Nuclear Instruments and Methods in Physics Research A 372 (1996) 289–292

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

ELSEVIER

Systematic errors in χ^2-fitting of Poisson distributions

Y. Jading[a,1], K. Riisager[b,*]

[a] *Institut für Kernchemie, Universität Mainz, D-55099 Mainz, Germany*
[b] *Institut for Fysik og Astronomi, Aarhus Universitet, DK-8000 Aarhus C, Denmark*

Received 31 March 1995; revised form received 3 November 1995

Abstract

Standard χ^2 fit procedures applied to Poisson-distributed data, e.g. count numbers, yield results that differ systematically from the "true values". A quantitative assessment is made of this bias that for a commonly used χ^2 method is about -1 per channel.

1. Introduction

A basic problem in data analysis is how to fit experimental data of low statistics. This can in principle be done in a variety of ways that differ with respect to consistency, biasedness, efficiency and robustness (see Ref. [1] for a detailed discussion and for a general account of the use of statistical methods in physics). In practice χ^2-minimization is often employed by physicists, whereas the maximum-likelihood principle prevails among statisticians. When data points follow the normal distribution these two methods are equivalent and give reliable results. Count numbers follow the Poisson distribution, but since this distribution converges for large numbers to the normal distribution, the question of how different fitting methods behave when applied to the Poisson distribution has received very little attention. However, for low count numbers this question must be considered seriously.

Before doing this a digression on what exactly is meant by "to fit" is needed. Three different steps can be implied, namely a goodness-of-fit test (called "test of hypothesis" in statistics), the determination of a parameter value ("point estimation") and determination of the uncertainty on the parameter ("interval estimation"). Although χ^2 methods can be used to obtain results for all three steps, one must keep in mind that they are in principle different. We shall throughout the paper be concerned mainly with the estimation of parameter values and only comment briefly on the determination of "error bars" or goodness-of-fit tests.

It is often tacitly assumed that χ^2 will work for Poisson-distributed data as long as the number of counts per channel is above 5. This seems to hold for goodness-of-fit tests [2], but as we shall show this is not true for a determination of the parameter value. That problems may occur has been known for some time [3–6], but no quantitative assessment has been published so far. Note, however, that one of our results mentioned below was obtained in the internal report [6]. Even though the systematic biases can become quite important, they are not mentioned in standard statistics textbooks for physicists [1,7]. The problem has apparently also not been discussed in the statistical literature. The simple solution is to use the maximum-likelihood method instead of χ^2, but due to the predilection of the latter method among physicists the following detailed account of its biases might nevertheless be useful.

2. The problem

In the following the notation will be introduced and the problem formulated precisely. The next section presents the analytical and numerical results and we shall end with a brief discussion.

The observed quantities are the number of counts n_i in the channels of a histogram or from a number of repetitions of the same experiment. The total number of channels (or repetitions) is N_0. The number of counts are assumed to obey the Poisson distribution [8,9]

$$P(n|\mu) = \mu^n e^{-\mu}/n!\,, \qquad (1)$$

where μ is the "true" value. The outcome of the analysis are the estimates θ_i for the values μ_i. One type of analysis is the minimization of the least-squares statistics

$$\sum_i w_i(n_i - \theta_i)^2\,, \qquad (2)$$

[1] E-mail jading@vkcmzd.chemie.uni-mainz.de.

* Corresponding author. Tel. +45 8942 3661, fax +45 8612 0740, e-mail kvr@dfi.aau.dk.

SSDI 0168-9002(95)01397-0

where different choices for the weights w_i are used. Here three explicit cases will be considered: the χ^2

$$\chi_P^2 = \sum_i (n_i - \theta_i)^2/\theta_i \,, \tag{3}$$

the modified χ^2

$$\chi_N^2 = \sum_i (n_i - \theta_i)^2/n_i \tag{4}$$

and the least-squares

$$S = \sum_i (n_i - \theta_i)^2 \,. \tag{5}$$

Since the names used for the former two statistics vary in the physics literature we shall, to avoid confusion, follow Ref. [10] and refer to them as "Pearson's χ^2" and "Neyman's χ^2". Another type of analysis consists in maximizing the likelihood function

$$L = \prod_i P(n_i|\theta_i) \,. \tag{6}$$

It is often convenient to consider the logarithm of the likelihood function instead of L itself. The two χ^2-methods differ in estimating the variance of a given data point as θ_i and n_i, respectively. In contrast, the likelihood function does not employ the variance, but rather the probability distribution itself.

Powerful computer routines are available for finding the extrema in the statistics as θ_i are varied according to the theoretical parametrization used in the fit, but it is important to be aware of possible biases when these routines are used. In order to expose clearly the biases present we consider here the case where all μ_i are identical (a "constant spectrum") and we shall be specially interested in the behaviour when the common value μ is small.

