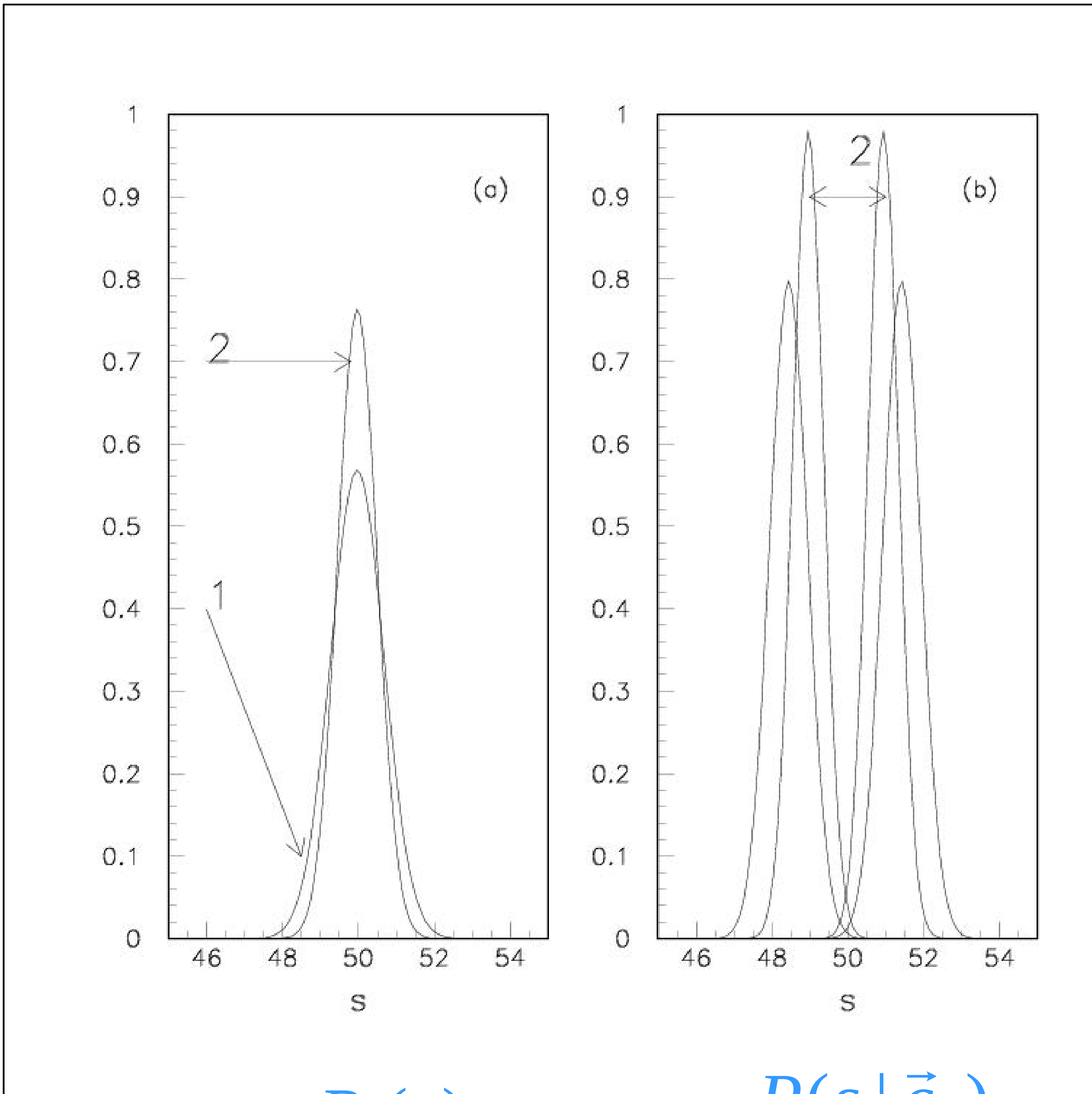
$$P^{iter}(s\,|\,\vec{c}_n) = \frac{P(\vec{c}_n\,|\,s)P_n(s)}{\int P(\vec{c}_n\,|\,s)P_n(s)ds}$$

