Nuclear Instruments and Methods in Physics Research 221 (1984) 437–442
North-Holland, Amsterdam

CLARIFICATION OF THE USE OF CHI-SQUARE AND LIKELIHOOD FUNCTIONS IN FITS TO HISTOGRAMS

Steve BAKER and Robert D. COUSINS

Department of Physics, University of California, Los Angeles, California 90024, USA

Received 18 July 1983

We consider the problem of fitting curves to histograms in which the data obey multinomial or Poisson statistics. Techniques commonly used by physicists are examined in light of standard results found in the statistics literature. We review the relationship between multinomial and Poisson distributions, and clarify a sufficient condition for equality of the area under the fitted curve and the number of events on the histogram. Following the statisticians, we use the likelihood ratio test to construct a general χ^2 statistic, χ^2_λ, which yields parameter and error estimates identical to those of the method of maximum likelihood. The χ^2_λ statistic is further useful for testing goodness-of-fit since the value of its minimum asymptotically obeys a classical chi-square distribution. One should be aware, however, of the potential for statistical bias, especially when the number of events is small.

1. Introduction

Standard results from the theory of statistics are often overlooked by scientists searching for sophisticated methods to analyze their data. Though good high-level statistics books have been written expressly for experimentalists [1], we believe that some misunderstanding remains in the professional physics literature. A case in point is the fitting of curves to histograms in which the data are distributed according to multinomial or Poisson statistics [2,3]. It is well-known, for example, that common methods of chi-square minimization suffer certain difficulties (such as the under- or over-estimation of the area under a peak) which can be traced to the implicit assumption of a Gaussian distribution of the errors. Hence, the method of maximum likelihood is often employed, explicitly incorporating the appropriate distribution from the start. Some authors [4] nonetheless revert to a common (Gaussian derived) chi-square test for goodness-of-fit. Certainly one should question the straightforward use of a statistic as a test for goodness-of-fit when it is known to be based on an inappropriate parent distribution. Our examination of the professional statistics literature during the course of a graduate seminar on data analysis underlined the need for clarification of these and other problems in the physics literature.

To begin with, we must carefully define some nomenclature which is unambiguous and consistent with usage in the statistical literature. Curve-fitting typically involves three tasks: (a) determining the "best fit" parameters of a curve, (b) determining the errors on the parameters, and (c) judging the goodness of the fit. In the language of the statisticians, these are known as (a) point estimation, (b) confidence interval estimation, and (c) goodness-of-fit testing. Because chi-square statistics are sometimes used for all three tasks, the distinction among them can become blurred. However, it is important to maintain this distinction. In general, one need not use the same statistic for all three purposes. Indeed, some of the most powerful tests of goodness-of-fit have little practical utility for point estimation [1]. Thus, if one uses the method of maximum likelihood for point and interval estimation, the choice of goodness-of-fit test(s) remains.

Since statisticians typically consider multinomial problems, we review the connection between multinomial and Poisson statistics. This connection is related to the conditions under which the maximum likelihood fit preserves the number of events under the curve. Next we highlight some of the historical controversy in the statistics literature over the assumed superiority of the principle of maximum likelihood. We then review how the likelihood ratio test for goodness-of-fit gives a prescription for constructing a general chi-square statistic directly from the likelihood function. This general chi-square statistic can then be used for point estimation, confidence interval estimation, and goodness-of-fit testing. Finally, we mention the important problem of biased versus unbiased estimation in the case of a finite sample size.

