/34

NATIONAL
OPTICAL
ASTRONOMY
OBSERVATORIES

Preprint Series

NOAO Preprint No. 835

PARAMETER ESTIMATION IN ASTRONOMY WITH POISSON-DISTIBUTED DATA. I. THE χ^2_γ STATISTIC

KENNETH J. MIGHELL

To Appear In: The Astrophysical Journal

March 1999

To appear in the Astrophysical Journal (accepted 1998 November 20)

PARAMETER ESTIMATION IN ASTRONOMY WITH POISSON-DISTRIBUTED DATA. I. THE χ^2_γ STATISTIC

KENNETH J. MIGHELL
Kitt Peak National Observatory, National Optical Astronomy Observatories[1],
P. O. Box 26732, Tucson, AZ 85726-6732
Electronic mail: mighell@noao.edu

ABSTRACT

Applying the standard weighted mean formula, $\left[\sum_i n_i \sigma_i^{-2}\right] / \left[\sum_i \sigma_i^{-2}\right]$, to determine the weighted mean of data, n_i, drawn from a Poisson distribution, will, on average, underestimate the true mean by $\sim$1 for all true mean values larger than $\sim$3 when the common assumption is made that the error of the ith observation is $\sigma_i = \max(\sqrt{n_i}, 1)$. This small, but statistically significant offset, explains the long-known observation that chi-square minimization techniques which use the modified Neyman's χ^2 statistic, $\chi^2_{\rm N} \equiv \sum_i (n_i - y_i)^2 / \max(n_i, 1)$, to compare Poisson-distributed data with model values, y_i, will typically predict a total number of counts that underestimates the true total by about 1 count per bin. Based on my finding that the weighted mean of data drawn from a Poisson distribution can be determined using the formula $\left[\sum_i \left[n_i + \min(n_i, 1)\right](n_i + 1)^{-1}\right] / \left[\sum_i (n_i + 1)^{-1}\right]$, I propose that a new χ^2 statistic, $\chi^2_\gamma \equiv \sum_i \left[n_i + \min(n_i, 1) - y_i\right]^2 / \left[n_i + 1\right]$, should always be used to analyze Poisson-distributed data in preference to the modified Neyman's χ^2 statistic. I demonstrate the power and usefulness of χ^2_γ minimization by using two statistical fitting techniques and five χ^2 statistics to analyze simulated X-ray power-law 15-channel spectra with large and small counts per bin. I show that χ^2_γ minimization with the Levenberg-Marquardt or Powell's method can produce excellent results (mean slope errors $\lesssim$3%) with spectra having as few as 25 total counts.

Subject headings: methods: numerical — methods: statistical — X-rays: general

[1]NOAO is operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

1. INTRODUCTION

The determination of the weighted mean is the fundamental problem for chi-square (χ^2) minimization methods. The goodness-of-fit between an observation of N data values, x_i, with errors, σ_i, and a model, m_i, can be determined by using the standard chi-square statistic:

$$\chi^2 \equiv \sum_{i=1}^{N} \left[\frac{x_i - m_i}{\sigma_i} \right]^2 . \tag{1}$$

The theory of least-squares states that the optimum value of all the parameters of the model are obtained when the chi-square statistic is minimized with respect to each parameter simultaneously. For example, the standard formula of the weighted mean can be derived by assuming that the model is a constant and then solving the equation,

$$\frac{\partial}{\partial \mu_{\mathrm{w}}} \sum_{i=1}^{N} \left[\frac{x_i - \mu_{\mathrm{w}}}{\sigma_i} \right]^2 = 0 \, , \tag{2}$$

for that constant:

$$\mu_{\mathrm{w}} \equiv \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \, . \tag{3}$$

The standard weighted-mean formula thus weights every data value, x_i, inversely by its own variance (i.e. σ_i^2).

Let us assume that all the data values come from a pure counting experiment where each data value, n_i, is a random integer deviate drawn from a Poisson (1837) distribution,

$$P(k;\mu) \equiv \frac{\mu^k}{k!} e^{-\mu} , \tag{4}$$

with a mean value of μ. Let us also make the common assumption that the error of each data value is the square root of the mean of the parent Poisson distribution. Using these transformations, $x_i \Rightarrow n_i$ and $\sigma_i \Rightarrow \sqrt{\mu}$, we see that Equation (3) becomes

$$\mu_{\mathrm{P}} \equiv \frac{\sum_{i=1}^{N} \frac{n_i}{\mu}}{\sum_{i=1}^{N} \frac{1}{\mu}} \, , \tag{5}$$

which reduces to become the definition of the sample mean:

$$\mu_{\mathrm{P}} \equiv \frac{1}{N} \sum_{i=1}^{N} n_i \, . \tag{6}$$

In the limit of a large number of observations of the Poisson distribution $P(k;\mu)$, we find that Equation (6) will, on average, determine the mean of the parent Poisson distribution for all true mean values μ:

$$
\begin{aligned}
\lim_{N\to\infty}[\mu_{\rm P}] &\equiv \lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} n_i\right] \\
&\approx \lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=0}^{\infty} k\left\{NP(k;\mu)\right\}\right] \\
&= \sum_{k=0}^{\infty} k\left\{P(k;\mu)\right\} \\
&= \sum_{k=0}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} \\
&= 0\frac{\mu^0}{0!}e^{-\mu} + e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^k}{(k-1)!} \\
&= e^{-\mu}\mu\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!} \\
&= e^{-\mu}\mu\sum_{j=0}^{\infty}\frac{\mu^j}{j!} \\
&= e^{-\mu}\mu e^{\mu} \\
&= \mu\ .
\end{aligned}
\tag{7}
$$

Applying the standard weighted mean formula, $\left[\sum_i n_i\sigma_i^{-2}\right] / \left[\sum_i \sigma_i^{-2}\right]$, *to determine the weighted mean of data,* n_i, *drawn from a Poisson distribution, will, on average, determine the mean of the parent Poisson distribution for all true mean values if a constant weight is assigned to all data values* (i.e. $\sigma^{-2} \equiv$ constant).

It is a common practice to assume that the error of a Poisson deviate n is $\sigma \equiv \sqrt{n}$. Unfortunately, this practice causes the standard weighted-mean formula to be undefined for data values of zero. A simple solution to this computational problem is to arbitrarily assign a non-zero constant error to all Poisson deviates with a value of zero. Let us make the common assumption that the error of each data value, n_i, is equal to $\sqrt{n_i}$ or 1 — whichever is greater. Using the following transformations, $x_i \Rightarrow n_i$ and $\sigma_i \Rightarrow \max(\sqrt{n_i}, 1)$, we see that Equation (3) becomes

$$
\mu_{\rm N} \equiv \frac{\displaystyle\sum_{i=1}^{N}\frac{n_i}{\max(n_i,1)}}{\displaystyle\sum_{i=1}^{N}\frac{1}{\max(n_i,1)}}\ . \tag{8}
$$

In the limit of a large number of observations of the Poisson distribution $P(k;\mu)$, we find that

$$
\begin{aligned}
\lim_{N\to\infty} [\mu_{\rm N}] \;&\equiv\; \lim_{N\to\infty} \left[\frac{\displaystyle\sum_{i=1}^{N} \frac{n_i}{\max(n_i,1)}}{\displaystyle\sum_{i=1}^{N} \frac{1}{\max(n_i,1)}} \right] \\
&\approx\; \lim_{N\to\infty} \left[\frac{\displaystyle\sum_{k=0}^{\infty} \frac{k}{\max(k,1)} \Big\{ N P(k;\mu) \Big\}}{\displaystyle\sum_{k=0}^{\infty} \frac{1}{\max(k,1)} \Big\{ N P(k;\mu) \Big\}} \right] \\
&=\; \frac{\displaystyle\sum_{k=0}^{\infty} \frac{k}{\max(k,1)} \Big\{ P(k;\mu) \Big\}}{\displaystyle\sum_{k=0}^{\infty} \frac{1}{\max(k,1)} \Big\{ P(k;\mu) \Big\}} \\
&=\; \frac{\displaystyle \frac{0}{1} P(0;\mu) + \sum_{k=1}^{\infty} \frac{k}{k} P(k;\mu)}{\displaystyle \frac{1}{1} P(0;\mu) + \sum_{k=1}^{\infty} \frac{1}{k} P(k;\mu)} \\
&=\; \frac{1 - e^{-\mu}}{\displaystyle e^{-\mu} + e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{k\,k!}} \\
&=\; \frac{e^{\mu} - 1}{\displaystyle 1 + \left[\sum_{k=1}^{\infty} \frac{\mu^k}{k\,k!} \right]} \qquad (9) \\
&=\; \frac{e^{\mu} - 1}{1 + \Big[\mathrm{Ei}(\mu) - \gamma - \ln(\mu) \Big]} \;, \qquad (10)
\end{aligned}
$$

where $\mathrm{Ei}(x)$ is the exponential integral of x, $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,dt = \int_{-\infty}^{x} \frac{e^{-t}}{t}\,dt$ (for $x > 0$), and γ is the Euler-Mascheroni constant: $\gamma \equiv \lim_{n\to\infty} \Big[\Big\{ \sum_{i=1}^{n} \frac{1}{n} \Big\} - \ln(n) \Big] = 0.5772156649\cdots$ (see, e.g., Abramowitz & Stegun 1964).

Let us now investigate the limit of Equation (10) with large Poisson mean values. The transformation of Equation (9) to Equation (10) used the power series of $\mathrm{Ei}(x)$,

$$
\mathrm{Ei}(x) = \gamma + \ln(x) + \frac{x}{1\cdot 1!} + \frac{x^2}{2\cdot 2!} + \frac{x^3}{3\cdot 3!} + \cdots \;, \qquad (11)
$$

which has the following asymptotic expansion:

$$
\mathrm{Ei}(x) \approx \frac{e^x}{x} \left(1 + \frac{1!}{x} + \frac{2!}{x^2} + \frac{3!}{x^3} + \cdots \right) \;. \qquad (12)
$$

From the following limit,

$$\lim_{x\to\infty}\left[x\left(1+\frac{1!}{x}+\frac{2!}{x^2}+\frac{3!}{x^3}+\cdots\right)^{-1}-x\right]=-1\ , \tag{13}$$

we see that $\mathrm{Ei}(x)$ asymptotically approaches the function $e^x/(x-1)$ for large values of x. For $x \geq 13$ this approximation has an error of $<1\%$; for $x \geq 33$ the error is $\leq 0.1\%$. In the limit of large mean Poisson values, we see that the numerator of Equation (10) is dominated by the e^μ term while the denominator is dominated by the $\mathrm{Ei}(\mu)$ term which asymptotically approaches the value of $e^\mu/(\mu-1)$. *We then have come to the surprising conclusion that for Poisson distributions with large mean values,* $\lim_{N\to\infty}[\mu_{\rm N}]$ *approaches the value of* $\mu - 1$ *instead of the expected value of* μ.

Equation (10) can also be investigated graphically. Figure 1a plots the difference between the weighted mean computed using Equation (8) and the true mean for Poisson-distributed data with true mean values between 0.001 and 1000. Each open square represents the weighted mean of 4×10^6 Poisson deviates at each given true mean value. The solid curve through the data [open squares in Fig. 1a] is the difference between Equation (10) and the true mean. Note that Equation (10) underestimates the true mean by $\sim$1 for large true mean values (as predicted above).

Applying the standard weighted mean formula, $\left[\sum_i n_i\sigma_i^{-2}\right]/\left[\sum_i \sigma_i^{-2}\right]$, *to determine the weighted mean of data,* n_i, *drawn from a Poisson distribution, will, on average, underestimate the true mean by* $\sim$1 *for all true mean values larger than* $\sim$3 *when the common assumption is made that the error of the ith observation is* $\sigma_i = \max(\sqrt{n_i}, 1)$.

2. THE WEIGHTED MEAN OF POISSON-DISTRIBUTED DATA

We will now develop a weighted-mean formula for Poisson-distributed data that will, on average, determine the true mean of the parent distribution for all true mean values.

Let us assume that the error of each data value, n_i, is equal to $\sqrt{n_i+1}$ instead of $\max(\sqrt{n_i}, 1)$. Using the following transformations, $x_i \Rightarrow n_i$ and $\sigma_i \Rightarrow \sqrt{n_i+1}$, we see that Equation (3) becomes

$$\mu_\alpha \equiv \frac{\displaystyle\sum_{i=1}^{N}\frac{n_i}{n_i+1}}{\displaystyle\sum_{i=1}^{N}\frac{1}{n_i+1}}\ . \tag{14}$$

In the limit of a large number of observations of the Poisson distribution $P(k;\mu)$, we find that

$$\lim_{N\to\infty}[\mu_\alpha] \equiv \lim_{N\to\infty}\left[\frac{\displaystyle\sum_{i=1}^{N}\frac{n_i}{n_i+1}}{\displaystyle\sum_{i=1}^{N}\frac{1}{n_i+1}}\right]$$

$$\approx \lim_{N\to\infty}\left[\frac{\sum\limits_{k=0}^{\infty}\frac{k}{k+1}\left\{NP(k;\mu)\right\}}{\sum\limits_{k=0}^{\infty}\frac{1}{k+1}\left\{NP(k;\mu)\right\}}\right]$$

$$= \frac{\sum\limits_{k=0}^{\infty}\frac{k}{k+1}\left\{P(k;\mu)\right\}}{\sum\limits_{k=0}^{\infty}\frac{1}{k+1}\left\{P(k;\mu)\right\}}$$

$$= \frac{\frac{1}{\mu}\left(\mu - 1 + e^{-\mu}\right)}{\frac{1}{\mu}\left(1 - e^{-\mu}\right)}$$

$$= \frac{\mu}{1-e^{-\mu}} - 1\ . \tag{15}$$

Figure 1b graphically confirms this finding. Increasing the error estimates from $\max(\sqrt{n_i}, 1)$ to $\sqrt{n_i + 1}$ has only yielded a minor improvement. Notice that the dip in the solid curve in Fig. 1a at $\mu \approx 6$ is not present in the solid curve in Fig. 1b. A more radical change appears to be required in order for us to develop a weighted-mean formula for Poisson-distributed data.