$$P_n^{iter}(s) = \frac{1}{N}\sum_{k=1}^{k=N} P_k^{iter}(s\,|\,\vec{c}_n)$$

Illustrative example

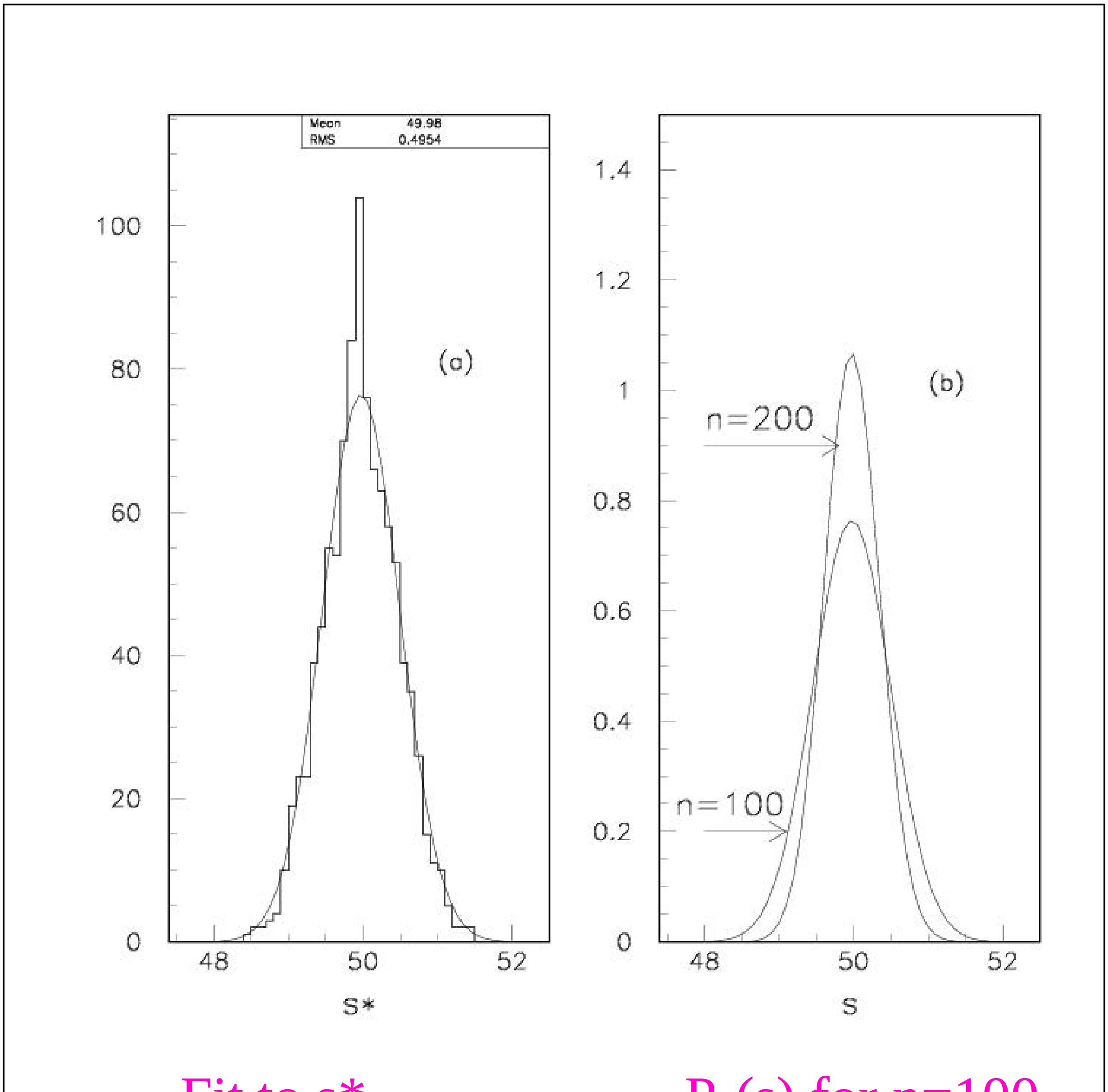


Illustrative example. Iterated functions



$P_n(s)$ $P(s \mid \vec{c}_n)$

Illustrative example



Fit to s* histogram

$P_n(s)$ for n=100, 200