2. Definitions and notation

We consider a histogram (one- or multi-dimensional) having k bins labelled by the index i running from 1 to

0167-5087/84/$03.00 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

k. We let
n_i = the number of events in the ith bin,
$\boldsymbol{n}$ = $(n_1, n_2, \ldots, n_k)$,
N = total number of events $= \Sigma_i\, n_i$.
Our task is to fit to the data a theoretical curve with J parameters labelled with the index j. We let
$\boldsymbol{\alpha}$ = $(\alpha_1, \alpha_2, \ldots, \alpha_j, \ldots, \alpha_J)$ = the set of parameters,
y_i = number of events predicted by the model to be in the ith bin,
$\boldsymbol{y}$ = $(y_1, y_2, \ldots, y_k)$,
N_0 = $\Sigma_i\, y_i$ = total number of events predicted by the model.
Note that $\boldsymbol{y}$ is a function of $\boldsymbol{\alpha}$. If the model gives a continuous probability distribution f instead of discrete bin contents, then rigorously we should compute

$$y_i = \int f, \tag{1}$$

where the integral is over the ith bin. In this case, N_0 is truly equal to the area under the fitted curve. In practice, y_i is often approximated by the product of the width of the ith bin and the value of f somewhere in bin i, so that N_0 is only approximately the area under the fitted curve.

At this point, we must distinguish between two different cases depending on the nature of the experimental data. If the total number of events N is *fixed* by the design of the experiment, then the distribution of the events among the bins is rigorously *multinomial*. On the other hand, if a counting experiment is designed to *measure* N, then independent *Poisson* statistics apply [5]. In more subtle in-between cases, one must decide which statistics are more applicable. Fortunately, we report below that in a large class of fitting problems, the multinomial and Poisson approaches give identical estimates of the parameters $\boldsymbol{\alpha}$.

Numerous statistics can be constructed from the quantities defined above. We let
S = weighted least-squares statistic.
= $\Sigma_i\, w_i(n_i - y_i)^2$ (neglecting correlations).
Depending on what one uses for the weights w_i, this often takes on one of the classical chi-square forms:

$$\chi^2_{\mathrm{P}} = \text{Pearson's } \chi^2 = \sum_i (n_i - y_i)^2 / y_i,$$

$$\chi^2_{\mathrm{N}} = \text{Neyman's } \chi^2 = \sum_i (n_i - y_i)^2 / n_i.$$

In multinomial problems, these definitions must always be supplemented by the constraint on $\boldsymbol{y}$ [6]:

$$N = N_0, \quad \text{i.e.,} \sum_i y_i = \sum_i n_i.$$

Part of the confusion about χ^2 statistics in the physics literature is due to a free use of the symbol χ^2 and various forms laballed "modified χ^2". When statisticians simply say "χ^2", they usually mean Pearson's χ^2. When they refer to a "modified χ^2", they usually mean Neyman's χ^2. To avoid any misunderstanding, we shall refer to each statistic by its full name or complete symbol.

It is important to note that these are not the only possible χ^2 statistics that one can construct. A general definition of a "chi-square statistic" is *any* function of $\boldsymbol{n}$ and $\boldsymbol{y}$ that is asymptotically distributed in the classical chi-square distribution [7]. This definition does not always include the weighted least-squares statistic S when the weights are chosen arbitrarily. For this reason, we have avoided the common practice of calling it "χ^2".

As is well-known, the principle of maximum likelihood may be used as a starting point to "derive" the statistic S with weights $w_i = 1/\sigma_i^2$ when the parent distributions are Gaussian with variances σ_i^2. One may equally well define statistics using the principle of maximum likelihood *directly* for Poisson- and multinomial-distributed histograms. The two relevant likelihood functions are:

$$L_{\mathrm{p}}(\boldsymbol{y};\boldsymbol{n}) = \text{likelihood function for Poisson histograms}$$
$$= \prod_i \exp(-y_i)\, y_i^{n_i}/n_i! \tag{2}$$

$$L_{\mathrm{m}}(\boldsymbol{y};\boldsymbol{n}) = \text{likelihood function for multinomial histograms}$$
$$= N! \prod_i p_i^{n_i}/n_i!$$
$$= N! N^N \prod_i y_i^{n_i}/n_i!, \tag{3}$$

where $p_i = n_i/N$ in the multinomial case. It is easy to verify that L_{p} and L_{m} are related by

$$L_{\mathrm{p}}(\boldsymbol{y};\boldsymbol{n}) = P(N) \times L_{\mathrm{m}}(\boldsymbol{y};\boldsymbol{n}),$$

where

$$P(N) = \exp(-N_0)\, N_0^N / N!,$$
$$N_0 = \sum y_i.$$

This is equivalent to the statement that the independent Poisson probability of observing a particular $\boldsymbol{n}$ is the product of the Poisson probability of observing the total number of events N and the multinomial probability of observing $\boldsymbol{n}$, *given* N [8].

