February 21, 2012

Note on Non-Detections Model

Joseph W. Richards

1. Formulating the Model

This note, based on the original note from JWR on Feb. 8, 2012, includes revisions from David Hogg, James Long, Nat Butler, and Josh Bloom.

As a reminder, we are modeling light curve data in the presence of non-detections, which are epochs of observation in which no detection of the source of interest was made.

1.1. Preliminaries

For each astronomical object, a photometric survey measures a multi-epoch light curve over N (typically unevenly spaced) epochs. At each epoch i, with associated time t_i, the survey takes an exposure at the location of the object and either (a) detects the object and records an estimate of its photon flux, f_i^* and the root variance of the statistical uncertainty in that estimate, s_i^*, or (b) fails to detect the object. In the latter case, most modern surveys either record a reference value to signify that no detection was made (as is done in ASAS) or an estimate of the upper detection limit, b_i, which is the brightest the object could have been given that it was not detected at a significant level by the software.

There are many reasons that a source might not show up in a catalog. These include:

- low S/N of the source, due either to a higher noise level or a fainter signal,
- the source falling outside of the detection window (e.g., near chip edge),
- occulting of the source by an artifact of the detector (e.g., hot pixels, masked out), or
- the source was out-shone by an intervening object (e.g., asteroid, comet, variable star, airplane, etc.).

Here, we present a statistical model that can be used to detect variable sources and model their variability using multi-epoch light curves containing epochs of non-detection. Previous efforts to detect and model variability using multi-epoch photometry have typically ignored non-detections (REFs) or used them in an ad-hoc manner lacking statistical rigor (REFs). An exception is Lang et al. (2009), who measure proper motions of sources in SDSS falling below the detection limit.

In this paper, we will assume that f_i^* and s_i^* take on real-valued numbers in epochs for which the source was detected and receive the reference value `NA` in epochs where no detection was made. For notational convenience we assemble all the light curve data for one source into a data set D given by

$$D \equiv \{D_i\} \tag{1}$$

$$D_i \equiv (f_i^*, s_i^*) \quad , \tag{2}$$

where (f_i^*, s_i^*) are the light curve measurements at t_i. The goal, for each astronomical object, is to construct the likelihood, f^n, for the data D, given a set of model parameters that describe the variability of the object and the characteristics of the observations, and either to maximize the likelihood with respect to the model parameters or to use it in further inference.

1.2. The Light Curve Model

To model the mean brightness of the light curve as a function of time, we use a multiple-harmonic Fourier model with angular oscillation frequency ω,

$$\mu_i = \sum_{k=1}^{K} A_k \sin(t_i\,\omega k) + B_k \cos(t_i\,\omega k) \quad , \tag{3}$$

where $\sqrt{A_k^2 + B_k^2}$ is the amplitude of the k^{th} harmonic of the frequency ω and $\tan^{-1}(B_k, A_k)$ is the relative phase offset of harmonic k. In this analysis, we will fix the number of harmonics to $K = 3$, though in principle we can use the data to choose the appropriate K to capture the complexity of the light curve. In addition to the expected brightness in Equation 3, we assume that there is uncertainty or inappropriateness in the model, leading to a *model variance* s_μ^2 at each point (assumed constant but easily generalized).

Additionally, we need to instantiate a latent variable, b_i, to represent the detection limit, in units of flux, at epoch i. The b_i parameter is essential because it constrains the possible values of the mean brightness, μ_i, of the light curve when there is a non-detection. By employing a hierarchical model for the distribution of b_i, we can fully utilize all of the information encoded in both the detections and non-detections when computing the data likelihood. At each epoch, we connect the observed flux, f_i^*, to the latent variable, f_i, signifying the true observable flux of the object, via the detection limit, b_i by

$$f_i^* \;=\; \begin{cases} f_i & \text{if } f_i \geq b_i \\ \texttt{NA} & \text{if } f_i < b_i \end{cases} \tag{4}$$

so that the observed flux is NA only when the true observable flux, f_i, is below the detection limit, b_i. And similarly, we introduce a latent variable, s_i, for the true observable standard error,

$$s_i^* \;=\; \begin{cases} s_i & \text{if } f_i \geq b_i \\ \texttt{NA} & \text{if } f_i < b_i \end{cases} \tag{5}$$

so that the true observable standard error of the flux measurement is NA only when the true observable flux is below the detection limit.

