Statistical tools for nuclear experiments, 4

Karsten Riisager

Aarhus University

March 21, 2013

Contents

Main topics

- Preliminaries
- Fitting
- Testing
- Other topics

Subtopics

- Low count numbers
- Combining results
- Robust methods
  - outliers
  - trimming, Winsorization
  - M-estimation
- Miscellaneous

Problems with low count numbers

- Poisson distribution clearly different from Gaussian approximation
- This may give changes in all three areas:
  - The fit results may be biased/wrong
  - The error bars/confidence intervals may become dubious
  - Chi-square not χ^2-distributed (not asymptotic behaviour)
- How to change from upper+lower error bar to upper limit ?

Our three fit methods for histograms

One can derive the result e.g. for a "constant background" μ. The results θ of the three fit methods show that areas from standard χ^2 are essentially always wrong !

Pearson's χ^2, $\chi^2_P = \sum_i (n_i - y_i)^2/y_i$: $\theta \simeq \mu + 1/2$

Neyman's χ^2, $\chi^2_N = \sum_i (n_i - y_i)^2/n_i$: $\theta \simeq \mu - 1$

Poisson likelihood χ^2, $\chi^2_\lambda = 2\sum_i (y_i - n_i + n_i \ln(n_i/y_i))$: $\theta = \mu$

(results are modified for μ below 10 or spectra with few channels).

Y. Jading and K. Riisager, NIM A372 (96) 289

See figures 1 and 3 in T. Hauschild and M. Jentschel, NIM A457 (01) 384

What to do ?

Option 1: rebin spectrum so that more than 80% of channels have more than 10 counts.

Option 2: change to χ^2_λ.
But also bias if total number of counts is less than 100, cf. U. Bergmann and K. Riisager, NIM A489 (02) 444

Option 3: stay with χ^2_N (or χ^2_P), but insist on background term to absorb bias (must allow it to be negative).
^{64}Cu halflife example, $\nu = 2228$, $t_{1/2} = 12.710(4)$ h, χ^2_λ/χ^2_N gives χ^2 2218/2222 and background 67.7(3)/66.7(3).

Option 4: do Monte Carlo simulation.

Construction of the confidence belt

The Neyman interval construction is the standard frequentist way of finding confidence intervals.

Constructed “horizontally” for each value of (theoretical) parameter.
Read out “vertically” for the measured value of the parameter.

Illustrated in figure 36.3 in **P**, see also **B**7.2, **C**9.2, **J**9.2

Modifications for low counts

The Poisson distribution is discrete
→ need overcoverage to always be right
→ fails to give lower limit if observed number of counts is too low (e.g. below background) → "flip-flopping"

Much activity among particle physicists on how best to tackle this
CERN Yellow Report, CERN 2000-005 + PHYSTAT conferences
P36.3.2.6, G.J. Feldman and R.D. Cousins, PRD 57 (98) 3873
"unified approach", uses likelihood ratio
Bayesian methods not "easy way out", you must read the literature

If you simulate with Monte Carlo: repeat with increasing statistics until results converge or read J.M. Juritz, J.W.F. Juritz and M.A. Stephens, J.Amer.Statist.Assoc. 78 (83) 441

Example: small signal with known background

Assume known background b and signal s, i.e. Poisson-distribution with parameter $\mu = s + b$.
Measure n counts, estimate signal (and limits) as $\hat{s} = n - b$ etc.

	90% conf limits		90% unified conf limits		
n	$b = 0$	$b = 1$	$b = 0$	$b = 1$	$b = 4$
0	0.00, 2.30	−1, 1.30	0.00, 2.44	0.00, 1.61	0.00, 1.01
1	0.11, 3.89	−0.89, 2.89	0.11, 4.36	0.00, 3.36	0.00, 1.39
2	0.53, 5.32	−0.47, 4.32	0.53, 5.91	0.00, 4.91	0.00, 2.33
10	6.22, 15.41	5.22, 14.41	5.50, 16.50	4.50, 15.50	1.94, 12.50

How to combine results ?

There are two situations where you may want to combine results:

The easy one: within one (your own) experiment.
Here you can often simply add data / do weighted averages.

The challenging one: combining different experiments.
— combining confidence intervals may give loss of information
— if the likelihood functions are available, you can proceed
For a Bayesian, use one result as prior distribution for the next.
For a frequentist, pretend you are analysing the experiments at the same time.

J9.6.2

PDG ideogram

The Particle Data Group procedure for averaging data.

