Treating the detection of transient phenomena as a statistical problem

G. Bélanger, ESA, ESAC

Presented on 2019 May 22 (Last updated: May 21, 2019)

This talk was presented on Wednesday May 30, 2019, at the 14th workshop of the IACHEC held from May 19 to 23 in Shonan Village, Japan. The subject is that of detecting transient phenomena, and introduces a powerful statistical method based on the use of the likelihood function that can be applied to a wide range of problems and data types, and it particularly well suited for real-time and machine learning applications. The details are presented in Belanger [2013].

Contents

Where do we start?

WHEN WE HEAR the word *transient*, what do we think of, what do we understand? For optical astronomers, it means a supernova. And that's because nothing else changes over time in the optical sky, really. For X-ray astronomers it may mean bursts, flares, or outbursts from sources that are below the detection threshold when they are in quiescence, and that become visible and often very bright during periods called outbursts. This is common in black hole binaries. For space γ-ray astronomers, transient has often just been a synonym for GRB.

This is not meant to be exhaustive, but just illustrative. And the point is that different people, with different backgrounds, and working in different fields have different—sometimes very different—notions of what transient means. I would like to generalise this notion of transient to anything that changes: anything that changes from an established expected behaviour to another in a way that can be quantified. In other words, to define criteria that can distinguish between states, or emission mechanisms, or behaviours, based on statistics in a rigorous and reliable way that allows us to quantify statements we may wish to make about the observed phenomena. Where do we start? Statistics!

How different?

THE VERY FIRST step is to be able to tell that something which is observed now is different than how it was before now. How do we do that? The only way it can be done is if we know what is expected, and can thus tell that was is observed is different than what is expected. Different how? Statistically different based on the threshold that has been determined appropriate for this purpose. Here is an example of what we mean:

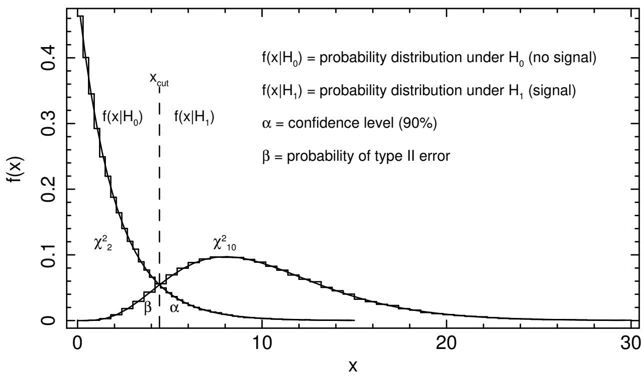

Figure 1: The expected density of the statistical test $f(x)$, in this case the Rayleigh periodogram statistic (χ^2_2), under the null hypothesis, which means when there is nothing special going on, is H_0. If there was a periodic signal of a particular strength (signal-to-noise ratio of 3.8), an alternative hypothesis H_1, we could expect another distribution as shown (χ^2_{10}). A threshold we could chose in this case to statistically distinguish one from the other is here labelled x_{cut}. The confidence level of the test is labelled α, and the probability to not detect the signal (type II error) is labelled β.

You all know this: we have a collection of measurements in here that represent the usual state of affairs, the expected behaviour, the background; once this is well defined and once we have established a threshold beyond which we consider a measurement to be distinct from this expected distribution, we can flag these outliers. And, that's that.

The next step is quantifying how different a measurement is from the expectation. Everyone here also know how to do this, right? The way we combine probabilities is through a likelihood function; a series of products where each term is the probability of an event, in this case, of a particular value having been measured. And the most critical element in this process is how to choose the most appropriate likelihood function. Of course it is! because how can we calculate a correct probability and correct likelihoods if we are not using the correct distribution function. And how do we know what to use? We let the data tell us!

Which distribution?

