Probability Density Functions for Positive Nuisance Parameters

Bob Cousins for the CMS Statistics Committee

May 19, 2010

Abstract

This note includes recommendations for probability density functions for a positive nuisance parameter in some common situations. We discourage the used of truncated Gaussians (except when it is known that the answer is insensitive to the choice), preferring instead lognormal or Gamma distributions. (Some plots will be added in the next version of this note.)

1 Introduction

Parameters which introduce systematic uncertainties into a high energy physics statistical analysis are often equivalent to what statisticians call (for good reason) *nuisance parameters*. There is no uniformly satisfactory way of dealing with them, as discussed in my review of the statistics and high energy physics literature at the Oxford PhyStat meeting in 2005 [1], and in Luc Demortier's review at PhyStat-LHC [2], and references within. This note pertains to one class of approaches, which in effect averages a result over a probability density function (pdf) in a nuisance parameter. This approach is natural within the Bayesian framework. It is also common within an otherwise-frequentist calculation, where it is sometimes referred to as a Cousins-Highland method [3], since we advocated its use in a particular special case where a purely frequentist calculation gave a result that we found unacceptable. (The math turned out to be the same as in the prior-predictive p-value advocated by Box [4].)

Letting ψ denote a nuisance parameter (e.g., the mean background, or integrated luminosity), one asserts that a pdf $p(\psi)$ encodes belief about its true value. As input to encoding $p(\psi)$, typically one has some physicist's "best estimate" denoted by $\hat{\psi}$, as well as some estimate of its standard error or proxy thereof, denoted here by $\delta\psi$.

In cases where $p(\psi)$ includes results from subsidiary (calibration) measurements, we can consider it to be the pdf which is posterior to those measurements (and hence in the Bayesian framework include a earlier prior pdf and the likelihood function of the subsidiary measurements). In the chain of reasoning it is then the prior pdf for the analysis at hand.

A common functional form for $p(\psi)$ is a Gaussian,

$$p(\psi) = \frac{1}{\sqrt{2\pi\sigma_\psi^2}} \exp\left(-\frac{\left(\psi - \hat{\psi}\right)^2}{2\sigma_\psi^2}\right), \tag{1}$$

in which case $\delta\psi$ is identified with σ_ψ.

Since ψ is typically a non-negative number (Poisson mean or luminosity), in computer codes one typically truncates this pdf at zero or higher. The effect of truncation is often negligible, but various issues can arise, conceptually and numerically. In at least one common case [5], the combination of such a truncated Gaussian with an improper prior for the signal mean can result in a improper posterior (or equivalently, an answer that depends on an arbitrary cutoff).

Meanwhile there are fairly well-motivated alternatives to the Gaussian, which are described in the rest of this note: the lognormal [3, 6, 7, 8, 9] and the Gamma distribution [6, 7, 9, 10, 11].

2 Lognormal Distribution

From Eadie et al., "The *log-normal distribution* represents a random variable whose logarithm follows a normal distribution. It provides a model for the error of a process involving many small multiplicative errors (from the Central Limit Theorem). It is also appropriate when the value of an observed variable is a random proportion of the previous observation." [6, 7].

A standard form is

$$p(\psi) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\frac{1}{\psi}\exp\left(-\frac{(\ln\psi-\mu)^2}{2\sigma^2}\right), \tag{2}$$

where μ and σ are parameters which in this form are perhaps a bit obscure, as neither corresponds to the Gaussian equivalent. (E.g., the expectation value of ψ is $\exp(\mu+\sigma^2/2)$.)

