Discussion of Akritas' paper by William H. Jefferys

There is little that I can add to Prof. Akritas' paper. Quoting from his paper, I would like to remind everyone of his essential point:

> It is emphasized that when the magnitude of the measurement error does not depend on the observation, more efficient procedures based on suitable weighting of the observations are possible. However, *when the magnitude of the measurement error depends on the observation, weighting biases the procedure* (Emphasis added).

In other words, the *obvious* thing to do can be the *wrong* thing to do. Oftentimes we find people using particular statistical procedures as "black boxes," without fully understanding them. It is important to understand what you are doing and why. Unthinking use of a particular procedure without understanding its properties is a recipe for disaster.

I also wish to call attention to William Wheaton's poster paper at this conference, "A Poisson parable: Bias in linear least squares estimation," which presents a simple example of the same phenomenon that Akritas has discussed. Wheaton considers the estimation of an unknown quantity which is measured with Poisson noise. For example, suppose we make k independent measurements of the luminosity L of a star by counting photons, obtaining for the ith observation N_i counts during the integration period t_i to yield an estimated luminosity $\hat{L}_i = N_i/t_i$. What is the correct method of combining the $\hat{L}_i$ to estimate L? Since the process being observed is Poisson, the variance of an individual observation is proportional to L_i, and one might be tempted to calculate a weighted average of the $\hat{L}_i$ with weights proportional to $1/N_i$; but this would be wrong, and the resulting estimator would be biased and in fact inconsistent. Wheaton shows that an unbiased estimator is given by $\sum N_i / \sum t_i$.

The example that Akritas discusses is somewhat more complex, but related. Here, we are measuring some property of a class of objects which has a natural cosmic scatter. For example, we could be measuring the luminosities of different members of a given class of stars in a cluster or galaxy, and the luminosity L_i of a given star i could be distributed according to some probability density $P(L_i \mid L, \ldots)$.

The individual L_i are unidentified; the best we can do is to estimate the L_i by $\hat{L}_i = N_i/t_i$, where N_i photons are counted during an integration period t_i. Again, Akritas shows that the "obvious" estimator for L obtained by computing a weighted average of the $\hat{L}_i$, weighting by the variance estimated in the obvious way from the Poisson nature of the photon counting process, yields a biased and even inconsistent estimator of L. Instead, the unweighted average is preferable, as it is manifestly unbiased. (In deriving his estimators, Akritas is following a moment method that has also been applied in astronomy by Deeming [Dee68].)

Since Akritas, Wheaton and I agree on these essential points, I thought it would be useful to use the remainder of my time to discuss an alternative, Bayesian approach to such problems. Thus, we observe counts $(N_i, \ldots, N_k)$ for stars $(1, \ldots, k)$ in some class. Each star is characterized by its actual luminosity L_i (expected counts/second) and integration time t_i. It follows that N_i follows a Poisson distribution:

$$P(N_i \mid L_i, t_i) = \frac{(L_i t_i)^{N_i} \exp(-L_i t_i)}{N_i!}$$

Because of cosmic scatter, each individual star in the class may have a different luminosity L_i. Assume, therefore, that the L_i are distributed according to some probability density, e.g., we might assume normality:

$$L_i \sim \mathcal{N}(L, \sigma^2)$$

so that

$$P(L_i \mid L, \sigma) \propto \frac{1}{\sigma} \exp\left(-\frac{(L_i - L)^2}{2\sigma^2}\right)$$

To approach this problem from a Bayesian viewpoint, we will also find it necessary to specify a prior distribution $P(L, \sigma)$ on L and σ. It follows from the definition of conditional probability that the prior distribution on (L_i, L, σ) is given by

$$P(L_i, L, \sigma) = P(L_i \mid L, \sigma) P(L, \sigma).$$

By Bayes' theorem, therefore, the posterior distribution of (L_i, L, σ), given the data, is proportional to the prior times the likelihood:

$$\begin{aligned} P(L_i, L, \sigma \mid N_i) &\propto P(L_i, L, \sigma) P(N_i \mid L_i, t_i, L, \sigma) \\ &= P(L_i, L, \sigma) P(N_i \mid L_i, t_i). \end{aligned}$$

Assuming independence, the complete posterior distribution for all observations is just the product of these over i:

$$P(L_i, \ldots, L_k, L, \sigma \mid N_i, \ldots, N_k) \propto \prod_i P(L_i, L, \sigma) P(N_i \mid L_i, t_i, L, \sigma).$$

Everything of interest is to be inferred from the posterior distribution. For example, we are interested in making inferences about L. The standard Bayesian prescription is to marginalize (integrate) with respect to the nuisance variables $(L_1, \ldots, L_k, \sigma)$, obtaining a posterior distribution in L alone. Thus

$$P(L \mid data) \propto \int \ldots \int P(L_1, \ldots, L_k, L, \sigma \mid data)\, dL_i \ldots dL_k\, d\sigma.$$

Once we have the posterior distribution of L in hand, we can compute Bayesian confidence intervals, posterior means, posterior medians, and so forth, for L.

