The Group Abundance Fraction: A statistically robust measure of particle composition and of spatial structure in images

Edmond C. Roelof
Johns Hopkins University/Applied Physics Laboratory, Laurel MD 20723-6099
E-mail: edmond.roelof@jhuapl.edu

Abstract. When two processes (A and B) are measured by counting instruments, we wish to present the results as an “abundance” along with an associated uncertainty. When the number of individual counts of either component is low, the usual ratio (A/B) is statistically ill-defined because there is a finite probability of zero counts in the denominator. However, the “Group Abundance Fraction” (F), e.g., A/(A+B), is a statistically robust parameter (under the condition that there must be at least one total count between A and B in order to constitute a measurement). We rigorously derive expectation values (over a conditional Poisson statistical ensemble) for <F> and its standard deviation \( \sigma_F \) that are valid for all values of total counts greater than one. We then derive useful estimators for these ensemble expectation values in terms of simple algebraic functions of the actual observed accumulated counts. We also incorporate the practical effects of instrument response functions and background rates. The results are not restricted to particle species composition (or spectra), but are also applicable when A and B represent counts in adjacent “pixels” of spatial images formed by the particles.

1. Introduction
The new physics of the boundary regions of our solar system is being revealed by in situ measurements of energetic charged particles from Voyagers 1 and 2 (from ~50 keV/nucleon up to galactic cosmic rays), as well as by remote sensing of energetic neutral atoms (ENAs) by IBEX (~0.5-6 keV) and Cassini/INCA (5-55 keV). After crossing the heliopause, the outer boundary of the heliosheath at 122 AU in August 2012, Voyager 1 now appears to be measuring the composition of galactic cosmic rays down to energies of a few MeV/nucleon, but at rather low counting rates. Consequently, there is a need for a statistically robust measure of composition “ratios” between two chemical species, e.g., A and B at specified energies. There is also a related need in the analysis of spatial distributions of ENA intensity in all-sky maps of the heliosheath. In this second case, A and B could represent counts in adjacent “pixels” of a spatial map, or counts in the same pixels over different time intervals. In the least intense regions of the images, the count rates are so low that individual pixels may not even accumulate 100 useable counts over one year. Nonetheless, it is scientifically important to be able to decide whether there are statistically significant spatial gradients in these regions within a given accumulated skymap, or whether there are statistically significant time variations in the same pixel in the skymaps from one year to another.

Consequently we are going to be discussing the Poisson counting statistics of a rather extreme (but nonetheless important) class of “one time” measurements of rare (or sparse) events. To frame this discussion in quite general terms, we imagine instruments that accumulate counts from two independent Poisson processes (A and B). We don’t require that these measurements be made simultaneously, or even with the same instrument. The sources could be cosmic ray particles, photons, or energetic neutral atoms. What counts are the counts. To emphasize this, we designate the counts obtained from process A by (M) and from process B by (N). One way of comparing the two processes is to form the “abundance ratio” \( R = M/N \), but this is obviously a poorly behaved statistic, because when the counts are sparse, the denominator (N) could easily be zero because of fluctuations in the counts. On the other hand, the “group abundance fraction” \( F = M/S \) where \( S = M+N \) (the total counts in the “group”) is much better behaved, as long as we can find a way to deal with the possibility that the denominator \( S = 0 \). We still have to allow
for the finite probability that there might be no counts at all (S=0, so M=N=0), and we will do that by introducing a properly defined conditional probability.

Here are a couple of comments before we get into the details of our approach. Firstly, physicists usually think in terms of expected rates, not counts, and they usually have some feeling for what the “true” rate should be, so a measurement with \( S=0 \) could possibly be used to estimate upper limits on the combined processes (A and B). This approach, if properly formulated, would lead us into Bayesian statistics and therefore beyond the scope of this paper. We will therefore not attempt the interpretation of the special case (S=0), but rather we will exclude it from our analysis by rigorously defining a conditional probability that requires \( S\geq 1 \). This conditional probability will provide the statistical description we need to extract quantitative information about the processes (A and B) from the one-time observation of the counts (M and N). Secondly, does it matter which species we choose for the numerator of the fraction \( F=M/(M+N) \)? It really doesn’t. If our observations indicated that \( F<1/2 \), then we would be analyzing the minor component (M<N), and we will be most interested in developing formulas in the limit as \( F\to 0 \). On the other hand, if \( F>1/2 \), we would be analyzing the major component (M>N), so we should demand that our formulas also be valid as \( F\to 1 \). However, in the latter case, what we are really interested in quantifying is how close (F) is to unity. The difference from unity is \( 1-F=1-M/(M+N)=N/(M+N)<1/2 \), so in this latter case we should just revert to the former analysis by simply switching our choices for the labels of M and N and then utilize our original formulas for \( F\to 0 \). If the counts (M,N) are comparable, then it doesn’t matter which component we analyze.

In order to proceed, we must be very clear in our nomenclature. Although it may be a bit redundant at times, we will indicate (when it is necessary to make the distinction) the actual one-time measured counts by \( M^* \) and \( N^* \). Unlike experiments usually analyzed in statistics textbooks (e.g., [1] which is particularly clear on nomenclature and illuminating examples), where one has (K) multiple samples of the same measureable quantity, like \( X_k \) (1\( \leq k\leq K \)), we have only a single sample of the pair of numbers (\( M^*, N^* \)) from our two different processes (A and B). Nonetheless, we can actually obtain information on A and B (even with this paucity of observational information), if we know beforehand that A and B are independent Poisson processes, because each Poisson process is completely specified by a single parameter, i.e., the expected number of accumulated counts (which we shall call \( C_A \) or \( C_B \), respectively).

1.1. The simple Poisson process

Let us regress to a brief discussion of a simple (uni-variate) Poisson process in order to clarify some issues relevant to our more general study and (in doing so) to illustrate some important terminology. Explicitly, this Poisson distribution is the well-known (but not always well-understood) probability that one will obtain \( n^* \) counts from a single actual measurement

\[
P_n(c) = \frac{e^{-c} c^n}{n!} \]

We will use \( n^* \) to indicate a particular measurement of the counts, while we will use \( n \) as the summation variable when calculating ensemble averages over a theoretical ensemble of all possible measurements weighted by the probability distribution \( P_n(c) \). Although by its definition as a real measurement, the observed number of counts (\( n^* \)) is an integer, the true number of counts (c) that specifies the Poisson distribution need not be an integer. This is because in its most familiar formulations, the “true” number of counts \( c=rT \) is defined in terms of an average true rate (r) of counts accumulated over a time T. Neither the true rate (r) nor the true number of counts (c) can ever be precisely measured; they can only be estimated from the data. Indeed, that estimation process is the challenge that we face when interpreting data.

One calculates the expectation value of any function \( x(n) \) of the number of counts (n) by averaging \( x(n) \) over the Poisson distribution function \( P_n(c) \). We will call \( x(n) \) a count function, because it may help us to estimate
(in some way) an ensemble parameter like \( c \). However, we’ll stick with the jargon-free terminology of “count function”. For the Poisson distribution, the expectation value \( y_x(c) \) is the ensemble average \( \{x(n)\} \) of count function \( x(n) \) given by the sum

\[
\{x(n)\} = \sum_{n=0}^{\infty} x(n)P_n(c) = y_x(c)
\]

Clearly the expectation value \( \{x(n)\} \) for the count function \( x(n) \) will be some corresponding function \( y_x(c) \) of the (sole) ensemble parameter \( c \), and the latter will depend functionally on the form of the mathematical form of count function \( x(n) \). However, more generally one needs to speak of a (possibly non-linear) ensemble function \( z(c) \) formed from a combination of the expectation values \( y_x(c), y_{x'}(c), \) etc., that are, in turn, the set of expectation values for the functions \( x(n), x'(n), \) etc. It is very important to note that if the relation defining the ensemble function \( z(c) \) is non-linear in the expectation values \( y_x(c), y_{x'}(c), \) etc., then (except under the extraordinary circumstance to be described immediately below), \( z(c) \) may not itself be calculable as a simple expectation value (average) over the distribution function.

Let us illustrate these potentially confusing relationships with some well-known examples from the simplest Poisson distribution \( P_n(c) \) of Eq. (1). For the count functions, let us take the number \( n^* \) itself, so that \( x_1(n)=n \), and its square \( (n^*)^2 \), so that \( x_2(n)=n^2 \). The familiar corresponding ensemble (unconditional) expectation values are \( \{n\}=y_{x_1}(c)=c \) and \( \{n^2\}=y_{x_2}(c)=c^2+c \). A widely used measure of the dispersion about the mean value of any count function is its variance, which we will write as \( \sigma^2=\langle (n-c)^2 \rangle \) in order to avoid confusion with \( \sigma=\langle n \rangle \) the familiar standard deviation (SD). In general, the variance is a non-linear combination of these expectation values: \( \sigma^2=\langle (n-\{n\})^2 \rangle=\{n^2\}-\{n\}^2=cy_{x_2}(c)-cy_{x_1}(c) \). Thus for a general probability distribution function, we see that the variance is not generally expressible itself as a simple expectation value (obtained by a simple weighted sum over the probability distribution), because it is a non-linear function of expectation values. However, in the special case of the Poisson distribution, the variance reduces to the surprisingly simple expression \( \sigma^2=cn \), where we already know that \( c=\{n\} \). In other words, the value of non-linear variance is just the expectation value of \( n^* \) itself. This peculiar (and defining) property of the Poisson distribution permits the deceptively simple estimation of the expectation value and SD of the mean from a single observation: we have to assume that the observed number of counts \( n^* \) is likely to be “close” to the expectation value \( \{n\}=c \) averaged over the Poisson ensemble specified by the ensemble parameter \( c \), the ”true” counts. It is the unique property of the Poisson distribution (\( \sigma^2=cn \)) that allows us to make the familiar estimate with its uncertainty (from just a single observed value \( n^* \) of the ”true” counts \( c \) as \( n^* \pm \sigma(n^*) / \sqrt{n} \), assuming the closeness of \( n^* \) to \( c \).