We are monitoring the number of photons hitting a particular pixel during each integration window, whatever that may be. Sometimes we see dozens, sometimes a few, sometimes none. When we take all these measurements—all of these photon counts—and look at how they are distributed, we find a Poisson distribution.

This is naturally the distribution with which we are most familiar in high-energy astronomy, especially here among scientists interested in calibration. Therefore we know that for any kind of analysis we might be doing with X-rays, whether it's working with images, with time series, or with spectra, that our data, whenever we are in some way counting photons, will be Poisson distributed. But what if we are interested in the count rates? We are no longer talking about a number of integer counts, we are now talking about fractional counts per second. What if, to illustrate the point as explicitly as we can, we are looking at the instantaneous count rate given by the inverse of the time between events. We no longer have a Poisson distribution, but what do we have?

And what if we are interested in studying, and monitoring, the evolution of the power spectral distribution of a source in order to gain insight about the different states of temporal variability it cycles between. In this case we would be monitoring the periodogram, which is a derived quantity calculated from the arrival time of events. The periodogram can be construction with a statistic like this one:

$$R^2 = 2N(C^2 + S^2) \tag{1}$$

where C and S are given by

$$C = \frac{1}{N}\sum_{i=1}^{N}\cos\phi_i \quad \text{and} \quad S = \frac{1}{N}\sum_{i=1}^{N}\sin\phi_i. \tag{2}$$

It's called the Rayleigh statistics, and, as you can see, it's the sum of two squared terms. Each of these terms is an average value: the

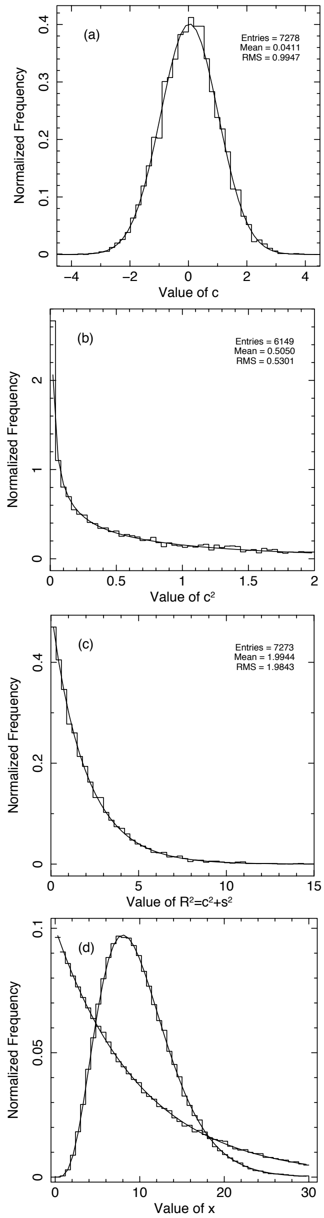

Figure 2: Illustration of the relationship between standard normal, χ^2 and exponential variables using the normalised frequency distributions and the analytical density functions: In panel (a) we see the variable $c = \sqrt{2N}C$ (standard normal); in panel (b) we see its square, $c^2 = 2NC$ (χ^2_1); and in panel (c) we see Rayleigh statistic, $R^2 = c^2 + s^2$ (χ^2_2: an exponential with $\tau = 2$). Panel (d) illustrates the difference between summing five χ^2_2 variables (χ^2_{10}), and scaling by five a χ^2_2 (exponential with $\tau = 10$).

first one is of cosines, and the second is of sines. Here, they are cosines and sines of the phase of times for a particular frequency. But it doesn't matter what they are sums of, because the central limit theorem tells us that the average of any given quantity tends to a normal distribution as the number of elements it is averaged over grows larger. Hence, each of these averages is a normal variable. Squaring a normal give us? A χ_1^2. Summing two χ_1^2 variables gives us? A χ_2^2. All of this is illustrated in in Figure 2. Finally, what is a χ_2^2 variable equal to? It's exactly equal to an exponential variable of mean 2.