Also, because we do not necessarily believe the reported measurement uncertainties, s_i^*, we choose to introduce a parameter, σ_i, to represent the true uncertainty variance for the brightness measurement at epoch i.

Hence, our initial model consists of the parameter vectors $\{f_1, ..., f_N\}$, $\{s_1, ..., s_N\}$, $\{b_1, ..., b_N\}$ and $\{\sigma_1, ..., \sigma_N\}$, and the model parameters

$$\theta \;\equiv\; (\omega, \{A_k\}, \{B_k\}, s_\mu^2) \quad , \tag{6}$$

along with prior information about observation times, $\{t_1, ..., t_N\}$ and other prior assumptions, which we make explicit by creating the prior information set

$$I \;\equiv\; (\{t_i\}, \text{assumptions}) \quad . \tag{7}$$

Our goal is to write down the form of the likelihood of the data, D, given θ and I. Then, for each light curve we can find the vector θ that maximizes the data likelihood.

1.3. Statistical Model for Light Curves with Non-Detections

We model the observed flux, f_i^*, as a Gaussian distribution with variance that has both measurement (σ_i) and model (s_μ) contributions. In the case that a detection is made, we require that the observed brightness be greater than the brightness of the detection limit ($f_i^* = f_i \geq b_i$) while in the case that no detection is made, we require that the brightness (if it could be measured) be less than the detection limit ($f_i < b_i$). To derive the likelihood of the observed f_i^*, given σ_i, θ and I, we must integrate over the prior distribution of the unknown b_i,

$$p(f_i^*|\sigma_i,\theta,I) = \begin{cases} N(f_i|\mu_i,\sigma_i^2+s_\mu^2)\int_0^{f_i} p(b_i|\theta)\,\mathrm{d}b_i & \text{if } f_i^* \neq \texttt{NA} \\ \int_0^\infty \int_0^{b_i} N(f_i|\mu_i,\sigma_i^2+s_\mu^2)\,p(b_i|\theta)\,\mathrm{d}f_i\,\mathrm{d}b_i & \text{if } f_i^* = \texttt{NA} \end{cases} \tag{8}$$

$$p(b_i|\theta) = N(b_i|B,V_B) \tag{9}$$

$$\theta \equiv (\omega,\{A_k\},\{B_k\},s_\mu^2,B,V_B) \quad , \tag{10}$$

where we have introduced the hyperparameters B and V_B for the Gaussian prior distribution of b_i. In Equation 8, in the epochs for which a detection was made ($f_i^* \neq \texttt{NA}$), we marginalize over the unknown b_i from 0 (low brightness limit) to f_i, enforcing that the detection limit be fainter than the observed brightness. Likewise, in the epochs for which no detection was made ($f_i^* = \texttt{NA}$), we integrate the joint (f_i, b_i) likelihood over all possible values of b_i and over the unknown f_i from 0 to b_i, ensuring that the brightness (if it were able to be observed) be fainter than the detection limit.

In the above, we have assumed no extra information on each of the b_i values besides the knowledge of whether a detection was made on that epoch. Hence, we draw, in Equation 9, each b_i value from a global prior distribution which is the same at all epochs. If instead, we are given an estimate of b_i plus its error distribution for each epoch (which, in principle can be inferred from the raw telescope images), we can replace Equation 9 with a different distribution per epoch. In the case that the b_i are assumed to be completely known (without error), the data likelihood of f_i^* becomes

$$p(f_i^*|\sigma_i,\theta,I,b_i) = \begin{cases} N(f_i|\mu_i,\sigma_i^2+s_\mu^2)\,I(f_i \geq b_i) & \text{if } f_i^* \neq \texttt{NA} \\ \int_0^{b_i} N(f_i|\mu_i,\sigma_i^2+s_\mu^2)\,\mathrm{d}f_i & \text{if } f_i^* = \texttt{NA} \end{cases} \tag{11}$$

where the boolean indictor function, $I(f_i \geq b_i)$, is 1 if $f_i \geq b_i$ and 0 otherwise.

Next, we model the reported standard error, s_i^*, on the uncertainty of the magnitude measurement. Instead of assuming that s_i^* is a perfect, error-free measurement, we probabilistically connect it to the true uncertainty standard error, σ_i through a truncated Gaussian likelihood, where the truncation enforces a non-negative variance for each observation. Our likelihood of observed s_i^*, given σ_i, θ and I, is

$$p(s_i^*|\sigma_i,\theta,I) = \begin{cases} N(s_i|\sigma_i,v_\sigma) & \text{if } s_i^* \neq \texttt{NA} \\ \int_0^\infty N(s_i|\sigma_i,v_\sigma)\,\mathrm{d}s_i & \text{if } s_i^* = \texttt{NA} \end{cases} \tag{12}$$

$$\theta \equiv (\omega,\{A_k\},\{B_k\},s_\mu^2,B,V_B,v_\sigma) \quad , \tag{13}$$

♣ **joey:** Need a distribution that is truncated at 0. Something better? ♣ where we have added a model parameter v_σ, which represents the variance in the distribution of reported uncertainties given the true uncertainty. The only purpose of including the likelihood in Equation 12 to the model, in practice, to keep the σ_i from drifting very far away from the s_i, as set by the hyperparameter v_σ.

Putting it all together, we can write down the likelihood of the data, D_i, for the parameters θ, on a single epoch, as

$$p(D_i|\theta, I) = \int_0^\infty p(f_i^*|\sigma_i, \theta, I)\, p(s_i^*|\sigma_i, \theta, I)\, p(\sigma_i|\theta)\, \mathrm{d}\sigma_i \quad (14)$$

$$p(\sigma_i|\theta) \propto \sigma_i^{-1} \quad (15)$$

$$\theta \equiv (\omega, \{A_k\}, \{B_k\}, s_\mu^2, B, V_B, v_\sigma) \quad , \quad (16)$$

where we have inserted the expressions from Equations (8) and (12), and integrated out the nuisance parameter, σ_i. We have assumed a Jeffrey's prior on σ_i, which is non-informative and invariant to reparametrization of the variance. We could alternatively use an inverse-gamma prior, which takes two hyperparameters (and is conjugate in the case of a normal likelihood with unknown variance). We have also assumed that f_i^* and s_i^* are conditionally independent given σ_i, θ, and I. This is a reasonable assumption in the case that a detection is made at epoch t_i.

Finally if we assume that the data collected at each epoch are independent given the model parameters, we have that

$$p(D|\theta, I) = \prod_i p(D_i|\theta, I) \quad . \quad (17)$$

This is the likelihood for the entire data set (all the measurements and non-detections of this star from all the epochs, as delivered by the survey) given the $2K + 5$ parameter vector θ. This model "correctly" or at least "justifiably" uses all of the information available, without making strong assumptions about the survey or its veracity.

1.4. Detecting Variability

We will describe here how we will do the likelihood ratio test of the model of constant flux to that of periodically varying flux.

1.5. Mixture Model for f_i^* or s_i^*

This may be too much for this paper, but will be useful in the presence of crappy data.

REFERENCES

Lang, D., Hogg, D. W., Jester, S., & Rix, H.-W. 2009, AJ, 137, 4400