We can quantify the “closeness” just mentioned above by the concepts of “unbiased” and “biased” estimators of an ensemble function \( z(c) \), using a single measurement \( n^* \) and assuming a Poisson distribution. For example, we call \( x_1(n^*)=n^* \) an unbiased estimator of the true counts \( c \) because for its expectation value, we have \( \{x_1(n)\}=y_{x_1}(c)=c \). By this definition, the observed value \( n^* \) is always an unbiased estimator of \( c \), the average of \( \{n\} \) over the ensemble. However \( x_1(n^*)=n^* \) is a biased estimator of the square of the true counts squared \( c^2 \), because \( \{x_1(n)\}=y_{x_2}(c) \). The bias can either be expressed additively, as written, or multiplicatively, as \( c^2(1-1/c) \). Either way, the biasing term becomes negligible for \( c>>1 \), so the biased estimator can still be useful in the limit of high counts. However, there is an unbiased estimator for \( c^2 \); it is a new count function \( x_2(n^*)=x_1(n^*)x_1(n^*)=n^*n^*-n^*/n^*n^*n^*(n^*-1) \), because \( \{n(n-1)\}=y_{x_2}(c) \). This works, by the way, because the expectation value of the estimator \( \{n(n-1)\} \) is a linear combination of expectation values, \( y_{x_2}(c) \). In general, the trick is to find an unbiased (or weakly biased) estimator like \( x_2(n^*) \) for the ensemble function of interest, like \( z(c)=c^2 \). That will be our approach in the following Sections where we will seek the mean and variance of the group abundance fraction for the full bivariant probability distribution. Please bear in mind that the estimator for the variance, being a function of the observed counts, will itself still be a random number that may differ significantly from its expectation value, but that’s often the best we can do. All such uncertainties are exacerbated when we are dealing with small numbers of counts.
2. Anticipation of results
All the foregoing discussion may seem to be formal overkill when applied to the simple case of a univariate unconditional Poisson distribution. However, the distinctions we have highlighted will prove to be crucial in deriving the properties of the group abundance fraction from a bivariate (conditional) Poisson distribution. The derivation involves a considerable amount of mathematical manipulation of infinite series, but surprisingly it leads to some results that are so simple and useful that we would like to anticipate them at the outset. If we observe $M^*$ counts from process A and $N^*$ counts from process B for a total of $S^* = M^* + N^*$ counts from both processes, then an unbiased estimator of the “true” fraction ($f$) of A-counts to the total (A and B)-counts is

$$f \rightarrow F^* = \frac{M^*}{S^*}$$

| Unconditional Joint Distribution $W(M,N)$ | Ensemble Expectation value | Estimator (unbiased) | Estimator (biased) |
|------------------------------------------|---------------------------|---------------------|-------------------|
| $\{M\}$                                  | $C_A = fC$                | $M^*$               | ---               |
| $\{N\}$                                  | $C_B = (1-f)C$            | $N^*$               | ---               |
| $\{S\} = \{M+N\}$                        | $C$                      | $S^* = M^* + N^*$   | ---               |
| $\sigma^2M = \{M^2\} - \{M\}^2$         | $C_A$                    | $M^*$               | ---               |
| $\sigma^2N = \{N^2\} - \{N\}^2$         | $C_B$                    | $N^*$               | ---               |

| Binomial Distribution $D(f|S)$ ($S \geq 1$) |                      |                     |                   |
|-----------------------------------------------|-----------------------|---------------------|-------------------|
| $<M>_S$                                      | $[f+(1-f)]^S = 1$     | ---                 | ---               |
| $<M>_S$                                      | $fS$                  | $M^*$               | ---               |
| $<M>_S$                                      | $fS + f^2S(S-1)$      | ---                 | ---               |
| $<M(S-M)>_S$                                 | $f(1-f)S(S-1)$        | ---                 | ---               |
| $\sigma^2 f_s = <F>_S - <F>_S^2$            | $f(1-f)/S$            | $F^*(1-F^*)/(S^*-1)$ |                   |

| Conditional Joint Distribution $U(M,S)$ ($S \geq 1$) |                      |                     |                   |
|---------------------------------------------------|-----------------------|---------------------|-------------------|
| $<F> = <M/S>$                                      | $f$                   | $M^*/S^*$           | ---               |
| $<M>$                                             | $fC/Q_0(C)$            | ---                 | $M^*$             |
| $<N>$                                             | $(1-f)C/Q_0(C)$       | ---                 | $N^*$             |
| $<S>$                                             | $C/Q_0(C)$             | ---                 | $S^*$             |
| $\sigma^2F = \sigma^2F^* - <F>_S^2$             | $f(1-f)g_1(C)$        | (?)                 | (?)               |
| $<\sigma^2F^* > = <F(1-F)/S>$                    | $\sigma^2[F(1-g_1(C)/g_1(C)]$ | --- | $F^*(1-F^*)/S^*$ |
| $<\sigma^2F^*> = <F(1-F)/(S-1)>$                 | $\sigma^2 F^* + [1-f(1-f)]C/\exp(C-1)$ | --- | $F^*(1-F^*)/(S^*-1)$ |

A biased estimator for the variance of the true group abundance fraction $\sigma^2F$ is

$$\sigma^2F \rightarrow \sigma^2F^* = \frac{F^*(1-F^*)}{(S^*-1)}$$

i.e., the appropriately defined ensemble expectation value of $<F(1-F)/(S-1)>$ is not exactly equal to $\sigma^2F$. Although biased (particularly for low total counts), $\sigma^2F^*$ on average will be greater than its ensemble expectation value $\sigma^2F$ and hence will on average give a conservative estimate of the true variance. These formulas (to be developed in detail in the rest of this paper) are valid in the context of a conditional joint probability distribution for $(M,N)$ wherein the occurrence of $S=0$ total counts is excluded.
Table 1 has been introduced at this point to aid the reader in following the inter-relationships among the formulas as they are introduced. The notation \( \{x(M,N)\} \) in the table is reserved for an average of a count function \( x(M,N) \) over the unconditional bivariate Poisson distribution \( W(M,N) \), while \( \langle x(M,S) \rangle \) is the average of \( x(M,S) \), where \( S=M+N \), over the binomial distribution \( D_M(f|S) \) for the group abundance fraction in which the total counts (\( S \)) are assigned a priori. Finally \( \langle x(M,S) \rangle \) will indicate the average of \( x(M,S) \), over the conditional Poisson distribution \( U(M,S) \) with the restriction \( S \geq 1 \) (to be developed later). And so we begin.

3. The unconditional bivariate Poisson distribution

We turn to the main calculations of this paper, valid for sparse data where \( M^* \) and/or \( N^* \) are only a few counts. If the Poisson processes are separate, they will be statistically independent. Thus, from the definition in Eq. (1), the joint probability distribution of the counts will be given by

\[
W(M,N) = P_M(C_A) P_N(C_B) = \frac{C_A^M}{M!} \exp(-C_A) \frac{C_B^N}{N!} \exp(-C_B)
\]  

where \( C_A \) and \( C_B \) are the expected number of counts from processes A and B, respectively. We are not using the \((M^*,N^*)\) notation in Eq. (3), because \((M,N)\) will be the running indices when we calculate ensemble averages over the bivariate (joint) Poisson distribution. As is well known, the expectation values averaged over the unconditional joint probability distribution \( W(M,N) \) of the counts \((M,N)\) and their variances are both equal to the “true” counts \((C_A \text{ and } C_B, \text{ respectively})\). See Table 1. At this point, we take leave of these familiar formulas, because our goal is to calculate exactly the expectation values of the group abundance ratio \( F=M/(M+N) \) over a conditional joint probability distribution. We introduce the ensemble parameter for the “group abundance fraction” \((f)\) into the notation by writing \( C_A=fC \) and \( C_B=(1-f)C \) for the expected counts from the A and B processes, respectively, and \( C=C_A+C_B \) for the total expected counts from the two processes. We also write \( S=M+N \) for the sum of the observed counts, and therefore set \( N=S-M \). Since we must have \( N \geq 0 \), this requires that \( 0 \leq M \leq S \). We thereby obtain a well-known transformation of the unconditional joint probability

\[
W(M,S-M) = f^M/M! \ (1-f)^N/N! \ C^S \exp(-C)
\]

where we have regrouped the remaining factors and introduced the notation \( B(M|S) \) for the binomial coefficient (that emphasizes the roles played by \( M \) and \( S \)).

\[
B(M|S)=S!/M!(S-M)!
\]

One then recognizes that \( W(M,S-M) \) is the product of a Poisson distribution (in \( S \)) and a binomial distribution (in \( M \)):

\[
W(M,S-M) = P_S(C) D_M(f|S) \quad 0 \leq M \leq S
\]

We have written \( P_S(C) \), as in Eq. (1), for the Poisson distribution of the total counts \((S)\) specified by the expected total counts \((C)\) for the ensemble, and \( D_M(f|S) \) for the binomial distribution of the A-counts \((M)\) specified by the ensemble fraction \( f \) of the total counts \((S)\), \( i.e. \), \( f=C_A/C \).

\[
P_S(C) = \exp(-C) \ C^S/S! \quad D_M(f|S)=B(M|S) \ f^M(1-f)^{S-M} \quad 0 \leq M \leq S
\]

Although we just derived the transformed unconditional bivariate distribution through a mathematical manipulation, we could have written it down directly, based on two well-known results from probability...
theory. Firstly, the additive combination (sum) of two independent Poisson processes is itself a Poisson process, i.e., the simple probability distribution (in terms of the total number of counts $S=M+N$) must be the Poisson distribution $P_S(C)$, since it represents the combination of the two Poisson distributions $P_M(C_A)$ and $P_N(C_B)$, where $M+N=S$ and $C_A+C_B=C$. Secondly, if we know beforehand that $S$ total counts occurred from the combined processes (A+B) and that the A counts are a fraction ($f$) of the total number of counts, then the A-process counts ($M$) selected at random must be distributed according to the binomial distribution $D_M(fS)$ generated by the individual coefficients of the expansion of $(f+(1-f))^S$. Thus the bivariate probability distribution of $W(M,S-M)$ must be the product of the conditional probability of obtaining ($M$) counts from the A process – if there were ($S$) counts from the combined A+B process (binomial distribution) – multiplied by the simple probability that ($S$) counts were actually obtained from the combined A+B process (Poisson distribution). That’s why Eq. (4) has the form of Eq. (6).

3.1. High-count limit

We can easily see how the joint distribution behaves in the high count limit ($S>>1$). As is well known, the simple Poisson distribution $P_S(C)$ approaches a sharply peaked Gaussian with mean value ($C$) and standard deviation ($C^{1/2}<C$). In the limit $C\to\infty$, the Gaussian becomes an expression for the delta function, $\delta(S-C)$, and, moreover, the observed total counts ($S^*$) very likely will give a good estimate for the expected total counts ($C$). Thus, in somewhat loose notation, we can consider that the high-count limit of the joint distribution must be

$$W(M,S-M)\to \delta(S-S^*)D_M(fS) \quad (8)$$

This approximation was the starting point of a statistical analysis of the group abundance ratio by Gehrels [2] in which he developed confidence limits on the estimate of ($f$) based on the observed counts ($M^*,N^*$). His basic two assumptions were that he could estimate $f$ by $M^*/S^*$ and $C$ by $S^*=M^*+N^*$. The confidence limits were then developed from the properties of the remaining binomial distribution in Eq. (8), and the results were presented in numerical tables. His analysis is summarized in Appendix A. Further work along this line [3] introduced Bayesian statistics.