If, following Korytov and Chen [12], we identify the physicist's best estimate $\hat{\psi}$ with the median (which equals $\exp(\mu)$) and introduce $\ln\kappa=\sigma$, then an equivalent form is

$$p(\psi) = \frac{1}{\sqrt{2\pi}\ln\kappa}\,\frac{1}{\psi}\exp\left(-\frac{\left(\ln(\psi/\hat{\psi})\right)^2}{2(\ln\kappa)^2}\right), \tag{3}$$

where $\kappa>1$ encodes the spread in the distribution, with $(\kappa-1)$ corresponding roughly to a physicist's multiplicative "relative uncertainty",

$$\kappa-1\equiv\sigma_{\rm rel}\equiv\delta\psi/\psi. \tag{4}$$

The pdf for $\ln\psi$ is a Gaussian with mean at $\ln\hat{\psi}$ and rms deviation $\ln\kappa$. The lognormal distribution assigns equal probabilities for ψ to be a factor of κ^n larger or smaller than the best estimate $\hat{\psi}$, i.e.,

$$P(\psi>\hat{\psi}\cdot\kappa^n)=P(\psi<\hat{\psi}/\kappa^n). \tag{5}$$

For n=1, these probabilities are 16% each, for $n=2$ they are 2.5%, and so on (following the usual Gaussian tail probabilities).

For small errors, κ is close to 1 and $\sigma_{\rm rel}<<1$, In this case, by expanding logarithms, one can see that the lognormal distribution becomes a familiar Gaussian with mean $\hat{\psi}$ and rms $\sigma_{\rm rel}\hat{\psi}$.

The lognormal distribution has a longer tail, goes to zero at $\psi=0$, and is a more meaningful way than a truncated Gaussian way to encode statements such as "factor of two uncertainty".

The ROOT [13] form of the lognormal function is more general:

$$p(\psi) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\frac{1}{(\psi-\theta)}\exp\left(-\frac{(\ln(\psi-\theta)-m)^2}{2\sigma^2}\right). \tag{6}$$

We reduce to the above by setting $\theta=0$ and $m=\mu$; we do not yet see a compelling need for $\theta\neq0$.

3 Gamma Distribution

From Eadie et al., "The gamma distribution is a basic statistical tool for describing variables bounded at one side, for example $0 < X < \infty$. ... The exponential, Erlangian, and chisquare distributions are all special cases of the gamma distribution" [6, 7]. Eadie et al.'s form is

$$p(\psi) = \frac{a(a\psi)^{b-1}e^{-a\psi}}{\Gamma(b)}. \tag{7}$$

The ROOT [13] form is again more general:

$$p(\psi) = \frac{\left(\frac{\psi-\mu}{\beta}\right)^{\gamma-1} \exp\left(-\frac{\psi-\mu}{\beta}\right)}{\beta\Gamma(\gamma)}, \tag{8}$$

corresponding to the above if we set $\mu = 0$, $\beta = 1/a$, and $\gamma = b$.

In the case where the estimate $\hat{\psi}$ is for a mean number of background events and is based on counting events in a control region (sideband), Linnemann has advocated the use of the gamma pdf for some time [9, 10], with an argument also known to the Gamma Ray Astronomy community[14]. Using the notation in Ref. [11], n_{off} events are observed in a Poisson sample from a control region with mean that is τ times that of the mean background μ_{b} in the signal region. (Here τ is assumed known exactly.) Then the likelihood function for the unknown mean μ_{b} in the signal region is

$$\mathcal{L}(\mu_{\mathrm{b}}) = \frac{(\tau\mu_{\mathrm{b}})^{n_{\mathrm{off}}} e^{-\tau\mu_{\mathrm{b}}}}{n_{\mathrm{off}}!}. \tag{9}$$

If one assumes a uniform prior for μ_{b} and applies Bayes's Theorem, then the (unnormalized) posterior pdf $p(\mu_{\mathrm{b}})$ is the same mathematical expression, which on comparison with Eqn. 7 is seen to be a (unnormalized) Gamma distribution. Ref. [11] discusses interesting identities between intervals based on this posterior and those from purely frequentist origin.

