SIGMA ONE

BARRY F. MADORE
Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA; barry@obs.carnegiescience.edu
Received 2006 December 25; accepted 2010 March 22; published 2010 April 8

ABSTRACT

We demonstrate that it is possible to calculate not only the mean of an underlying population but also its dispersion, given only a single observation and physically reasonable constraints (i.e., that the quantities under consideration are non-negative and bounded). We suggest that this counterintuitive conclusion is in fact at the heart of most modeling of astronomical data.

Key words: methods: statistical

“Probability theory is nothing but common sense reduced to calculation.”

P. S. Laplace (1819) as quoted by E. T. Jaynes (2003)

1. INTRODUCTION

In their discussion of optimizing search strategies for the simultaneous discovery and characterization of variable stars using predetermined but highly non-uniform sampling, Madore & Freedman (2005) investigated the small-sample limits. Their emphasis was on situations where only a handful of observations (e.g., 2–20) might be available from which it would be necessary to determine periods, measure mean magnitudes, and derive full amplitudes. In this paper, we discuss how it is possible to use a single observation to derive confidence intervals (CIs) on the mean (i.e., limits on variability, or estimates of the population variance, or its full amplitude, etc.).

Shortly after the original manuscript was completed, it was pointed out that similar arguments have been made much earlier by Gott (1993). Gott was interested in the estimation of ages and lifetimes of events based on single instances; and he presented CIs for the longevity of an object given a single observation of its present age. He assumed a flat prior (of finite duration), and took the stance that there cannot be anything special about any given time that a random observation is made of an object that has a finite lifetime. We return to a contextual discussion of this example, and a generalization of it to modeling, at the end of this paper.

2. ONE OBSERVATION

What information can be derived about the parent population from a single, isolated (non-negative) observation? Here, we are clearly working at the observational information limit, and so a number of plausible assumptions (“priors”) will of necessity enter the calculations. A tacit assumption is that the observation in question is drawn from an underlying population having some dispersion\(^1\) that we are trying to estimate. We also assume that this is a physical object and that all values (observed or true mean, \(\langle Y \rangle\), etc.) of the data point from the true population mean (i.e., \(\epsilon_i = |(Y) - y_i|\)). By this definition, \(g_i \times y_i = \epsilon_i\) and so \(g_i = |(Y)/y_i - 1.0|\). With the g-factor distribution in hand, for any given prior distribution, it is then possible to calculate specific multipliers (g(50.0), g(68.5), g(90.0), etc., by integrating over the g-factor distribution function) that will guarantee that set fractions (50.0%, 68.5%, 90.0%, etc.) of the data set will fall within the range \(y_i \pm g(50.0) \times y_1, y_1 \pm g(68.5) \times y_1\), etc., where \(y_i\) is again our estimate for \((Y)\).

3. PRIORS

We now consider five realizations of four plausible underlying population distribution functions—a Poisson distribution, a uniform distribution, an exponential distribution, and two Gaussians.

3.1. Poisson Distribution

In this first example, invoking the Poisson distribution as being the appropriate form for the underlying parent distribution, it is almost trivially simple to demonstrate that a single observation can deliver both an average and a meaningful value for the dispersion of the underlying distribution. We first recall that the Poisson distribution for rare events has the form \(P(y) = \lambda y^e / y!\) where it is well known that the distribution is determined by a single variable \(\lambda\), and that \(\lambda\) is numerically equal to both the mean and the variance of this distribution. Therefore, if we have a single observation \(y_1\), which we assume comes from a Poisson process then, as a first approximation we can equate \(y_1\) with the mean value for that population \((Y) = \lambda = y_1\), and by invoking the known (definitional) equality of the mean and the variance for a Poisson process, we have therefore also measured the variance,\(^2\) with only a single observation.

---

\(^1\) Of course, if we assumed that the underlying population had no dispersion, our task would already be completed. To first order most observers make this default assumption, modulo observational error.