3.2. The properties of the binomial distribution

However, using the same Eq. (8), we could take an approach mathematically independent from the confidence limit formalism. Rather than computing confidence limits on $F=M/S$ for the binomial distribution $D_M(fS)$, we can calculate the expectation values $\langle F \rangle_S$ and the variance $\sigma^2_F = \langle F^2 \rangle_S - \langle F \rangle_S^2$ from the same binomial distribution $D_M(fS)$. We have introduced the notation $\langle X \rangle_S$ for the expectation value for a count function $X(S)$ over the binomial distribution for a pre-assigned value of $S$ (such as the observed total counts $S^*$). Then we will be able to express the uncertainties in our result as a “one-sigma” expression $\langle F \rangle_S \pm \sigma_F S$. Consider the expectation values for the binomial moments of the form ($M^k$).

$$\langle M^k \rangle_S = \sum_{M=0,S} M^k D_M(fS) = \sum_{M=0,S} M^k B(M|S) f^M (1-f)^{S-M}$$

For the reader’s convenience, these familiar results are re-derived in Appendix B and summarized in Table 1. The immediately relevant ones are $\langle M^1 \rangle_S = fS$ and $\langle M^2 \rangle_S = (fS)^2 + Sf(1-f)$.

$$\langle F \rangle_S = \langle M/S \rangle_S = \langle M \rangle_S/S = f \quad (9)$$

$$\sigma^2 F_S = \langle (M/S)^2 \rangle_S - \langle M/S \rangle_S^2 = \langle M^2 \rangle_S/S^2 - \langle M \rangle_S^2/S^2 = [f^2 + f(1-f)/S] - f^2 = f(1-f)/S \quad (10)$$
Figure 1. Comparison of expectation values for the group abundance fraction $F^* = M^*/(M^*+N^*)$ and its uncertainty for very low counts $(M^*,N^*)$. Solid lines: $F^*_S \pm \sigma_{F^*_S}$ from the “one-sigma” binomial estimators, Eq. (12) of this paper. Dashed lines: $F \pm CL$ from binomial distribution confidence limits, from [2]. The agreement between these two quite different estimation techniques is quite good.

It turns out that there is an unbiased estimator $(\sigma^2 F^*_S)$ for $\sigma^2 F_S$ in terms of the observed counts. Curiously, it is not $F^*(1-F^*)/S^*$, but rather it is $\sigma^2 F^*_S = F^*(1-F^*)/(S^*-1)$, because (as we can see from the binomial moments in Table 1)

$$\langle \sigma^2 F^*_S \rangle_S = \langle F(1-F)/(S-1) \rangle_S = \langle M_2 \rangle_S / S(S-1) - \langle M^2 \rangle_S / S^2(S-1)$$

$$= (fS)/S(S-1) - [(fS)^2 + Sf(1-f)]/S^2(S-1) = [fS^2 - f^2S^2 - Sf + Sf^2]/S^2(S-1) = f(1-f)/S \quad (11)$$

We can then represent the estimated results for the binomial distribution in terms of a two-sided “one-sigma” uncertainty

$$F^*_S \pm \sigma_{F^*_S} = (M^*/S^*) \pm [M^*(S^*-M^*)/S^2(S^*-1)]^{1/2} \quad (12)$$

In the six panels of Figure 1, we compare this simple algebraic formula for the estimator of $F^*_S \pm \sigma_{F^*_S}$ with the confidence limits from the tables of Gehrels [2] for single-sided confidence levels of 0.8413.
(corresponding to one-sigma for a Gaussian). Each of the six panels of Fig. 1 plots the \( M^*/(M^*+N^*) \), the estimator for \( f \), in diamonds) for a fixed values of the \( A \)-counts \( (M^*=1,2,3,4,5, \text{ and } 10) \). Each plot (whose abscissa covers the \( B \)-count range \( 1 \leq N^* \leq 10 \)), also compares the confidence limits (CL from Table 6 of Gehrels [2], dashed curves) with the algebraic estimator for \( (f \pm \sigma_F) \) from Eq. (12) (solid curves). Figure 1 reveals that the one-sigma estimates of \( F^* \pm \sigma_F \) using our biased algebraic estimator \( \sigma^2 F^* = F^*(1-F^*)/(S^*-1) \) agree rather well for even very small values of \( (M^*,N^*) \) with the confidence limits tabulated by Gehrels [2] in his Table 6. Our one-sigma upper limit almost always falls below his upper confidence level estimate (for \( N^*>1 \)) and our one-sigma lower limit almost always lies above his lower confidence level estimate (except for \( M^*=1 \) where our lower limit is always identically equal to zero).

We would like to obtain similar results that are valid for the entire range of expected total counts \( (C\geq0) \), unencumbered by the high-count approximation \( S^*>>1 \) that approximates the joint probability distribution \( W(M,S-M) \) by the delta function and the binomial distribution \( D_M(f|S) \), as in Eq. (8). Gehrels [2] anticipated that his confidence levels should “give somewhat conservative limits” when \( S^* \) was a small integer. We shall see that this is indeed borne out by the following analysis in which we will obtain a one-sigma estimator valid for all values of total observed counts \( S^*\geq1 \). We will also demonstrate that this more general estimator \( (\sigma^2 F^*) \) will turn out to be identically the same as that just obtained for the \textit{unbiased} binomial estimator, namely \( \sigma^2 F^* = F^*(1-F^*)/(S^*-1) \) in Eq. (11). However, for small values \( (1\leq S^* \leq 5) \), our algebraic estimator \( \sigma^2 F^* \) will give a \textit{biased overestimate} of \( \sigma^2 F \). Thus, the “true” expectation values of “one-sigma” uncertainties will be even smaller than indicated in Fig. 1, \textit{i.e.}, they would lie between the solid curves and further inside the dashed curves (from Table 6 of [2]). Consequently, \( \sigma^2 F^* \) will be a \textit{conservative overestimate} of the true expectation value of the variance \( (\sigma^2 F) \) for all allowed values (\( S^*\geq1 \)) of the total counts. It will require the following two Sections to establish quantitatively this qualitative result.

4. The conditional bivariate Poisson distribution \( (S\geq1) \)

We would like to find good estimators (in terms of the observed counts \( M^* \) and \( N^* \)) not only the expectation value \( (f) \) for \( F \), but also for its \textit{variance} \( (\sigma^2 F) \) about \( f \). However, we immediately can see a difficulty in calculating the required moments of \( F=M/S \). Any moment \( \{F^k\} \) with \( k>0 \) over the unconditional bivariate Poisson distribution \( W(M,S-M) \) will be undefined when \( S=0 \) because it will include the term \( M=N=S=0 \) for which \( F \) is indeterminate. The way out of the difficulty lies in acknowledging (see the related comment in Section 1), that we gain very little physically useful information on the abundance fraction if no counts at all were accumulated from either process \( (i.e., \ 0/0 \text{ is indeterminate}) \). Fortunately, within the basic laws of probability, there is a mathematical method for rigorously \textit{excluding} the special case \( (S=0) \) from our analysis by re-casting it in terms of a \textit{conditional probability} \( U(M, S) \) which properly describes all the remaining cases \( (S\geq1,0\leq M \leq S) \). We don’t ignore the possibility that \( S=0 \); rather we just do proper “book-keeping” in terms of probabilities. Such “book-keeping” of zeros is essential to the treatment of sparse data, \textit{e.g.}, [4].

Since these two cases \( (S=0 \text{ and } S\geq1) \) are mutually exclusive, their probabilities must be additive. The \textit{unconditional} joint probability \( W(M,N) \) – with the variable change \( N=S-M \) so that \( 0\leq M \leq S \) -- must be the sum of the simple probability that \( S=0 \) plus the \textit{conditional} probability \( U(M, S) \) multiplied by the simple probability that \( S\geq1 \).

\[
W(M, S-M) = \delta_{S,0} \text{ prob}(S=0) + (1-\delta_{S,0}) \text{ prob}(S\geq1) U(M, S)
\]

However, in our notation given in Eq. (6),

\[
W(M, S-M) = \delta_{S,0} P_0(C) D_0(f, 0) + (1-\delta_{S,0}) [1- P_0(C) D_0(f, 0)] U(M, S)
\]  

(13)
because \( \text{prob}(S=0) = P_0(C) = \exp(-C) \) and \( \text{prob}(S \geq 1) = 1 - \text{prob}(S=0) = 1 - \exp(-C) \). We have used the Kronecker logical symbol \( \delta_{S,0} \) to clarify the relationship of the terms. By definition, \( D_0(f|0) = 1 \). Also, because the combination will occur so often, we will designate

\[
Q_0(C) = \text{prob}(S \geq 1) = 1 - P_0(C) = 1 - \exp(-C)
\]

Eq. (13) then defines the conditional joint probability distribution function \( U(M,S) \) for the condition \( S \geq 1 \).

\[
U(M,S) = (1 - \delta_{S,0}) W(M,S-M)/Q_0(C) = (1 - \delta_{S,0})/Q_0(C) P_S(C) D_M(f|S) \quad S \geq 1 \text{ and } 0 \leq M \leq S \tag{14}
\]

Henceforward, we will suppress the factor \( (1 - \delta_{S,0}) \), since is understood to be an intrinsic part of the definition of \( U(M,S) \), because it exists if and only if \( S \geq 1 \).

4.1. Conditional expectation value of \( \langle F \rangle \) the group abundance fraction (\( S \geq 1 \))

We are now prepared to calculate the conditional expectation value of the group abundance fraction \( \langle F \rangle = \langle M/S \rangle \) under the condition \( (S \geq 1) \).

\[
\langle F \rangle = \langle M/S \rangle = 1/Q_0 \sum_{S=1}^\infty \sum_{M=0,S} (M/S) U(M,S) = 1/Q_0 \sum_{S=1}^\infty P_S(C) \sum_{M=0,S} (M/S) D_M(f|S)
\]

Because the separability of the conditional probability \( U(M,S) \), the inner summation will be over moments of the form \( \langle M^k \rangle_S \) from the binomial distribution \( D_M(f|S) \). These moments have been discussed above in Eqs. (9) and (10), calculated in Appendix B, and summarized in Table 1. For example the first conditional bivariate moment is

\[
\langle F \rangle = \langle M/S \rangle = P_0/Q_0 \sum_{S=1}^\infty C^S/S! \times (1/S) <M>_S = (P_0/Q_0) f \sum_{S=1}^\infty C^S/S!
\]

This elementary summation will occur so often that we will assign it the symbol

\[
I_0(C) = \sum_{S=1}^\infty C^S/S! = \exp(C) - 1
\]

In fact, the other commonly occurring factor is

\[
P_0(C)/Q_0(C) = \exp(-C)/[1-\exp(-C)] = 1/I_0(C)
\]

Substituting these functions, we have the important result that \( F^* \) itself is an unbiased estimator of the ensemble parameter \( f \) over the bivariate conditional probability distribution \( U(M,S) \) because

\[
\langle F \rangle = f \tag{15}
\]

4.2. Discussion of \( \langle F \rangle \)

Eq. (15) contains our first important result: the Poisson expectation value \( \langle M/S \rangle \) (under the condition \( S \geq 1 \)) exactly equals the ensemble fraction parameter \( f \). In other words, the abundance fraction of the observed counts (\( F^* = M^*/S^* \)) is an unbiased estimator of the ensemble fraction parameter \( f \). This lack of bias is a non-trivial result, as can be seen from the expectation values of the counts \( (M, N, \text{and } S) \). These are quickly obtained from Table 1.

\[
\langle M \rangle = fC/Q_0 \quad \langle N \rangle = \langle S-M \rangle = (1-f)C/Q_0 \quad \langle S \rangle = C/Q_0
\]
In other words, when we exclude the possibility that M=N=S=0, the conditional expectation value of \( \langle M \rangle \) is no longer simply the abundance fraction \( f \) multiplied by the number of counts \( C \). It, as well as the expectation values for \( \langle N \rangle \) and \( \langle S \rangle \) are all biased by the factor \( 1/Q_0(C) \). This highlights the important fact that the lack of bias in \( \langle F \rangle = f \) is actually a consequence of our use of the conditional probability. Interestingly enough, \( \langle F \rangle = \langle M \rangle/\langle S \rangle \), i.e., the unbiased expectation value \( \langle F \rangle \) equals the ratio of the biased expectation values for \( M \) and \( S \).

