Statistical tools for nuclear experiments, 2

Karsten Riisager

Aarhus University

March 19, 2013

Contents

Main topics

- Preliminaries
- Fitting
- Testing
- Other topics

Subtopics

- Least squares, χ^2 etc
- Maximum likelihood
- Parameter fit
- Finding error bars

χ^2 as standard tool

χ^2 gives:

- value of fit parameter estimate
- error bar of parameter confidence interval
- goodness-of-fit hypothesis test

must treat these three aspects separately...

Ingredients to build a model (3)

Need to choose a "fit method" = an estimator
Desirable properties:

consistent — unbiased — efficient

In principle many possible estimators/fit method.
In practice, two:

Maximum Likelihood (ML)

Least Squares (LS)

- Non-weighted least squares
- χ^2, exp errors
- χ^2, theory errors

J7, **P**36.1

Maximum Likelihood

Definition: probability distribution $P(x, a)$ with parameter(s) a

$$L(x_1, \ldots, x_N; a) = \prod_i P(x_i, a)$$

Fit/estimate: ML = maximize L = maximize $\ln L$
maximum gives parameter(s) that "makes data most probable"

Two nice properties:

- (asymptotically) efficient = reaches MVB (minimum variance bound)
- invariant under parameter transformation

i.e. most efficient method, but in general biased...
(expectation value corresponds to mean, not mode)

B5,**C**6

Parameter invariance in ML

ML is invariant under "smooth" transformations of parameters a
(since, for $a = a(a')$, $0 = \partial L/\partial a \cdot \partial a/\partial a'$)

Chose parameter transformation to make L Gaussian in a' !

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(a'-a'_0)^2/2\sigma^2}$$

Can now read off error bars at 1σ, 2σ etc:

$a' - a'_0$	1σ	2σ	$n\sigma$
factor on L_{max}	$e^{-1/2}$	e^{-2}	$e^{-n^2/2}$
$\Delta(-2\ln L)$	1	4	n^2

NB! Same $\Delta(-2\ln L)$ after transforming back to a.
Can be justified asymptotically! Use the same rules for χ^2 ?

ML $\rightarrow \chi^2$

When probability distributions *are* Gaussian

$$P(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

the log-likelihood turns into

$$\begin{aligned} -2\ln L &= -2\ln\left(\prod_i P(x_i, \mu, \sigma)\right) \\ &= 2N\ln(\sigma\sqrt{2\pi}) + \sum_i \frac{(x_i - \mu)^2}{\sigma^2} \end{aligned}$$

NB! Not a derivation of χ^2 !

Different chi-squares

Chi-square distribution with ν dof defined as $\sum_{i=1}^{\nu} X_i^2$ where X_i are independent and Gaussian distributed (mean 0, $\sigma = 1$).

So we expect $S = \sum_i \frac{(x_i - \mu)^2}{\sigma^2}$ to be χ^2 distributed.

The likelihood ratio, $\lambda = -2\ln(L/L_0)$, has approximately a χ^2 distribution.

Use S as estimator also for non-Gaussian variables.

Estimate by minimizing S. Minimum gives μ parameters.

σ known / estimated from data / estimated from theory / $= 1$ (non-weighted LS)

Example: count numbers

Poisson distributed count numbers n_i in histogram ($i = 1, 2, \ldots, N$), fit-function y_i (with parameters)

Pearson's χ^2, $\chi_P^2 = \sum_i (n_i - y_i)^2 / y_i$

Neyman's χ^2, $\chi_N^2 = \sum_i (n_i - y_i)^2 / n_i$

Poisson likelihood χ^2,
$\chi_\lambda^2 = 2 \sum_i (y_i - n_i + n_i \ln(n_i / y_i))$

S. Baker and R.D. Cousins, NIM 221 (84) 437

What about $n_i = 0$ — or negative y_i ?

Summary for one parameter

Procedure for χ^2 and $-2 \ln L$ is the same !
Fit function depends on one parameter a, need to calculate χ^2 numerically.

