A proposal for a different chi-square function for Poisson distributions

F.M.L. Almeida Jr.[a],*, M. Barbi[a],[1], M.A.B. do Vale[b]

[a]*Instituto de Física, Universidade Federal do Rio de Janeiro, Ilha do Fundão, Rio de Janeiro 21945-970, RJ, Brazil*
[b]*Centro Brasileiro de Pesquisas Físicas, Rua Xavier Sigaud 150, Urca, Rio de Janeiro 22290-180, RJ, Brazil*

Abstract

We obtain an approximate Gaussian distribution from a Poisson distribution after doing a change of variable. A new chi-square function is obtained which can be used for parameter estimations and goodness-of-fit testing when adjusting curves to histograms. Since, the new distribution is approximately Gaussian we can use it even when the bin contents are small. The corresponding chi-square function can be used for curve fitting. This chi-square function is simple to implement and presents a fast convergence of the parameters to the correct value, especially for the parameters associated with the width of the fitted curve. We present a Monte Carlo comparative study of the fitting method introduced here and two other methods for three types of curves: Gaussian, Breit–Wigner and Moyal, when each bin content obeys a Poisson distribution. It is also shown that the new method and the other two converge to the same result when the number of events increases. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

Nowadays, there is an intensive experimental effort to search for new phenomena in many fields of particle physics. These signals should be rare events where one needs to treat low statistics data and non-Gaussian errors. In order to handle these data one needs special care, since all the usual Gaussian procedures and techniques are no longer valid. On the other hand, it turns out that the minimization of the chi-square functions is normally simple, relatively fast and a very familiar way of fitting data, estimating parameters and their errors so as the goodness-of-fit testing. The χ^2 of fitting a theoretical curve to experimental data in order to estimate some parameter values is to minimize the function

$$\chi_G^2 = \sum_{i=1}^{N} \frac{(f(x_i, \boldsymbol{\alpha}) - n_i)^2}{\sigma_i^2} \tag{1}$$

where $f(x_i, \boldsymbol{\alpha})$ is a known function(model prediction) calculated at the point x_i and $\boldsymbol{\alpha}$ is a vector of the parameters one wants to obtain; n_i is the measured experimental value associated with the bin located at x_i and N is the number of bins in the range of interest. This method has been largely used but presents some limitations such as the assumption that each bin content must obey a Gaussian distribution of spread σ_i or the measurement errors are at least almost "Gaussian". If the contents of each

*Corresponding author.

E-mail addresses: marroqui@if.ufrj.br (F.M.L. Almeida Jr.), barbi@mail.desy.de (M. Barbi), aline@lafex.cbpf.br (M.A.B. do Vale).

[1] Now at McGill University, Canada.

bin obeys a Poisson distribution some authors [1] recommend that each bin content must have at least a statistically significant number of entries n_i such as that $\sigma_i \approx \sqrt{n_i}$, where the asymptotically limiting case for large number of measurements starts, and the square root of the variance can be considered as a good interval estimation. For values smaller than these, there are suggestions that one must use bins of variable sizes such that each bin content be greater than the statistically significant number of entries n_i. The disadvantage of these suggestions is that one can lose important information about the structure of the studied distribution. There is no rule of thumb for the bin width or for the ideal number of bins in which the region of interest should be divided. There is yet another hint that the ideal case is to use bins of variable width of equal probability contents [2].

In a very interesting paper, Baker and Cousins [3], call attention and discuss some topics such as point estimation, confidence interval estimation, goodness-of-fit testing, biased estimation, etc., when fitting curves to histograms using chi-square function.

They presented a χ^2_{BC} function for fitting histograms when the bin contents obey a Poisson distribution. They defined a Poisson likelihood chi-square which is given by

$$\chi^2_{BC} = 2 \sum_{i=1}^{N} \left[f(x_i, \boldsymbol{\alpha}) - n_i + n_i \log(n_i/f(x_i, \boldsymbol{\alpha})) \right]. \quad (2)$$

This function behaves asymptotically as a chi-square distribution and then can be used for estimation and goodness-of-fit testing.

The chi-function presented in this paper has also asymptotically a behavior like the classical chi-square function, a fast convergence to the correct value, much less fluctuations when compared with χ^2_G and it works also when one has distributions with long tails where bin contents can be very low. It is of easy implementation in any minimization program.

After this introduction, we demonstrate in Section 2 how one can transform a non-Gaussian pdf into an approximate Gaussian pdf. Section 3 is devoted to obtain the chi-square function for the approximate Gaussian pdf from a Poisson distribu-

tion and discuss some of its characteristics. In Section 4, we compare with the results obtained using Eqs. (1) and (2), and the new chi-square function for different number of entries in the histograms. This comparison is made using Monte Carlo events generated according to some distributions with known parameters, as suggested by Baker and Cousins [3] and in Section 5 we present the conclusions. In the appendix we obtain the equivalent chi-square expression for a binomial distribution.

2. Obtaining an approximate Gaussian distribution

The basic motivation is to transform, via a variable transformation, a non-Gaussian probability density function (pdf) into an approximately Gaussian pdf preserving the probability even when one has small number of events [2,4]. Let us consider a non-Gaussian pdf $p(x)$ and one wants to obtain a transformation such that the probability is preserved Eq. (3) and that the new pdf $q(z)$ should be approximately Gaussian:

$$q(z)dz = p(x)\,dx. \quad (3)$$

Then, $q(z)$ can be written as

$$q(z) = p(x)\left|\frac{dx}{dz}\right|. \quad (4)$$

When a pdf is unimodal and obeys some regularity conditions [4], the logarithm of it is approximately quadratic so that

$$\log(p(x)) \approx \log(p(\hat{x})) - \frac{1}{2}\left(-\frac{\partial^2 \log(p(x))}{\partial x^2} \right)_{\hat{x}} (x - \hat{x})^2 \quad (5)$$

where $\hat{x}$ is the point associated to the maximum of $p(x)$ and one can define the following quantity:

$$J(\hat{x}) = \left(-\frac{\partial^2 \log(p(x))}{\partial x^2} \right)_{\hat{x}}. \quad (6)$$

On the other hand, the logarithm of a normal pdf $g(x)$, with mean $\hat{\mu}$ and standard deviation σ, is of the form

$$\log(g(x)) = \text{const} - \frac{1}{2\sigma^2}(x - \hat{\mu})^2 \quad (7)$$

so that given the location parameter $\hat{\mu}$, it is completely determined by its standard deviation σ.