The formulas reveal why there must be biasing for very few counts. Consider their limiting values as the expected counts \( C \to 0 \). Without the biasing denominator, we would have \( \langle M \rangle, \langle N \rangle, \text{and} \langle S \rangle \) all three going to zero, in contradiction to our condition \( S \geq 1 \). However, because \( C/Q_0 \to 1 \) when \( C \to 0 \), i.e., for very rare events, we have instead.

\[
\lim_{C \to 0} \langle M \rangle = f \\
\lim_{C \to 0} \langle N \rangle = 1-f \\
\lim_{C \to 0} \langle S \rangle = 1
\]

These limits make perfect sense. When the events are very rare, we must then have \( S=1 \), because we are imposing the condition \( S \geq 1 \), i.e., we must still measure at least a single count. Under this circumstance \( (S=1) \), the conditional expectation value for \( \langle M \rangle \) should be the ensemble abundance fraction \( f \) itself, and that of \( \langle N \rangle \) should be \( (1-f) \), because the probability for a single total count coming from species A must be \( f \), and for a single total count coming from species B must be \( (1-f) \).

### 4.3. Conditional expectation value: variance \((\sigma^2 F)\) of the group abundance fraction \((S \geq 1)\)

The variance of the abundance fraction is \( \sigma^2 F = \langle F^2 \rangle - \langle F \rangle^2 \), where

\[
\langle F^2 \rangle = \langle M^2/S^2 \rangle = (1/Q_0) \sum_{S=1}^{\infty} P_S(C) \sum_{M=0,S} (1/S^2) \langle M^2 \rangle_S
\]

\[
= (1/I_0) \sum_{S=1}^{\infty} C^S/S! \left( (S^2)/(S-1) \right) = (1/I_0) \sum_{S=1}^{\infty} C^S/S! \left( f(1-f)/S + f^2 \right)
\]

\[
= f(1-f)/I_0 \sum_{S=1}^{\infty} C^S/S! + f^2 (I_0/I_0) = f(1-f)g_1(C) + f^2
\]

\[
\sigma^2 F = \langle F^2 \rangle - \langle F \rangle^2 = f(1-f)g_1(C)
\]

(16)

An important new function \( g_1(C) \) has appeared in our formula for the variance of \( F \). It is actually only one of a more general set defined by the normalized sums \( I_k(C) \)

\[
g_k(C) = I_k(C)/I_0(C) \\
I_k(C) = \sum_{S=1}^{\infty} C^S/S!S^k \\
k \geq 0
\]

(17)

where it so happens that for \( k=0 \), the summation reduces to \( \exp(C)-1 = I_0(C) \); in fact, that is why we chose that symbol for that function. Please note that, despite the similar notation, the infinite sums \( I_k(x) \) are not modified Bessel functions. We need to know how \( g_1(C) \) varies with \( C \), but it will be just as easy to examine the properties of the general set of functions \( g_k(x) \) for \( k \geq 1 \) as it is for just \( k=1 \) alone. Some useful properties of these functions are derived, and they are evaluated in the two extreme ranges \( C<<1 \) and \( C>>1 \) in Appendix C. The functions \( g_1(C) \) actually behave quite simply, decreasing monotonically away from \( g_1(0)=1 \) for \( C<<1 \) and asymptotically towards zero as \( 1/C^k \) for \( C>>1 \). The behaviors of \( g_1(C) \) and \( g_2(C) \) are graphed in Figure 2, along with \( 1-g_2(C)/g_1(C) \), over the range \( 0 \leq C \leq 20 \).

### 4.4. Statistical significance: “n-sigma” ratios for the group abundance fraction

We can now introduce “n-sigma” statistical significance properties of the group abundance fraction \( F \) in terms of the ensemble expectation mean \( \langle F \rangle = f \) and its expected variance \( (\sigma^2 F) \). Let us designate the “n-sigma” ratio as \( (\eta) \) and define its square by
Various probability functions that appear in the expectation values of the variance of the Group Abundance Fraction ($\sigma^2 F$) and its estimators.  See Table 1 for a summary and specific discussions in the text.

$$\eta^2 = \frac{<F>^2}{\sigma^2 F} = \frac{f^2}{f(1-f)g_1(C)} \approx \frac{C}{1/f-1} \quad (18)$$

We could use function $g_1(C)$ in the subsequent discussion, but it is much easier to use the approximation $g_1(C) \approx 1/C$ which is good within $\pm 20\%$ for $1<C<3.5$ and then becomes $\approx +10\%$ for $C>10$, i.e. over our practical range of interest. See Fig. 2. As examples of its utility, we can quickly answer two important questions (relevant to making a low-count observation) in terms of the ensemble parameters. First, how many true counts ($C$) are required to measure a true abundance fraction (say $f=0.1$) to a significance ratio of $\eta=1, 2, \text{ or } 3$ sigma? From Eq. (18), $C=\eta^2(1/f-1) = \eta^2(9)$, so the answer is $C>9$, 36, or 81 total counts, respectively. Second, what is the smallest true abundance fraction ($f$) that can be measured (statistically significantly) with $C=9$ counts for different significance ratios ($\eta=1, 2, \text{ or } 3$)? Also from Eq. (18), $f=1/(1+C/\eta^2) = 1/(1+9/\eta^2)$, so the answers are $f=0.10(0.90)$, $0.31(0.69)$, or $0.50(0.50)$ respectively. There are double values for $f$, because inspection of Eq. (18) reveals that if $f$ is a solution, so is its complement $(1-f)$. See the related comment in Section 1. The degree to which these answers concerning ($\eta$) are practically meaningful obviously depends on how close our data-based estimators ($F^*$ and $\sigma^2 F^*$, to be derived below) are, on average, to the ensemble expectation values for $f$ and $\sigma^2 F$, calculated in Eqs. (15) and (16), respectively. For now, a crude estimate on the uncertainty range of ($\eta_{\text{est}}$) can be obtained from Eq. (18) by ignoring any uncertainty in $\sigma^2 F$ and writing $\eta < \eta_{\text{est}} < \eta + 1$

$$(f-\sigma f)/\sigma f < \eta_{\text{est}} < (f+\sigma f)/\sigma f$$

or

$$\eta - 1 < \eta_{\text{est}} < \eta + 1$$

5. Estimator ($\sigma^2 F^*$) for the variance ($\sigma^2 F$)

In order to make best use of our observed values ($M^*, N^*$), we need to obtain an estimator ($\sigma^2 F^*$) from the data for the expectation value of the variance ($\sigma^2 F$) given in Eq. (16). We already know from Eq. (15) that we can use $F^* = M^*/S^*$ as an unbiased estimator of ($f$). What we therefore need for practical
applications is a “statistic” for $\sigma^2 F$, i.e., a count function (call it $\sigma^2 F^\ast$) of the observed counts $(M^\ast, S^\ast)$ whose expectation value (call it $<\sigma^2 F^\ast>$), will perform be “close” to the function $\sigma^2 F = f(1-f)g_1(C)$, i.e., we seek an acceptably biased estimator ($\sigma^2 F^\ast$) for $\sigma^2 F$ in terms of the observed counts $(M^\ast$ and N$^\ast$). If found, we could present our measurement of $F^\ast=M^\ast/S^\ast$ with a “one-sigma” uncertainty as $F^\ast\pm\sigma F^\ast$, where $\sigma F^\ast=(\sigma^2 F^\ast)^{1/2}$.

We were encouraged that it might be possible to find a simple algebraic expression (in terms of $M^\ast$ and $N^\ast$) for the estimator $\sigma^2 F^\ast$ by the fact that, for the binomial distribution (in which $S$ is set equal to $S^\ast$), we found in Eq. (11) an unbiased estimator for $\sigma^2 F_S = f(1-f)/S$ in terms of the observed counts, namely $\sigma^2 F^\ast_S = F^\ast(1-F^\ast)/(S^\ast-1)$, because $<\sigma^2 F^\ast>_S = \sigma^2 F_S$. See Table 1. At first, the form of $\sigma^2 F = f(1-f)g_1(C)$ suggested that we try a statistic of the form

$$\sigma^2 F^\ast = (1-\sigma^2 F^\ast)/S = M^\ast(S^\ast-M^\ast)/S^\ast$$

because we have shown in Appendix C (and Fig. 2) that $g_1(C)=1/C$ as long as $C>1$, i.e., over a useful range of “small” numbers of total counts. However, there is an immediate difficulty for our conditional probability distribution (requiring $S \geq 1$) at the smallest limit $S^\ast=1$, because this estimator ($\sigma^2 F^\ast$) would yield exactly zero for either $M^\ast=0$ or $M^\ast=1$ (where $N^\ast=S^\ast-M^\ast=0$). This is an unacceptable limit for an estimator, because it says that there is no uncertainty at all in the estimate of $\sigma^2 F$ for very small counts, and that is nonsense. In fact, we can quickly compute its expectation value for all values of the ensemble parameters ($f$ and $C$) by substituting the binomial moments $<M^k>_S$ in Table 1 into our summation over $(S \geq 1)$ used above.

$$<\sigma^2 F^\ast> = <F(1-F)/S> = <M(S-M)/S^3> = (1/I_0) \sum_{s=1}^{\infty} C^S/S!S^3 f(1-f)S(S-1) = f(1-f)/I_0 \sum_{s=1}^{\infty} C^S/S! (1/S-1/S^2)$$

The resulting multiplicative aliasing factor $(1-g_2/g_1)$ is plotted in Fig. 2 and its limiting behavior is discussed in Appendix C. It asymptotically approaches unity as $(1-1/C)$ for $C>>1$ and zero as $(C/8)$ for $C<<1$. Thus it is always $\leq 1$ for finite values of $(C)$, so $<\sigma^2 F^\ast>$ always underestimates the true variance. Even though (as just mentioned above) its zero-trending behavior in the low-count limit is unacceptable, we shall see below that it still yields a useful lower bound for the true value $\sigma^2 F$ for $C>1$.