Let us now add one to all data values and assume that the error of each data value is the square root of the new data value. Using these transformations, $x_i \Rightarrow n_i + 1$ and $\sigma_i \Rightarrow \sqrt{n_i + 1}$, we see that Equation (3) becomes

$$\mu_\beta \equiv \frac{\sum\limits_{i=1}^{N}\frac{n_i+1}{n_i+1}}{\sum\limits_{i=1}^{N}\frac{1}{n_i+1}}\ . \tag{16}$$

In the limit of a large number of observations of the Poisson distribution $P(k;\mu)$, we find that

$$\lim_{N\to\infty}[\mu_\beta] \equiv \lim_{N\to\infty}\left[\frac{\sum\limits_{i=1}^{N}\frac{n_i+1}{n_i+1}}{\sum\limits_{i=1}^{N}\frac{1}{n_i+1}}\right]$$

$$\approx \lim_{N\to\infty}\left[\frac{\sum\limits_{k=0}^{\infty}\frac{k+1}{k+1}\left\{NP(k;\mu)\right\}}{\sum\limits_{k=0}^{\infty}\frac{1}{k+1}\left\{NP(k;\mu)\right\}}\right]$$

$$
\begin{aligned}
&= \frac{\sum_{k=0}^{\infty} \left\{ P(k;\mu) \right\}}{\sum_{k=0}^{\infty} \frac{1}{k+1} \left\{ P(k;\mu) \right\}} \\
&= \frac{1}{\frac{1}{\mu}\left(1 - e^{-\mu}\right)} \\
&= \frac{\mu}{1 - e^{-\mu}} \, .
\end{aligned} \tag{17}
$$

Figure 1c graphically confirms this finding. We have now made significant progress towards our goal of developing a weighted-mean formula for Poisson-distributed data. Applying Equation (16) to determine the weighted mean of Poisson-distributed data, will, on average, estimate the true mean with $\lesssim 1\%$ errors for true Poisson mean values $\mu \gtrsim 5$.

The deviation of the solid curve in Figure 1c from zero can be eliminated by making just a minor change to our transformations. Using the same errors as above, $\sigma_i \Rightarrow \sqrt{n_i + 1}$, but now adding one to only those data values that are initially greater than zero, $n_i \Rightarrow n_i + \min(n_i, 1)$, we see that Equation (3) becomes

$$
\mu_\gamma \equiv \frac{\sum_{i=1}^{N} \frac{n_i + \min(n_i, 1)}{n_i + 1}}{\sum_{i=1}^{N} \frac{1}{n_i + 1}} \, . \tag{18}
$$

In the limit of a large number of observations of the Poisson distribution $P(k;\mu)$, we find that

$$
\begin{aligned}
\lim_{N\to\infty} \left[\mu_\gamma\right] &\equiv \lim_{N\to\infty} \left[\frac{\sum_{i=1}^{N} \frac{n_i + \min(n_i, 1)}{n_i + 1}}{\sum_{i=1}^{N} \frac{1}{n_i + 1}} \right] \\
&\approx \lim_{N\to\infty} \left[\frac{\sum_{k=0}^{\infty} \frac{k + \min(k, 1)}{k+1} \left\{ N P(k;\mu) \right\}}{\sum_{k=0}^{\infty} \frac{1}{k+1} \left\{ N P(k;\mu) \right\}} \right] \\
&= \frac{\sum_{k=0}^{\infty} \frac{k + \min(k, 1)}{k+1} \left\{ P(k;\mu) \right\}}{\sum_{k=0}^{\infty} \frac{1}{k+1} \left\{ P(k;\mu) \right\}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\frac{0}{1}P(0;\mu) + \sum_{k=1}^{\infty} \frac{k+1}{k+1} P(k;\mu)}{\sum_{k=0}^{\infty} \frac{1}{k+1} P(k;\mu)} \\
&= \frac{1 - e^{-\mu}}{\frac{1}{\mu}\left(1 - e^{-\mu}\right)} \\
&= \mu \ . \qquad (19)
\end{aligned}
$$

Figure 1d graphically confirms this finding. We have now achieved our goal of developing a weighted-mean formula for Poisson-distributed data. *Applying Equation (18) to determine the weighted mean of Poisson-distributed data, will, on average, estimate the true mean for all true Poisson mean values* $(\mu \geq 0)$.

3. THE χ^2_γ STATISTIC

Based on my finding that the weighted mean of data drawn from a Poisson distribution can be determined using the formula $\left[\sum_i [n_i + \min(n_i, 1)] (n_i + 1)^{-1}\right] / \left[\sum_i (n_i + 1)^{-1}\right]$, I propose that, given N observations (n_i) and a model (m_i), a new χ^2 statistic,

$$
\chi^2_\gamma \equiv \sum_{i=1}^{N} \frac{[n_i + \min(n_i, 1) - m_i]^2}{n_i + 1} \ , \qquad (20)
$$

should always be used to analyze Poisson-distributed data in preference to the modified Neyman's χ^2 statistic,

$$
\chi^2_{\mathrm{N}} \equiv \sum_{i=1}^{N} \frac{(n_i - m_i)^2}{\max(n_i, 1)} \ , \qquad (21)
$$

because the weighted-mean formula for the modified Neyman's χ^2 statistic [μ_{N}: Equation (8)] systematically underestimates the true mean value of Poisson-distributed data with true mean values $\mu \gtrsim 0.5$ (see Fig. 1a).

For Poisson-distributed data, it has long been observed that, in many cases, chi-square fits using the modified Neyman's χ^2 statistic and the Pearson's χ^2 statistic,

$$
\chi^2_{\mathrm{P}} \equiv \sum_{i=1}^{N} \frac{(n_i - m_i)^2}{m_i} \ , \qquad (22)
$$

will underestimate and overestimate the total area, respectively, while the usage of the maximum likelihood ratio statistic for Poisson distributions,

$$
\chi^2_\lambda \equiv 2 \sum_{i=1}^{N} \left[m_i - n_i + n_i \ln\left(\frac{n_i}{m_i}\right) \right] \ , \qquad (23)
$$

preserves the total area (e.g., Baker & Cousins 1984 and references therein).

It has been known for decades that chi-square minimization techniques using the modified Neyman's χ^2 statistic to analyze Poisson-distributed data will typically predict a total number of counts (total area) that underestimates the true total counts by about 1 count per bin (e.g., Bevington 1969, Wheaton et al. 1995). The reason why this underestimation occurs is now obvious: the application of the modified Neyman's χ^2 statistic to Poisson-distributed data causes the fitted model value at each bin, m_i, to be, on average, underestimated by $\sim$1 count for all true Poisson model mean values $\gtrsim$3. The underestimation of the true mean by one count gives a very large 20% error when the true mean of the data is 5 but only a 1% error when the true mean of the data is 100. It would clearly be difficult to detect such a small systematic error with *small samples* of Poisson-distributed data with *large true mean values.* Figure 1a shows that this underestimation is real and is easily measurable with *large samples* of Poisson-distributed data.

The number of degrees of freedom, commonly represented with the symbol ν, of a chi-square minimization problem is the difference between the number of observations (sample size) and the number of free parameters (M) of the model: $\nu \equiv N - M$.

The reduced chi-square of the Pearson's χ^2 statistic is, by definition, the value of Pearson's χ^2 statistic divided by the number of degrees of freedom:

$$\frac{\chi^2_{\rm P}}{\nu} \equiv \frac{1}{N-M}\sum_{i=1}^{N}\frac{(n_i - m_i)^2}{m_i} \; . \tag{24}$$

On average, the expected reduced chi-square value of a proper χ^2 statistic with a perfect model is one — given a large number of observations. Now let us assume that our data comes from a Poisson distribution with a mean value of μ. In this case, the model m_i will be a constant, $\mu_{\rm P}$ [Equation (5)], which will, on average, have a value, $\mu_{\rm P'}$, given by Equation (7) in the limit of a large number of observations (N.B. $\mu_{\rm P'} \equiv \mu$). The model is a constant and therefore there is only one degree-of-freedom: $M = 1$. Given these assumptions, we find that, in the limit of a large number of observations, the reduced chi-square of the Pearson's χ^2 statistic with the model μ_P is

$$\begin{aligned}
\lim_{N\to\infty}\left[\frac{\chi^2_{\rm P}}{\nu}\right] &\equiv \lim_{N\to\infty}\left[\frac{1}{N-M}\sum_{i=1}^{N}\frac{(n_i - m_i)^2}{m_i}\right] \\
&= \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{i=1}^{N}\frac{(n_i - \mu_{\rm P})^2}{\mu_{\rm P}}\right] \\
&\approx \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{k=0}^{\infty}\frac{(k - \mu_{\rm P'})^2}{\mu_{\rm P'}}\left\{NP(k;\mu)\right\}\right] \\
&= \sum_{k=0}^{\infty}\frac{(k - \mu_{\rm P'})^2}{\mu_{\rm P'}}\left\{P(k;\mu)\right\} \\
&= \sum_{k=0}^{\infty}\frac{(k - \mu)^2}{\mu}\left\{P(k;\mu)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mu}\left(\left[\sum_{k=0}^{\infty}k^2P(k;\mu)\right] - 2\mu\left[\sum_{k=0}^{\infty}kP(k;\mu)\right] + \mu^2\left[\sum_{k=0}^{\infty}P(k;\mu)\right]\right) \\
&= \frac{1}{\mu}\left(\left[\mu^2+\mu\right] - 2\mu\left[\mu\right] + \mu^2\left[1\right]\right) \\
&= 1\ .
\end{aligned} \tag{25}
$$

The reduced chi-square of the modified Neyman's χ^2 statistic is, by definition, the value of the modified Neyman's χ^2 statistic divided by the number of degrees of freedom:

$$
\frac{\chi^2_{\mathrm{N}}}{\nu} \equiv \frac{1}{N-M}\sum_{i=1}^{N}\frac{(n_i-m_i)^2}{\max(n_i,1)}\ . \tag{26}
$$

Now let us assume that our data comes from a Poisson distribution with a mean value of μ. In this case, the model m_i will be a constant, μ_{N} [Equation (8)], which will, on average, have a value, $\mu_{\mathrm{N}'}$, given by Equation (10) in the limit of a large number of observations. Given these assumptions, we find that, in the limit of a large number of observations, the reduced chi-square of the modified Neyman's χ^2 statistic with the model μ_N is

$$
\begin{aligned}
\lim_{N\to\infty}\left[\frac{\chi^2_{\mathrm{N}}}{\nu}\right] &\equiv \lim_{N\to\infty}\left[\frac{1}{N-M}\sum_{i=1}^{N}\frac{(n_i-m_i)^2}{\max(n_i,1)}\right] \\
&= \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{i=1}^{N}\frac{(n_i-\mu_{\mathrm{N}})^2}{\max(n_i,1)}\right] \\
&\approx \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{k=0}^{\infty}\frac{(k-\mu_{\mathrm{N}'})^2}{\max(k,1)}\left\{NP(k;\mu)\right\}\right] \\
&= \sum_{k=0}^{\infty}\frac{(k-\mu_{\mathrm{N}'})^2}{\max(k,1)}\left\{P(k;\mu)\right\} \\
&= \mu_{\mathrm{N}'}^2e^{-\mu} + \sum_{k=1}^{\infty}\frac{(k-\mu_{\mathrm{N}'})^2}{k}P(k;\mu) \\
&= \mu_{\mathrm{N}'}^2e^{-\mu} + \left[\sum_{k=1}^{\infty}kP(k;\mu)\right] - 2\mu_{\mathrm{N}'}\left[\sum_{k=1}^{\infty}P(k;\mu)\right] + \mu_{\mathrm{N}'}^2\left\{\sum_{k=1}^{\infty}\frac{1}{k}P(k;\mu)\right\} \\
&= \mu_{\mathrm{N}'}^2e^{-\mu} + \left[\mu\right] - 2\mu_{\mathrm{N}'}\left[1-e^{-\mu}\right] + \mu_{\mathrm{N}'}^2\left\{e^{-\mu}\left[\mathrm{Ei}(\mu)-\gamma-\ln(\mu)\right]\right\} \\
&= \mu_{\mathrm{N}'}^2e^{-\mu}\left[1+\mathrm{Ei}(\mu)-\gamma-\ln(\mu)\right] - 2\mu_{\mathrm{N}'}\left[1-e^{-\mu}\right] + \left[\mu\right] \\
&= \mu_{\mathrm{N}'}^2e^{-\mu}\left[\frac{e^{\mu}-1}{\mu_{\mathrm{N}'}}\right] - 2\mu_{\mathrm{N}'}\left[1-e^{-\mu}\right] + \left[\mu\right] \\
&= \{\mu_{\mathrm{N}'}\}\left[e^{-\mu}-1\right] + \mu \\
&= \left\{\frac{e^{\mu}-1}{1+\mathrm{Ei}(\mu)-\gamma-\ln(\mu)}\right\}\left[e^{-\mu}-1\right] + \mu \\
&= \frac{2-e^{\mu}-e^{-\mu}}{\mathrm{Ei}(\mu)-\gamma-\ln(\mu)+1} + \mu\ .
\end{aligned} \tag{27}
$$

In the limit of large mean Poisson values, we see that the numerator of the first term of Equation (27) is dominated by the $-e^{\mu}$ term while the denominator of the first term is dominated by the $\mathrm{Ei}(\mu)$ term which asymptotically approaches the value of $e^{\mu}/(\mu-1)$. We then conclude that the reduced chi-square of the χ^2_{N} statistic applied to a Poisson distribution [Equation (27)] approaches the value of one for large true Poisson mean values. Figure 2 graphically confirms this finding; we see that Equation (27) reaches a value of $\sim$1 only for very large true Poisson mean values ($\mu \gtrsim 100$).