A comparison between Eqs. (5) and (7) shows that the variance of the pdf $p(x)$ is approximately equal to $J^{-1}(\hat{x})$. Let us suppose now that $z(x)$ is a one-to-one transformation between x and z, then using the chain rule for derivatives one gets the relation

$$
\begin{aligned}
J(\hat{z}) &= \left(-\frac{\partial^2 \log(p(x))}{\partial x^2} \right)_{\hat{x}} \left| \frac{dx}{dz} \right|_{\hat{z}}^2 \\
&= J(\hat{x}) \left| \frac{dx}{dz} \right|_{\hat{z}}^2 .
\end{aligned}
\tag{8}
$$

Let us choose $z(x)$ such that

$$
\left(\frac{dx}{dz} \right)_{\hat{z}} = J^{-1/2}(\hat{x}).
\tag{9}
$$

This choice is made so as to make $J(\hat{z})$ independent of $\hat{z}$, the standard deviation equal to one and the new distribution $q(z)$ approximately translation invariant along the z-axis. Thus the metric can be obtained from the relationship obeying the above conditions

$$
\frac{dz}{dx} = J^{1/2}(x)
\tag{10}
$$

$$
z = \int^x J^{1/2}(t)\, dt.
\tag{11}
$$

3. The chi-square function

As an example, let us apply the above prescription to a Poisson pdf given by

$$
p_P(x) = \frac{x^k e^{-x}}{\Gamma(k+1)}
\tag{12}
$$

which means that after observing k events one has a pdf of the estimated mean parameter x. Now one wants to find a Gaussian-like pdf through a transformation of x. The location of the $p_P(x)$ maximum is easily shown to be at $\hat{x} = k$ and the term associated to the second derivative is

$$
-\frac{\partial^2 \log(p(x))}{\partial x^2} = \frac{k}{x^2}.
\tag{13}
$$

Using the fact that $\hat{x} = k$, one obtains

$$
J(\hat{x}) = \frac{1}{k}
\tag{14}
$$

then

$$
J^{1/2}(x) = \frac{1}{\sqrt{x}}.
\tag{15}
$$

Using Eq. (9), the one-to-one transformation is obtained as

$$
\frac{dz}{dx} = \frac{1}{\sqrt{x}}
\tag{16}
$$

$$
z = \int^x \frac{1}{\sqrt{t}}\, dt
\tag{17}
$$

$$
z = 2\sqrt{x}
\tag{18}
$$

and the inverse transformation is

$$
x = \left(\frac{z}{2} \right)^2.
\tag{19}
$$

Reusing the above expression in Eq. (12) and using Eq. (4), one obtains an approximately Gaussian expression $q_P(z)$ associated to the Poisson pdf $p_P(x)$, which is

$$
q_P(z) = \frac{(z/2)^{2k+1} e^{-(z/2)^2}}{\Gamma(k+1)}
\tag{20}
$$

It is not difficult to show that the above expression is normalized, $J(\hat{z}) = 1$ and is translation invariant by construction. The approximate Gaussian and exact Gaussian distributions, $q_P(z)$ and $g(z)$, respectively, are shown in Fig. 1 for different values of k ($k = 0, 1, 2, 5$ and 10). The worst approximation occurs at $k = 0$ but it gets better very fast as k increases.

This pdf $q_P(z)$ has a maximum at $\hat{z} = \sqrt{4k+2}$ which corresponds to $x_{\hat{z}} = k + 1/2$. It is interesting to note that $x_{\hat{z}}$ is between $\hat{x} = k$ and the median of the Poisson pdf, $x_m \approx k + \frac{2}{3}$, i.e., the maximum of $p_P(x)$ is not directly related to the maximum of $q_P(z)$ via Eqs. (18) and (19).

It is not difficult either to obtain an analytical expression for the confidence intervals for roughly 68.3% of confidence level since the obtained $q_P(z)$ has a maximum at $\hat{z} = \sqrt{4k+2}$ and a standard
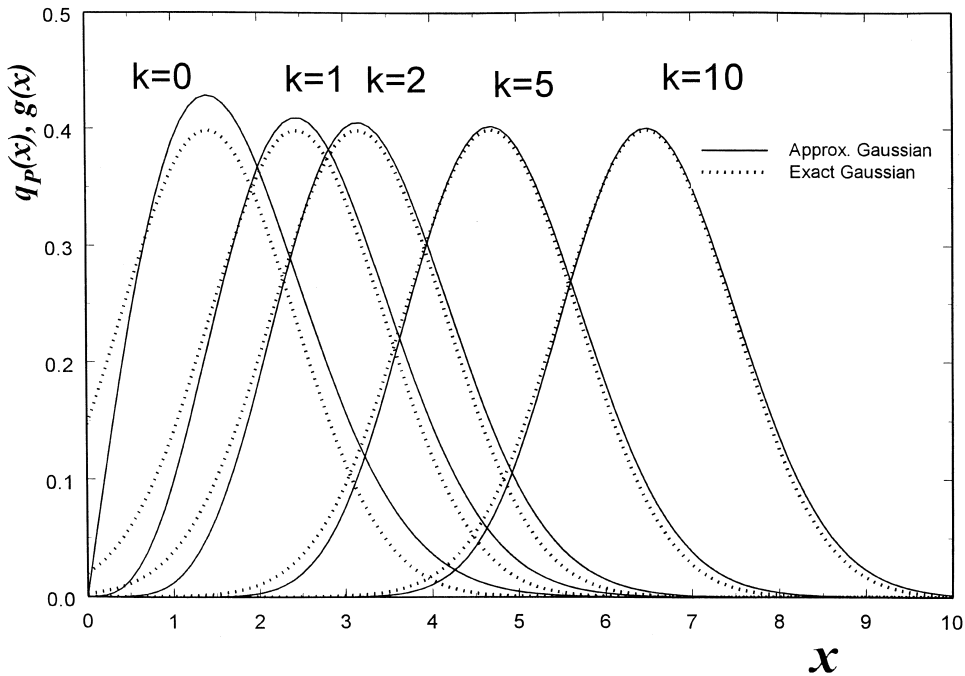
Fig. 1. The approximate Gaussian probability density function (pdf) associated to a Poisson pdf for $k = 0, 1, 2, 5$ and 10 and the exact Gaussian distribution with variance equal to one.

deviation equal to unit, one gets

$$[z_{min}, z_{max}] = [\sqrt{4k + 2} - 1, \sqrt{4k + 2} + 1]. \quad (21)$$

Taking the inverse transformation one gets the corresponding interval associated with the original Poisson pdf

$$[x_{min}, x_{max}] = [(z_{min}/2)^2, (z_{max}/2)^2]. \quad (22)$$

The confidence level calculated according to Eqs. (21) and (22) is shown in Table 1 for different values of k, and so are the probability contents of the calculated interval. One can see that these intervals overestimate the 68.27% confidence level but converge to it as k increases.

Let us suppose that one wants to fit a set of data when the contents of each bin obeys a Poisson distribution. If the contents of each bin has small numbers of events, one can no longer use as its standard deviation $\sigma_i = \sqrt{n_i}$, since the errors are asymmetrical and consequently one can not use the least-squares fit method, Eq. (1), either since it

works only for Gaussian pdf. If one insists one could get large deviations for the estimated parameters as is shown in the figures of the next section. After taking the above transformation, the contents of each bin is approximately Gaussian.

Using the likelihood ratio test theorem [2,3], one gets the following expression in terms of bin contents n_i:

$$\chi_P^2 = -2 \sum_{i=1}^{N} \log(\lambda_i) \quad (23)$$

where

$$\lambda_i = \frac{(\sqrt{4f(x_i, \boldsymbol{\alpha}) + 2}/2)^{2n_i + 1} e^{-(\sqrt{4f(x_i, \boldsymbol{\alpha}) + 2}/2)^2}}{(\sqrt{4n_i + 2}/2)^{2n_i + 1} e^{-(\sqrt{4n_i + 2}/2)^2}} \quad (24)$$

which gives

$$\chi_P^2 = \sum_{i=1}^{N} \left[2(f(x_i, \boldsymbol{\alpha}) - n_i) + (2n_i + 1) \log\left(\frac{2n_i + 1}{2f(x_i, \boldsymbol{\alpha}) + 1}\right) \right] \quad (25)$$

Table 1
Approximate 68.3% confidence limits, $[z_{min}, z_{max}]$ and $[x_{min}, x_{max}]$, for the almost Gaussian and Poisson distributions, respectively, for different k values

k	$z_{min} = $ $\sqrt{4k+2} - 1$	$z_{max} = $ $\sqrt{4k+2} + 1$	$x_{min} = (z_{min}/2)^2$	$x_{max} = (z_{max}/2)^2$	$\int_{y_{min}}^{y_{max}} q_P(z)\,dz$ $\int_{x_{min}}^{x_{max}} p_P(x)\,dx$
0	0.414213562	2.414213562	0.04289321873	1.457106781	0.7251045244
1	1.449489743	3.449489743	0.5252551288	2.974744873	0.6990903884
2	2.162277660	4.162277660	1.168861170	4.331138830	0.6926877287
3	2.741657387	4.741657387	1.879171307	5.620828695	0.6898630257
4	3.242640687	5.242640687	2.628679658	6.871320343	0.6882789039
5	3.690415760	5.690415760	3.404792120	8.095207880	0.6872666311
6	4.099019514	6.099019514	4.200490245	9.299509758	0.6865643067
7	4.477225575	6.477225575	5.011387213	10.48861279	0.6860486195
8	4.830951895	6.830951895	5.834524053	11.66547595	0.6856539615
9	5.164414003	7.164414003	6.667792998	12.83220700	0.6853422256
10	5.480740698	7.480740698	7.509629650	13.99037035	0.6850897795
20	8.055385138	10.05538514	16.22230743	25.27769258	0.6839191522
30	10.04536102	12.04536102	25.22731950	36.27268053	0.6835159909
40	11.72792206	13.72792206	34.38603895	47.11396103	0.6833119043
50	13.21267040	15.21267040	43.64366478	57.85633518	0.6831886429
60	14.55634919	16.55634919	52.97182543	68.52817463	0.6831061362
70	15.79285562	17.79285562	62.35357215	79.14642778	0.6830470354
80	16.94435844	18.94435844	71.77782073	89.72217918	0.6830026194
90	18.02629759	20.02629759	81.23685120	100.2631488	0.6829680213
100	19.04993766	21.04993766	90.72503123	110.7749689	0.6829403087

which asymptotically behaves like a chi-square distribution. This expression is similar to Eq. (2) obtained by Baker and Cousins [3].

One can also derive an equivalent expression for a binomial distribution. This is shown in the appendix.

4. Comparing the different chi-square functions

Let us now compare the results obtained by minimizing Eqs. (1), (2) and (25), when the bin contents obeys a Poisson distribution. This comparison was made for three different curves: Gaussian, Breit–Wigner and Moyal, Eqs. (26), (27) and (28), respectively.

$$f_G(x) \propto e^{-(x-\mu_G)^2/2\sigma_G^2} \tag{26}$$

$$f_{BW}(x) \propto \frac{1}{(2\sigma_{BW}^2 + (x-\mu_{BW})^2)} \tag{27}$$

$$f_M(x) \propto e^{-h-e^{-h}} \tag{28}$$

where $h = (x - \mu_M)/\sigma_M$.

The first two are symmetrical curves with "short" and long "tails", respectively, while the last one is an asymmetrical function with a left "short" and a right "long" tails.

The parameters $\{\mu_j\}$ are associated to the maximum value of the distribution and $\{\sigma_j\}$ are related to the spread of the distribution, where $j = G, BW, M$.

One generates random points with known $\bar{\mu}_j$ and $\bar{\sigma}_j$ according to each of the above distributions and filling histograms with 100 bins (see Table 2).

The number of entries ranges from 20 to 10^4 and for each fixed number of entries, 10^4 sets of points were generated.

For each fixed number of entries we calculate the average value of the fitted parameters μ_j and σ_j which are the estimators of $\bar{\mu}_j$ and $\bar{\sigma}_j$, respectively, using Eqs. (1), (2) and (25) and the MERLIN optimization package [5].

The fit was done from the first bin to the last bin content different from zero, although there could exist bins of contents equal to zero in between, except for Eq. (1) where the bins of contents

Table 2
Distributions, their parameters and the range divided in 100 bins according to which the random numbers were generated

Distribution	$\bar{\mu}_j$	$\bar{\sigma}_j$	Range
Gauss	0.0	1.0	$[-5.0, 5.0]$
Breit–Wigner	50.0	2.0	$[0.0, 100.0]$
Moyal	2.0	0.5	$[0.0, 8.0]$

equal to zero were excluded in order to avoid singularities. One also calculates the expected mean errors defined as $\Delta\sigma_j = \sqrt{\langle\sigma_j - \bar{\sigma}_j\rangle^2}$ and $\Delta\mu_j = \sqrt{\langle\mu_j - \bar{\mu}_j\rangle^2}$ of the fitted μ_j and σ_j with respect to the "true" known $\bar{\mu}_j$ and $\bar{\sigma}_j$, fixing the number of entries.