Because it must enter into the averaging summation over all $S \geq 1$ (including $S=1$), we need an estimator that makes “sense” and predicts a finite variance, even for the extreme value $S^\ast=1$ (one total count) when either $M^\ast=0$ or $M^\ast=1$. Remarkably, an estimator that does work has precisely the form of the binomial estimator $\sigma^2 F_S$ from Eq. (11) and Table 1.

$$\sigma^2 F^\ast = \sigma^2 F_S = (1-M^\ast)/S^\ast \times [(S^\ast-1)/(S^\ast-1)] = 1$$

At first glance, the estimator would appear to be singular at $S^\ast=1$ due to the factor $(S^\ast-1)$ in the denominator. However, when $S^\ast=1$, $M^\ast$ can only take on the values $M^\ast=0$ or $M^\ast=1$. In either case, the numerator is also zero. Actually, we can write $M^\ast=S^\ast-1$ in the first case, and $S^\ast-M^\ast=S^\ast-1$ in the second. Thus in both cases the zeros in the numerator and denominator form the ratio $(S^\ast-1)/(S^\ast-1)=1$. One could make this argument more rigorously in terms of “limits” $S^\ast \rightarrow 1$ and $M^\ast \rightarrow 0$ or $M^\ast \rightarrow 1$, but it doesn’t really matter because, being an estimator, we are free to define the value of the fraction in these limiting cases, as long as we treat it consistently in our summations over the distribution function.
\[ \sigma^2 F^*|S^*=1, M^*=1 = M^*/S^*^2 \frac{(S^*-1)/(S^*-1)} = 1 \]  

(21)

Then, according to our convention, the variance \( \sigma^2 F^*=1 \) for \( S^*=1 \) (a single count) for either \( M^*=0 \) and \( N^*=1 \) or \( M^*=1 \) and \( N^*=0 \). We argue that this makes sense, because for these values, either \( F^*=0 \) or \( F^*=1 \) and we have no way of knowing how correct either value is. The uncertainty \( \sigma F^*=(\sigma^2 F^*)^{1/2} \) is therefore well represented by unity.

5.1. Calculation of the expectation value for the estimator \( \sigma^2 F^* \)

We can now apply this convention to an exact calculation of the expectation value for the estimator over our conditional probability distribution. In this case, though, it is necessary to split off the first term (with \( S=1 \)) from the outer summation.

\[ <\sigma^2 F^*>= \frac{1}{I_0} \sum_{s=1, \infty} C^S/S!S^2 \sum_{M=0,5} M(S-M)/(S-1) D_M(f|S) \]

\[ = C/I_0 \sum_{M=0,1} M(1-M)/(S-1) D_M(f|1) + \frac{1}{I_0} \sum_{s=2, \infty} C^S/S!S^2 \sum_{M=0,5} M(S-M)/(S-1) D_M(f|S) \]

By the arguments given above, the ratio \( M(S-M)/(S-1) \) should be set to \( 1 \) for either \( M=0 \) or \( M=1 \), whereas \( D_0(f|1)=f(1-f)^0=1-f \) and \( D_1(f|1)=f1(1-f)^0=f \). Thus our convention leads to

\[ \sum_{M=0,1} M(1-M)/(S-1) D_M(f|S) = (1-f) + (f) = 1 \]

The remaining summation (\( S\geq2 \)) is easily evaluated from the formula in Table 1 and Appendix B.

\[ 1/I_0 \sum_{s=2, \infty} C^S/[S! S^2(S-1)] <M(S-M)>S/ = f(1-f)/I_0 \sum_{s=2, \infty} C^S/S!S = f(1-f) [g_1(C)-f(1-f)C/I_0] \]

which, when re-united with the \( S=1 \) term, yields our desired result.

\[ <\sigma^2 F^*> = f(1-f) g_1(C) + [1-f(1-f)C/I_0(C)] \]

(22)

Since the term \( f(1-f)\leq1/2 \) and \( C/[\exp(C)-1]\leq1 \), the bracketed quantity in Eq. (22), which constitutes an additive bias for the estimator of \( \sigma^2 F \), is always between 0 and 1/2. In other words, the expectation value of the estimator \( <\sigma^2 F^*>=<F(1-F)/(S-1)> \) always overestimates the exact value \( \sigma^2 F=\sigma^2 F^* \). In fact, decreasing to a zero bias (as \( C\rightarrow\infty \)). The bias reflects the fact (demonstrated above) that the estimator for the abundance fraction \( \sigma^2 F^*=M^*/N^*/S^2(S^*-1) \) is observationally quite uncertain as \( C\rightarrow0 \) under the condition \( S^*\geq1 \), because then \( S^* \) must take the value \( S^*=1 \) in this limit of very low counts, where \( F^*=M^*/S^* \) is then going to have to take either the value \( F^*=0 \) or \( F^*=1 \).

The behavior of the multiplicative bias in the two estimators \( \sigma F^*/\sigma F \) and \( \sigma F^{*+}/\sigma^2 F \) is graphed in Figure 3 as the square root of the ratio of the biased estimator to the true value. Using the somewhat sloppy notation \( \sigma F^*=<\sigma^2 F^*>^{1/2} \) and \( \sigma F^{*+}=<\sigma^2 F^{*+}>^{1/2} \)

\[ \sigma F^*/\sigma F = [<\sigma^2 F^*>/\sigma^2 F]^{1/2} \quad \text{and} \quad \sigma F^{*+}/\sigma F = [<\sigma^2 F^{*+}>/\sigma^2 F]^{1/2} \]

The curves demonstrate that the former ratio always exceeds unity (for any value of \( f \)), while the latter ratio is always less than unity (independent of the value of \( f \)). Another interpretation of the plot is that the true expectation value \( \sigma^2 F \) must lie in between the two biased expectation values \( \sigma^2 F^* \) and \( \sigma^2 F^{*+} \).
Therefore we recommend the estimator

$$\sigma F^* = \left[ F^*(1-F^*)/(S^*-1) \right]^{1/2} \quad (23)$$

Figure 3. Ratios $<\sigma F^*>/\sigma F = \left[ <\sigma^2 F^*>/\sigma^2 F \right]^{1/2}$ and $<\sigma F^*>/\sigma F = \left[ <\sigma^2 F^*>/\sigma^2 F \right]^{1/2}$ as a function of the number of counts (C) and selected values for the group abundance fraction (f). The quantities $<\sigma^2 F^*>$ and $<\sigma^2 F'^*>$ are biased estimators of the “true” variance ($\sigma^2 F$). See Table 1 and the discussion in the text.

as a conservatively biased estimator for the uncertainty ($\sigma F$) in the estimated value $F^* = M^*/S^*$ for the group abundance fraction (f):

$$f \pm \sigma F \rightarrow M^*/S^* \pm \left[ F^*(1-F^*)/(S^*-1) \right]^{1/2} \quad (24)$$

Having obtained in Eq. (24) our desired result, we next take up two issues that are essential in the analysis of actual observations. These go beyond the mathematics of Poisson probability, because they involve the inescapable physical characteristics of the instruments that obtain the counts.

6. Relationship between counts and intensities

Our statistical analysis concerned itself with the “true” counts ($C_A$ and $C_B$) or equivalently (f and C), because those are the sole parameters that specify the Poisson distributions for species A and B. However, the physically significant quantities are the incident uni-directional differential intensities being measured ($j_A$ and $j_B$). If the counts were both accumulated by the same detector over the same time interval, then the abundance fraction of the true counts (f) is also that of the intensities (h)

$$f = C_A/(C_A+C_B) \quad h = j_A/(j_A+j_B)$$
The general relation between the counts and intensities is \( C_A = \rho A j_A \), where the proportionality constant \( \rho A = (TG \Delta E)_A \). Here \( T \) is the accumulation time, \( \Delta E \) is the energy pass-band, and \( G \) is the efficiency-weighted geometry factor (e.g., in units of cm\(^2\)sr). Then there is generally a somewhat more complicated relation between \( f \) and \( h \) since \( C_A = \rho A j_A \) and \( C_B = \rho B j_B \)

\[
f = \frac{1}{1 + \frac{\rho B j_B}{\rho A j_A}}
\]

Figure 4. Solid curves: The group intensity abundance fraction \( h = j_A / (j_A + j_B) \) vs. the group count abundance fraction \( f = C_A / (C_A + C_B) \). Dashed curves: approximate one-sigma uncertainties in the fractions owing to the variance in the count fraction \( \sigma^2 F \). The four panels are all for \( C = 5 \) total counts, but the detector response difference \( (d_{AB}) \) takes different values.

Since \( h = 1 / (1 + j_B / j_A) \) we have \( j_B / j_A = 1 / h - 1 \) and

\[
1/f = 1 + (\rho_B/\rho_A)(1/h-1)
\]

This is easily solved for the intensity group abundance ratio \( h = j_A / (j_A + j_B) \).

\[
h = f / [f + (\rho_A/\rho_B)(1-f)]
\]

The solution is well behaved, in that that \( f = 0 \) implies \( h = 0 \) and \( f = 1 \) implies \( h = 1 \). For intermediate values of \( f \), we always have \( h < 1 \) because the term in the denominator with the factor \( (1-f) \) is always positive.

But, what is the uncertainty in intensity fraction \( (h) \) due to the uncertainty in the group abundance fraction \( f \)? The most straightforward (although crude) two-sided “one-sigma” estimate is \( h < h < h \), where

\[
h = f / [f + (\rho_A/\rho_B)(1-f)]
\]

\[
f = \pm \sigma F(C)
\]

\[
\sigma^2 F = f(1-f)g_1(C)
\]  

(26)
These formulas are all expressed in terms of the ensemble parameters \( f \) and \( C \). They are plotted \( h \) vs. \( f \) in Figure 4 for choices of the total number of counts \( C \) and the channel detector response ratio \( d_{AB} = \rho_A/\rho_B \). To estimate from the actual data the group abundance ratio of the incident intensities \( h = j_A/(j_A+j_B) \), we would use our estimators (based on the observed counts \( M^* \) and \( N^* \)) for the ensemble parameters

\[
f \rightarrow F^* = M^*/(M^*+N^*) = M^*/S^* \quad \text{and} \quad \sigma^2 F^* = F^*(1-F^*)/(S^*-1)
\]

7. **Inclusion of counting backgrounds**

Experimentally, there is still an additional complication, that of the backgrounds in actual detectors. We can easily include Poisson counting backgrounds (individually) into each of the processes \((A \text{ and } B)\), since the sum of two Poisson processes is a Poisson process. We simply add background counts \( k_A \) and \( k_B \) to the foreground counts \( c_A \) and \( c_B \) for two the two species \((A \text{ and } B)\). This will again yield a Poisson process. We will assume that the average rates for the backgrounds are known with high precision previous calibration measurements. Then, by the foregoing arguments, the ensemble group abundance fraction is

\[
f = \frac{(c_A+k_A)}{(c+k)} \quad \text{so that} \quad 1-f = \frac{(c_B+k_B)}{(c+k)} \quad (27)
\]

where \( c = c_A + c_B \) is the total of the foreground counts and \( k = k_A + k_B \) is the total of the background counts. We will assume that the expectation values for the background counts \((k_A \text{ and } k_B)\) are well-estimated by previous calibration measurements.