The reduced chi-square of the new χ^2_γ statistic is, by definition, the value of the χ^2_γ statistic divided by the number of degrees of freedom:

$$\frac{\chi^2_\gamma}{\nu} \equiv \frac{1}{N-M}\sum_{i=1}^{N}\frac{[n_i+\min(n_i,1)-m_i]^2}{n_i+1}\ . \tag{28}$$

Now let us assume that our data comes from a Poisson distribution with a mean value of μ. In this case, the model m_i will be a constant, μ_γ [Equation (18)], which will, on average, have a value, $\mu_{\gamma'}$, given by Equation (19) in the limit of a large number of observations (N.B. $\mu_{\gamma'} \equiv \mu$). Given these assumptions, we find that, in the limit of a large number of observations, the reduced chi-square of the new χ^2_γ statistic with the model μ_γ is

$$\begin{aligned}
\lim_{N\to\infty}\left[\frac{\chi^2_\gamma}{\nu}\right] &\equiv \lim_{N\to\infty}\left[\frac{1}{N-M}\sum_{i=1}^{N}\frac{[n_i+\min(n_i,1)-m_i]^2}{n_i+1}\right]\\
&= \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{i=1}^{N}\frac{[n_i+\min(n_i,1)-\mu_\gamma]^2}{n_i+1}\right]\\
&\approx \lim_{N\to\infty}\left[\frac{1}{N-1}\sum_{k=0}^{\infty}\frac{[k+\min(k,1)-\mu_{\gamma'}]^2}{k+1}\Big\{NP(k;\mu)\Big\}\right]\\
&= \sum_{k=0}^{\infty}\frac{[k+\min(k,1)-\mu_{\gamma'}]^2}{k+1}\Big\{P(k;\mu)\Big\}\\
&= \sum_{k=0}^{\infty}\frac{[k+\min(k,1)-\mu]^2}{k+1}\Big\{P(k;\mu)\Big\}\\
&= \mu^2e^{-\mu}+\sum_{k=1}^{\infty}\frac{[k+1-\mu]^2}{k+1}\Big\{P(k;\mu)\Big\}\\
&= \mu^2e^{-\mu}+\left[\sum_{k=1}^{\infty}kP(k;\mu)\right]+(1-2\mu)\left[\sum_{k=1}^{\infty}P(k;\mu)\right]+\mu^2\left[\sum_{k=1}^{\infty}\frac{1}{k+1}P(k;\mu)\right]\\
&= \mu^2e^{-\mu}+\Big[\mu\Big]+(1-2\mu)\Big[1-e^{-\mu}\Big]+\mu^2\left[\frac{1}{\mu}\left(1-e^{-\mu}\right)-e^{-\mu}\right]\\
&= 1+e^{-\mu}\left(\mu-1\right)\ .
\end{aligned} \tag{29}$$

Figure 2 shows that the reduced chi-square of the χ^2_γ statistic applied to a Poisson distribution [Equation (29)] approaches the value of one for small true Poisson mean values (i.e. $\mu \gtrsim 7$).

Figure 3 shows the variance of the reduced chi-square of the χ^2_{P}, χ^2_{N}, χ^2_γ, and χ^2_λ statistics as a function of the true Poisson mean. This figure was derived by analyzing the data used in Figure 1.

4. SIMULATED X-RAY POWER-LAW SPECTRA

I now demonstrate the new χ^2_γ statistic by using it to study a dataset of simulated X-ray power-law spectra. This dataset is based on my duplication of the simple numerical experiment of Nousek & Shue (1989). The number of X-ray photons per energy interval (bin) of a X-ray power-law spectrum is

$$dN = N_0 E^{-\gamma}\, dE \ . \tag{30}$$

Over an energy range, $E_{\min} \le E \le E_{\max}$ keV, the expectation value for the total number of counts can be determined as follows

$$N = N_0 \int_{E_{\min}}^{E_{\max}} E^{-\gamma}\, dE \ , \tag{31}$$

which implies that

$$N_0 = \frac{N}{E_{\min}^{1-\gamma} - E_{\max}^{1-\gamma}} \ . \tag{32}$$

Following Nousek & Shue, I chose the slope value of $\gamma \equiv 2.0$ and used the energy range of 0.095–0.845 keV which was split into 15 equal bins of 0.050 keV per bin. I simulated 10^4 X-ray spectra for each of the theoretical N values used by Nousek & Shue: 25, 50, 75, 100, 150, 250, 500, 750, 1000, 2500, 5000, and 10^4 photons per spectrum. Figure 4 shows four of the simulated X-ray power-law spectra.

4.1. Powell's Method: Solving for γ and N using χ^2_{N}, χ^2_{P}, χ^2_γ

I determined the best-fit model parameters γ_{calc} and N_{calc} for each simulated spectrum with Powell's function minimization method[2] using the modified Neyman's χ^2 statistic (χ^2_{N}), Pearson's χ^2 statistic (χ^2_{P}), and the new χ^2_γ statistic. I used the following crude initial guesses: $\gamma = 0.0$ and $N = 1.3\sum_i^{15} n_i$, where n_i is the observed number of photons in the ith channel (bin). I computed the robust mean (average) and robust standard deviation[3] of the ratios $\gamma_{\mathrm{calc}}/\gamma$ and N_{calc}/N for the 10^4 simulated spectra of each dataset. The results of Powell's method with two free parameters (γ, N) using the χ^2_{N}, χ^2_{P}, χ^2_γ statistics are presented in Table 1 and Figure 5 . The first column,

[2]The primary reference for Powell's minimization method is Powell (1964). More accessible descriptions may be found in the numerical-methods literature (e.g., Acton 1970, Gill, Murray & Wright 1981, and Press et al. 1986)

[3]The robust mean given in all the tables is the mean of all values within two average deviations of the standard mean value. The robust standard deviation given in all the tables is 1.55σ where σ is the standard deviation of all values within two average deviations of the standard mean values.

N, of Table 1 corresponds to the total theoretical number of counts in the spectrum. The columns "$\gamma_{\rm calc}/\gamma$" and "$N_{\rm calc}/N$" are the robust mean values of the ratios of the best-fit parameters divided by the original value that was used to create the datasets. The parenthetical numbers are the robust standard deviations which can be used to determine the significance of the deviation from the perfect ratio value of one. For example, the first value of the 2nd column of Table 1 is 1.002(11) which represents the value of 1.002 ± 0.011. The deviation of this value from one (i.e. 0.002) is statistically significant since the error of the mean is only $\sim 0.011/\sqrt{10^4}$ or $\sim$0.00011.

Figure 5 indicates that the new χ^2_γ statistic gives the best results. Using a 5% criteria for both fitted parameters (γ, N), we see that the χ^2_γ statistic gives good results for spectra with $\gtrsim$50 photons. By comparison, Pearson's χ^2 statistic requires $\gtrsim$250 photons and the modified Neyman's χ^2 statistic requires $\gtrsim$750 photons in order to get the same quality of results. Baker & Cousins (1984) noted that, in many cases, χ^2 fits using the the modified Neyman's χ^2 statistic will underestimate the total number of counts while χ^2 fits using Pearson's χ^2 statistic will overestimate the total number of counts; both systematic errors are clearly seen in the bottom panel of Figure 5. I stated in the previous section that the usage of the modified Neyman's χ^2 statistic with Poisson-distributed data will typically underestimate the total counts by one count per bin. My results for the $\chi^2_{\rm N}$ statistic clearly exhibit this systematic error: the results of the ratio $N_{\rm calc}/N$ for spectra with $N \gtrsim 250$ photons (squares in the bottom panel of Fig. 5) are well modeled by the function $(N - 15)/N$ where 15 is the number of bins (channels) in our spectra [see the dashed curve in the bottom panel of Fig. 5].

A comparison of my analysis of $\gamma_{\rm calc}/\gamma$ using the modified Neyman's χ^2 statistic (2nd column of Table 1) with the analysis of Nousek & Shue for Powell's method (3rd column of their Table 3) shows nearly identical results. In my version of this numerical experiment, I used the two parameters N and γ while Nousek & Shue used N_0 and γ. A comparison of my analysis of $N_{\rm calc}/N$ (3rd column of Table 1) with their Powell's method analysis of $N_{\rm calc}/N_0$ (2nd column of their Table 3) shows that my analysis with $N_{\rm calc}/N$ has produced better estimates. This should not be surprising because the parameter N_0 is not an independent parameter – N_0 depends on both the slope of the spectrum and the theoretical number of photons in the spectrum. As a general rule, one gets better results by solving for independent parameters instead of dependent parameters.

4.2. Levenberg-Marquardt Method: Solving for γ and N using $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

I determined the best-fit model parameters $\gamma_{\rm calc}$ and $N_{\rm calc}$ for each simulated spectrum with Levenberg-Marquardt method[4] using the modified Neyman's χ^2 statistic ($\chi^2_{\rm N}$), Pearson's χ^2 statistic ($\chi^2_{\rm P}$), and the new χ^2_γ statistic. I used the previous crude initial guesses: $\gamma = 0.0$ and

[4]The primary references for Levenberg-Marquardt method are Levenberg (1944) and Marquardt (1963). More accessible descriptions may be found in the numerical-methods literature (e.g., Bevington 1969, Gill, Murray & Wright 1981, and Press et al. 1986)

$N = 1.3 \sum_i^{15} n_i$. I computed the robust mean and robust standard deviation of the ratios $\gamma_{\rm calc}/\gamma$ and $N_{\rm calc}/N$ for the 10^4 simulated spectra of each dataset. The results of Levenberg-Marquardt method with two free parameters (γ, N) using the $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ statistics are presented in Table 2 and Figure 6 .

Figure 6 indicates that the new χ^2_γ statistic gives the best results. Using a 5% criteria for both fitted parameters (γ, N), we see that the χ^2_γ statistic gives good results for all the spectra ($N \gtrsim 25$ photons). By comparison, Pearson's χ^2 statistic requires $\gtrsim$100 photons and the modified Neyman's χ^2 statistic requires $\gtrsim$500 photons in order to get the same quality of results.

The results for the χ^2_γ and $\chi^2_{\rm N}$ statistics are nearly identical with either Powell's method (Table 1) or the Levenberg-Marquardt method (Table 2). This finding refutes the determination by Nousek & Shue (1989) that Powell's method gives more accurate results than the Levenberg-Marquardt method.

The results for Pearson's χ^2 improved significantly by using the Levenberg-Marquardt method instead of Powell's method. An inspection of the individual fits showed that the Levenberg-Marquardt method with the $\chi^2_{\rm P}$ statistic produced a best-fit value for N that was within a one-tenth of one percent of the total number of photons in the spectrum. Needless to say, with such an improvement in the determination of N, a much better estimate for the slope γ could be determined.

This peculiar result tells us something important about this particular minimization problem: an excellent estimate of the total number of photons in the best-fit spectrum is the total number of photons in the actual spectrum. Thus by setting N to be a constant, $N \equiv \sum_i^{15} n_i$, we can eliminate one parameter and solve for γ alone.

4.3. Powell's Method: Solving for γ using $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

I determined the best-fit model parameter $\gamma_{\rm calc}$ for each simulated spectrum with Powell's function minimization method using the modified Neyman's χ^2 statistic ($\chi^2_{\rm N}$), Pearson's χ^2 statistic ($\chi^2_{\rm P}$), and the new χ^2_γ statistic. I set $N \equiv \sum_i^{15} n_i$ and used the crude initial guess of $\gamma = 0.0$. I computed the robust mean and robust standard deviation of the ratios $\gamma_{\rm calc}/\gamma$ for the 10^4 simulated spectra of each dataset. The results of Powell's method with two free parameters (γ, N) using the $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ statistics are presented in Table 3 and Figure 7 .

Figure 7 indicates that the new χ^2_γ statistic gives the best results. Using a 5% criteria, we see that the χ^2_γ statistic gives good results for all the spectra ($N \gtrsim 25$ photons). By comparison, Pearson's χ^2 statistic requires $\gtrsim$250 photons and the modified Neyman's χ^2 statistic requires $\gtrsim$750 photons in order to get the same quality of results.

Fitting only for the slope γ has improved the results for the new χ^2_γ statistic and the modified Neyman's χ^2 statistic. The results for Pearson's χ^2 show no improvement over the two free parameter result.

4.4. Levenberg-Marquardt Method: Solving for γ using $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

I determined the best-fit model parameter $\gamma_{\rm calc}$ for each simulated spectrum with the Levenberg-Marquardt minimization method using the modified Neyman's χ^2 statistic ($\chi^2_{\rm N}$), Pearson's χ^2 statistic ($\chi^2_{\rm P}$), and the new χ^2_γ statistic. I set $N \equiv \sum_i^{15} n_i$ and used the crude initial guess of $\gamma = 0.0$. I computed the robust mean and robust standard deviation of the ratios $\gamma_{\rm calc}/\gamma$ for the 10^4 simulated spectra of each dataset. The results of the Levenberg-Marquardt method with one free parameter (γ) using the $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ statistics are presented in Table 4 and Figure 8 .