These results are summarized in Figs. 2–13 in terms of the number of entries. We can see clearly that the minimization of χ_P^2 shows faster or equal convergence to "true" value and smaller or equal expected mean errors as the number of entries increase than when using χ_{BC}^2. Both these functions

are systematically much better than χ_G for the convergence and expected mean errors of the parameters. For small number of entries, we can also note that the parameters σ_i are systematically overestimated for all the shown cases while μ_M for the Moyal case is underestimated. All the chi-square functions here presented converge systematically to the correct value as the number of entries increase. One can observe in all figures that all three methods coincide as the number of entries increase. We can clearly see the advantage of using the chi-square function introduced here. It converges equally well or faster and has equal or smaller expected mean errors than the other two methods. Besides, Eq. (25) is also of easy implementation in optimization programs.

5. Conclusions

This article presented an improvement over the usual minimization of chi-square function
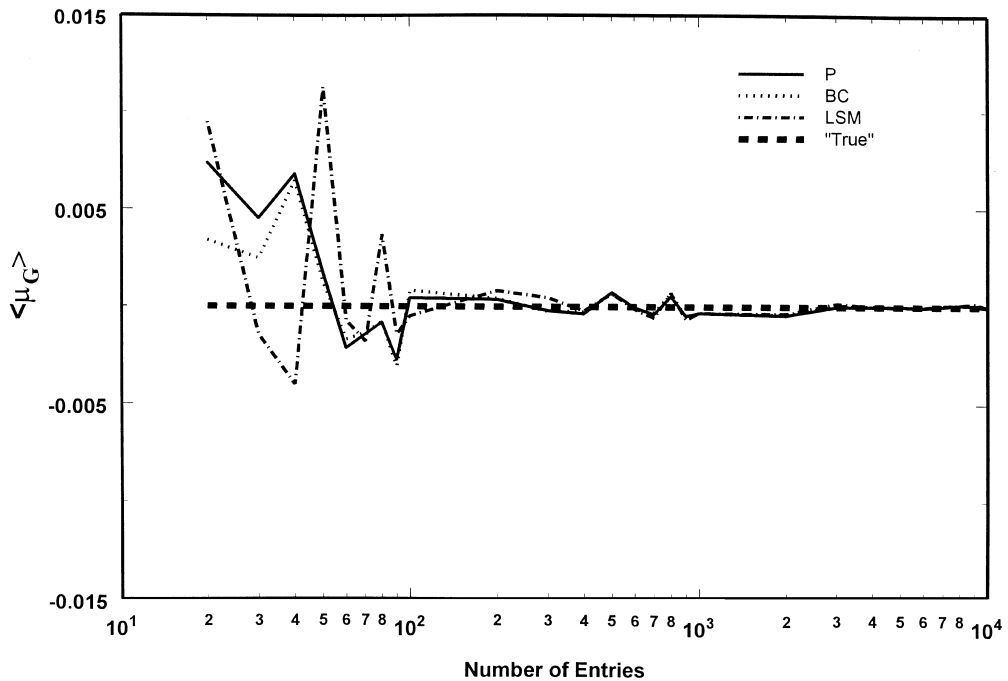

Fig. 2. The average value of the fitted Gaussian parameter μ_G as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ_P^2, χ_{BC}^2 and χ_G^2, respectively.
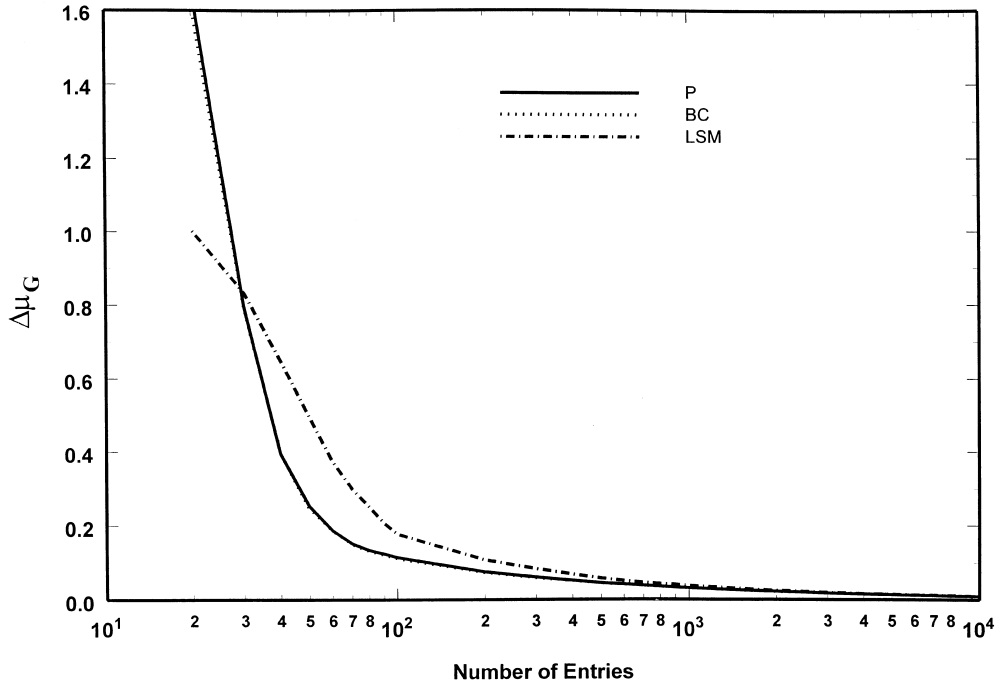
Fig. 3. The expected mean error $\Delta\mu_G$ of the fitted parameter μ_G as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ^2_P, χ^2_{BC} and χ^2_G, respectively.
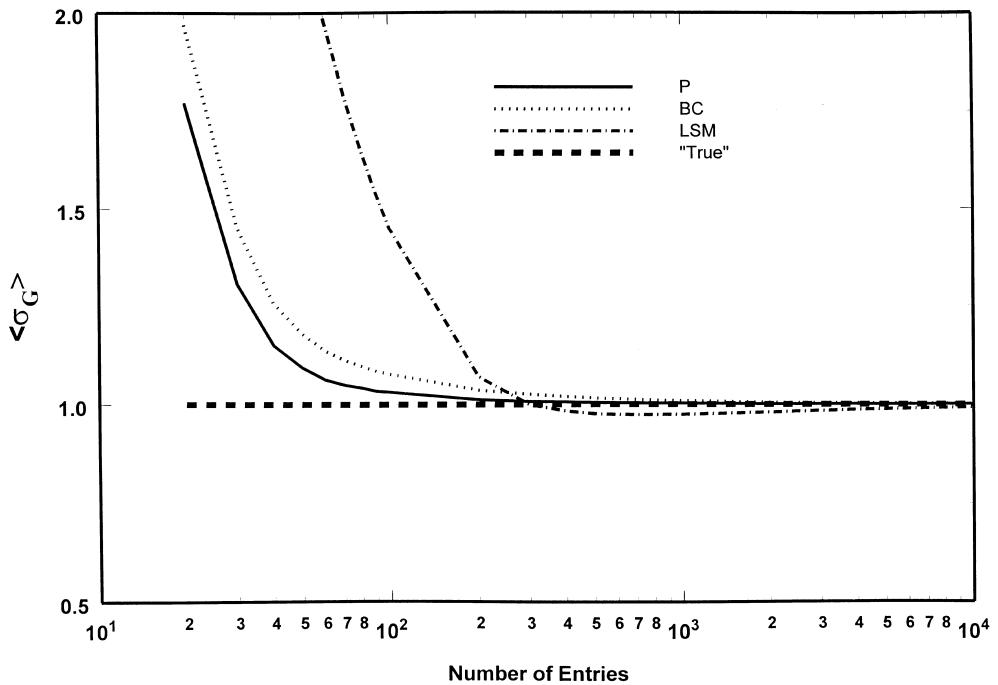