One can make use of the powerful theorem on the likelihood ratio test for goodness-of-fit, found in elementary statistics textbooks [9], to construct another statistic. This theorem, perhaps not as widely appreciated by physicists as it might be, enables one to convert the likelihood function into the form of a general χ^2 statistic. We let $\boldsymbol{m}$ be the true (unknown) values of $\boldsymbol{n}$ that one would get if there were no errors. Then one forms the likelihood ratio λ defined by

$$\lambda = L(\boldsymbol{y};\boldsymbol{n}) / L(\boldsymbol{m};\boldsymbol{n}).$$

The likelihood ratio test theorem says that the "likelihood χ^2", defined by

$$\chi^2_\lambda = -2\ln\lambda = -2\ln L(\boldsymbol{y};\boldsymbol{n}) + 2\ln L(\boldsymbol{m};\boldsymbol{n}),$$

asymptotically obeys a chi-square distribution. One notes that the second term is independent of $\boldsymbol{y}$, so that minimization of χ^2_λ is entirely equivalent to maximization of the likelihood function L. The χ^2_λ statistic may thus be useful for both estimation *and* goodness-of-fit testing.

For the Poisson- and multinomial-distributed histograms, we may replace the unknown $\boldsymbol{m}$ by its bin-by-bin model-indpendent maximum likelihood estimation which is just $\boldsymbol{n}$ in both cases [10]. With the details given in the appendix, this leads to the Poisson likelihood chi-square,

$$\chi^2_{\lambda,\mathrm{p}} = 2\sum_i y_i - n_i + n_i \ln(n_i/y_i),$$

and the multinomial likelihood chi-square,

$$\chi^2_{\lambda,\mathrm{m}} = 2\sum_i n_i \ln(n_i/y_i).$$

3. Historical background

With all these possible test statistics available, one has naturally asked which is "best" in some sense, and perhaps, which is more "fundamental"? Interestingly, these "classical" statistics problems continue to arouse some controversy. In the 1981 opinion of Efron [11], the superiority of maximum likelihood as a device for *summarizing* data with a probability density function has never been seriously challenged. However, he distinguishes this from maximum likelihood *estimation* of parameters, where it is less well founded.

Before one is tempted to try to choose a "best" statistic, it is instructive to look at the historical developments in the theory of statistics. The maximum likelihood method was in fact considered by Gauss, and studied extensively by Fisher and others [12]. We have already noted its familiar use to "derive" some chi-square statistics.

The chi-square statistics can, however, stand on their own merits. Statisticians have considered them primarily in the multinomial context, including Pearson's classic studies of χ^2_P [13]. In a landmark 1949 study of the multinomial problem, Neyman [14] studied the merits of χ^2_P, χ^2_N, and L_m as point estimators in the asymptotic (large N) region. He found that all three estimators had a set of properties which he considered optimal. He called them "best asymptotically normal" (BAN) estimators.

In 1962, Rao [15] introduced the concept of "second order efficiency" to try to describe the speed (as N increases) with which the various estimators approach their asymptotic properties. In a paper greeted enthusiastically by many statisticians, Rao found that L_m was "best" according to his criteria, followed by χ^2_P, with χ^2_N being the worst of the three.