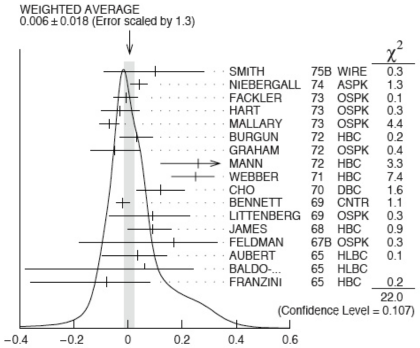

Figure 1: A typical ideogram. The arrow at the top shows the position of the weighted average, while the width of the shaded pattern shows the error in the average after scaling by the factor S. The column on the right gives the χ^2 contribution of each of the experiments. Note that the next-to-last experiment, denoted by the incomplete error flag ($\perp$), is not used in the calculation of S (see the text).

figure **P** p. 15 introduction
Illustrated e.g. with SNO data, figure **P**13.3

Robust methods — why and when ?

Two cases where **robust methods** may apply:

- Insufficient knowledge of the underlying distribution
- Indications of contaminated data

Several ways to proceed:

- Change what you extract from data (e.g. mean $\rightarrow$ median)
- Change analysis method
- "Throw away data"

Will typically end up with less efficient methods.

Could be very useful in an exploratory stage.

J8.7, **B**9

Outliers

Outlier = data point from a contamination or data point that is surprising to the experimentalist.

Never use "outlier" as an easy excuse for not understanding your experiment.

"The rejection of outliers on a purely statistical basis is and remains a dangerous procedure. Its very existence may be a proof that the underlying population is, in reality, not what it was assumed to be."
Gumbel in W. Kruskal et al, Technometrics 2 (1960) 157-166

Throw away points (1): trimming

Mean value is sensitive to outliers, median is not. Alternatives are:

trimmed mean, remove k points from both upper and lower tail, take the mean of the remaining

Winzorized mean, replace k highest points with the higest remaining value, ditto for the k lowest points, take the mean of the remaining

Simple rule of thumb:

- $n \leq 6$, use median
- $n = 7$, trim two observations from each tail
- larger n, trim 25% from each tail

Throw away points (2): M-estimation

Both ML and LS on Gaussian data is minimization of

$$\sum_i \rho(t_i)\ , \quad \rho(t) = \frac{(x - f(a))^2}{\sigma^2} = t^2, \quad t = \frac{x - f(a)}{\sigma}$$

M-estimation is similar, but with less weight on extreme points

$$\text{Huber} \quad \rho(t) = \begin{array}{ll} \frac{1}{2} t^2 & \text{for } |t| \leq k \\ k|t| - \frac{1}{2} k^2 & \text{for } |t| > k \end{array} \quad k = 1\text{–}2$$

$$\text{biweight} \quad \rho(t) = \begin{array}{ll} \frac{B^2}{6} \left(1 - \left[1 - \left(\frac{t}{B}\right)^2 \right]^3 \right) & \text{for } |t| \leq B \\ \frac{B^2}{6} & \text{for } |t| > B \end{array} \quad B = 1$$

$\sigma/\text{MAD} = 1 \ / \ 6\text{–}9$, see also Numerical Recipes, chap 14/15

Look-elsewhere effect

Finding a new peak "somewhere" is different from finding it at a previously given position. **C**4.6, **J**11.5.1

From the Wikipedia entry: A Swedish study in 1992 tried to determine whether or not power lines caused some kind of poor health effects. The researchers surveyed everyone living within 300 meters of high-voltage power lines over a 25-year period and looked for statistically significant increases in rates of over 800 ailments. The study found that the incidence of childhood leukemia was four times higher among those that lived closest to the power lines, and it spurred calls to action by the Swedish government. The problem with the conclusion, however, was that they failed to compensate for the look-elsewhere effect; in any collection of 800 random samples, it is likely that at least one will be at least 3 standard deviations above the expected value, by chance alone. Subsequent studies failed to show any links between power lines and childhood leukemia, neither in causation nor even in correlation.

Some subjects not covered here

- Monte Carlo techniques **P**37, F. James, Rep.Prog.Phys. 43 (80) 1145
- unfolding **C**11
- Student's t etc **B**7.3+8.4 and many standard statistics textbooks
- blind analysis J.R. Klein and A. Roodman, Ann.Rev.Nucl.Part.Sci. 55 (05) 141
- other new particle physics methods — P.C. Bhat, Ann.Rev.Nucl.Part.Sci. 61 (11) 281