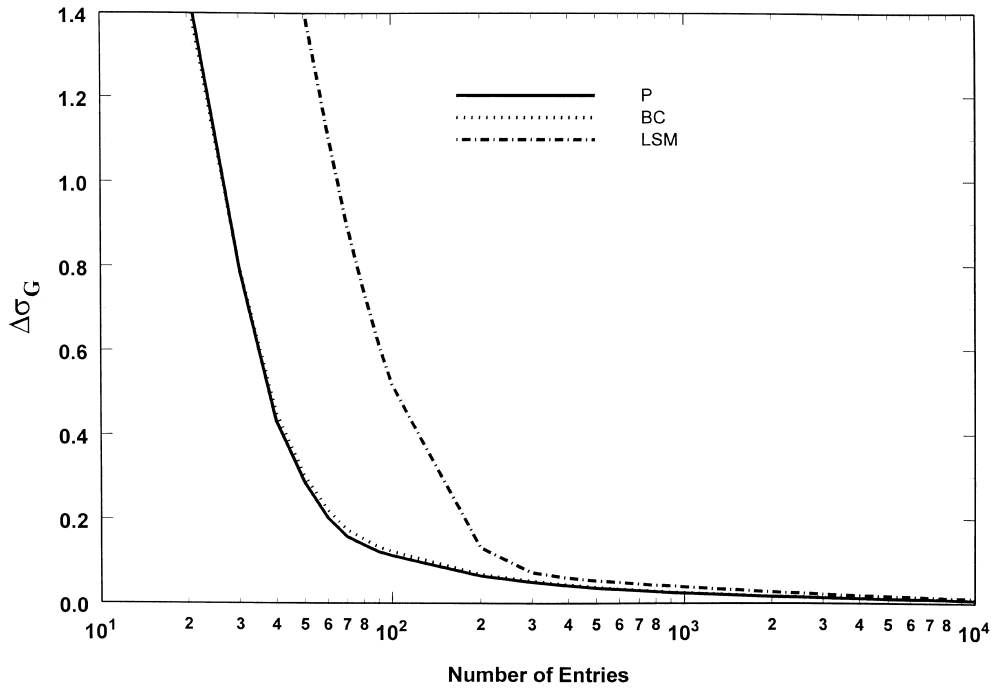Fig. 4. Same as Fig. 2 but for the parameter σ_G.
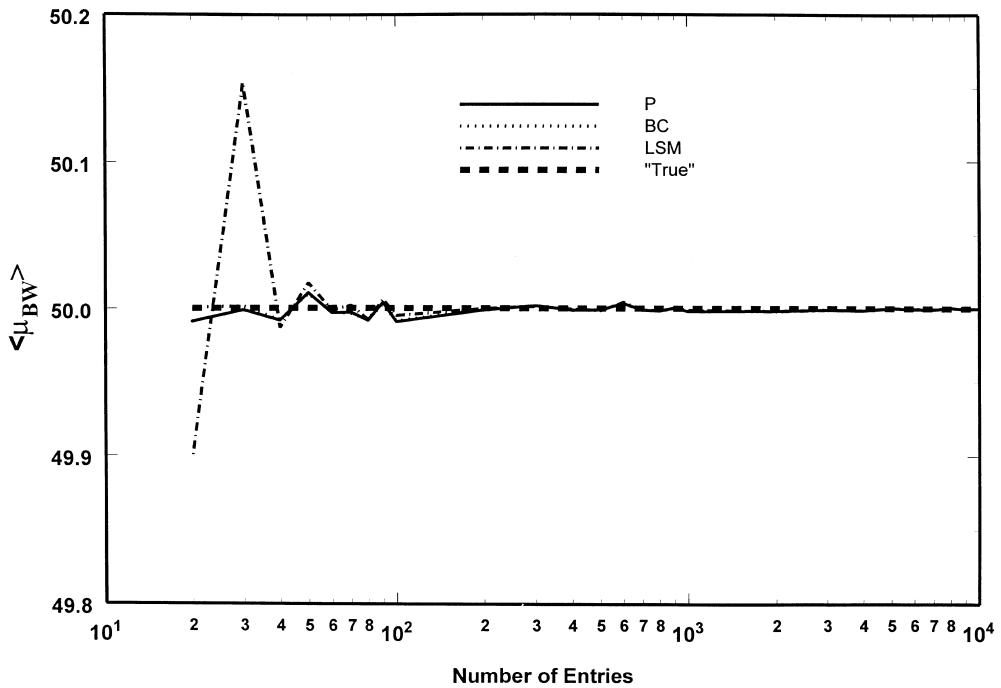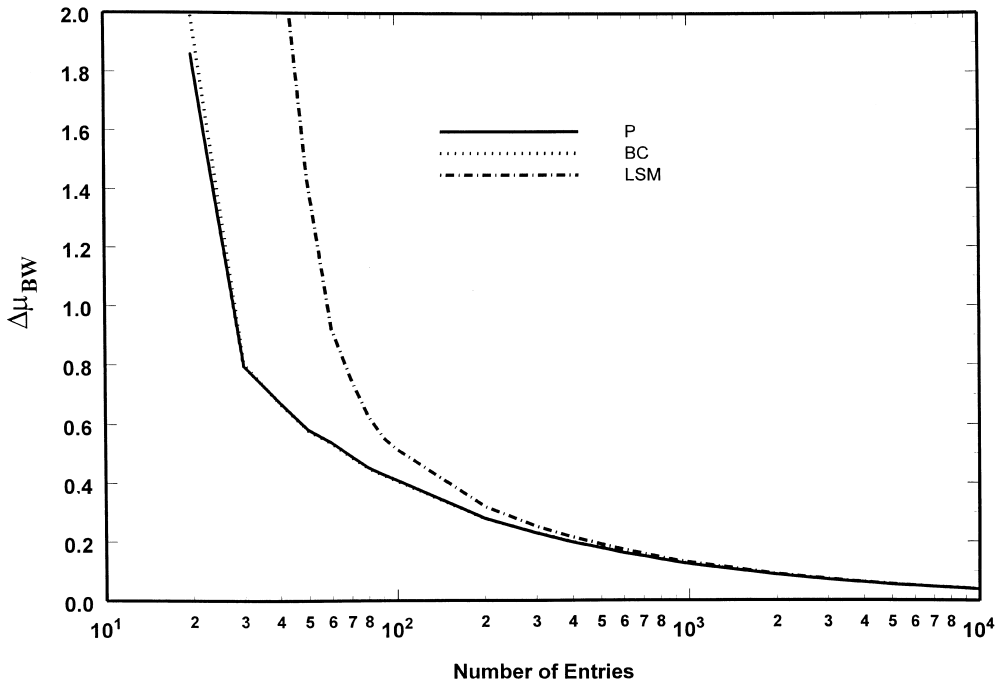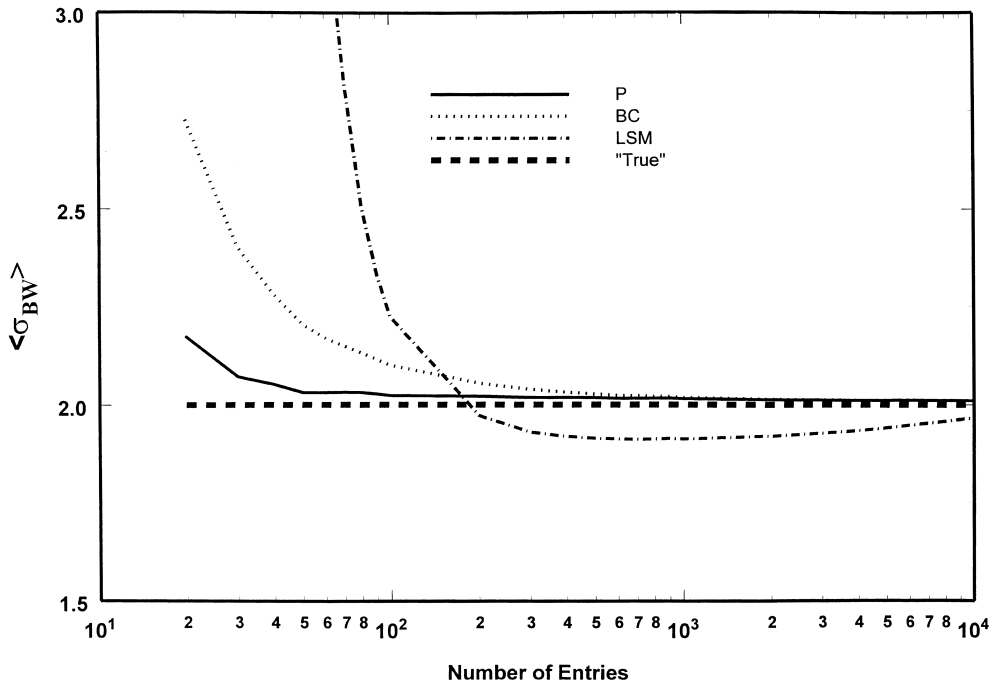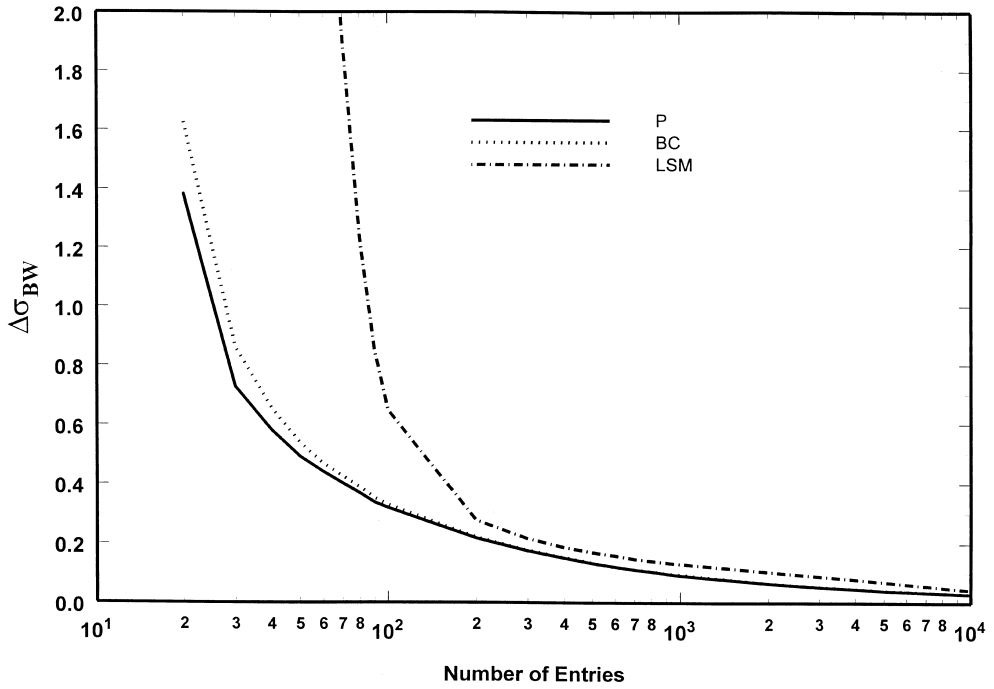
Fig. 5. Same as Fig. 3 but for the parameter σ_G.