As emphasized by Neyman and Rao, the class of BAN estimators is quite large. The conjecture that maximum likelihood estimators are preferable to all other BAN estimators has not been proven generally. Indeed, there are plenty of examples in the literature where the likelihood method performs poorly. For a recent example where the likelihood function has bad problems, with many references to earlier examples, see the paper by Ferguson [16].

A small but persistent minority (most visibly Berkson [7]) has repeatedly argued in favor of a general χ^2 statistic as the most "fundamental" statistic. They note that one can, for example, consider L_m to be "derived" from $\chi^2_{\lambda,\mathrm{m}}$. Thus, the question of which is the more fundamental statistic is, on one level, a semantic question, since various χ^2 estimators are completely equivalent to corresponding likelihood estimators, and vice versa. On a deeper level, however, one would like to know which of different estimators to use, particularly since Berkson can give no general prescription, except to choose the easiest one to compute. This criterion of ease of computation, suggested in many contexts by various classical statisticians, seems somewhat artificial and rather obsolete for today's physicist armed with high-speed computers. For a lively 1980 paper presentation of Berkson, followed by discussion from eminent statisticians, see ref. [7].

4. Preservation of the area

While statisticians continue to debate the fine points of estimation theory, physicists continue to accumulate vast practial experience in fitting histograms. One topic which has attracted great interest is discovering which statistics, when minimized, preserved the number of events:

$$N_0 \equiv \sum y_i = \sum n_i \equiv N. \tag{4}$$

Often this condition is described as "preserving the area". This nomenclature is rigorously valid when $\boldsymbol{y}$ is actually computed from eq. (1), but is a slight misnomer when one uses approximate values of $\boldsymbol{y}$ in constructing the test statistic. Of course, in multinomial problems, the area is preserved for *any* statistic minimized, since the constraint $N_0 = N$ is imposed. For Poisson-distributed data, it has long been observed that in many cases maximum likelihood fits using L_p preserves the area, while chi-square fits using χ^2_P and χ^2_N over-estimate and under-estimate the area, respectively. One can clarify a sufficient condition for a Poisson maximum likelihood fit to preserve the area. We prefer the following approach.

For well-behaved likelihood functions, the extrema can be found by solving the "normal equations" (vanishing first derivatives with respect to each parameter):

$$\partial L_P/\partial \alpha_j = 0 \Rightarrow \partial \ln L_p/\partial \alpha_j = 0; \quad j = 1,\dots,J. \tag{5}$$

Plugging in L_p from eq. (2), we obtain

$$0 = \sum_i (n_i - y_i)(1/y_i)(\partial y_i/\partial \alpha_j); \quad j = 1,\dots,J. \tag{6}$$

Suppose that one can find a new set of J parameters $\boldsymbol{\alpha}' = (\alpha'_1,\dots,\alpha'_J)$ which are functions of the old parameters with non-vanishing Jacobian

$$\partial(\alpha'_1,\dots,\alpha'_J)/\partial(\alpha_1,\dots,\alpha_J) \neq 0,$$

such that the functions $\boldsymbol{y}(\boldsymbol{\alpha})$ can be rewritten as

$$y_i = \alpha'_1 \cdot g_i(\alpha'_2,\alpha'_3,\dots,\alpha'_J), \tag{7}$$

where the g_i are functions which do not depend on α'_1. One can maximize L_p with respect to $\boldsymbol{\alpha}'$:

$$\partial L_P/\partial \alpha'_j = 0 \Rightarrow \partial \ln L_p/\partial \alpha'_j = 0; \quad j = 1,\dots,J. \tag{8}$$

That is,

$$0 = \sum_i (n_i - y_i)(1/y_i)(\partial y_i/\partial \alpha'_j); \quad j = 1,\dots,J. \tag{9}$$

Since the Jacobian is non-zero, the same extrema of L_p are found whether one uses eq. (6) or eq. (9). Using eq. (9) with $j = 1$, we obtain

$$0 = \sum_i (n_i - y_i)(1/\alpha'_1 g_i)(g_i),$$

whence $N = N_0$. Hence, if the above parameter redefinition is possible, the maximization of L_p preserves the area [17]. Of course, one need not actually use the new parameter set when fitting.