You'll also remember that the exponential is a single parameter distribution, and that this parameter defines both the mean and standard deviation. Therefore, in this case as in the previous case of looking at count rates, we are no longer working with the Poisson distribution.

In the case of the periodogram there is an additional complication: the reasoning we just went through only applies when the cosines and sines, or actually, when the phases are independent of one another; and this is only the case for white noise, i.e., no variability. But how often do we ever see white noise? Not very often. It turns out that we can generalise our result by recognising that the variability will introduce a correlation between the phases, and thus between the Fourier components of cosines and sines, but that the structure of the statistic, the way we compute it, makes it that the values of power in the periodogram are all exponential variables scaled by the underlying shape of the power spectral distribution. Therefore, at each test frequency, power estimates are exponential variables, and if we remove the underlying power spectral shape, we recover the initial χ_2^2 distribution as we can see here:

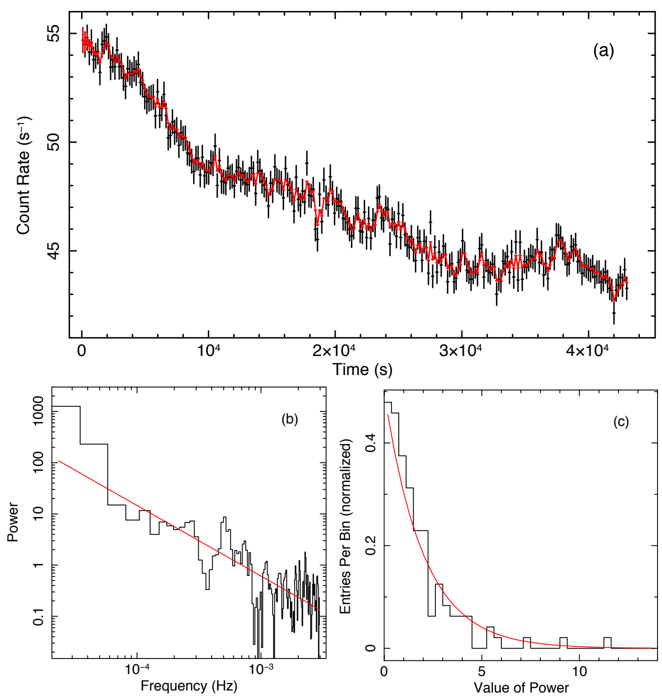

Figure 3: Illustration of periodogram powers of astrophysical red noise as scaled χ_2^2 (exponential) variables using an *XMM-Newton* observation of Mkn 421: Panel (a) shows the RGS time series in rates (0.3–2 keV with 85 s bins); panel (b) shows the periodogram with the best fit power-law model; and panel (c) is the distribution of de-trended periodogram powers overlaid with the analytical form of the χ_2^2 density function, the exponential density with mean of 2 (decay constant of 1/2): $f(x) = \frac{1}{2}e^{-x/2}$.

Which likelihood?

I REALISE WE'VE spent a long time on this, and this because I want to emphasise how important this point really is: we must absolutely determine how the random variable we are working with is actually distributed in order to use the correct probability density and likelihood function appropriate for the problem. So, for example, the Poisson likelihood is expressed as

$$L(\boldsymbol{\nu}|\boldsymbol{n}) = \prod_i \frac{{\nu_i}^{n_i} e^{-\nu_i}}{n_i!}. \tag{3}$$

The normal likelihood is expressed as

$$L(\boldsymbol{\mu},\boldsymbol{\sigma}|\boldsymbol{x}) = \prod_i \frac{1}{\sqrt{2\pi{\sigma_i}^2}} e^{-(x_i-\mu_i)^2/2\sigma_i^2}. \tag{4}$$

The χ^2 likelihood is expressed as

$$L(\boldsymbol{k}|\boldsymbol{x}) = \prod_i \frac{1}{2^{k_i/2}\Gamma(k_i/2)} x_i^{k_i/2-1} e^{-x_i/2}. \tag{5}$$