\(^2\) While not putting too fine a point to it, we also note that this single observation also determines several higher moments including the skew = \(\lambda^{-1/2}\) and the kurtosis = \(\lambda^{-1}\).
3.2. Uniform Parent Population

We now consider a uniform distribution spanning the non-negative interval $[0, 2]$ (Figure 1, upper panel). The distribution of $g$-factors is expected to be highly skewed: points near zero will require very large factors to make their error bars overlap with the true mean; however, all points greater than the mean (that is, fully half of the ensemble for a uniform prior) will have calculated error bars that overlap the mean for multiplicative $g$-factors that are all less than 0.5.

We first solved for the $g$-factor distribution function by simulation. We successively drew 100,000 observations from a uniform distribution, calculated the factor that needed to be applied to that number such that $y_1 \pm g_a \times y_1$ just overlaps the true mean ($\bar{y}_a$). In Figure 1, the results of that computer simulation are shown, where the distribution function of multiplicative $g$-factors is found in the lower panel, and the parent distribution of individual observations going into the simulation is given in the upper panel. The lower panel also shows where the 50%, 68.5%, and 90% CIs are found for the $g$-factor distribution corresponding to this uniform prior.

If the underlying population, from which a single data point $y_1$ is drawn, is itself uniformly populated and non-negative, then $y_1 \pm 0.41 \times y_1$ will contain the mean of the parent population 50% of the time. Other selected CIs are given in Table 1.

After the computer simulations were completed, it made possible for us to derive a closed-form analytic solution for $g$-factors, $g_u$ as a function of the CI associated with this case of a uniform prior. We give these solutions below; they are based on a simple mapping of the uniform distribution into its $g$-factors (using the absolute values of the $g$-factors accounts for the curious shape of the distribution in the lower panel of Figure 1). The CIs are then found by integrating the normalized functional form of that mapping up to the required value of CI:

$$ g_u = \frac{\sqrt{1.0 + 4(CI)^2} - 1.0}{2(CI)}, \quad [CI \leq 2/3] $$

$$ g_u = \frac{1.0 - 2CI}{2(CI - 1.0)}, \quad [CI \geq 2/3]. $$

Exemplary values for common CIs for the uniform prior are given in the second column of Table 1. For the uniform prior, they were calculated from the analytic solution and are confirmed by the simulations; all other solutions for other priors were derived from the computer simulations.

3.3. Exponential Parent Population

Another plausible parent population is the exponential distribution. Here, we investigate the $g$-factors corresponding to that distribution function. We have chosen to simulate an exponential distribution with mean $1$ and a width of 2. The lower panel shows the cumulative distribution of the multiplicative $g$-factors (as a function of the $g$-factor itself) needed to convert random samples taken from the upper panels into the mean. The $g$-factors required to give 50%, 68.5% (1σ), and 90.0% CIs for this underlying (exponentially distributed) sample are given as vertical bars and labeled accordingly.

![Figure 1. Uniform prior. The upper panel shows the input distribution of points having a uniform distribution with a mean of 1 and a width of 2. The lower panel shows the cumulative distribution of the multiplicative $g$-factors (as a function of the $g$-factor itself) needed to convert random samples taken from the upper panels into the mean. The $g$-factors required to give 50% and 68.5% (1σ) CIs for this underlying (uniformly distributed) sample are given as vertical bars and labeled accordingly.](image)

![Figure 2. Exponential prior. The upper panel shows the input distribution of points having an exponential distribution with a mean of 1. The lower panel shows the cumulative distribution of the multiplicative $g$-factors (as a function of the $g$-factor itself) needed to convert random samples taken from the upper panels into the mean. The $g$-factors required to give 50%, 68.5% (1σ), and 90.0% CIs for this underlying (exponentially distributed) sample are given as vertical bars and labeled accordingly.](image)

| Confidence Interval | Uniform $g$-Factor | Exponential $g$-Factor | Gaussian (0.50) $g$-Factor | Gaussian (0.25) $g$-Factor |
|---------------------|-------------------|-----------------------|---------------------------|---------------------------|
| 50.0%               | ±0.414            | ±0.67                 | ±0.31                     | ±0.16                     |
| 68.5%               | ±0.563            | ±1.60                 | ±0.45                     | ±0.24                     |
| 90.0%               | ±4.000            | ±8.5                  | ±1.78                     | ±0.48                     |
| 95.0%               | ±9.000            | ±19.0                 | ±4.0                      | ±0.70                     |
| 99.8%               | ±32.333           | ±65.0                 | ±14.0                     | ±1.2                      |

Table 1

Table 1 confirms this quantitatively for all CIs listed.
the upper panels into the mean. The function of the lower panel shows the cumulative distribution of the multiplicative points having a Gaussian distribution with a mean of 1 and a sigma of 0.5. The Figure 3.

2054 MADORE Vol. 139

in Figures 3 and 4, and in Columns 5 and 6 of Table 1. any dispersion less than 0.5. Gaussian distributions that have the same (unit) mean and certainly one will also encompass the CIs derived for a variety of Gaussian priors as well. That is, by adopting the g-factor itself needed to convert random samples taken from the upper panels into the mean. The g-factors required to give 50%, 68.5% (1σ), and 90.0% (2σ) CIs for this underlying (normally distributed) sample are given as vertical bars and labeled accordingly.

3.4. Two Gaussian Distributions

Since the uniform and the exponential priors are characterized by a single parameter, the mean, they are far more easily constrained by the observation(s). But, for completeness and for illustrative purposes, we consider here two Gaussian distributions each with a mean of unity, but having differing dispersions. We make no claim that these two parameters can be constrained by a single observation. We simply want to illustrate quantitatively that the uniform prior at least and the exponential prior in its extreme both encompass the results for these Gaussian priors as well. That is, by adopting the g-factors for a uniform prior, one will also encompass the CIs derived for a variety of Gaussian distributions that have the same (unit) mean and certainly any dispersion less than 0.5.

The results for these additional priors are given graphically in Figures 3 and 4, and in Columns 5 and 6 of Table 1.

4. DISCUSSION AND CONCLUSIONS

Our interest in this formalism started with a desire to characterize luminosity variations of astronomical objects (in the first instance through their first two moments, chosen to be their means and amplitudes) having a highly restricted number of observations. Here we have taken that thought experiment to its limit of a singular observation. For an underlying distribution of finite extent (or duration), the dispersion and the full amplitude are equivalent, differing only by a constant scale factor in any given case. Accordingly, our derivation of the variance of a population based on a single observation is conceptually equivalent to Gott’s (1993) derivation of future longevity (that is, Gott’s longevity is a one-directional semi-amplitude) for an object based on a “single” observation of its present age.

When dealing with physical observations, certain assumptions are tacitly taken for granted. The obvious assumptions are that those quantities are non-negative, and that the underlying population from which they are drawn does not have an infinite range (in time, mass, energy, size, etc.) However, those same assumptions carry additionally useful (prior) quantitative information that can be used to constrain limits on observations as they are obtained. This paper had the intent of making those assumptions explicit and then formalizing the mathematical consequences expressed as CIs on the underlying population as derived from one observation.

While it was our hope that this intentionally short contribution might stimulate others into finding applications not obvious to the author, the referee requested that examples be given of how this formalism might be applied to astronomy. In keeping with the spirit of this paper, we offer a modest example from the past, and predict that there will be future examples; all of this is based on a sample of N = 1.

Laplace developed his “rule of succession” when confronted with a question as to the mathematical (not the physical) probability that the Sun will rise tomorrow given its past (statistical) performance. After observing N events, Laplace derived that the probability of the next occurrence was (N + 1)/(N + 2). This would suggest that on the first day (N = 0) the probability of the Sun rising was actually 50:50. On the second day (N = 1), the probability would have gone up to 66%, and so on. Gott (1993) asked a similar question, not just about the next occurrence of something that has a past persistence, but about the sum of all future occurrences. How long will a thing last, given that we know how old it is now? Gott was reformulating Laplace’s question to be, if we have a single measurement (N = 1) of the age of something, what can one say about its total lifetime (its full amplitude) or rather its future longevity (a semi-amplitude). Of course, any
such prediction is best described in terms of probabilities, and so rather than predicting a firm lifetime based on a precise age, Gott predicted a forward-looking probability distribution (expressed as a variance) based on a backward-looking age.