Fit by finding minimum of χ^2 as function of a (e.g. MINUIT).
Parameter value at minimum, a_0, is the fit result.

For uncertainties ($n\sigma$ error bars): find points where

$$\chi^2(a_\pm) = \chi^2(a_0) + n^2,$$

the (asymmetric) error bars are then $a_+ - a_0$ and $a_0 - a_-$.
(The points are often estimated by a local parabolic approximation. . .)

Will return to goodness-of-fit later.

Error bars for more parameters

If we only look at each parameter separately: forget about the rest (integrate them out).

What if we wish combined limits for two or more parameters ?
Easy way through: note that (by definition) for one parameter

$P_{Gaussian}$(within 1 sigma) = $P_{\chi^2(\nu=1)}$(less than 1)

This generalizes to more parameters, say k.
Decide on your "confidence level" (1 sigma = 68.27%, 2 sigma = 95.45%).
Look up this probability in tables of chi-square with k dof → Δ.
Find the contours where $\chi^2 = \chi^2_{min} + \Delta$.

B7.2.7, **C**9.7, **J**9.1.2, F. James, Comput.Phys.Commun. 20 (80) 29

Examples from particle physics

CPT test in neutral kaon decay (**P** PDG figure p. 841)
The two axes in the figure are the difference in mass and the width of K^0 and $\bar{K}^0$. One needs combined uncertainties.

B_s mixing phase: CERN Courier, october 2011, vol 51, no 8, p. 8
= Story on LHCb measurement (improved since then) of B_s oscillation.
Two B_s mass eigenstates, the figure shows $\Delta\Gamma$ their difference in width, ϕ_s the mixing phase.
CL 68%, 95%: $\Delta = 2.28, 5.99$

Many other examples in PDG...

Linear least squares

Special case — sufficiently frequent to include — when fit function is linear in parameters, i.e. can write it in matrix form y=Ca.

In general (including covariance): $S = (n^T - y^T)V^{-1}(n - y)$

The solution is:

$$a_0 = V(a)C^T V^{-1} n \text{ where } V(a) = (C^T V^{-1} C)^{-1}$$

Notation: a vector of fit parameters, $V(a)$ its covariance matrix,
n vector of input data, V corresponding covariance matrix,
C matrix in fit function (need not be a square matrix)

B6, **C**7.2

Interjection: fit to polynomials

Fit to a power series will typically give large covariances between fit parameters.
One solution is to change fit functions.

Example: use $x - \bar{x}$ and c rather than x and c.

A linear calibration $E = aCh + b$ gives an error
$s(E)^2 = s(a)^2 Ch^2 + s(b)^2 + 2\mathrm{cov}(a, b)Ch$.

Systematic prescription: R. Barlow, SLUO lectures 7
http://www-group.slac.stanford.edu/sluo/lectures/Stat_Lectures.html

If frequent situation: consider using e.g. Chebyshev polymonials.

Confidence levels

Formal theory on error bars → confidence levels / interval estimation uses Neyman construction.

skip this for the moment

Important conceptual point: "95% confidence interval" means that 19 out of 20 times you construct the interval it will contain the true value — frequentists cannot talk about the probability of the true value being somewhere

Trivial point: no unique way to cover 95% (central/lower/upper interval. . .)
Illustration of symmetric 90% confidence interval in **P** figure 36.4

B7.2, **C**9, **J**9, **P**36.3

Reminder on Bayesian methods

Need prior distribution $\pi(\theta)$, use the same likelihood $L(x|\theta)$ as above. Deduce the posterior $p(\theta|x)$ distribution:

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int L(x|\theta')\pi(\theta')\mathrm{d}\theta'}$$

Can choose to interpret most likely θ as “fit value” and can choose a confidence interval from $p(\theta|x)$ $\leftrightarrow$ same solution as frequentist, if prior distribution is uniform.

(1) Do not have to interpret in this way, p has more information
(2) Interpretation anyway different !
(3) Prior may not be uniform...

C6.13, **J**7.5+9.6