The exponential likelihood is expressed as

$$L(\boldsymbol{\tau}) = \prod_i \frac{1}{\tau_i} e^{-x_i/\tau_i}. \tag{6}$$

And the likelihood to use for count rates, which is equal to the inverse of the time between events, it turns out, is the inverse exponential, is given by

$$L(\boldsymbol{t}) = \prod_i \frac{t_i}{{x_i}^2} e^{-t_i/x_i}. \tag{7}$$

Now we have what we need to, each time we get a new measurement, ask the question of how likely it is to have measured that value given what we expect based on the collection of previous measurement, and answer this question in a clear and quantitative manner through the use of the correct probability density function incorporated into the formulation of the likelihood function. What these likelihood functions mentioned above actually look like for a measured data value of 2 is shown in Figure 4.

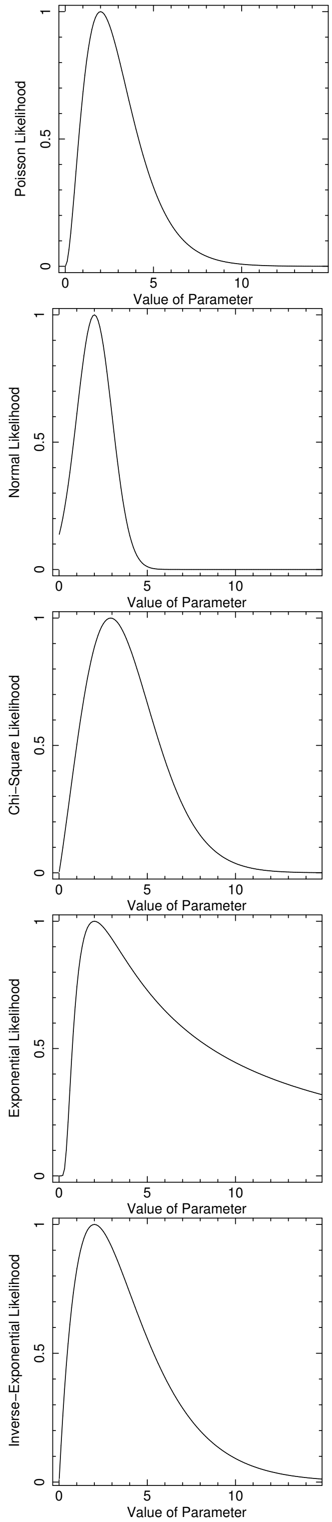

Figure 4: Illustration of different shapes of ikelihood functions for different kinds random variables all based on a measured data value of 2.

Show me how it works!

LET'S SEE HOW all of this works in practice for the case of count rate monitoring:

Video of monitoring event rate (explain while watching).

If we are looking for a transient QPO, for example, like in a black hole system like GX 339–4, then we would do something like this:

Video of monitoring periodogram (explain while watching)

So, did we accomplish what we set out to do? Did we succeed in establishing a way to treat the detection of transient phenomena as a statistical problem? I believe we have. And what were the essential point critical to being able to do it? The first is establishing very precisely what is expected in the absence of transients, and from it establish the detection threshold. The second is determining the way the data we are working with are distributed, and from this define the correct likelihood function we will use to quantify how likely or unlikely a given measurement is with respect to the expectation.

Did we forget anything?

We are nearing the end. But let me anticipate what some of you will have been thinking maybe from the very beginning: what if what we expect in the absence of transients is itself changing in time? How do we deal with that? The answer is: in exactly the same way, except that the baseline distribution against which we compare the incoming measurements evolves in time, and with it, the likelihood function and detection thresholds. Like this:

Video of event rate on variable background (explain while watching)

I leave you here. You can find more details on this in Belanger [2013]. Thank you for listening.

References

G Belanger. On Detecting Transient Phenomena. *ApJ*, 773(1):66, August 2013.