By the theorem for combining independent Poisson processes, the observed counts from process \(A\) \((\text{before imposing the condition } S=M+N \geq 1)\) will have the expectation value \( C_A = c_A+k_A \), and those from process \(B\) will have \( C_B = c_B+k_B \), with \( C = c+k \) expected for the total observed counts. We therefore can directly write

\[
<F> = f = \frac{(c_A+k_A)}{(c+k)} \quad (28)
\]
Let us rewrite the expectation value of the background-affected abundance fraction \( f \) in terms of the (unobservable) ensemble foreground fraction \( f_0 = c_{A}/c \). Then

\[
f = \frac{(f_0 + k_{A}/c)}{1 + k/c}
\]

so that

\[
f_0 = \frac{(1 + k/c)f - k_{A}/c}{c} = f + \frac{(kf - k_{A})}{c}
\]

(29)

The linear relation \( f \) vs. \( f_0 \) reveals the effect of backgrounds on the expectation value of the group abundance fraction (as a function of the “true” foreground fraction \( f_0 \)). The relation depends only on the ratios of the A and B backgrounds \( (k_{A} \) and \( k_{B} \)) to the total foreground counts \( c \). It is plotted in Figure 5 (left panel) for the particular background/foreground ratios \( k_{A}/c=0.3 \) and \( k_{B}/c=0.5 \). One sees immediately that the expectation value \( f \) cannot be smaller than \( k_{A}/(k+c) \), nor can \( f \) exceed the ratio \( (k_{A}+c)/(k+c)<1 \).

What about the expectation value of the variance \( \sigma^2 \)? By the just-cited properties of the sums of independent Poisson processes, the formula in the presence of backgrounds remains

\[
\sigma^2 = f(1-f)g_1(C) = \frac{(c_{A}+k_{A})(c_{B}+k_{B})}{(c+k)^2} g_1(C) = \frac{[Cg_1(C)](f_0c+k_{A})[(1-f_0)c+k_{B}]/(c+k)^3}{(f_0+x_{A})(1-f_0+x_{B})/(1+x)^3}
\]

(30)

We see that after multiplying the equation through by the number of foreground counts \( c \), the RHS involves only background/foreground ratios: \( x_{A} = k_{A}/c, x_{B} = k_{B}/c, x = x_{A} + x_{B} = k/c, \) and \( \Delta x = k_{B} - k_{A} \). The divisor on the LHS \( [Cg_1(C)] \) is nearly unity (within ±20%, see Fig. 1), so if we set to unity we get the convenient approximation

\[
\sigma^2 \approx \frac{(f_0^2+(1+\Delta x)f_0 + x_{A}(1+x_{B}))/(1+x)^3}{(f_0+x_{A})(1-f_0+x_{B})}
\]

(31)

We will call the RHS of this approximation the “variance factor”, and its square root the “SD factor”. The LHS contains the only explicit dependence on the foreground counts \( c \). The SD factor is plotted in Figure 5 (center panel) for the specific parameter values \( x_{A} = k_{A}/c = 0.3 \) and \( x_{B} = k_{B}/c = 0.5 \). Again, as for \( f \) vs. \( f_0 \), the RHS depends only upon the ratios of the backgrounds \( (k_{A} \) and \( k_{B} \)) to the total foreground counts \( c \). The approximate value of \( \sigma^2 \) itself is obtained (to good approximation) by dividing the plotted “SD factor”, the square root of the RHS of Eq. (31), by the square root of the total number of foreground counts \( c^{1/2} \). Considered as a function of the true fraction \( 0 \leq f_0 \leq 1 \), the numerator is the square root of a downward-opening parabola whose value at \( f_0 = 0 \) is \( x_{A}(1+x_{B}) \) and at \( f_0 = 1 \) is \( x_{B}(1+x_{A}) \). Thus the variance is always positive over the interval \( 0 \leq f_0 \leq 1 \). The maximum ordinate is attained at \( f_0 = (1+\Delta x)/2 \) (which is not necessarily within the valid interval \( 0 \leq f_0 \leq 1 \)), where its value is \((1+\Delta x)^2/4 + x_{A}(1+x_{B})\).

We now can revisit our statistical significance “n-sigma” discussion (Section 4.3), with instrument background effects included. Although \( \eta^2 = f^2/\sigma^2 \), it is algebraically more convenient to first consider its inverse.

\[
\eta^2 = \sigma^2 f^2 = (1/f-1)g_1(C) = [Cg_1(C)] [(1+x)/(f_0+x_{A})-1]/(c+k)
\]

Multiplying the equation through by the foreground counts \( c \) and re-arranging terms as we just did in obtaining \( f \) and \( \sigma^2 \), we get

\[
c\eta^2/[Cg_1(C)] = (1-f_0+x_{B})/[(f_0+x_{A})(1+x)]
\]

(32)
Again, the RHS depends only upon the background/(total foreground) ratios. Then the square root of the RHS of the inverse we call the n-sigma factor. Since \( \sqrt{Cg(C)} \approx 1 \) is usually a good approximation, we take the RHS as an approximation for \( \sqrt{\eta/c} \). The RHS is plotted vs. \( f_0 \) in Figure 5 (right panel).

\[
\eta/c^{1/2} = (f_0 + x_A)^{1/2}/((1-f_0 + x_B)(1+x))^{1/2}
\]

(33)

The n-sigma value itself (\( \eta \)) is obtained by multiplying the ordinate by the square root of the total foreground counts (\( c^{1/2} \)).

All of the foregoing background analysis was conducted in terms of “true” ensemble parameters, but how can we estimate (\( f_0 \)) and (\( \eta \)) using the observed counts (\( M^*, N^* \))? It may not be worth the effort for very low counts, because unless \( S^* \gg k \) (total observed counts exceeding the total background counts), the Poisson fluctuations in the background contamination (\( \sigma^2k_A = k_A \) and \( \sigma^2k_B = k_B \)) will likely be severe enough to preclude obtaining a useful estimate (\( f_0^* \)) of the ensemble value for the foreground abundance fraction (\( f_0 \)) from the observed counts (\( M^* \)) and \( S^* = M^* + N^* \).

8. Numerical simulations of the Group Abundance Fraction

It may be helpful to the reader to relate all of the foregoing calculations and relationships to some numerical simulations of the Group Abundance Fraction \( F^* = M^*/S^* \). Figure 6 illustrates clearly the nature of the Poisson fluctuations of the group abundance fraction when we suppress occurrences of \( S^* = 0 \). It presents the result of 100 Poisson (random-number-generated) simulations of individual observations of \( M_k^* \) and \( N_k^* \), with \( S_k^* = M_k^* + N_k^* \) (1\( \leq k \leq 100 \)). This was a conditional simulation in which we excluded trials with \( M^* = N^* = S^* = 0 \). The results then should be consistent with our conditional bivariate Poisson probability (\( S^* \geq 1 \)) distribution given by Eq. (14). The ensemble parameters for these 100 trials were always set at \( c = 5 \) counts (for the combined A and B processes) and \( f = 0.2 \) (which implies that the ensemble value \( fC = 1 \) count for A-process). In the upper panel, the results of the 100 trials are presented vs. the trial index (\( k \)) as the group abundance fraction. Its estimated uncertainty shown by error bars is calculated from our algebraic estimators \( F^* = M^*/(M^* + N^*) \) and \( \sigma^2 F^* = (F^*(1-F^*)/(S^*-1))^{1/2} \). The lower panel plots the simulated counts \( M_k^* \) (lower points) and \( S_k^* \) (upper points), and it is evident that about 1/3 of the \( M^* \)-counts equal zero. It is also clear that the \( S^* \) (total counts) fluctuate significantly about ensemble value (\( C = 5 \)), justifying our use of the full bivariate Poisson distribution, as opposed to assuming that \( S^* = C \) (exactly).

Returning to the upper panel of Figure 6, the ensemble value of the group ensemble fraction (\( f = 0.2 \)) is indicated by a horizontal dashed line, as are the ensemble expectation values of \( F^* \) for the combined A and B processes and \( f = 0.2 \) (demonstrating some consistency with the estimated uncertainties). On the other extreme, almost 1/3 of the trials had \( F^* = 0 \) because \( M^* = 0 \) (compare with the lower panel). It is precisely because of these \( F^* = 0 \) trials that there also had to be a set of \( F^* \)-values lying above \( F^* = f + \sigma f \) (and thus compensating for the many \( F^* = 0 \) values), since our estimator demands that the over-all average expectation value \( \langle F^* \rangle = f \).

Thus Fig. 6 illustrates the relationships among our three classes of numbers: firstly, the ensemble parameters (\( f, C \)), secondly the expectation values \( \langle F^* \rangle = f \) and \( \sigma^2 F = f(1-f)g(C) = 0.04112 \) for \( C = 5 \) giving the ensemble expectation value for the SD of \( f = 0.2828 \). The average values for the 100-trial run agree with the expectation values within a fraction of 1%. A few dozen trials have fractions (\( F^* \)) above the one-sigma level, but the majority of their lower error bars reach down to the line \( f = 0.2 \) (demonstrating some consistency with the estimated uncertainties). On the other extreme, almost 1/3 of the trials had \( F^* = 0 \) because \( M^* = 0 \) (compare with the lower panel). It is precisely because of these \( F^* = 0 \) trials that there also had to be a set of \( F^* \)-values lying above \( F^* = f \) (and thus compensating for the many \( F^* = 0 \) values), since our estimator demands that the over-all average expectation value \( \langle F^* \rangle = f \).