3. Analytical solution

3.1. Finite N_0

We first attempt to describe in detail the process of fitting a constant spectrum and thus consider N_0 Poisson-distributed random variables X_i, all with expectation value $E(X) = \mu$ and variance $\mathrm{Var}(X) = \mu$. The advantage of using X_i rather than the count numbers n_i (the "function values" of X_i) is that one can then not only look at average properties, but also analyse the statistical fluctuations in the fitted values. The log likelihood statistics is now

$$\ln L = \sum_{i=1}^{N_0} (X_i \ln\theta - \theta - \ln(X_i!)) \,, \tag{7}$$

with derivative

$$\frac{\partial \ln L}{\partial\theta} = \sum_{i=1}^{N_0} (X_i/\theta - 1) \,. \tag{8}$$

The result of the fit is thus given by the random variable

$$\theta_L = \frac{1}{N_0}\sum_{i=1}^{N_0} X_i \,, \tag{9}$$

and is a simple average of Poisson distributions. Since the sum of Poisson distributions is a (new) Poisson distribution, one can immediately write down the expectation value $E(\theta_L) = \mu$ and the variance $\mathrm{Var}(\theta_L) = \mu/N_0$. (It is easy to show that the same result is obtained from the unweighted least-squares statistics, S. This statistics can, however, not be expected to perform well for the general case where μ_i varies with i.)

Pearson's χ^2 has the derivative

$$\frac{\partial\chi_P^2}{\partial\theta} = \sum_{i=1}^{N_0}\left(\frac{2(\theta - X_i)}{\theta} - \frac{(\theta - X_i)^2}{\theta^2}\right) \tag{10}$$

and is minimized for

$$0 = \sum_{i=1}^{N_0} [2(\theta - X_i)\theta - (\theta - X_i)^2] = \sum_{i=1}^{N_0} (\theta^2 - X_i^2) \,, \tag{11}$$

which gives

$$\theta_P = \sqrt{\frac{1}{N_0}\sum_{i=1}^{N_0} X_i^2} \,. \tag{12}$$

For Neyman's χ^2 one has the derivative

$$\frac{\partial\chi_N^2}{\partial\theta} = \sum_{i=1}^{N_0}\frac{2(\theta - X_i)}{X_i} = 2\left[\theta\sum_{i=1}^{N_0}\frac{1}{X_i} - N_0\right] , \tag{13}$$

which is minimal for

$$\theta_N = \left[\frac{1}{N_0}\sum_{i=1}^{N_0}\frac{1}{X_i}\right]^{-1} . \tag{14}$$

The expectation value and variance for θ_P and θ_N are given in Tables 1 and 2 for the lowest values of N_0. In both cases a bias is clearly seen in the expectation value: the deviations $|E(\theta) - \mu|$ increase with N_0 and are largest for θ_N. The variance decreases with increasing N_0, it is largest for θ_N and in both cases larger than the one for L. The best results are thus obtained when using L, and χ_P^2 is favoured

Table 1
Pearson's χ^2. Expectation values and variances

N_0	$\mu = 5$		$\mu = 10$		$\mu = 20$	
	$E(\theta)$	$\mathrm{Var}(\theta)$	$E(\theta)$	$\mathrm{Var}(\theta)$	$E(\theta)$	$\mathrm{Var}(\theta)$
2	5.234	2.602	10.242	5.112	20.246	10.118
3	5.314	1.761	10.323	3.435	20.328	6.773
4	5.354	1.331	10.364	2.587	20.369	5.090
5	5.379	1.070	10.389	2.074	20.394	4.077
6	5.395	0.894	10.405	1.732	20.411	3.400
∞	5.477	–	10.488	–	20.494	–

Table 2
Neyman's χ^2. Expectation values and variances

N_0	$\mu=5$		$\mu=10$		$\mu=20$	
	$E(\theta)$	Var(θ)	$E(\theta)$	Var(θ)	$E(\theta)$	Var(θ)
2	4.522	2.783	9.500	5.471	19.500	10.487
3	4.323	1.998	9.310	3.852	19.322	7.150
4	4.212	1.572	9.208	3.001	19.230	5.434
5	4.141	1.299	9.143	2.470	19.174	4.385
6	4.090	1.106	9.099	2.106	19.136	3.678
∞	3.780	–	8.844	–	18.940	–

to χ^2_N (this ordering of the three methods is rather general [1,10]).