Figure 8 indicates that Pearson's χ^2 statistic gives the best results. Using a 5% criteria, we see that both the new χ^2_γ statistic and the $\chi^2_{\rm P}$ statistic give good results for all the spectra ($N \gtrsim 25$ photons). By comparison, the modified Neyman's χ^2 statistic still requires $\gtrsim$750 photons in order to get the same quality of results. Once again, we note that the results for the χ^2_γ and $\chi^2_{\rm N}$ statistics are nearly identical with either Powell's method (Table 3) or the Levenberg-Marquardt method (Table 4).

4.5. Error Estimates

One expects the quality of the slope determination to degrade as the total number of photons in the X-ray spectra decline. Figure 9 shows the distribution of the best-fit values for the slope γ for the faintest spectra with a theoretical total of 100, 50, and 25 photons. As expected, the range of best-fit slope values measured for spectra with only $N \equiv 25$ photons is considerably larger than the range of values for spectra with $N \equiv 100$ photons.

The Levenberg-Marquardt method not only provides best-fit values for parameters but it also provides an error estimate (approximately $1\,\sigma$ errors) of those fitted parameters. How believable are these error estimates? Figure 10 shows an analysis of the errors estimated by the Levenberg-Marquardt method when the new χ^2_γ statistic was used to analyze spectra with theoretical totals of 100, 50, and 25 photons.

The top panel of Figure 10 shows the error analysis of spectra with $N \equiv 100$ photons. The median slope value is 1.989 and the median error estimate is 0.194 . A total of 15.87% of the spectra have estimates of $\gamma \leq 1.789$ and 15.87% of the spectra have estimates of $\gamma \geq 2.211$. For a normal distribution, one expects 68.26% of the deviates to be found within one standard deviation of the mean. Assuming that the distribution of best-fit γ values approximates a normal distribution, then half of the difference between the 84.13 and 15.87 percentile values of γ can be used as an estimate for the slope error: $\sigma_\gamma \approx (\gamma_{84.13\%} - \gamma_{15.87\%})/2 = (2.211 - 1.789)/2 = 0.211$.

This value is 8.8% larger than the median Levenberg-Marquardt error estimate; a fractional error of 10.6% instead of the predicted 9.8%.

The middle panel of Figure 10 shows the error analysis of spectra with $N \equiv 50$ photons. The median slope value is 2.009 and the median error estimate is 0.301. A total of 15.87% of the spectra have estimates of $\gamma \leq 1.732$ and 15.87% of the spectra have estimates of $\gamma \geq 2.334$. This gives an estimated slope error of $\sigma_\gamma \approx (\gamma_{84.13\%} - \gamma_{15.87\%})/2 = 0.301$. This value is exactly equal to the median Levenberg-Marquardt error estimate.

The bottom panel of Figure 10 shows the error analysis of spectra with $N \equiv 25$ photons. The median slope value is 2.071 and the median error estimate is 0.484. A total of 15.87% of the spectra have estimates of $\gamma \leq 1.692$ and 15.87% of the spectra have estimates of $\gamma \geq 2.570$. This gives an estimated slope error of $\sigma_\gamma \approx (\gamma_{84.13\%} - \gamma_{15.87\%})/2 = 0.439$. This value is 9.3% less than the median Levenberg-Marquardt error estimate; a fractional error of 21.2% instead of the predicted 23.4%.

The errors estimated by the Levenberg-Marquardt method are seen to be reasonable. Figure 11 shows the simulated X-ray spectra of Fig. 4 now plotted with χ^2_γ fits produced by the Levenberg-Marquardt method with one free parmater. The Levenberg-Marquardt method has done a good job even with the two faintest spectra which have actual totals of only 28 and 101 photons.

4.6. The χ^2_λ and Cash's C statistics

For the sake of completeness, I determined the best-fit model parameter $\gamma_{\rm calc}$ for each simulated spectrum with Powell's function minimization method using the maximum likelihood ratio statistic for Poisson distributions, χ^2_λ [Equation (23)], and Cash's C statistic,

$$C \equiv 2\sum_{i=1}^{N}\left[m_i - n_i \ln\left(m_i\right)\right] \tag{33}$$

[Equation (6) of Cash 1979]. I set $N \equiv \sum_i^{15} n_i$ and used the crude initial guess of $\gamma = 0.0$. I computed the robust mean and robust standard deviation of the ratios $\gamma_{\rm calc}/\gamma$ for the 10^4 simulated spectra of each dataset. The results of Powell's method with one free parameter (γ) using the χ^2_λ statistic and Cash's C statistic are presented in Table 5 and Figure 12 .

Table 5 and the right panel of Figure 12 shows that Cash's C statistic and the maximum likelihood ratio statistic for Poisson distributions, χ^2_λ, give identical results. This is not surprising because Cash's C statistic is a variant of the more well-known χ^2_λ statistic which has been discussed in the literature for over 70 years (e.g., Neyman & Pearson 1928).

I also determined the best-fit model parameter $\gamma_{\rm calc}$ for each simulated spectrum with the Levenberg-Marquardt minimization method using the maximum likelihood ratio statistic for

Poisson distributions, χ^2_λ. I set $N \equiv \sum_i^{15} n_i$ and used the crude initial guess of $\gamma = 0.0$. I computed the robust average and robust standard deviation of the ratios $\gamma_{\rm calc}/\gamma$ for the 10^4 simulated spectra of each dataset. The results of the Levenberg-Marquardt method with one free parameter (γ) using the χ^2_λ statistic is presented in Table 6 and Figure 12 . The maximum likelihood ratio statistic for Poisson distributions, χ^2_λ, produces nearly identical results with either Powell's method or the Levenberg-Marquardt minimization method.

Of the two statistics, χ^2_λ and the new χ^2_γ, which is better? Although Tables 6 and 4 indicate that the χ^2_λ is slightly better, we see that the actual differences between the distributions presented in Figure 12 are really quite negligible when compared with the overall uncertainty caused by simple sampling errors (counting statistics) of the simulated X-ray spectra.

5. SUMMARY

I have demonstrated that the application of the standard weighted mean formula, $\left[\sum_i n_i\sigma_i^{-2}\right] / \left[\sum_i \sigma_i^{-2}\right]$, to determine the weighted mean of data, n_i, drawn from a Poisson distribution, will, on average, underestimate the true mean by $\sim$1 for all true mean values larger than $\sim$3 when the common assumption is made that the error of the ith observation is $\sigma_i = \max(\sqrt{n_i}, 1)$. This small, but statistically significant offset, explains the long-known observation that chi-square minimization techniques which use the modified Neyman's χ^2 statistic, $\chi^2_{\rm N} \equiv \sum_i (n_i - y_i)^2/\max(n_i, 1)$, to compare Poisson-distributed data with model values, y_i, will typically predict a total number of counts that underestimates the true total by about 1 count per bin. Based on my finding that the weighted mean of data drawn from a Poisson distribution can be determined using the formula $\left[\sum_i [n_i + \min(n_i, 1)]\,(n_i+1)^{-1}\right] / \left[\sum_i (n_i+1)^{-1}\right]$, I proposed that a new χ^2 statistic, $\chi^2_\gamma \equiv \sum_i [n_i + \min(n_i, 1) - y_i]^2 / [n_i + 1]$, should always be used to analyze Poisson-distributed data in preference to the modified Neyman's χ^2 statistic.

I demonstrated the power and usefulness of χ^2_γ minimization by using two statistical fitting techniques (Powell's method and the Levenberg-Marquardt method) and five χ^2 statistics ($\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ, χ^2_λ, and Cash's C) to analyze simulated X-ray power-law 15-channel spectra with large and small counts per bin. I showed that χ^2_γ minimization with the Levenberg-Marquardt or Powell's method can produce excellent results (mean slope errors $\lesssim$3%) with spectra having as few as 25 total counts.

This analysis shows that there is nothing inherently wrong with either the Levenberg-Marquardt method or Powell's method in the low-count regime — provided that one uses an appropriate χ^2 statistic for the type of data being analyzed. Given Poisson-distributed data, one should always use the new χ^2_γ statistic in preference to the modified Neyman's χ^2 statistic because that statistic produces small, but statistically significant, systematic errors with Poisson-distributed data.

While the new χ^2_γ statistic is not perfect, neither is the more well-known χ^2_λ statistic (e.g., see Figures 2 and 3). Both statistics have problems in the very-low-count regime. The new χ^2_γ statistic complements but does not replace the older χ^2_λ statistic. Which statistic is "best" will generally depend on the particular problem being analyzed. An important difference between these two statistics is that the χ^2_λ statistic assumes that all data is perfect. With data from perfect counting experiments, the χ^2_λ statistic may give slightly better results than the new χ^2_γ statistic. However, data is typically obtained under less-than-perfect circumstances with multiple imperfect detectors. The χ^2_γ statistic, by definition, is a *weighted* χ^2 statistic which makes it easy to assign a *lower* weight to data from poor detectors. Thus in the analysis of real data obtained with noisy and imperfect detectors, the χ^2_γ statistic may well outperform the classic χ^2_λ statistic because low-quality data can be given a lower weight instead of being completely rejected.

Finally, I note in passing that two simple transformations may make it possible to retrofit many existing computer implementations (i.e. executable binaries) of χ^2_N minimization algorithms to do χ^2_γ minimization through the simple expedient of *changing the input data* from $[n_i]$ to $[n_i + \min(n_i, 1)]$, and error estimates, σ_i, from $[\max(\sqrt{n_i}, 1)]$ to $[\sqrt{n_i + 1}]$.

I would like to thank my former colleagues at Mount Stromlo and Siding Spring Observatories and current colleages at the National Optical Astronomy Observatories with whom I have had many useful discussions about various topics related to this research. Keith Arnaud, Martin Elvis, and Andy Ptak provided valuable practical information about the implementation details of the χ^2_λ statistic within the XSPEC X-ray spectral-fitting program. I would also like to thank the anonymous referee whose thoughtful comments and suggestions improved the final article. I am grateful to the anonymous referee of the previous version of this paper (based on the χ^2_β statistic) for providing a copy of the Baker & Cousins (1984) article. Special thanks are due to Christopher Burke who checked all the proofs in an early draft of the manuscript and to Ian Dell'Antonio who carefully read the second-to-last revision of the manuscript. Mary Guerrieri has my heartfelt thanks for cheerfully helping me acquire copies of the many articles in statistical journals which are not available in the NOAO library collection. KJM was supported by a grant from the National Aeronautics and Space Administration (NASA), Order No. S-67046-F, which was awarded by the Long-Term Space Astrophysics Program (NRA 95-OSS-16). Publication costs were partially paid by NASA through Grant No. STSCI GO-5386 from the Space Telescope Science Institute, which is operated by the by the Association of Universities for Research in Astronomy, Inc., under NASA Contract No. NAS5-26555. This research has made use of NASA's Astrophysics Data System Abstract Service which is operated by the Jet Propulsion Laboratory at the California Institute of Technology, under contract with NASA.

REFERENCES

Abramowitz, M., & Stegun, I. 1964, "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", National Bureau of Standards, Applied Mathematics Series # 55, edited by M. Abramowitz & I. Stegun (Washington D.C.: U.S. Government Printing Office)

Acton, F.S. 1970, Numerical Methods That Work, (New York: Harper and Row)

Baker, S., & Cousins, R.D. 1984, Nuclear Instruments and Methods in Physics Research, 221, 437

Bevington, P.R. 1969, Data Reduction and Error Analysis for the Physical Sciences (New York: McGraw-Hill)

Cash, W. 1979, ApJ, 228, 939

Gehrels, N. 1986, ApJ, 303, 336

Gill, P.E., Murray, W., & Wright, M.H. 1981, Practical Optimization (New York: Academic Press)

Levenberg, K. 1944, Quarterly of Applied Mathematics, 2, 164

Marquardt, D. 1963, J. SIAM, 11, 431

Neyman, J., & Pearson, E.S. 1928, Biometrica, 20A, 263

Nousek, J.A., & Shue, D.R. 1989, ApJ, 342, 1207

Poisson, S.-D. 1837, Recherches sur la Probabilité des Jugements en Matière Criminell et en Matière Civile, pp. 205 et seq.

Powell, M.J.D. 1964, Computer Journal, 7, 155

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. 1986, Numerical Recipes (Cambridge: Cambridge University Press)

Wheaton, W.A., et al. 1995, ApJ, 438, 322

This preprint was prepared with the AAS LaTeX macros v4.0.