The simple interpretation of the above result is that if the parametrization allows complete freedom to adjust the overall scale of the fit, then the area is preserved. This will be true, for example, when the fitting function is a sum of arbitrarily complicated terms, each of which is multiplied by an independently adjustable parameter.

5. Point estimation

Now, suppose that the functions y_i can be transformed to the form given by eq. (7). Then at the extrema of L_p the constraint which distinguishes the multinomial distribution from the Poisson distribution is automatically satisfied. In this case we have the important result that point estimation of the parameters will be the *same* whether one starts with Poisson or multinomial likelihood functions.

This result can be envisioned in two ways. Geometrically, maximizing L_p finds the most likely parameters $\boldsymbol{\alpha}_0$ in the entire J-dimensional parameter space, while maximizing L_m finds the most likely parameters in the $(J-1)$-dimensional subspace constrained by the equation $N = N_0$. If the form (7) applies, then $\boldsymbol{\alpha}_0$ lies in the $(J-1)$-dimensional subspace, so that both methods are able to reach the same point. One arrives at the same conclusion algebraically by noting that $\chi^2_{\lambda,p}$ (equivalent to L_p) and $\chi^2_{\lambda,m}$ (equivalent to L_m) differ only by the term $\Sigma(n_i - y_i)$. If we know that this term vanishes at the extrema of L_p, we can impose this from the beginning, so that $\chi^2_{\lambda,p}$ and $\chi^2_{\lambda,m}$ are identical.

When the point estimates from using L_m and L_p are identical, then we may apply the results of many statistical studies of L_m (which have generally ignored L_p) to our Poisson-distributed histograms. Thus, the question of "Poisson or multinomial?" becomes moot in many common cases of maximum likelihood point estimation.

6. Confidence interval estimation

In the asymptotic region, the likelihood ratio test again allows us to find confidence intervals using χ^2_λ as a general chi-square statistic. Confidence interval estimation is discussed in detail by Eadie et al. [1], and summarized clearly by James [18]. Both give a prescription for tracing out the boundaries of confidence regions by finding contours of constant L_p (or equivalently, $\chi^2_{\lambda,p}$). For finite N, one must in general study the specific problem (often by Monte Carlo methods [19]) in order to obtain the best results.

We note that the question of "Poisson or multinomial?" can be relevant here, since the interval estimates are not identical even if eq. (7) is satisfied.

7. Goodness-of-fit

Traditionally, L_p has been used as a point- and interval-estimator, but not as often formulated as a goodness-of-fit test. For example, one has sometimes used L_p for point estimation, but reverted to the statistic χ^2_P or χ^2_N for goodness-of-fit testing [4]. However, L_p can be used in a goodness-of-fit test simply by using the likelihood ratio test to define χ^2_λ, as was done in section 2. Thus, for those who desire a three-in-one (point estimation, interval estimation, and goodness-of-fit) test statistic, χ^2_λ gives gives estimation results identical to maximum likelihood, and behaves asymptotically in the way that physicists expect a chi-square statistic to behave.

Even though χ^2_λ has been discussed in the statistics literature for many years [20], one still finds papers in the physics literature which re-discover the maximum likelihood method of point estimation and call it "new" [3], while failing to take advantage of the likelihood ratio test of goodness-of-fit.

8. Small sample size: biased vs. unbiased estimation

It would seem desirable to have the true value of a parameter equal to the mean of the distribution of estimates obtained from repeated independent identical experiments. Estimators with this property are called unbiased. Estimators from least squares and maximum likelihood methods are often biased, even though they are consistent (converge asymptotically to the true value). Correction for bias may be possible by direct evaluation of the expectation value of the estimator when the form of the parent distribution is known. However, in most cases the bias can easily be removed only to order $1/N$, and then usually at the expense of increased variance [21]. The method of weighted least-squares always produces unbiased parameter estimates provided the fitting function is linear in the parameters and the weights are independent of the parameters [22].