Thus Fig. 6 illustrates the relationships among our three classes of numbers: firstly, the ensemble parameters (\( f, C \)), secondly the expectation values \( \langle F^* \rangle = f \) and \( \sigma^2 F = f(1-f)g(C) \), which are functions of the ensemble parameters (\( f, C \)), and thirdly, the estimators \( F^* \) and \( F^* = (1-F^*)/(S^*-1) \) for \( \langle F^* \rangle \) and \( \sigma^2 F \), respectively. It is remarkable that even with the large fluctuations (\( \delta S^* \)) in the total counts (\( S^* \)), wherein \( \delta S^* \approx C = 5.0 \) counts, the biased estimator (\( \sigma^2 F^* \)) for the variance of the group abundance fraction (\( f \)) from the conditional bivariate probability distribution has turned out to be \( i(\sigma^2 F^*) \) identical to the unbiased estimator (\( \sigma^2 F^* \)) for the binomial probability distribution with wherein \( C \) estimated by \( S^* \) itself. See Eqs. (8) and (20) and the discussions thereof.
Figure 6. Results of 100 Poisson conditional random-number “trials” (excluding those with $S^*=0$). 
Upper panel. The observed group abundance fraction $F^*=M^*/S^*=M^*/(M^*+N^*)$ (diamonds) with twosided error bars $F^*\pm \sigma^2F^*$ estimated by $\sigma^2F^*=F^*(1-F^*)/(S^*-1)$. The horizontal lines are the ensemble values $f\pm \sigma_f$ for $\sigma^2f=f(1-f)g(1)$ with $f=0.2$ and $C=5$. Note the numerous points with $F^*=0$ (corresponding to $M^*=0$) along the bottom axis. These zero counts are a consequence of the Poisson fluctuations about $M^*=fC=1$ in this low-count regime. Lower panel. The randomly-generated values of $M^*$ and $S^*$ for the ensemble parameters: $M$-counts=$(0.2)(5)=1.0$ and total counts $C=5$.

8.1. Cumulative distribution function for $F$: Poisson realizations specified by $(f,C)$
We use Figure 7 make a final point. One might imagine that there is a well-behaved probability distribution function for the group abundance fraction. Indeed, one does exist, but only in the strict limit $C\to\infty$. However, it may not be generally realized that for small values of $C$, the probability distribution is a rather peculiar and ill-behaved mathematical object. We can demonstrate this with numerical simulations of the cumulative unconditional distribution function $Z(F;f,C)$. This is the probability that the observed fraction ($F^*$) does not exceed the value $F$, i.e., that $F^*\leq F$. As is well-known, $Z(F;f,C)$ may be simulated by using a Poisson random number generator to produce a large number ($K$) of pairs $(M^*_k,N^*_k)$
using the same ensemble parameters $C_A = fC$ and $C_B = (1-f)C$ for $1 \leq k \leq K$. We ran K=30,000 trials. One then sorts the corresponding fractions ($F^*_k$) in ascending order from zero to one and plots them vs. their ordering index (divided by K). Such a plot is shown in the left-hand panel of Figure 7 for C=5 counts and an ensemble fraction $f=0.2$. Since the abscissa gives the fraction of all of the realizations where $F^*$ was less than or equal to the ordinate (F), the cumulative distribution function (CDF), the probability that $F^* \leq F$ as a function of F, is immediately obtained by interchanging the ordinate and the abscissa, as displayed in the right-hand panel of Figure 7.

**Figure 7.** Poisson simulation (constructed from 30,000 trials of $M^*$ and $N^*$) of the cumulative probability distribution function $Z(F|f,C)$ for the Group Abundance Fraction $F=M/(M+N)$ for ensemble parameters $f=0.2$ and $C=5$ total counts. **Left panel:** Sort of the K=30,000 values of $F^*$ plotted by the normalized realization number ($k/30,000$). **Right panel:** Cumulative probability distribution $Z(F|f,C)$, the fraction of occurrences with $F^* < F$. The discontinuities in the function (marked by dashed lines) are produced by different pairs of counts $(M,S)$ “near” $M=1$ and $S=5$ having the same ratio $F=M/S$. Consequently, for low numbers of total counts ($S$), the ordinary probability density (normally given by $dZ/dF$) will contain “spikes” at the discontinuities of $Z$.

What immediately strikes the eye is the plethora of discontinuous jumps in the CDF. These discontinuities are not an artifact of the simulation, because it was performed with K=30,000 Poisson realizations. The jumps are an intrinsic property of the CDF, and their cause is revealed by the vertical dashed lines (which correspond to the horizontal dashed lines in the original plot in the left-hand panel). They have been drawn where $F=M/S$ takes the values of a ratio of small integers: 0.16667=1/6, 0.20=1/5, 0.25=1/4, 0.33333=1/3, 0.40=2/5, 0.5=1/2, etc. The vertical jumps occur at particular values of F in the right-hand plot, because there several realizations that give the same value of F for different pairs $(M,N)$ in the left-hand plot. Thus pairs with the same $F$ may have a different number total counts $S=M+N$. Such multiplicities cannot appear in the binomial approximation wherein S is set to the specific observed value $(S^*)$. See Eq. (8). The jumps in $F=M/S$ due multiplicity in S can only appear in the mathematical form of the CDF when the exact distribution function $P_S(C)$ is retained in the conditional distribution function $U(M|S)$ given by Eq. (14).

For the case simulated, we chose $C=5$ and $f=0.2$, so the most likely value of $M^*$ is $fC=1$. This means that the realizations $F^*=M^*/S^*$ are quite likely to take values with $M^*$ near 1 and $S^*$ near 5, like 1/6, 1/5 (or 2/10), 1/4 (or 2/8), 1/3 (or 2/6 or 3/9), 2/5, 1/2 (or 2/4 or 3/6 or 4/8). The higher likelihood for multiplicities of $F=1/3$ or $F=1/2$ yields larger discontinuities than at $F=2/5$. Note that this effect was not produced by our choosing integer values for the ensemble parameters $C=5$ and $fC=1$. The discontinuities would appear just as strongly if we had chosen $C=5.35$ and $fC=0.8$ or 1.2. The strong discontinuities always appear at values of $F=M/N$ where M and N are small numbers (as long as $C$ itself is a small
number). Of course, the discontinuities become undetectably small for $C \gg 1$ because then there are many rational fractions (no longer restricted to the ratio of small integers) in the vicinity of any value of $F$.

When one has a CDF for a continuous process (like a Gaussian), then one can obtain the usual probability “density” (or probability distribution function, PDF) by differentiating the CDF with respect to its argument. However, in our situation with $C$ a small integer (as revealed by the RH plot of Fig. 7), such a differentiation will produce a series of delta-function-like “spikes” in the PDF at each of the jumps. In other words, one cannot develop a “smooth” approximation to the CDF by any fitting of smooth (finite-order) polynomials in $Z(F)$, because it is effectively not differentiable.

Despite the intrinsic “roughness” of the CDF, our expressions for the ensemble expectation values $\langle F \rangle = f$ and $\sigma^2 F = f(1-f)g_1(C)$ are exact, because no approximations were made in the summation calculations over the joint Poisson probability distribution. The numbers presented within the RH panel of Fig 7 are the computed values from the 30,000 realizations of $\langle F \rangle = 0.2006$ and its SD is $\sigma F = 0.2052$. Our ensemble expectation values are $f = 0.2000$ and $\sigma^2 F = f(1-f)g_1(C) = (0.2)(0.8)(0.287) = 0.04112$ (taken from the exact formulas), giving the ensemble value for the SD of $\sigma F = 0.2028$. The agreement of both is within 0.1%.

9. Summary and Conclusion
As we anticipated from the outset in Section 2, the results of this paper are quite simply summarized by the algebraic expression, from Eq. (24), for the estimator (in terms of the measured counts $M^*$ and $N^*$ with total $S^*$ from two independent Poisson processes) for the ensemble expectation value ($f$) of the group abundance fraction $F = M/(M+N)$ and its one-sigma uncertainty ($\sigma F$)

$$f \pm \sigma F \rightarrow M^*/S^* \pm [F^*(1-F^*)/(S^*-1)]^{1/2}$$

The estimator for ($f$) is statistically unbiased, while that for the standard deviation ($\sigma F$) is conservatively biased, in that (on average over the ensemble), it always overestimates the “true” value ($\sigma F$) for very low total counts, but approaches it asymptotically beyond $S^* > 5$. See Fig. 3. These results were obtained by introducing an exactly valid conditional Poisson joint probability distribution function $U(M,S)$ satisfying the requirement of non-zero total measured counts ($S^* \geq 1$). See Eq. (14). The interesting relationship to earlier results [1] based on an approximately valid binomial distribution is traced in Section 3 and Appendices A and B. The final three sections (6, 7, and 8) of the paper are devoted to topics not (to our knowledge) addressed similarly in the literature: the relationship (due to instrument response functions) between the measured group abundance fraction for the counts and that for the incident intensities; the effect of backgrounds in the counting detectors on the estimate of the foreground group abundance fraction; and the peculiar mathematical properties of the cumulative probability distribution function for the group abundance fraction.

Acknowledgements This research was partially supported by the Voyager Interstellar Mission under NASA Contract NNX07AB02G, by the Cassini Mission under Task Order 003 of NASA Contract NAS5-97271, and by the ACE Mission under NASA Grant NNX13AR20G (all to the Johns Hopkins University). I thank Dr. Tycho Von Rosenvinge of NASA/GSFC and Dr. Alan Cummings of the Space Radiation Laboratory/CalTech for directing my attention to the confidence limit analyses of Niel Gehrels [2] and De Angelis and Iori [3], respectively. I am particularly indebted to the Editor, Dr. Gary P. Zank, for allotting me sufficient pages for a full development and explanation of the ideas in this paper.

References
[1] Lindgren, B W 1976 Statistical Theory (3rd edition), Macmillan (New York)
[2] Gehrels, N 1986 Confidence limits for small number of events in astrophysical data, Astrophys. J., 303, 336-346
[3] De Angelis, A and Iori, M 1987 Errors on ratios of small numbers, Nucl. Inst. Methods Physics. Res.
Appendix A. Confidence limits approach

The joint probability distribution \( W(M,S-M) \) given in Eq. (6) was the starting point for a thorough confidence limit analysis of the group abundance fraction by Gehrels [2]. However, he then avoided using the Poisson distribution of the total counts \( P_S(C) \) by setting the population parameter (C) precisely equal to the observed number of total counts \( (S^*) \). He noted that the remaining binomial distribution for \( (M) \) was a function solely of the ensemble abundance fraction \( (f) \). He then tabulated single-sided uncertainties \( (f_U \) and \( f_L \)) for a specific confidence limit (CL) using the following equations for a binomial distribution (re-expressed in our current notation)

\[
\sum_{M=0}^{M^*-1} B(M|S^*) f_L^M (1-f_L)^{S^*-M} = CL
\]

\[
\sum_{M=0}^{M^*} B(M|S^*) f_U^M (1-f_U)^{S^*-M} = 1-CL
\]  

(A.1)

where \( M^* \) and \( S^* \) are the observed counts. The latter equation is equivalent to

\[
\sum_{M=M^*+1}^{S^*} B(M|S^*) f_U^M (1-f_U)^{S^*-M} = CL
\]

for any value of the ensemble fraction parameter \( (f) \). These equations needed to be solved numerically for the parameters \( f_L \) and \( f_U \) as functions of \( M^* \) and \( S^* \). He also took \( F^*=M^*/(M^*+N^*) \) as the estimator for the ensemble fraction \( (f) \), so his confidence limits were \( f_L<F^*<f_U \). He states that when both \( M^* \) and \( N^* \) “are both small, these equations will give somewhat conservative upper limits”, as a result of fixing the ensemble value for the total counts \( (C) \) equal to the observed number of counts \( (S^*) \). Extensive tables were presented of the computations. In a later paper [3], other authors took a similar approach to the binomial distribution, but introduced Bayesian estimation in order to also incorporate the Poisson distribution \( P_S(C) \). Again, tables of computed quantities were presented.