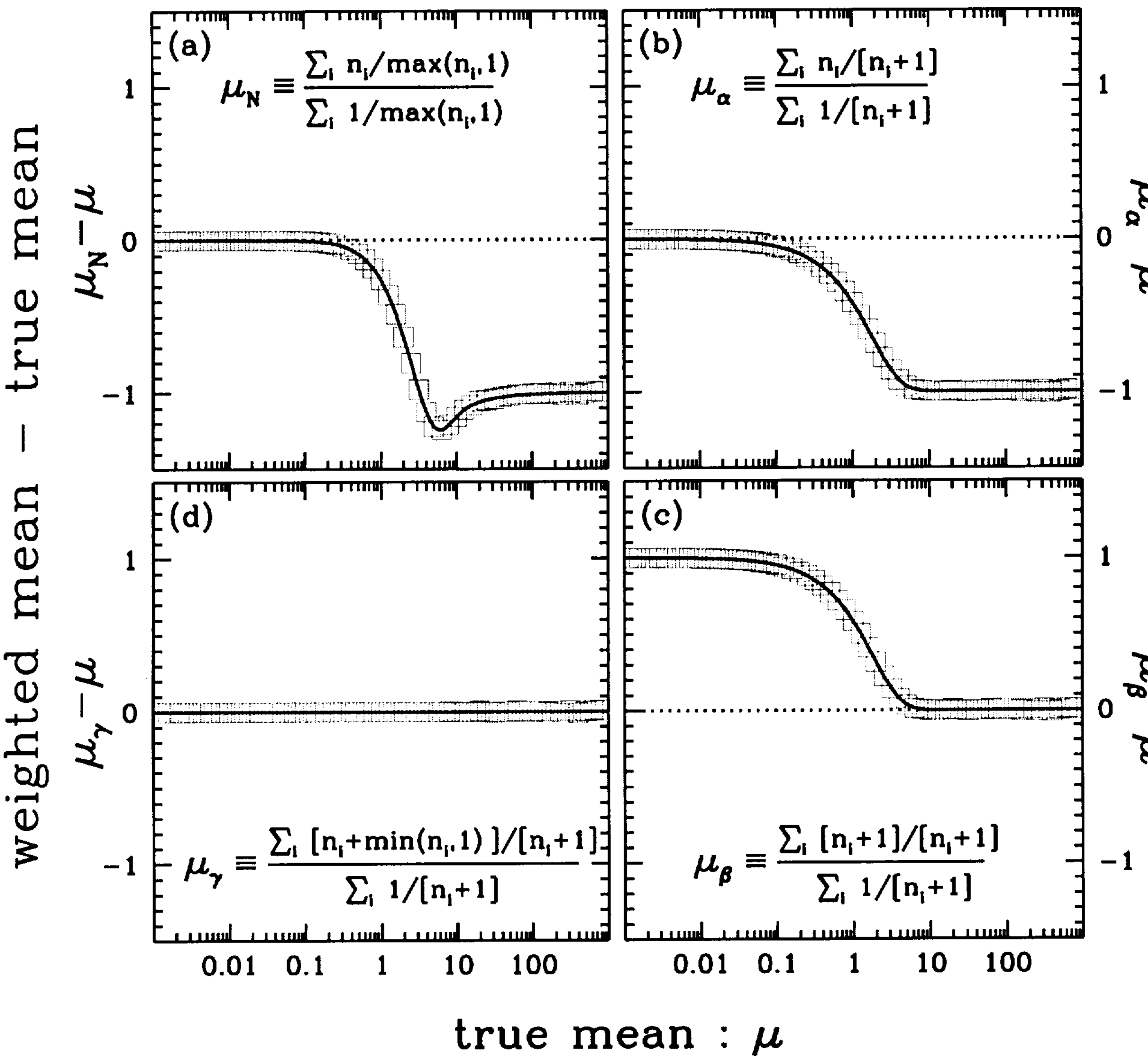

Fig. 1.— Analysis of four weighted-mean formulae applied to Poisson-distributed data. Each open square represents the weighted mean of 4×10^6 Poisson deviates at each given true mean value: $0.001 < \mu < 1000$.

(a) The difference between the weighted mean computed using Equation (8), μ_N, and the true mean, μ. The solid curve is the difference between Equation (10) and the true mean: $\{[e^\mu - 1][1 + \mathrm{Ei}(\mu) - \gamma - \ln(\mu)]^{-1}\} - \mu$.

(b) The difference between the weighted mean computed using Equation (14), μ_α, and the true mean, μ. The solid curve is the difference between Equation (15) and the true mean: $\{\mu[1 - e^{-\mu}]^{-1} - 1\} - \mu$.

(c) The difference between the weighted mean computed using Equation (16), μ_β, and the true mean, μ. The solid curve is the difference between Equation (17) and the true mean: $\{\mu[1 - e^{-\mu}]^{-1}\} - \mu$.

(d) The difference between the weighted mean computed using Equation (18), μ_γ, and the true mean, μ. The solid curve is the difference between Equation (19) and the true mean. The difference is zero because μ_γ is the weighted-mean formula for Poisson-distributed data.

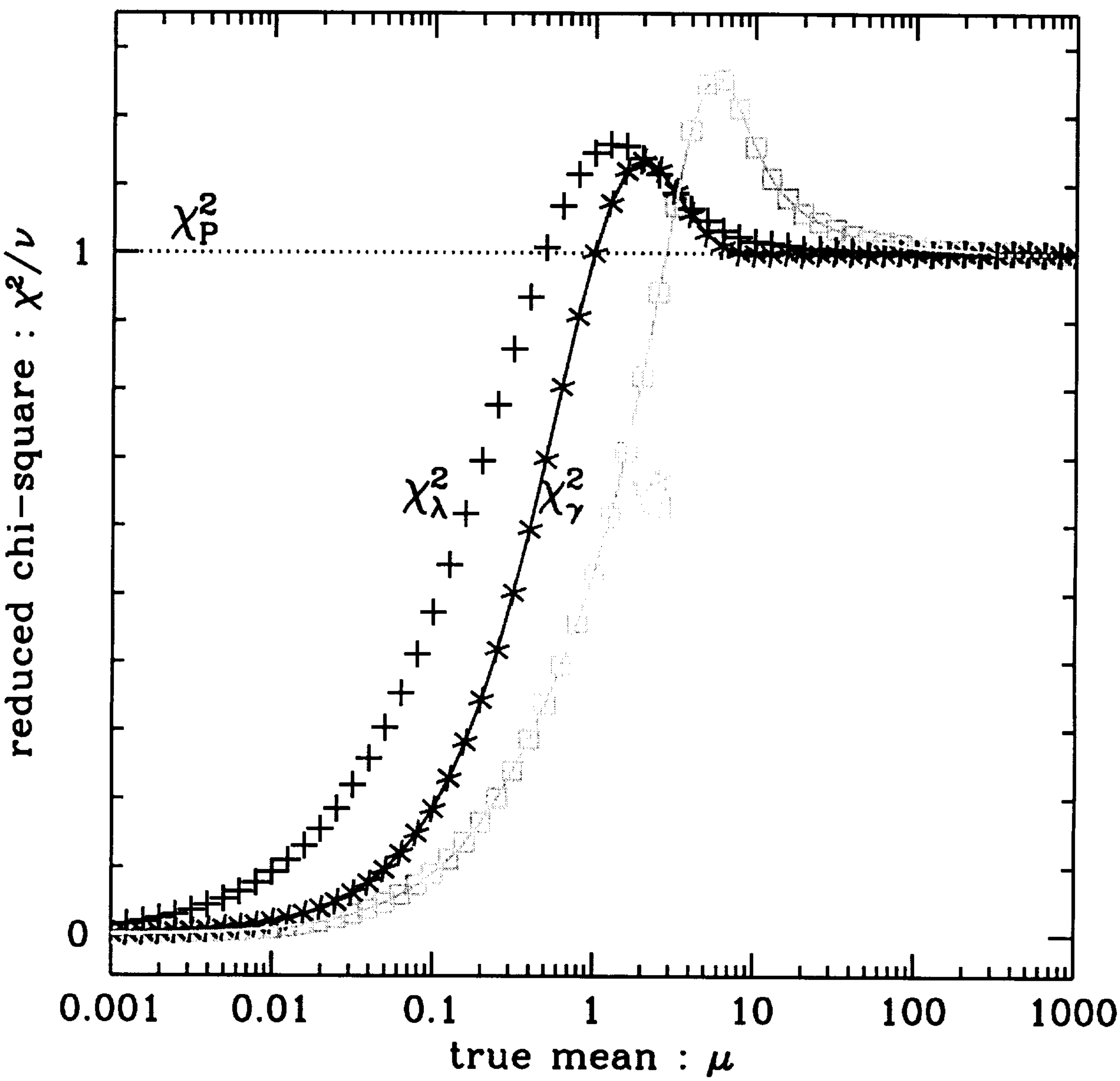

Fig. 2.— Reduced chi-square (χ^2/ν) as a function of true Poisson mean, μ, for 4 χ^2 statistics: Pearson's χ^2 [$\chi^2_{\rm P} \equiv \sum_{i=1}^{N}(n_i - m_i)^2/m_i$], the modified Neyman's χ^2 [$\chi^2_{\rm N} \equiv \sum_{i=1}^{N}(n_i - m_i)^2/\max(n_i, 1)$], the new χ^2_γ statistic [$\chi^2_\gamma \equiv \sum_{i=1}^{N}(n_i + \min(n_i, 1) - m_i)^2/(n_i+1)$], and the maximum likelihood ratio statistic for Poisson distributions $\left[\chi^2_\lambda \equiv 2\sum_{i=1}^{N}\left(m_i - n_i + n_i \ln\left(n_i/m_i\right)\right)\right]$. The Poisson distributions of Figure 1 were analyzed to produce this plot. The formula for the curve connecting the values for modified Neyman's χ^2 statistic ($\chi^2_{\rm N}$) is given in Equation (27). The formula for the curve connecting the values for new χ^2_γ statistic is given in Equation (29). The dotted line shows the ideal value of one.

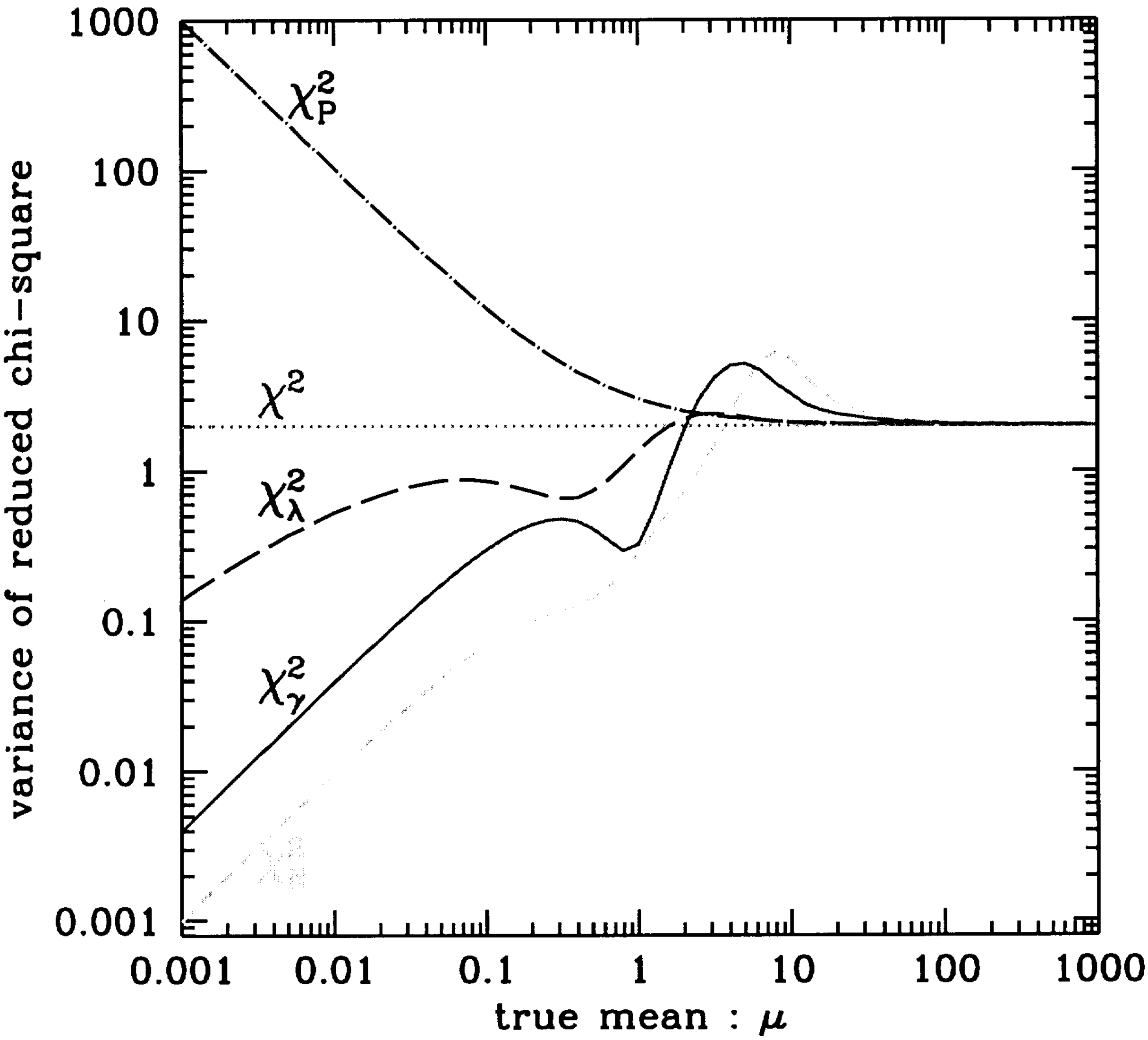

Fig. 3.— The variance of the reduced chi-square ($\sigma^2_{\chi^2/\nu}$) as a function of true Poisson mean, μ, for 5 χ^2 statistics: the standard χ^2, Pearson's χ^2 ($\chi^2_{\rm P}$), the modified Neyman's χ^2 ($\chi^2_{\rm N}$), the new χ^2_γ statistic, and the maximum likelihood ratio statistic for Poisson distributions (χ^2_λ). The Poisson distributions of Figure 1 were analyzed to produce this plot. The formula for the variance of the reduced Pearson's χ^2 statistic is $2 + \mu^{-1}$. The dotted line shows the ideal value of two.