It is important to remember that the properties of the various χ^2 statistics that make them useful for confidence interval estimation and goodness-of-fit testing are *asymptotic*. Unfortunately, for a finite sample size the probability distribution function of the minimum value of a test χ^2 statistic may deviate from a χ^2 distribution, resulting in biased estimates and tests. Bias in confidence interval estimation means that statements made about expected frequency of occurrence will be misleading; bias in goodness-of-fit testing means that the probability of accepting a bad fit is higher than the probability of accepting a good fit. We would like to avoid both of these situations if possible. Sometimes a change of variable in the parameters can remove the estimation bias, but for the likelihood ratio the bias is invariant under such transformations [23]. The likelihood ratio test can be made unbiased by correcting for the bias in the parameters [24]. In any case, an unequivocal improvement can be made to χ^2_λ by applying a scale factor correction which adjusts the expectation value of χ^2_λ so as to coincide with the expectation value of a χ^2 variate with the same number of degrees of freedom [25]. Probably the safest thing one can do when N is small is to study the probability distribution function of a test statistic by Monte Carlo simulation.

Though it has considerable intuitive appeal, unbiasedness is by no means a necessary condition for a good estimation rule [26]. One can argue that it is more desirable for an estimator to have the smallest possible expected mean-square-error (i.e., the smallest possible variance about the true value). It is quite possible for a biased estimate to have a smaller expected mean-square-error than does an unbiased one [27]. To quote a 1975 article by Efron, ...*in more complicated situations involving the estimation of several parameters at the same time, statisticians have begun to realize that biased estimation rules have definite advantages over the usual unbiased estimators* [28]. This is a current topic of research in the field of statistical estimation. For a further discussion of this and other controversial topics in statistics, see ref. [11] and especially ref. [26] and references therein.

9. Conclusion

Of the many statistics available for fitting curves to Poisson-distributed histograms, the likelihood statistic L_p is popular because under condition (7) it guarantees that $N_0 = N$, i.e., that the area is preserved. Under these conditions, L_p and L_m yield, upon maximization, the same point estimates of the parameters. Alternatively, one may obtain identical results by minimizing the chi-square statistics $\chi^2_{\lambda,\mathrm{p}}$ and $\chi^2_{\lambda,\mathrm{m}}$ in place of maximizing L_p and L_m, respectively. These chi-square statistics are also useful in interval estimation and tests of goodness-of-fit. For finite sample size (small N) general results are lacking; one must carefully study the problem at hand in order to choose and interpret a test statistic.

Appendix

One obtains the form of $\chi^2_{\lambda,\mathrm{p}}$ and $\chi^2_{\lambda,\mathrm{m}}$ from the likelihood ratio test as follows. We estimate the true values $\boldsymbol{m}$ by $\boldsymbol{n}$ [10]. Then

$$\lambda_\mathrm{p} = L_\mathrm{p}(\boldsymbol{y}; \boldsymbol{n})/L_\mathrm{p}(\boldsymbol{n}; \boldsymbol{n}),$$

where $L_\mathrm{p}(\boldsymbol{y}; \boldsymbol{n})$ is given by eq. (2), from which

$$L_\mathrm{p}(\boldsymbol{n}; \boldsymbol{n}) = \prod_i \exp(-n_i) n_i^{n_i}/n_i!.$$

Thus

$$\lambda_\mathrm{p} = \prod_i \exp(-y_i + n_i)(y_i/n_i)^{n_i},$$

from which

$$\chi^2_{\lambda,\mathrm{p}} = -2\ln\lambda_\mathrm{p} = 2\sum_i y_i - n_i + n_i \ln(n_i/y_i).$$