3.2. Asymptotic behaviour

For large N_0 the statistical fluctuations become unimportant (the relative variances go to zero) and the expectation values for the χ^2 statistics can be derived analytically. This is done by noting that the number of channels in the sum with the value k is $N_0P(k|\mu)$. By reordering the summation one obtains for Pearson's χ^2

$$\chi^2_P \rightarrow N_0 \sum_{k=0}^{\infty} P(k|\mu)(k-\theta)^2/\theta . \tag{15}$$

The derivative with respect to θ is easily calculated

$$\begin{aligned}\frac{1}{N_0}\frac{\partial \chi^2_P}{\partial \theta} &= \sum_{k=0}^{\infty} P(k|\mu)(1-k^2/\theta^2) \\ &= 1-\theta^{-2}\sum_{k=0}^{\infty} P(k|\mu)(k^2-k+k) \\ &= 1-\frac{\mu^2+\mu}{\theta^2} .\end{aligned} \tag{16}$$

The minimum in χ^2_P is thus obtained for

$$\theta_P = \mu\sqrt{1+1/\mu} \simeq \mu + 1/2 . \tag{17}$$

Fits with Pearson's χ^2 will systematically yield too large values.

The result for Neyman's χ^2 is slightly more complicated to derive since one here must safeguard against division by zero. We follow the standard procedure of attributing a variance of one to count numbers that are zero. From

$$\chi^2_N \rightarrow N_0 \sum_{k=0}^{\infty} P(k|\mu)(k-\theta)^2/\max(k,1) \tag{18}$$

one obtains the derivate

$$\frac{1}{N_0}\frac{\partial \chi^2_N}{\partial \theta} = 2\left[P(0|\mu)\theta + \sum_{k=1}^{\infty} P(k|\mu)(\theta/k-1)\right] . \tag{19}$$

The minimum is thus obtained for

$$\theta_N = \frac{1-P(0|\mu)}{P(0|\mu)+\sum_{k=1}^{\infty} P(k|\mu)/k} , \tag{20}$$

where for large μ one can approximate

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{P(k|\mu)}{k} &= \sum_{k=1}^{\infty} P(k|\mu)\left(\frac{1}{k+1}+\frac{1}{k}-\frac{1}{k+1}\right) \\ &\simeq \frac{1}{\mu}+\sum_{k=1}^{\infty}\frac{P(k|\mu)}{k(k+1)} \simeq \frac{1}{\mu}+\frac{1}{\mu^2}\end{aligned} \tag{21}$$

so that

$$\theta_N \simeq \frac{\mu^2}{\mu+1} \simeq \mu - 1 . \tag{22}$$

Fits with Neyman's χ^2 will yield too small values [6]. Qualitatively the biases in the two χ^2 statistics arise as follows. For Pearson's χ^2 the $(k-\theta)^2$ factor would give the same minimum as for S, but the factor θ^{-1} clearly favours larger values of θ and gives the upward shift. For Neyman's χ^2 the k-values less than μ are weighted stronger than the ones larger than μ and this shifts the minimum downwards [4–6]. In Ref. [5] a relation between the total areas of fit function and data was derived (for simple fit functions) that is similar to the asymptotic results just described. The present derivation is, however, more general and shows that the bias stems directly from the treatment of Poisson distributions with χ^2 statistics.

The asymptotic values just derived are also given in Tables 1 and 2. The results are also plotted as a function of μ in Fig. 1 for all three statistics considered. The asymptotic deviations of -1 and $+\frac{1}{2}$ for χ_N and χ_P are seen to be good approximations down to μ about 10. Neyman's χ^2 is again the most deviant. In the figure numerical results from Monte-Carlo calculations are also given. Poisson-distributed values were obtained by using random numbers generated with the CERNLIB routine RANMAR, and fits with the three different statistics were performed by MINUIT [11]. Data sets ("spectra") of varying length N_0 were used. For each value of μ and N_0 the average over several data sets were taken so that the final error on θ in all cases lie in the range 0.005 to 0.03. For small N_0 the minimization sometimes gave problems for very low values of μ, the lowest plotted points for $N_0=5$ therefore probably carry an additional systematic uncertainty. One should note that χ^2_N convergences much slower to the asymptotic values, both analytically and numerically, than χ^2_P.