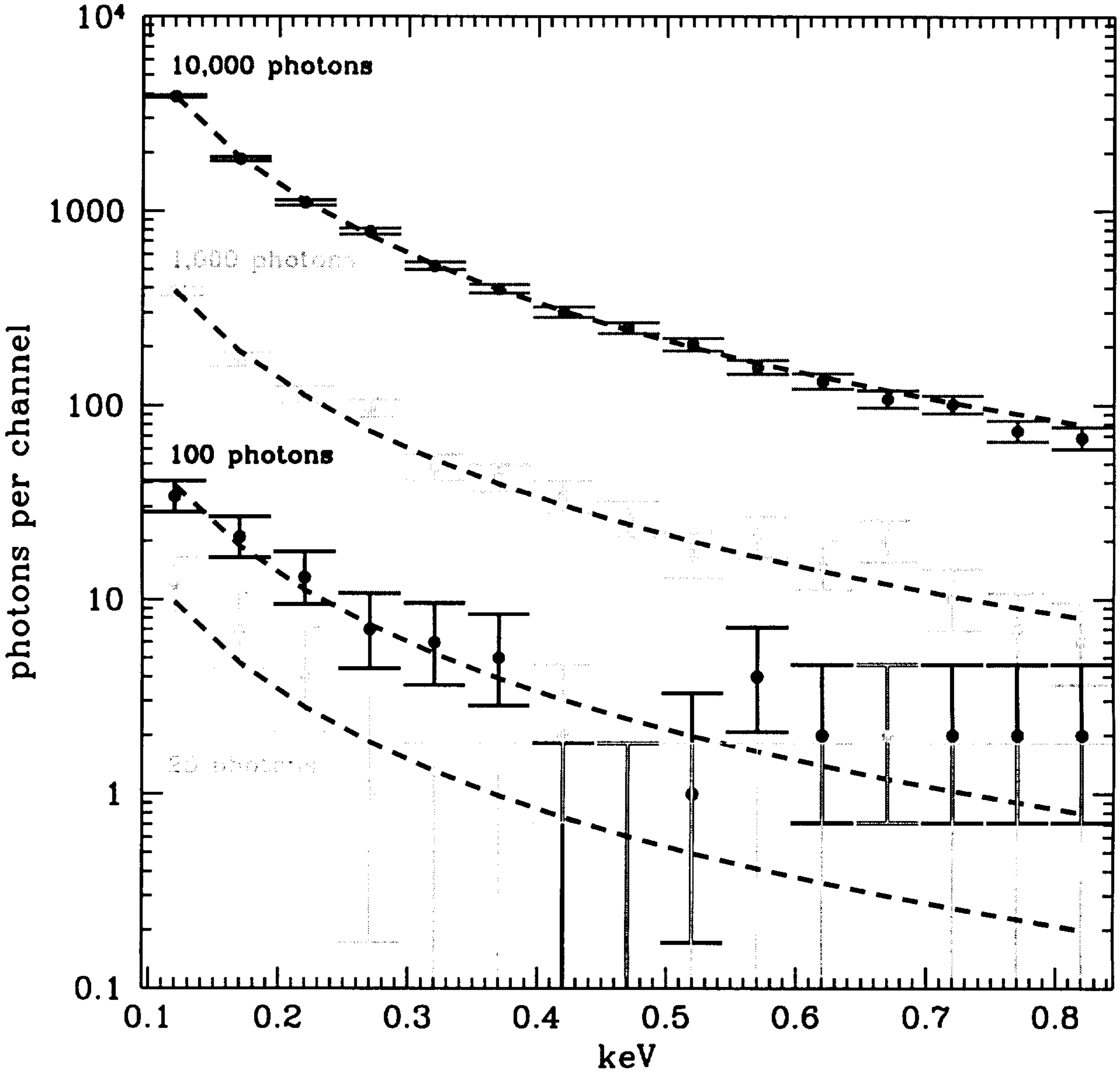

Fig. 4.— The dashed lines show 4 ideal X-ray power-law spectra with a total of 25, 100, 1000, and 10000 photons. Four simulated X-ray spectra with totals of 28, 101, 1015, and 9938 photons are shown with 1σ error bars estimated with Equations (9) and (14) of Gehrels (1986). (N.B. Some errorbars overlap and the bottom two spectra have identical data values at the 0.47 and 0.67 keV bins.)

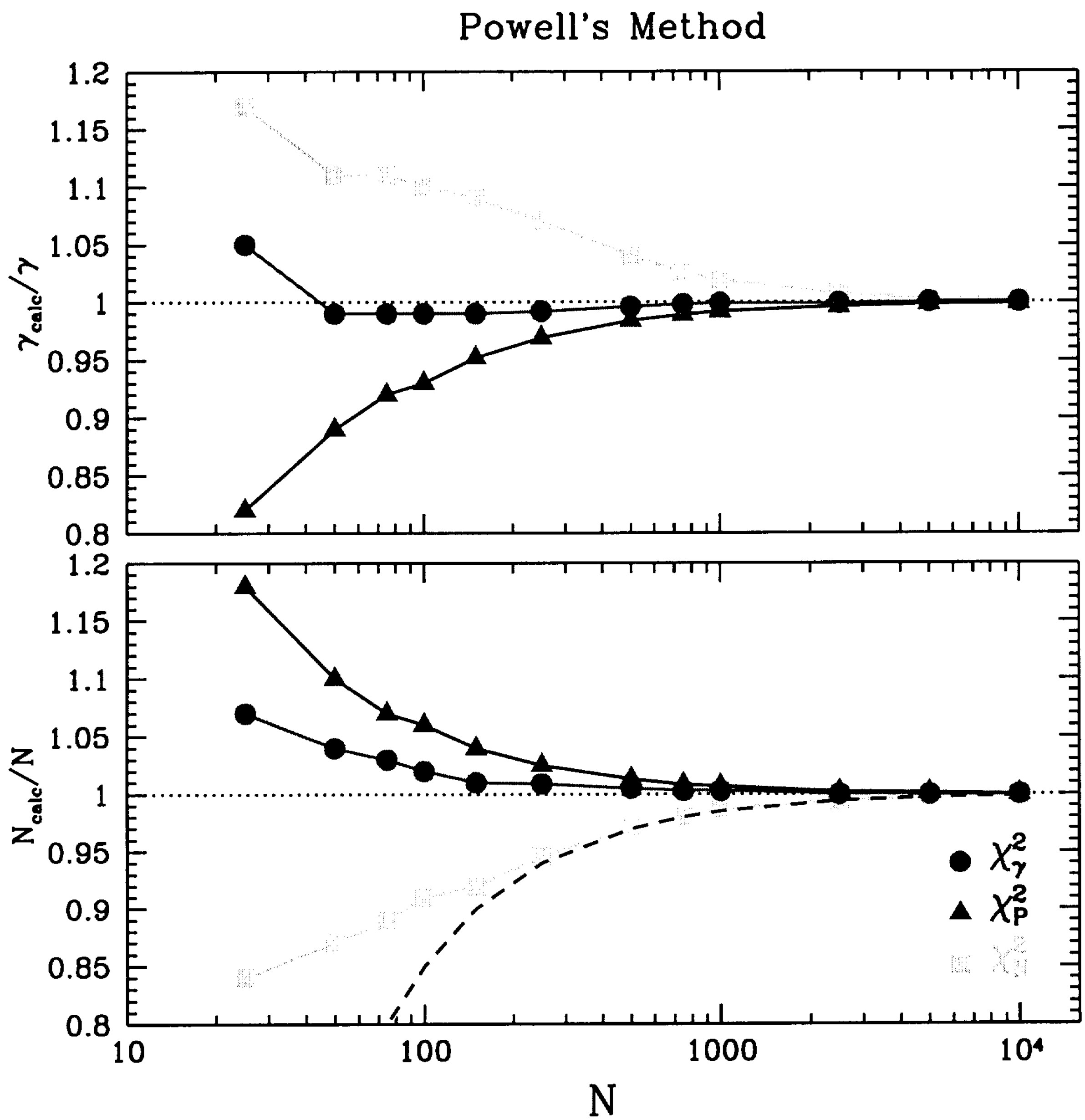

Fig. 5.— Results of Powell's method with two free parameters (γ, N) for three statistics: χ^2_γ (circles), χ^2_N (squares), and χ^2_P (triangles). This figure uses the data given in Table 1. The dotted lines show the ideal ratio value of one. The dashed curve in the bottom panel shows the function $(N - 15)/N$ which is a good model for the N_{calc}/N results of the χ^2_N statistic for all spectra with $N \gtrsim 250$ photons.

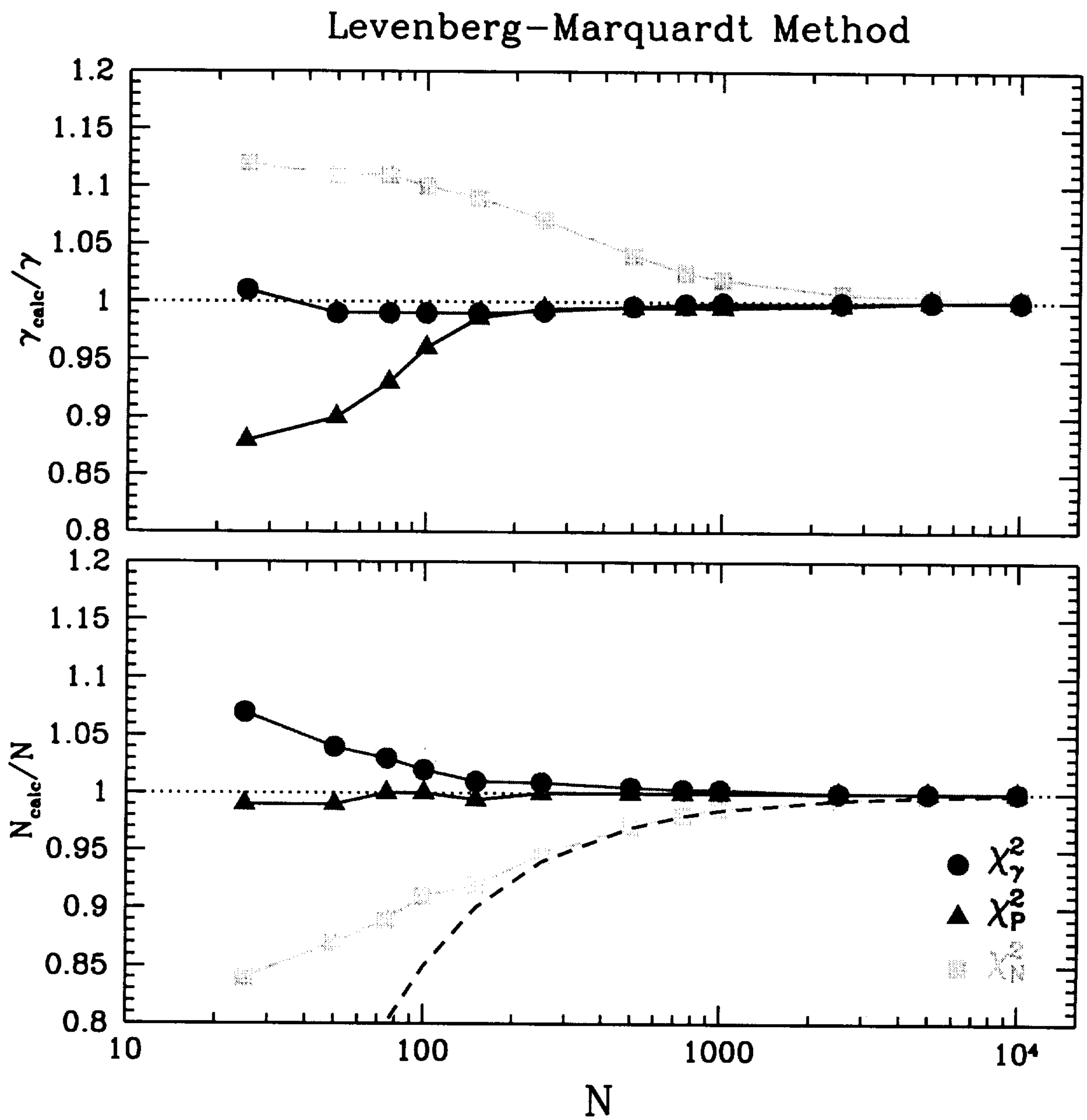

Fig. 6.— Results of the Levenberg-Marquardt method with two free parameters (γ, N) for three statistics: χ^2_γ (circles), χ^2_N (squares), and χ^2_P (triangles). This figure uses the data given in Table 2. The dotted lines show the ideal ratio value of one. The dashed curve in the bottom panel shows the function $(N - 15)/N$ which is a good model for the N_{calc}/N results of the χ^2_N statistic for all spectra with $N \gtrsim 250$ photons.

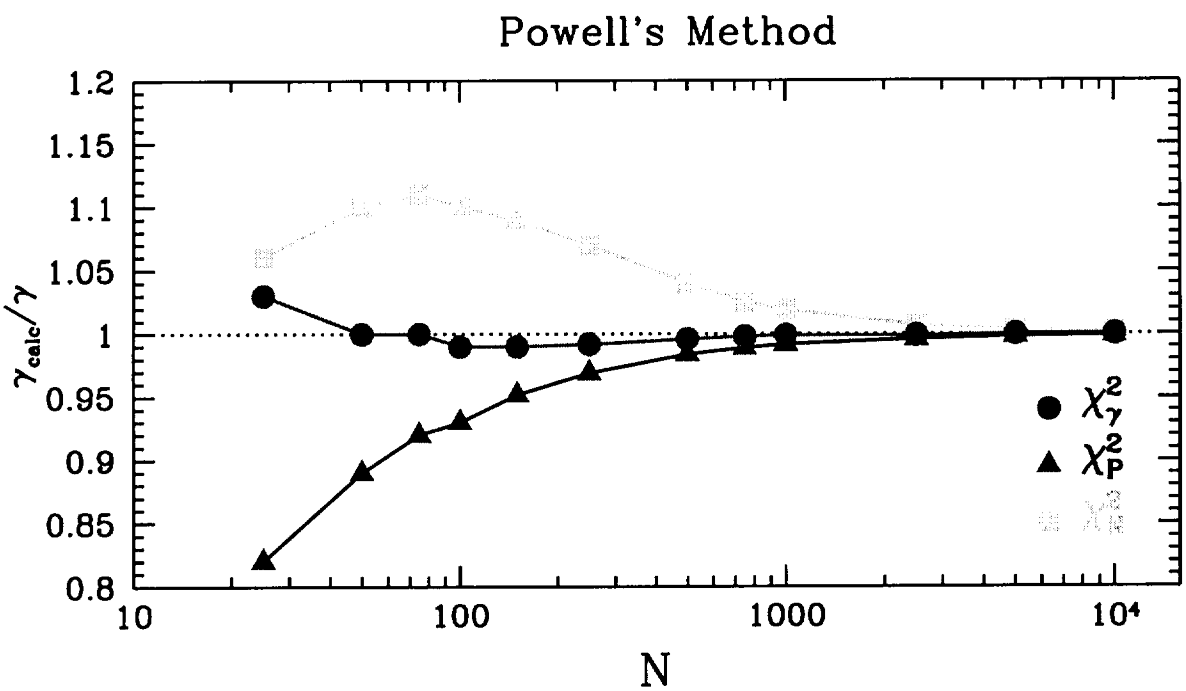

Fig. 7.— Results of Powell's method with one free parameter (γ) for three statistics: χ^2_γ (circles), χ^2_N (squares), and χ^2_P (triangles). This figure uses the data given in Table 3. The dotted lines show the ideal ratio value of one.