Fig. 6. The average value of the fitted Breit–Wigner parameter μ_{BW} as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ^2_P, χ^2_{BC} and χ^2_G, respectively.

Fig. 7. The expected mean error $\Delta\mu_{\text{BW}}$ of the fitted parameter μ_{BW} as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ^2_{P}, χ^2_{BC} and χ^2_{G}, respectively.

Fig. 8. Same as Fig. 6 but for the parameter σ_{BW}.

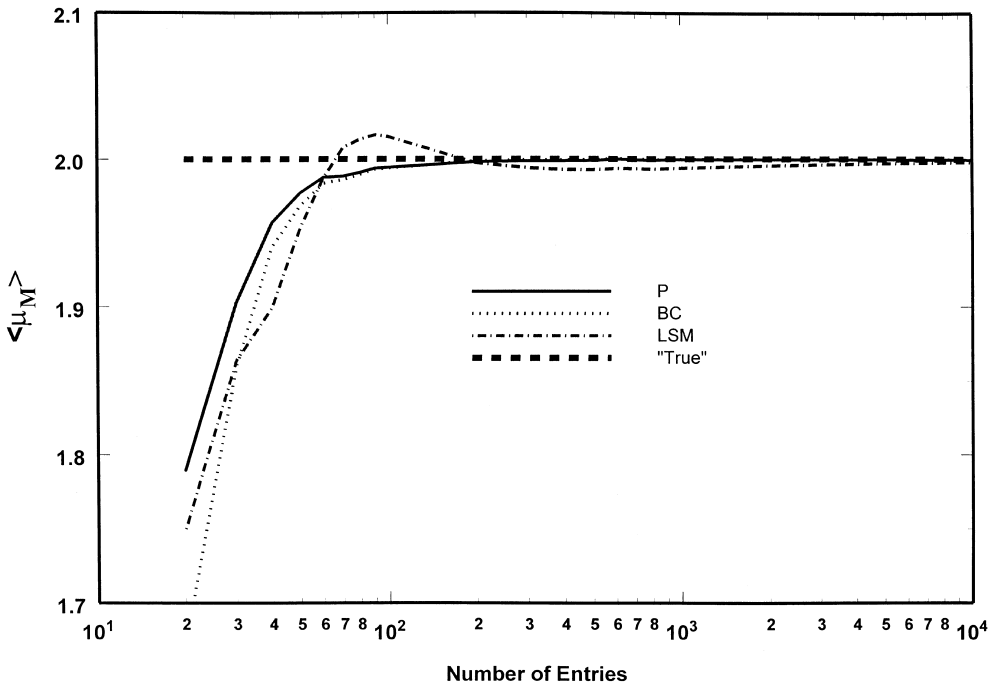Fig. 9. Same as Fig. 7 but for the parameter σ_{BW}.

Fig. 10. The average value of the fitted Moyal parameter μ_M as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ^2_P, χ^2_{BC} and χ^2_G, respectively.
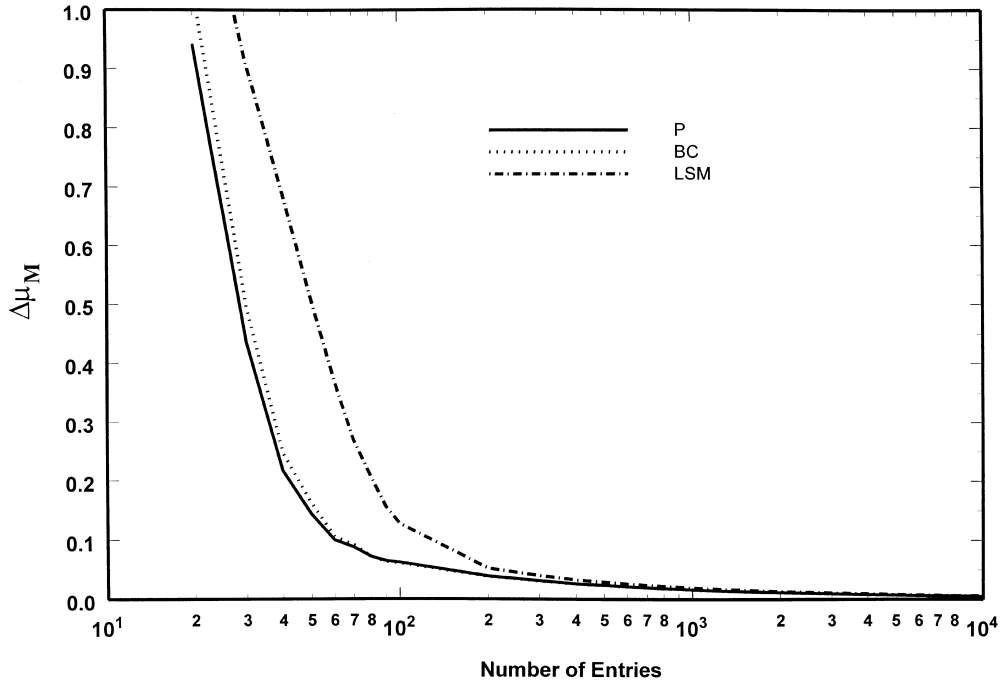
Fig. 11. The expected mean error $\Delta\mu_M$ of the fitted parameter μ_M as a function of the number of entries. The curves P, BC and LSM correspond to chi-square functions χ^2_P, χ^2_{BC} and χ^2_G, respectively.
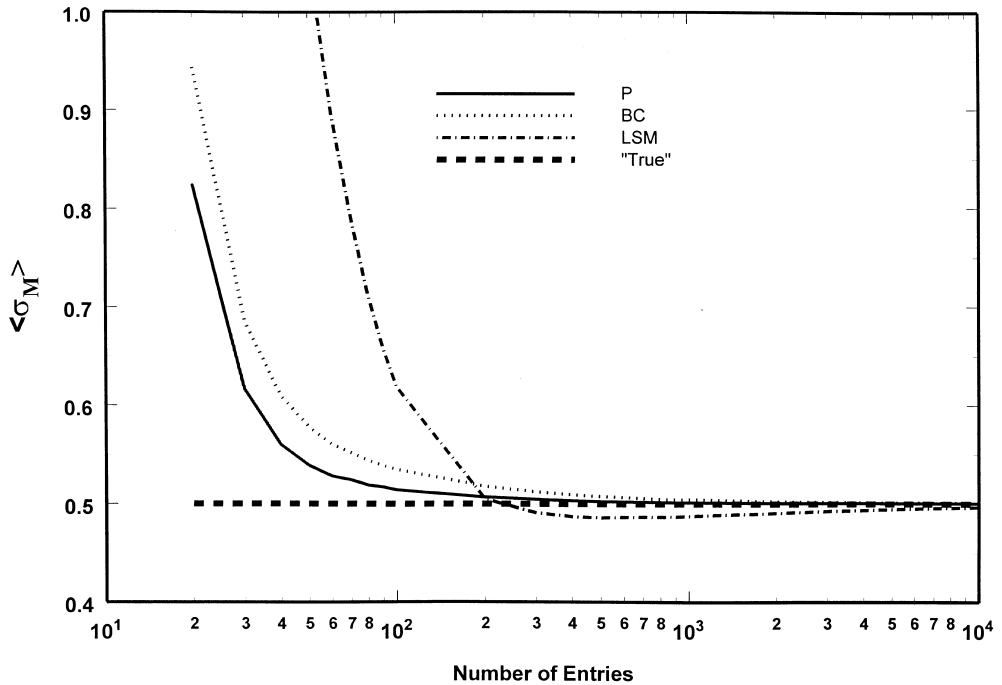