One could argue that Gott actually required two observations: one of the time at which something began and another of the time at which the prediction was being made. This may be seen as quibbling but it is, in fact, equivalent to our physical prior, stated above, that none of the quantities to which this method applies can drop below zero or become infinite in amplitude (mass, luminosity, time, etc.).

Grounded with postdictions on the longevity of the Soviet Union and the Berlin Wall, Gott went on to predict the longevity of a variety of things astronomically big and small: from the expected demise of Nature magazine itself (somewhere between 3.15 and 4800 years) to the probability that we will end as a civilization (in 5100–8 million years with 95% probability), or colonize the Galaxy (the odds are against it). Each of these predictions were based on a single observation, \( N = 1 \) and a uniform prior.

It is quite clear that the uniform, non-negative prior distribution of data points discussed above is at one extreme (of simplicity, or of ignorance.) However, this extreme is also rather inclusive. If the true range of the underlying distribution is smaller, more centrally peaked, or more skewed toward the upper bound of the distribution function, than a uniform distribution, then their CIs will also be smaller than those calculated for a uniform prior; under those conditions, the uniform prior is likely to provide a conservative upper bound on the uncertainty.

4.1. Is This Just Another Name for Modeling?

Finally, we suggest that aspects of the scientific enterprise as a whole, as practiced by many astronomers in interpreting observational data, might simply be a generalization of the Sigma One methodology discussed here. Seen in that retrospective light, Sigma One becomes a fairly benign and low-level form of what would otherwise be called “modeling.”

Consider the observation of a color–magnitude diagram for a composite stellar population, in a nearby galaxy, say. On the basis of that single observation, one could ask what the magnitude and color of any given star might be on the next exposure (whenever that may be). Depending on the amount of prior knowledge about the underlying distribution function for that star, one could make a prediction. Indeed, we do this all the time. It is known that intrinsic variables (Cepheids, Miras, RR Lyrae stars, etc.) occupy fairly well-delineated regions of the color–magnitude diagram. Armed with known amplitudes and timescales, one could invoke those distribution functions with their specific means and variances to predict the expected variance in those selected stars. Stars in regions not known to be variable on those same timescales would have different priors used to predict their means and variances.

But all of this could also be recast into a very different form of the underlying distribution function in the case where extremely long (astronomically long) timescales are being considered. The predictive prior for a single observation would then become stellar evolution theory itself. That is, given a star observed (once) today at a given place in the color–magnitude diagram, what is its color and magnitude distribution function integrated over its projected future existence? And then how might the ensemble change with time? We apparently have no problem in undertaking population synthesis modeling, for example, taking a single integrated spectrum and/or a single color–magnitude diagram and extrapolating it to encompass the entire life history (backward and forward in time) of a given star and/or its associated contemporary population (an entire galaxy). So our point here is that if we are comfortable extracting very complex “moments” (in time and composition, etc.) from single (but admittedly very rich) observations by invoking very complicated priors (i.e., models), then it should come as no surprise that it is possible to extract more than just one moment (i.e., a mean and a variance at least) from a single data point by assuming very simple priors (i.e., models), in the form of well known, commonly invoked, but simple, distribution functions.

I thank Wendy Freedman for pointing out both the prior nature and the a posteriori relevance of Gott’s (1993) Nature study. Discussions with David Hogg were both stimulating and illuminating, as always. And finally, I thank both the editor-in-chief, Jay Gallagher, for his patience in dealing with this paper as it slowly evolved, and the anonymous referee, who carefully read the paper and gave many helpful suggestions.

REFERENCES

Gott, J. R., III. 1993, Nature, 363, 315
Jaynes, E. T. 2003, in Probability Theory: The Logic of Science, ed. G. L. Bretthorst (Cambridge: Cambridge Univ. Press)
Madore, B. F., & Freedman, W. L. 2005, ApJ, 630, 1054