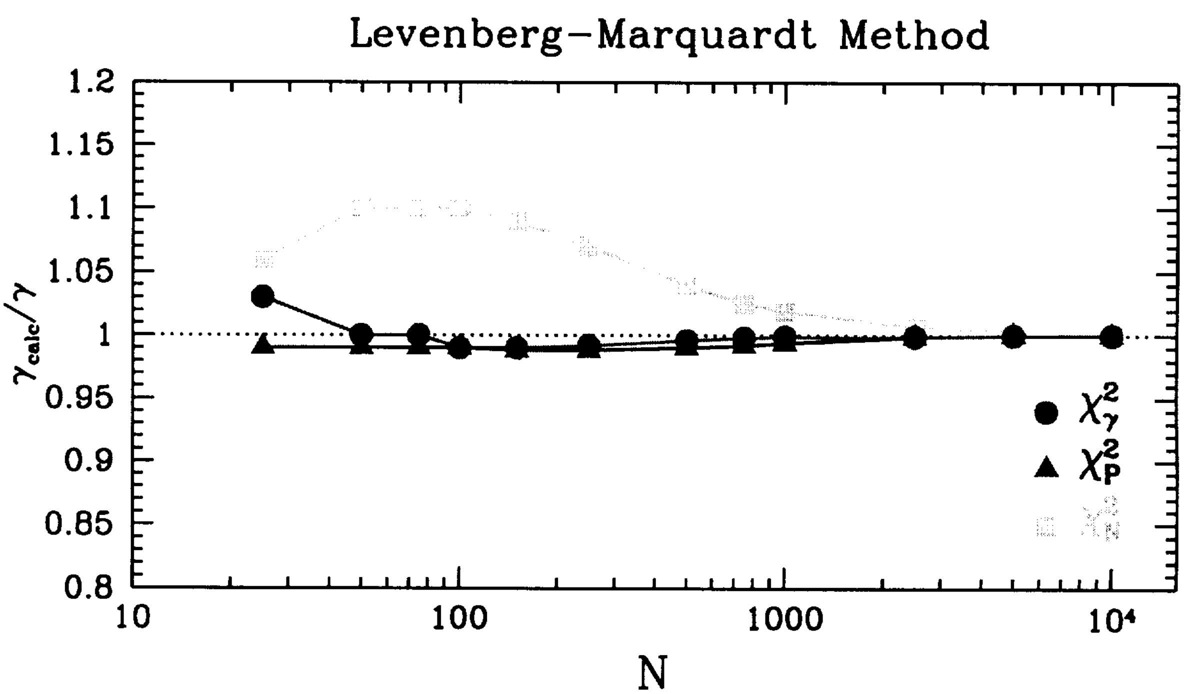

Fig. 8.— Results of the Levenberg-Marquardt method with one free parameter (γ) for three statistics: χ^2_γ (circles), χ^2_N (squares), and χ^2_P (triangles). This figure uses the data given in Table 4. The dotted lines show the ideal ratio value of one.

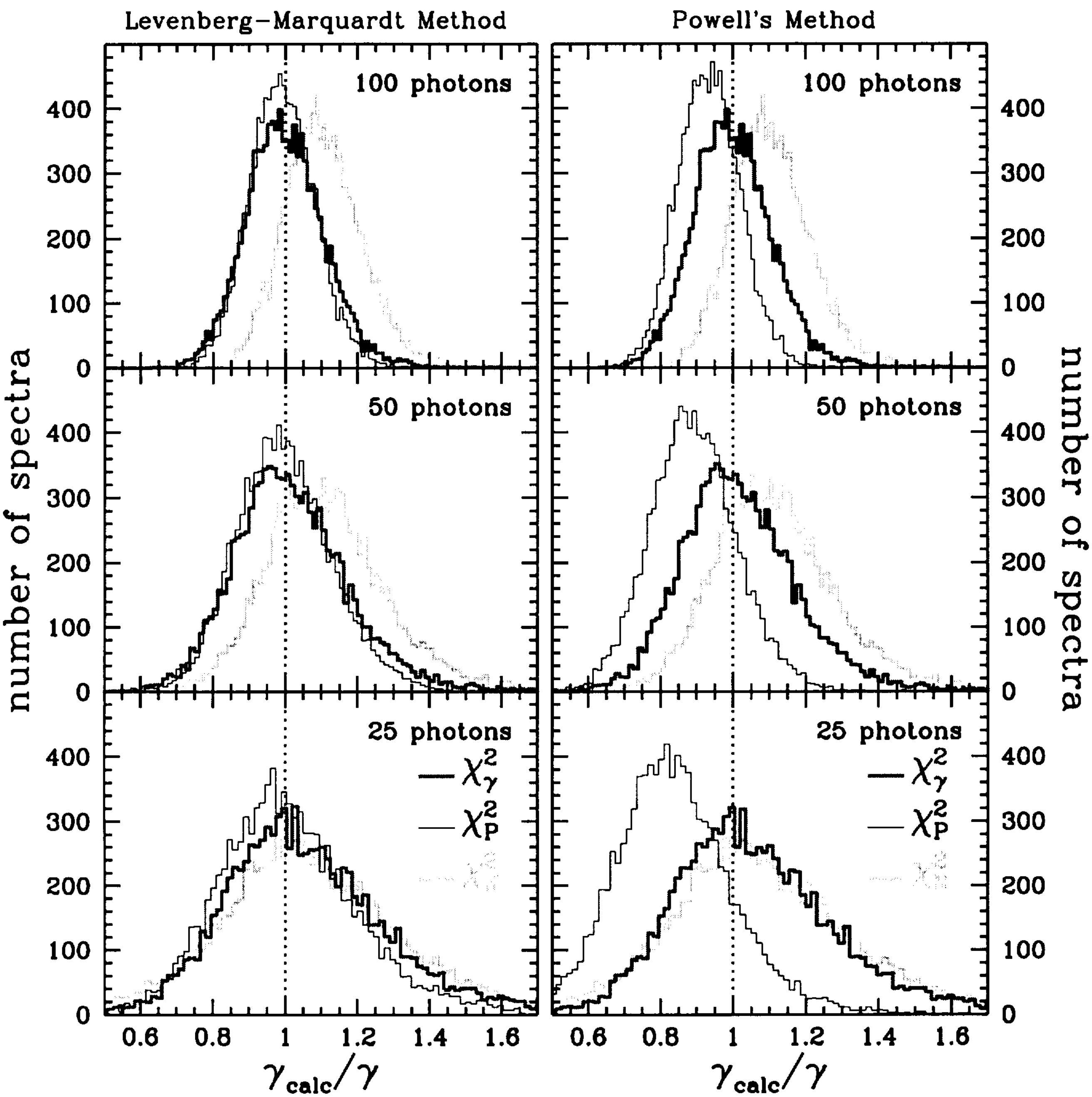

Fig. 9.— A comparison of the results of the analysis of the simulated X-ray spectra with theoretical totals of 100, 50, and 25 photons using the Levenberg-Marquardt method with 1 free parameter (left) and Powell's method with 1 free parameter (right). Note that the histograms for the χ^2_γ and χ^2_N are nearly identical for both methods. The statistical analysis of this data is presented in Tables 4 and 3.

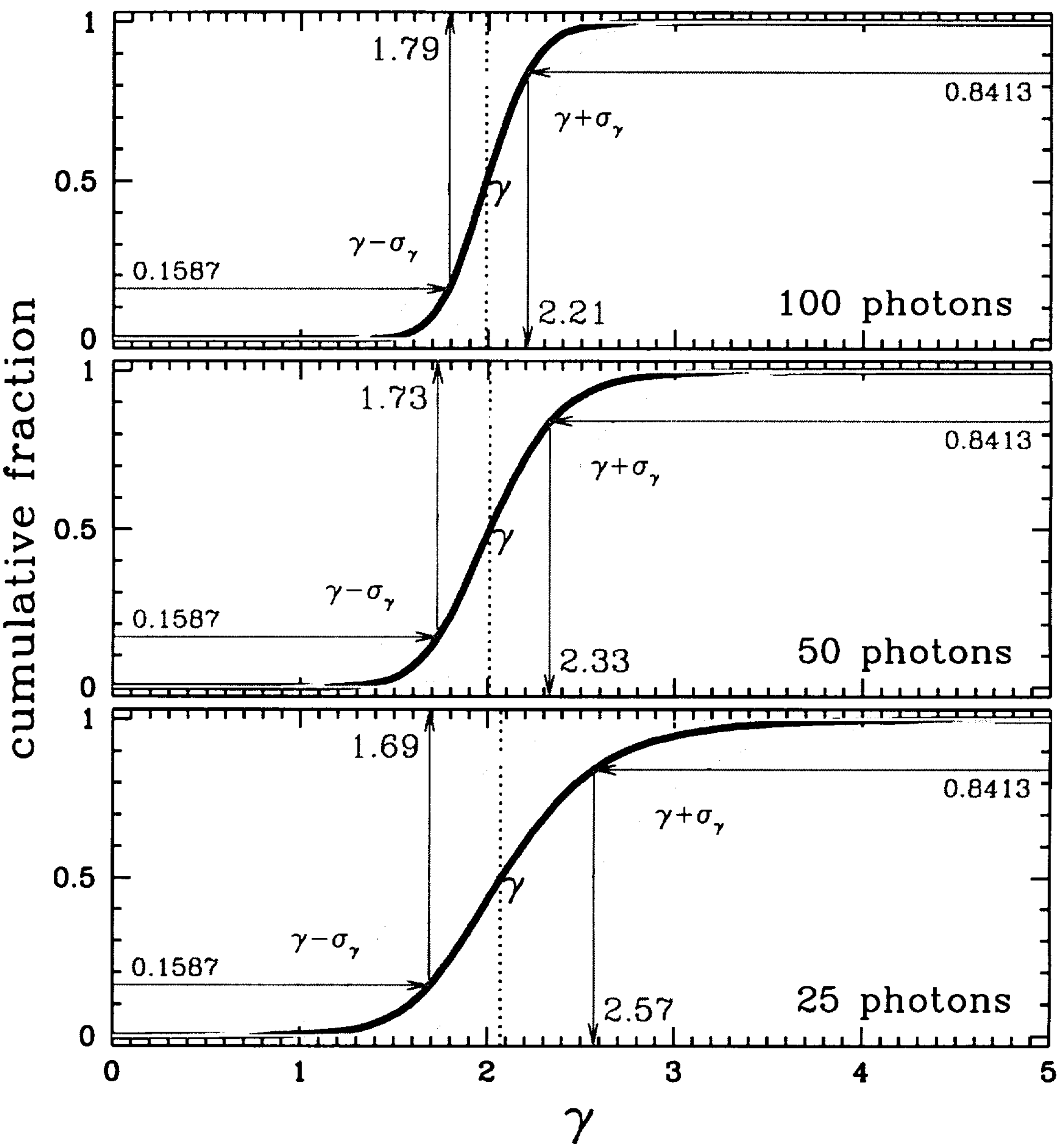

Fig. 10.— Error analysis of the Levenberg-Marquardt method results using the χ^2_γ statistic with one free parameter. The thick curve in each panel shows the cumulative distribution of the best-fit estimates of the slope γ. The right (left) thin curve in each panel shows the cumulative distribution of γ plus (minus) σ_γ which is the error estimate of the best-fit slope value. The statistical analysis of this data is presented in Table 4.