Similarly,

$$\lambda_\mathrm{m} = L_\mathrm{m}(\boldsymbol{y}; \boldsymbol{n})/L_\mathrm{m}(\boldsymbol{n}; \boldsymbol{n}),$$

where $L_\mathrm{m}(\boldsymbol{y}; \boldsymbol{n})$ is given by eq. (3), from which

$$L_\mathrm{m}(\boldsymbol{n}; \boldsymbol{n}) = N!N^N \prod_i n_i^{n_i}/n_i!.$$

Thus [29]

$$\lambda_\mathrm{m} = \prod_i (y_i/n_i)^{n_i},$$

from which

$$\chi^2_{\lambda,\mathrm{m}} = -2\ln\lambda_\mathrm{m} = 2\sum_i n_i \ln(n_i/y_i).$$

References

[1] Our favorite is W.T. Eadie, D. Drijard, F.E. James, M. Roos and B. Sadoulet, Statistical methods of experimental physics (North-Holland/American Elsevier, Amsterdam/New York, 1971).

[2] G.W. Phillips, Nucl. Instr. and Meth. 153 (1978) 449.

[3] T. Awaya, Nucl. Instr. and Meth. 165 (1979) 317.

[4] A recent example is M.E. Nelson et al., Phys. Rev. Lett. 50 (1983) 1542. See also R.C. Eggers and L.P. Somerville, Nucl. Instr. and Meth. 190 (1981) 535.

[5] See Eadie, p. 46, for the precise definitions.

[6] Note that the addition of this constraint removes one degree of freedom from χ^2_P and χ^2_N.

[7] J. Berkson, Ann. Stat. 8 (1980) 457, and references therein.

[8] M.G. Kendall and A. Stuart, The advanced theory of statistics, vol. 2, 4th ed. (Griffin, London, 1979) p. 449. See also R.A. Fisher, Proc. Camb. Phil. Soc. 22 (1925) 700.

[9] For example, P.G. Hoel, Introduction to mathematical statistics, 4th ed. (Wiley, New York, 1971) p. 211. Some original papers are listed in ref. [20] below.

[10] Eadie, p. 256.

[11] B. Efron, Ann. Stat. 10 (1982) 340.

[12] The original papers are reproduced in R.A. Fisher, Contributions to mathematical statistics (Wiley, New York, 1950).

[13] K. Pearson, Phil. Mag. 1 (1900) 157.

[14] J. Neyman, Proc. 1st Berkeley Symp. on Mathematical statistics, ed., J. Neyman (1949) p. 239. See also a review by C. Rao, Bull. Inter. Inst. Stat. 35(2) (1957) 25.

[15] C. Rao, J. Royal Stat. Soc. 24B (1962) 46.

[16] T.S. Ferguson, J. Am. Stat. Assoc. 77 (1982) 831.

[17] Occasionally one reads that the area is *always* preserved if the maximum likelihood method is used [3]. This is not true.

[18] F. James, Comp. Phys. Comm. 20 (1980) 29.

[19] For a recent primer on Monte Carlo methods, see F. James, Rep. Prog. Phys. 43 (1980) 1145.

[20] Some early and later studies include: J. Neyman and E.S. Pearson, Biometrika 20A (1928) 263; S.S. Wilks, Ann. Math. Stat. 6 (1935) 190; S.S. Wilks, Ann. Math. Stat. 9 (1938) 60; R.A. Fisher, Biometrics 6 (1950) 17; J. Berkson, Biometrics 28 (1972) 443.

[21] Kendall, p. 5; Eadie, p. 181.

[22] Eadie, p. 163.

[23] Eadie, p. 212.

[24] Kendall, p. 252.

[25] Eadie, pp. 212, 236; Kendall, p. 250.

[26] B. Efron, Am. Math. Month. 85 (1978) 231. See p. 234.

[27] See Stein's phenomenon in ref. [26], p. 243.

[28] B. Efron, Adv. Math. 16 (1975) 259. See p. 260.

[29] This is the same as eq. (12) in the 1928 article by Neyman and Pearson cited above [20].