4. Discussion

Stated briefly, our main result is that χ^2-fits will yield a wrong value for the parameter of a Poisson distribution. The values obtained by two standard analysis methods, χ^2_P and χ^2_N, will asymptotically deviate by $+\frac{1}{2}$ and -1,

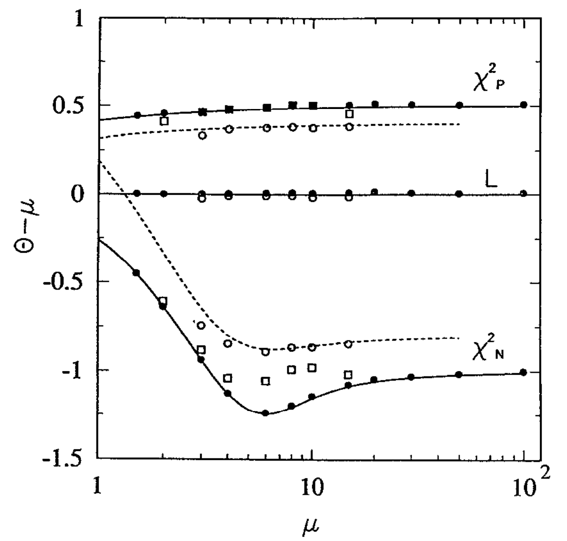

Fig. 1. The difference between the fit estimate θ and the Poisson parameter μ is plotted versus μ for the two χ^2 statistics and for the maximum-likelihood statistics (note the logarithmic scale). Solid lines are the asymptotic theoretical values for N_0 very large, dashed lines the ones for $N_0 = 5$. The points are results from Monte-Carlo calculations with $N_0 = 5$, 20, 10 000 (open circles, open squares, filled circles).

respectively. We shall now discuss some of the implications of this result.

Since the relative size of the bias goes as μ^{-1} it appears to be favourable to sum channels (rebin the spectrum) in order to reduce the effect. This will of course work perfectly well for a constant background, but there are limits to the applicability of this procedure for more realistic spectra since features present will gradually disappear due to the "lumping together". For more details on the loss of information in binning and a discussion of the optimal bin size, see Ref. [2].

For the case considered so far, namely a constant background, the much simpler procedure of taking the average of all count numbers gives the correct result. So why worry about more elaborate fitting methods? An advantage is that they do allow for goodness-of-fit tests, i.e. one can check that the data actually are distributed as one believes them to be, but the constant background indeed mainly serves to illustrate the fit biases clearly, and we shall now consider briefly what happens in more realistic cases.

For a spectrum with structure we would again expect χ^2_P and χ^2_N fits to give estimates for the count numbers in individual channels that are off by about $+\frac{1}{2}$ and -1. The area of a component in the spectrum can therefore come out wrong, but also other parameters – the half life for an exponential decay [12,13] or the width of a peak – can be affected. This holds in particular if the fit is performed without a free background term, as sometimes is necessary to obtain reasonable results when the total number of counts is low. If a freely varying background term is included in the fit, it will often be able to include most of the bias and will thus give appreciably more reliable results for the other parameters. We have verified this with numerical simulations both for an exponential decay and for a Gaussian peak, with and without a background term, but since these cases now contain much more parameters, general results are more complicated to write down. To estimate the bias for a specific case one should do Monte-Carlo studies. It is of course preferable to avoid the bias by employing maximum-likelihood methods, and the Poisson likelihood chi-square $\chi^2_{\lambda,P}$ from [10] is here particularly attractive since it also allows for goodness-of-fit tests.

In ending, it is perhaps worth stressing that the bias in the χ^2 methods depends on the number of counts per channel and *not* on the total number of counts in the spectrum or of any component in the spectrum (if one disregards the slight N_0-dependence). Also, it should be noted that the χ^2_N statistics, which seems to be the most widely used, gives the largest discrepancies. It should therefore be avoided for the determination of parameter values.

Acknowledgement

We thank the referee for drawing our attention to Prof. Müller's work.

References

[1] W.T. Eadie, D. Drijard, F.E. James, M. Roos and B. Sadoulet, Statistical Methods in Experimental Physics (North-Holland, Amsterdam, 1971).
[2] ibid., chap. 11.
[3] M.G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 2, 2nd Ed. (Griffin, London, 1967) p. 94.
[4] G.W. Philips, Nucl. Instr. and Meth. 153 (1978) 449.
[5] T. Awaya, Nucl. Instr. and Meth. 165 (1979) 317.
[6] J.W. Müller, A propos d'un abus de poids statistiques, Rapport BIPM-84/7 (1984).
[7] L. Lyons, Statistics for Nuclear and Particle Physicists (Cambridge University Press, 1986).
[8] F.A. Haight, Handbook of the Poisson Distribution (Wiley, New York, 1967).
[9] N.L. Johnson, S. Kotz and A.W. Kemp, Univariate Discrete Distributions, 2nd Ed. (Wiley, New York, 1992) Chap. 4.
[10] S. Baker and R.D. Cousins, Nucl. Instr. and Meth. 221 (1984) 437.
[11] F. James and M. Roos, MINUIT, Function Minimization and Error Analysis, CERN D506 (Long Writeup).
[12] T.Awaya, Nucl. Instr. and Meth. 174 (1980) 237.
[13] A.H. Kalantar, Nucl. Instr. and Meth. 211 (1983) 556.