In actuality, the particular case we have probably can't be integrated in closed form, so some approximate method such as Markov Chain Monte Carlo (MCMC) would have to be used. However, we can consider Wheaton's limiting case, obtained by letting $\sigma \to 0$, which results in a simplified problem that can be solved in closed form. We use the usual "automatic" (improper) prior

$$P(L) \propto \frac{1}{L},$$

which yields the posterior distribution

$$\begin{aligned} P(L \mid t_i, N_i) &\propto \frac{1}{L} \prod_i (Lt_i)^{N_i} \exp(-Lt_i) \\ &\propto L^{N-1} \exp(-LT) \end{aligned}$$

where

$$T = \sum_i t_i, \; N = \sum_i N_i$$

are sufficient statistics. The full normalized posterior distribution is therefore

$$P(L \mid data) = \frac{T(LT)^{N-1} \exp(-LT)}{(N-1)!}$$

The procedure at this point would be to derive whatever is desired from the posterior distribution. For example, if we want an estimator for L, we can compute the posterior mean or mode. The posterior mean is

$$\hat{L}_{mean} = \int L \, P(L \mid data) \, dL = \frac{N}{T}$$

The mode would be $(N-1)/T$, which is biased but consistent.

The interesting thing about this is that the Bayesian prescription automatically tells us not to use weighted averages, and instead leads us to estimators similar to (and in this simple case even identical to) the estimator advocated by Akritas and Wheaton. In more complex cases such as those with cosmic scatter, however, it is to be expected that the Bayesian estimators would not end up being a simple unweighted average, but would in general be nonlinear and computable only by numerical means. Nonetheless, they may well turn out to be better than the simple unweighted averages advocated by Akritas.

The discussion can be extended to the problem where the background count must also be considered [Lor92]. Whereas a straightforward approach using classical estimators can run into the problem in low signal situations of yielding unphysical negative luminosities, the natural Bayesian solution to the same problem cannot result in such anomalies. At the same time, the Bayesian solution is typically simple and to set up, while at the same time handling the problem of unphysical parameters quite automatically.

Response to Professor Rao

Professor Rao asks about maximizing the likelihood function instead of calculating marginal distributions. A Bayesian would be more likely to ask about maximizing the posterior distribution, obtaining the so-called maximum a posteriori (MAP) estimate, but in the situation at hand either method may run into difficulty. It is well-known that in situations where some variables are unidentified, like errors-in-variables problems, maximum likelihood gives the wrong answer for the variance—it is off by a factor of two [KS79]. Similarly, in nonlinear errors-in- variables problems, maximum likelihood may similarly produce inconsistent estimators for other interesting parameters [Ful87]. In contrast, the standard Bayesian prescription is to marginalize, that is, to integrate over the unidentified and nuisance variables. This does not result in an inconsistent estimator, if the prior is chosen properly. For a discussion of the Bayesian approach to this problem in the context of linear regression, see [Zel87].

1 References

[Dee68] Terence J. Deeming. The analysis of linear correlation in astronomy. In Arthur Beer, editor, *Vistas in Astronomy*, volume 10, pages 125–142. Pergamon Press, New York, 1968.

[Ful87] Wayne A. Fuller. *Measurement Error Models*. John Wiley & Sons, Inc., New York, 1987.

[KS79] Maurice Kendall and Alan Stuart. *The Advanced Theory of Statistics, Volume 2*. Charles Griffin & Co., Ltd., London, 1979.

[Lor92] T.J. Loredo. The promise of Bayesian inference for astrophysics. In E.D. Feigelson and G.J. Babu, editors, *Statistical Challenges in Modern Astronomy*, pages 275–306. Springer-Verlag, New York, 1992.

[Zel87] Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics*. Krieger Publishing Co., Florida, 1987.