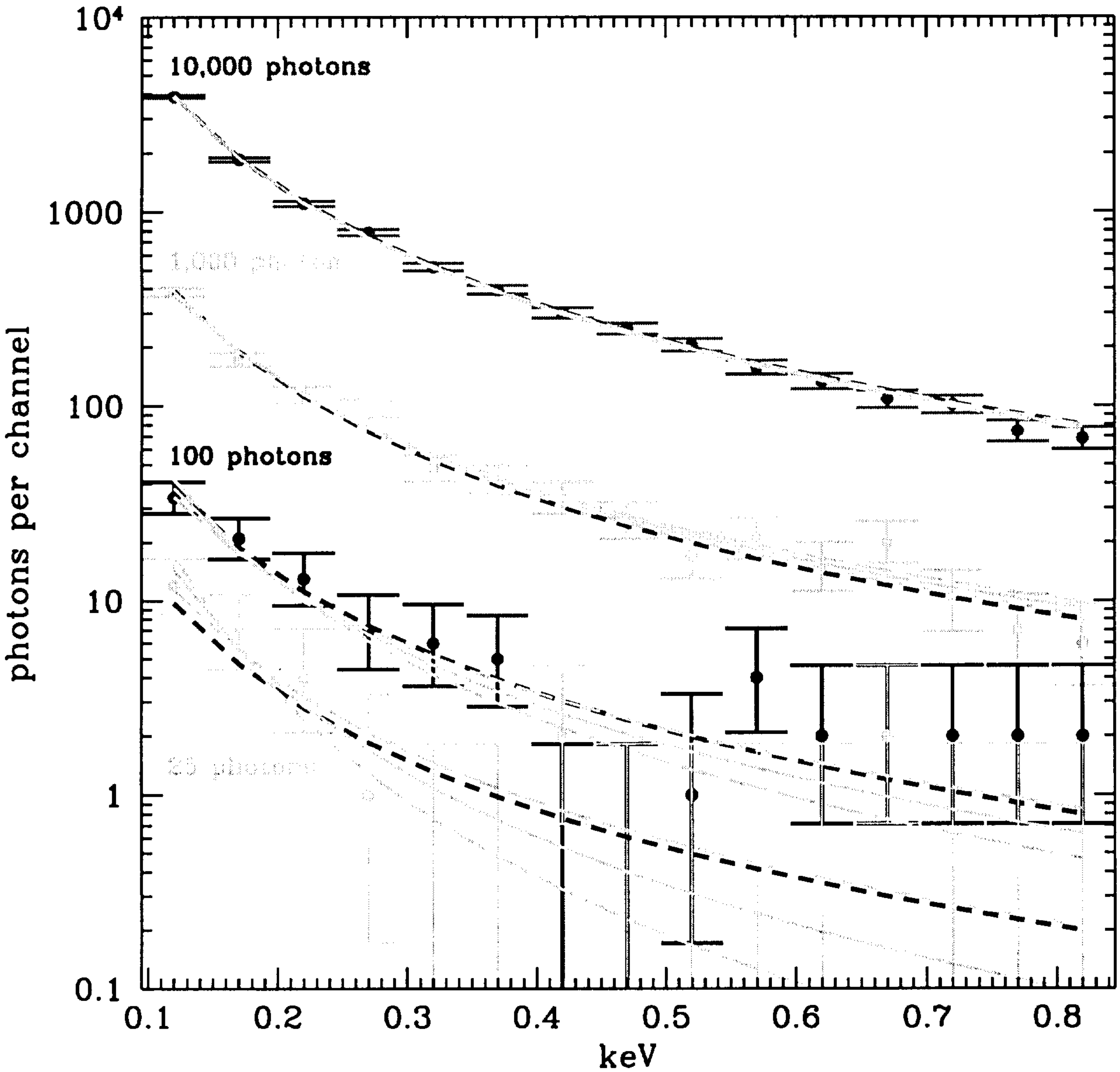

Fig. 11.— The simulated X-ray spectra of Fig. 4 now plotted with χ^2_γ fits. The best fits are shown with solid curves. The upper and lower $1\,\sigma$ slope estimates are shown with long dashed curves.

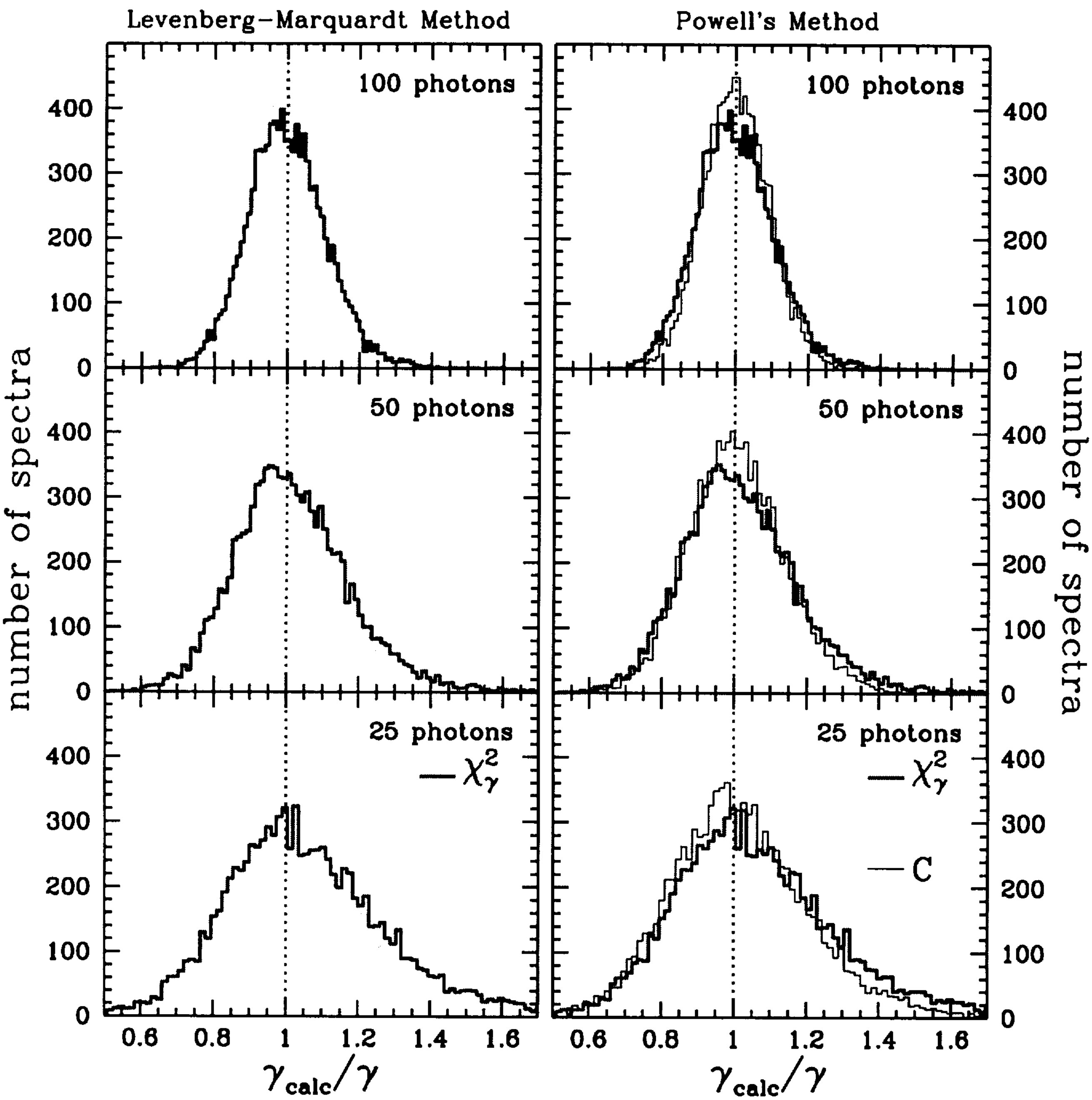

Fig. 12.— A comparison of the results of the Levenberg-Marquardt method with 1 free parameter (left) and Powell's method with 1 free parameter (right) for three statistics: χ^2_γ, χ^2_λ, Cash's C. The statistical analysis of this data is presented in Tables 3, 5, and 6.

TABLE 1.
Results of Powell's method with 2 free parameters (γ,N) for 3 statistics: $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

N	$\chi^2_{\rm N}$		$\chi^2_{\rm P}$		χ^2_γ	
	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$
10000	1.002(11)	0.999(12)	0.999(11)	1.000(12)	1.000(11)	1.000(12)
5000	1.003(15)	0.998(17)	0.998(15)	1.001(17)	1.000(15)	1.000(17)
2500	1.007(23)	0.994(24)	0.996(22)	1.002(24)	0.999(22)	1.000(24)
1000	1.019(37)	0.987(39)	0.992(34)	1.007(39)	0.999(35)	1.003(39)
750	1.025(45)	0.981(45)	0.989(39)	1.008(43)	0.998(41)	1.003(44)
500	1.040(58)	0.971(56)	0.984(47)	1.013(54)	0.996(51)	1.005(55)
250	1.071(82)	0.946(79)	0.969(67)	1.025(77)	0.992(80)	1.009(81)
150	1.09(10)	0.92(10)	0.952(84)	1.04(10)	0.99(10)	1.01(11)
100	1.10(13)	0.91(13)	0.93(10)	1.06(12)	0.99(13)	1.02(13)
75	1.11(15)	0.89(14)	0.92(12)	1.07(14)	0.99(15)	1.03(15)
50	1.11(20)	0.87(17)	0.89(14)	1.10(18)	0.99(19)	1.04(19)
25	1.17(50)	0.84(24)	0.82(19)	1.18(26)	1.05(40)	1.07(28)

TABLE 2.
Results of the Levenberg-Marquardt method with 2 free parameters (γ,N) for 3 statistics: $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

N	$\chi^2_{\rm N}$		$\chi^2_{\rm P}$		χ^2_γ	
	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$	$\gamma_{\rm calc}/\gamma$	$N_{\rm calc}/N$
10000	1.002(11)	0.999(12)	1.000(11)	1.000(12)	1.000(12)	1.000(11)
5000	1.004(15)	0.998(17)	1.000(15)	1.000(17)	1.000(15)	1.000(17)
2500	1.007(23)	0.994(24)	0.997(22)	1.000(24)	0.999(22)	1.000(24)
1000	1.019(37)	0.987(39)	0.995(34)	1.000(38)	0.999(35)	1.003(39)
750	1.025(45)	0.981(45)	0.995(38)	1.000(43)	0.998(41)	1.003(44)
500	1.040(58)	0.971(56)	0.995(46)	1.000(54)	0.996(51)	1.005(55)
250	1.071(82)	0.946(79)	0.994(67)	1.000(76)	0.992(80)	1.009(81)
150	1.09(10)	0.92(10)	0.986(91)	0.994(97)	0.99(10)	1.01(11)
100	1.10(13)	0.91(13)	0.96(12)	1.00(12)	0.99(13)	1.02(13)
75	1.11(15)	0.89(14)	0.93(13)	1.00(14)	0.99(15)	1.03(15)
50	1.11(20)	0.87(17)	0.90(14)	0.99(17)	0.99(19)	1.04(19)
25	1.12(35)	0.84(24)	0.88(17)	0.99(23)	1.01(29)	1.07(28)

TABLE 3.
Results of Powell's method with 1 free parameter (γ) for 3 statistics: $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

N	$\chi^2_{\rm N}$	$\chi^2_{\rm P}$	χ^2_γ
	$\gamma_{\rm calc}/\gamma$	$\gamma_{\rm calc}/\gamma$	$\gamma_{\rm calc}/\gamma$
10000	1.002(11)	0.999(11)	1.000(11)
5000	1.003(15)	0.998(15)	1.000(15)
2500	1.007(23)	0.996(22)	0.999(22)
1000	1.019(37)	0.992(34)	0.999(35)
750	1.025(45)	0.989(39)	0.998(41)
500	1.040(58)	0.984(47)	0.996(51)
250	1.070(82)	0.969(67)	0.992(80)
150	1.09(11)	0.952(84)	0.99(10)
100	1.10(13)	0.93(10)	0.99(13)
75	1.11(15)	0.92(12)	1.00(15)
50	1.10(19)	0.89(14)	1.00(18)
25	1.06(31)	0.82(19)	1.03(27)

TABLE 4.
Results of the Levenberg-Marquardt method with 1 free parameter (γ) for 3 statistics: $\chi^2_{\rm N}$, $\chi^2_{\rm P}$, χ^2_γ

N	$\chi^2_{\rm N}$	$\chi^2_{\rm P}$	χ^2_γ
	$\gamma_{\rm calc}/\gamma$	$\gamma_{\rm calc}/\gamma$	$\gamma_{\rm calc}/\gamma$
10000	1.002(11)	1.000(11)	1.000(11)
5000	1.004(15)	1.000(15)	1.000(15)
2500	1.007(23)	0.999(22)	0.999(22)
1000	1.019(37)	0.994(34)	0.999(35)
750	1.025(45)	0.992(39)	0.998(41)
500	1.040(58)	0.990(48)	0.996(51)
250	1.070(82)	0.988(68)	0.992(80)
150	1.09(10)	0.988(87)	0.99(10)
100	1.10(13)	0.99(11)	0.99(13)
75	1.10(15)	0.99(12)	1.00(15)
50	1.10(19)	0.99(16)	1.00(18)
25	1.06(31)	0.99(22)	1.03(27)

TABLE 5.
Results of Powell's method with 1 free parameter (γ) for 2 statistics: χ^2_λ, Cash's C

N	χ^2_λ	Cash's C
	$\gamma_{\rm calc}/\gamma$	$\gamma_{\rm calc}/\gamma$
10000	1.000(11)	1.000(11)
5000	1.000(15)	1.000(15)
2500	1.000(22)	1.000(22)
1000	1.000(34)	1.000(34)
750	1.000(39)	1.000(39)
500	1.000(48)	1.000(48)
250	0.999(68)	0.999(68)
150	0.999(88)	0.999(88)
100	1.00(11)	1.00(11)
75	1.00(12)	1.00(12)
50	1.00(16)	1.00(16)
25	1.00(22)	1.00(22)

TABLE 6.
Results of the Levenberg-Marquardt method with 1 free parameter (γ) for the χ^2_λ statistic

N	χ^2_λ
	γ_{calc}/γ
10000	1.000(11)
5000	1.000(15)
2500	1.000(22)
1000	1.000(34)
750	1.000(39)
500	1.000(48)
250	0.999(68)
150	0.999(88)
100	1.00(11)
75	1.00(12)
50	1.00(16)
25	1.00(22)