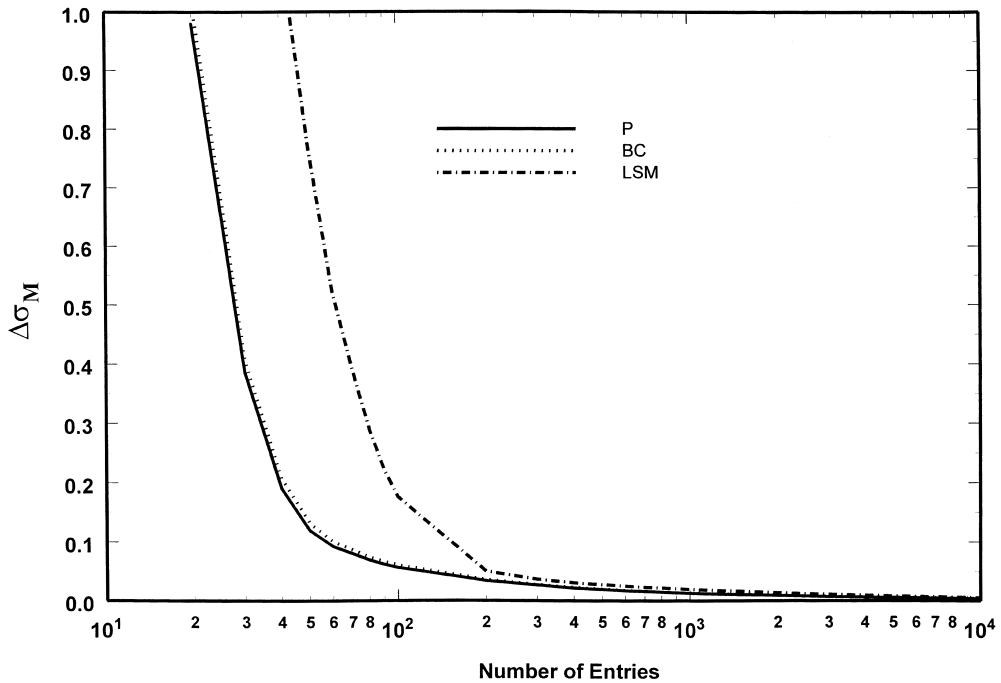Fig. 12. Same as Fig. 10 but for the parameter σ_M.

Fig. 13. Same as Fig. 11 but for the parameter σ_M.

technique for fitting functions in order to extend its applicability to low statistics data when one has asymmetrical errors. This new method is obtained through the change of variables such that one gets approximately Gaussian pdf when transforming the original pdf. The approximate Gaussian pdf is obtained associated to a Poisson pdf and a chi-square function is adapted for this new Gaussian like pdf. Monte Carlo generated events show an improvement in low statistics region, as expected, although the results converge to the "true" value as the number of events increases. This method has shown a fast convergence to the correct parameter values especially to the parameter associated to the curve spread, σ_j. The proposed chi-square function is consistent since the fitted parameters converge to the true value of the parameters and their expected mean errors decreases as the number of observations increases. The results are simple and of easy use in standard optimization procedures. A similar result was obtained in the appendix for binomial distributions.

Appendix. Chi-square function for binomial distribution

Let us now derive a χ_B^2 for the case when each bin content obeys a binomial distribution, this is the usual case when dealing with uncorrelated normalized data. The binomial distribution $p_B(x)$, normalized with respect to x, can be written as

$$p_B(x) = \frac{\Gamma(n+2)x^k(1-x)^{n-k}}{\Gamma(k+1)\Gamma(n-k+1)} \tag{A.1}$$

where $x \in [0,1]$, then

$$J(\hat{x}) = \left(-\frac{\partial^2 \log(p_B(x))}{\partial x^2} \right)_{\hat{x}}$$

$$= \frac{1}{\hat{x}(1 - \hat{x})}. \tag{A.2}$$

Supposing again that $z(x)$ is a one-to-one transformation, then

$$J(\hat{z}) = -\left(\frac{\partial^2 \log(p_B(x))}{\partial x^2} \right)_{\hat{x}} \left| \frac{dx}{dz} \right|^2_{\hat{x}}. \tag{A.3}$$

Let us choose dx/dz such that

$$\frac{dx}{dz} = J^{-1/2}(x)$$

$$= \sqrt{x(1 - x)} \tag{A.4}$$

then inverting the above equation one gets

$$\frac{dz}{dx} = J^{1/2}(x). \tag{A.5}$$

Integrating the latter expression

$$z(x) = \int^x \frac{1}{\sqrt{t(1 - t)}} \, dt$$

$$= \arcsin(2x - 1). \tag{A.6}$$

Then the approximately Gaussian pdf associated to the binomial pdf above is

$$q_B(z) = \frac{\Gamma(n + 2)}{2^{n+1}\Gamma(k + 1)\Gamma(n - k + 1)}$$

$$\times (1 + \sin(z))^k (1 - \sin(z))^{(n-k)} \cos(z) \tag{A.7}$$

where $z \in [-\pi/2, \pi/2]$. This function has a variance $\sigma^2 = 1/(n + 1)$ which is independent of k and its

maximum is located at

$$\hat{z} = \arcsin\left(\frac{2k - n}{n + 1} \right)$$

$$= \arcsin\left(\frac{2k/n - 1}{1 + 1/n} \right). \tag{A.8}$$

The approximate 68.27% confidence limits are given by the analytical expression

$$\arcsin\left(\frac{2k/n - 1}{1 + 1/n} \right) \mp \frac{1}{\sqrt{n + 1}} \tag{A.9}$$

and after using the likelihood ratio test theorem, one obtains the chi-square function for a binomial pdf

$$\chi_B^2 = \sum_{i=1}^{N} \left[(2n_i + 1)\log\left(\frac{2n_i + 1}{2nf(x_i, \alpha) + 1} \right) \right. \tag{A.10}$$

$$\left. + (2(n - n_i) + 1)\log\left(\frac{2(n - n_i) + 1}{2n(1 - f(x_i, \alpha)) + 1} \right) \right] \tag{A.11}$$

where $f(x_i, \alpha)$ is the fitted function with estimated parameters α.

References

[1] S. Brandt, Statistical and Computational Methods in Data Analysis, North-Holland, Amsterdam, 1970.
[2] W.T. Eadie, D. Drijard, F.E. James, M. Ross, B. Sadoulet, Statistical Methods in Experimental Physics, North-Holland, Amsterdam, 1971.
[3] S. Baker, R.D. Cousins, Nucl. Instr. and Meth. 221 (1984) 437.
[4] G.E.P. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, Wiley, New York, 1992.
[5] D.G. Papageorgiou, I.N. Demetropoulos, I.E. Lagaris, Comput. Phys. Commun. 109 (1998) 227.