We saw (in Section 5) that our new results, which take quite a different approach that exactly incorporates the Poisson distribution \( P_S(C) \), are nonetheless consistent with Gehrels’ [2] confidence limits to quite good accuracy. However, our simple algebraic formula of Eq. (24) is much simpler to apply than the computed tables.

Appendix B. Moments of the binomial distribution

We begin with a basic iterative identity for the binomial coefficient \( B(M|S) \)

\[
M \ B(M|S) = M \ S!/M!/(S-M)! = S!/(M-1)!/(S-M)! = S \ B(M'|S')
\]

However, we must first ensure that \( M \geq 1 \) so that the factorial \( (M-1)! \) in the denominator has a non-negative argument. Then, since we already require \( S \geq 1 \) for the conditional probabilities, both \( M'=M-1 \) and \( S'=S-1 \) in the identity. A similar useful identity is

\[
M(M-1) \ B(M|S) = M(M-1) \ S!/M!/(S-M)! = S!/S-2)!/(S-M)! = S(S-1) \ B(M''|S'')
\]

in which \( M''=M-2 \) and \( S''=S-2 \) requires that \( M \geq 2 \) and \( S \geq 2 \). We now evaluate several moments containing \( M^k \), with \( k=0,1,2 \) over the binomial distribution with \( (S) \) total counts, as indicated by the notation \( <M^k>_S \) introduced with Eq. (8).
\[
\langle M^f \rangle_S = \sum_{M=0,S} B(M|S) f^M (1-f)^{S-M} = [f+(1-f)]^S = 1 \tag{B.1}
\]

For the first moment (M itself), the first term (M=0) in the summation vanishes, so the original range \(0 \leq M \leq S\) becomes \(1 \leq M \leq S\) which (because \(S \geq 1\)) can be shifted to \(0 \leq M' \leq S'\).

\[
\langle M^1 \rangle_S = \sum_{M=1,S} MB(M|S) f^M (1-f)^{S-M} = \sum_{M=1,S} S B(M'|S') f^M (1-f)^{S-M} = Sf \sum_{M'=0,S'} B(M'|S') f^{M'+1} (1-f)^{S'-M'} = Sf \tag{B.2}
\]

Similarly, for the compound moment \(M(M-1)=M^2-M\), the first two terms in the summation vanish, so the original range \(0 \leq M \leq S\) becomes \(2 \leq M \leq S\) which (because \(S \geq 1\)) can be shifted to \(0 \leq M'' \leq S''\) because \(2 \leq M \leq S\) requires that that \(S \geq 2\).

\[
\langle M(M-1) \rangle_S = \sum_{M=2,S} M(M-1) B(M|S) f^M (1-f)^{S-M} = \sum_{M=2,S} S(S-1) B(M'|S') f^M (1-f)^{S-M} = S(S-1)f^2 \tag{B.3}
\]

Therefore

\[
\langle M^2 \rangle_S = \langle M(M-1) \rangle_S + \langle M \rangle_S = S(S-1)f^2 + Sf \tag{B.4}
\]

and we also obtain the compound moment \(M(S-M)=SM-M^2\).

\[
\langle M(S-M) \rangle_S = S\langle M \rangle_S - \langle M^2 \rangle_S = S(fS) - (fS)^2 - (Sf)(1-f) = f(1-f)S(S-1) \tag{B.5}
\]

in which the first term in the expanded factors is restricted by \(1 \leq M \leq S\) and \(S \geq 1\), while the second and third are restricted by \(2 \leq M \leq S\) and \(S \geq 2\).

**Appendix C. Properties of the functions \(I_k(C)\) and \(g_k(C)\)**

Consider the infinite sum (absolutely convergent for \(k\) non-negative)

\[
I_k(x) = \sum_{r=1,\infty} x^r/r!^k \quad \text{where} \quad I_0(x) = \exp(x)-1
\]

The normalized summations then define our set of probability functions

\[
g_k(x) = I_k(x)/I_0(x) \tag{C.1}
\]

For the range \(x<<1\), we can make use of the generating function for the Bernoulli number (\(B_n\))

\[
x/[\exp(x)-1] = x/I_0(x) = \sum_{n=0,\infty} B_n x^n/n! \quad B_n=1,-1/2,1/6,\ldots \text{for } n=0,1,2,\ldots
\]

with Abel’s summation formula for the product of two convergent power series.

\[
g_k(x) = I_k(x)/I_0(x) = (1/x)I_k(x) \sum_{n=0,\infty} B_n x^n/n! = \sum_{r=0,\infty} x^r/(r+1)!x^k/(r+1)^k \sum_{n=0,\infty} B_n x^n/n! = \sum_{r=0,\infty} x^{r+k}/(r+1)! \sum_{n=0,\infty} B_n x^n/n!
\]
\[ \sum_{s=0}^{\infty} x^s \sum_{n=0,s} B_n/n!(s-n+1)!(s-n+1)^k \quad s=r+n \]

\[ B_0 + (B_0/2!2^k + B_1)x + (B_0/3!3^k + B_1/2!2^k + B_2/2!)x^2 + O(x^3) \]

Since the first few Bernoulli numbers are \(B_0=1, B_1=-1/2, B_2=1/6\), we have the behavior near zero

\[ g_k(x) = 1 - \frac{x}{2}(1-1/2^k) + \frac{x^2}{12}(1-3/2^k+2/3^k) + O(x^3) \]

\[ g_0(x) = 1 \quad \text{identically} \]

\[ g_1(x) = 1 - \frac{x}{2}(1-1/2) + \frac{x^2}{12}(1-3/2+2/3) + \ldots = 1 - \frac{x}{4} + \frac{x^2}{72} + \ldots \]

\[ g_2(x) = 1 - \frac{x}{2}(1-1/4) + \frac{x^2}{12}(1-3/4+2/9) + \ldots = 1 - \frac{3x}{8} - \frac{17x^2}{432} + \ldots \]

\[ g_3(x)/g_2(x) = x/8 + 23x^2/432 + \ldots \]

\[ 1 - g_2(x)/g_2(x) = x/8 + 73x^2/864 + \ldots \quad \text{(C.2)} \]

At the other extreme of the range \(x>>1\), we can make use of the identity (obvious from the definition)

\[ \frac{dI_k(x)}{dx} = \left( \frac{1}{x} \right) I_{k-1}(x) \quad \text{(C.3)} \]

combined with l’Hospital’s rule (because the limits of the derivatives of the numerator and denominator exist).

\[ \lim_{x \to \infty} x^k g_k(x) = \lim_{x \to \infty} x^k I_k(x)/I_0(x) \]

\[ = \lim_{x \to \infty} \left[ k x^{k-1} I_k + x^k (1/x) I_k(x) \right]/dI_0(x)/dx \]

\[ = \lim_{x \to \infty} \left[ (k/x)(x^k g_k) + x^{k-1} I_{k-1} \right]/(dlnI_0/dx) \]

The first limit on the RHS is \(O(k/x)\) of the limit on the LHS, and hence for \(x>>k\) it can be neglected. As for the denominator, since \(lnI_0(x) = ln[exp(x)] + ln[1-exp(-x)]\), we have

\[ dlnI_0/dx = 1 + 1/[exp(x)-1] = 1 + O(exp(-x)) \]

and we are left with the recursive relation

\[ \lim_{x \to \infty} x^k g_k(x) = \lim_{x \to \infty} x^{k-1} g_{k-1}(x) \]

which can be solved by induction, since for \(k=1\), we have

\[ x^{k-1} g_{k-1}(x) = g_0(x) = I_0(x)/I_0(x) = 1 \]

for all \(x\) (by its definition). Thus the general limiting relation (for all \(k\geq0\)) is

\[ \lim_{x \to \infty} x^k g_k(x) = 1 \quad \text{(C.4)} \]

A similar analysis of the function \(x(x^k g_k-1)\) reveals that
\[ \lim_{x \to \infty} x(x^{k-1}g_k) = \lim_{x \to \infty} \frac{x(x^kI_k-I_0)/I_0}{\lim_{x \to \infty} (x^kI_k-I_0)/I_0} \]

\[ = \left[ \lim_{x \to \infty} \frac{d(x^{k+1}I_k)/dx-d/dx(xI_0)}{dI_0/dx} \right] \]

\[ = \lim_{x \to \infty} \left[ (k+1)x^kI_k+x^kI_{k-1}-x(I_0-1)/[\lim_{x \to \infty} (I_0-1)] \right] \]

Since \( \lim_{x \to \infty} x^k g_k = 1 \) for all \( k \geq 1 \) and \( x \gg k \), we obtain the recursion relation

\[ \lim_{x \to \infty} x(x^{k-1}g_k) = \lim_{x \to \infty} x(x^{k-1}g_{k-1}) + 1 \]

Again we proceed by induction. Starting with \( k=1 \) and noting that \( x^0 g_0 = 1 \) identically for all \( x \), we have

\[ \lim_{x \to \infty} x(x^{k-1}g_{k-1}) = 2 \]

because \( x^0 g_0(x) = 1 \) identically for all values of \( x \). Then for arbitrary \( k \)

\[ \lim_{x \to \infty} x(x^{k-1}g_{k-1}) = \sum_{k'=1,k} (k'+1) = k(k+1)/2 + k = k(k+3)/2 \quad (C.5) \]

Therefore the behavior of \( g_k(x) \) for \( x \gg 1 \) is given by what appears to be an asymptotic expansion

\[ g_k(x) \sim 1/x^k + k(k+3)/2x^{k+1} + \ldots \quad (C.6) \]

Consequently, the functions \( g_k(x) \) behave quite simply, decreasing monotonically away from \( g_k(0) = 1 \) for \( x << 1 \) then decreasing monotonically towards zero as \( 1/x^k \) for \( x \gg 1 \). In particular, for the function of immediate interest, \( g_1(C) = 1/C \) is quite a good approximation over the open interval \( C > 1 \). While \( g_1(1) = 0.76684 \) and \( g_1(1.5) = 0.99887 \), \( g_1(C) \) never exceeds \( 1/C \) by more than \( \sim 32\% \) for \( 2 \leq C \leq 5 \). The ratio has decreased to only \( 10\% \) higher by \( C = 12 \) and then asymptotically approaches \( 1/C + 2/C^2 \) for higher values (see Fig. 1).