If the estimates $\hat{\psi}$ and $\delta\psi$ are in fact summarizing such a sideband-based situation, with ψ corresponding to the mean background μ_{b} in the signal region, then typically one has obtained these estimates by simple Poisson formula, $\hat{\psi} = n_{\mathrm{off}}/\tau$ and $\delta\psi == \sqrt{n_{\mathrm{off}}}/\tau$. From these one can reverse-engineer the original "raw" data $\tau = \hat{\psi}/(\delta\psi)^2$ and $n_{\mathrm{off}} = \hat{\psi}\tau$, and hence the values of a and b to use in Gamma posterior of the form in Eqn. 7:

$$a = \tau = \hat{\psi}/(\delta\psi)^2; \qquad b = n_{\mathrm{off}} + 1 = (\hat{\psi}/\delta\psi)^2 + 1. \tag{10}$$

4 Summary

Given $\hat{\psi}$ and $\delta\psi$, the lognormal distribution (Eqn. 3 with κ defined from Eqn. 4) and Gamma distribution (Eqn. 7 with a and b defined from Eqn. 10) are preferable to a truncated Gaussian if the respective arguments are relevant. This requires some care, and ultimately (as with any Bayesian prior) an analysis of the sensitivity to the prior; the Gamma prior above rests on the arbitrary flat prior before the sideband measurement, and the lognormal prior may or may not capture the essence of a theoretical systematic uncertainty. (For the case in which the lognormal is chosen as an approximation to the effect of a product of many terms, the parameters might be better identified from the characteristics of the functions forming the product.) In any case, trying out all three can identify cases where the truncated Gaussian has numerical issues, or where there is sensitivity to the form of $p(\psi)$. In the latter case, of course further analyses is required.

These simple suggestions are a beginning, not the end, of the story regarding Bayesian priors. Various studies have shown the sorts of pitfalls one can fall into when combining Bayesian priors for more than one parameter. Only recently have some first results [15] been obtained by high energy physicists trying to employ some more sophisticated methods from the statistics literature, in an effort to take us out of the "pseudo-Bayesian" era.

References

[1] Robert D. Cousins, "Treatment of Nuisance Parameters in High Energy Physics, and Possible Justifications and Improvements in the Statistics Literature", Proceedings of PhyStat 05: Statistical Problems in Particle Physics, Astrophysics and Cosmology (Oxford, Sept. 12-15, 2005), `http://www.physics.ox.ac.uk/phystat05/proceedings/`. See also the response following by N. Reid.

[2] Luc Demortier, "P values and nuisance parameters," Proceedings of the PhyStat 2007 conference (CERN, Geneva, Switzerland, June 2007), `http://phystat-lhc.web.cern.ch/phystat-lhc/2008-001.pdf`, pp. 23-33.

[3] R.D. Cousins and V.L. Highland, "Incorporating systematic uncertainties into an upper limit," Nucl. Instrum. Meth. A **320** (1992) 331.

[4] George E.P. Box, "Sampling and Bayes' Inference in Scientific Modelling and Robustness", Journal of the Royal Statistical Society. Series A (General), **143** (1980) 383.

[5] Luc Demortier, "Bayesian treatments of systematic uncertainties," Proceedings of Conference on Advanced Statistical Techniques in Particle Physics (Durham, England, 18-22 Mar 2002). `http://www.ippp.dur.ac.uk/old/Workshops/02/statistics/proceedings//demortier.pdf`

[6] W.T. Eadie, D. Drijard, F.E. James, M. Roos, B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam, 1971), pp. 76, 79-80.

[7] Frederick James, *Statistical Methods in Experimental Physics*, 2nd Edition, (World Scientific, Singapore, 2006), pp. 83-87.

[8] J. Aitchison and J.A.C. Brown, *The Lognormal Distribution* (Cambridge University Press, 1957).

[9] J. Linnemann, "Upper Limits and Priors", talk at Workshop on Confidence Limits, Fermilab (2000) `http://conferences.fnal.gov/cl2k/copies/linnemann1.pdf`

[10] James T. Linnemann, "Measures of significance in HEP and astrophysics," Proceedings of PhyStat 2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology (SLAC, Stanford, California USA Sept. 8-11, 2003) [arXiv:physics/0312059]. `http://www.slac.stanford.edu/econf/C030908/papers/MOBT001.pdf`

[11] Robert D. Cousins, James T. Linnemann, Jordan Tucker, "Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into

a test of the background-only hypothesis for a Poisson process," