Occular's Successors:
Overview of Statistical Underpinnings

Robert L. Anderson, Jr

October 18, 2012

Introduction:

Occular is a program that has been used, particularly within the IOTA community, to process asteroid occultation signals. It was first released in July 2007 with a major upgrade in February 2009. The statistical approach taken in Occular would be described as 'frequentist' and used Monte Carlo techniques and hypothetical data in producing its results. To put that program on a more solid statistical foundation, the author, prodded (unmercifully) by Tony George, undertook a study of how Bayesian Inference might be applied to the analysis of occultations.

That study was successful in developing relevant mathematical procedures based on Bayesian Inference and/or Maximum Likelihood estimation. Those studies were done using the R language, an interpreted language that is particularly well suited for statistical calculations, but not for production use. We are now at the point where the procedures that have been developed are ready to be made accessible to the IOTA community in an easy to use program. Hristo Pavlov, an active IOTA member, has agreed to take on the task of producing two programs currently identified as OTE (Occultation Time Extractor) and LCSA (Light Curve Statistical Analyser). The OTE program will deal with occultations where the 'event' is obvious. LCSA will deal with more problematic observations where the occultation is not readily apparent.

Mathematical background:

We start with a parametized model of a light curve

$$LC_i = model(x_i, \theta_1 \ldots \theta_n)$$

where x_i is the time of the reading and $\theta_1 \ldots \theta_n$ are the parameters of the light curve. For example, a star disk intersecting an asteroid disk model will have up to 7 parameters such as star and asteroid diameter, asteroid speed, track offset, etc.

Given an occultation observation y_i (i=1 to m), we want to determine the values for $\theta_1 \ldots \theta_n$ that best 'explains/fits' the observed data using the selected light curve model.

In order to solve this problem using the equivalent approaches of Bayesian Inference (BI) or Maximum Likelihood Estimation (MLE), we must also select a noise model. For star/asteroid occultations it is reasonable to assume that the readings are affected by noise that has Gaussian distribution.

Given these two explicit models (explicit means we know how to produce numbers from the models), we can now calculate the probability of each observation point relative to the theoretical light curve as follows:

$$p(y_i|\theta_1 \ldots \theta_n) = \frac{d\theta_1 d\theta_2 \ldots d\theta_n}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-(y_i - model(x|i,\theta_1 \ldots \theta_n))^2}{2\sigma_i^2}}$$

where the right hand side of the above equation is simply the Gaussian probability density function. All we are saying is that the observed data points differ from the theoretical value given by our light curve model by the addition of gaussian noise characterized by σ_i (the noise at that point) and furthermore that points that lie off the expected light curve are less probable than those that lie on or near to it.

Note: $p(y_i | \theta_1 \ldots \theta_n)$ is the usual notation for conditional probability and is verbalized as 'the probability of y_i given $\theta_1 \ldots \theta_n$'.

The probability of a series of independent measurements is simply the product of the individual probabilities, so the conditional probability of the complete observation can be calculated as:

$$p(obs|\theta_1 \ldots \theta_n) = \prod_{i=1}^{m} p(y_i|\theta_1 \ldots \theta_n)$$

At this point, we could take the MLE approach and state that the values of $\theta_1 \ldots \theta_n$ that maximize the complete observation probability are the 'solution' to our problem.

Or, we could apply Bayes rule that links the conditional probability of a model, given the data, to the conditional probability of the data, given the model.

The linkage that Bayes rule gives us is:

$$p(\theta_1 \ldots \theta_n|obs) = \frac{p(obs|\theta_1 \ldots \theta_n)p(\theta_1 \ldots \theta_n)}{p(obs)}$$

The term $p(\theta_1 \ldots \theta_n)$ on the right hand side is called (in Bayes-speak) the prior. It encodes what we know about the values of the parameters before we make a measurement. Since we have no reason to favor any particular value of a parameter over another, we choose a 'uniform' prior (equal probabilities for all parameter values) which results in $p(\theta_1 \ldots \theta_n)$ being a constant.

The term p(obs) is also a constant. It is also one that is very difficult to compute, but fortunately we don't have to because of the fact that

$$p(\theta_1 ... \theta_n | obs) \propto p(obs | \theta_1 ... \theta_n)$$

is sufficient for our needs as it says that if you maximize the right-hand side (which we know how to calculate), the left side is maximized as well, and importantly, the shape of the probability distribution of the left hand side is the same as the shape of the right hand distribution --- multivariate normal; i.e., a product of gaussians. Note that the term $d\theta_1 d\theta_2 ... d\theta_n$ also becomes irrelevant.

So, both the MLE and the BI approaches maximize the same equation, and that's why we earlier asserted that the approaches are equivalent. The BI approach gives us easily the additional important information about the parameter distributions that allows us to confidently compute error bars.

In principal, we're done --- the problem is solved. But in practice, maximizing $p(obs | \theta_1 \ldots \theta_n)$ is non-trivial. For example, suppose one tried a brute force approach where each parameter was allowed to range over a hundred values, a value was calculated at each point, and the parameter set that corresponded to the highest value was selected as the solution. If the model included 6 variable parameters, one would need to calculate 10^12 points. That would take too long, even in today's computing environment. The MCMC algorithm (Markov Chain Monte Carlo) was developed to solve such problems and is the technique that will be used in the OTE program. It is straightforward to apply this computational tool when the SNR is favorable, and that is the regime in which the OTE program will be used.

We have also succeeded in giving a solid statistical treatment of low SNR observations. That work will be incorporated into the LCSA program and makes extensive use of the AIC (Akaike Information Criterion) to compare a straight line model (i.e., no event) with a square wave model. The parameter region(s) where the probability of a square wave model is at least 0.99 is then explored using the MCMC algorithm. It is necessary to perform this pre-location work because low SNR events can be troublesome for the MCMC algorithm unless it is given a good starting point. We have demonstrated that this is possible with both live and simulated data.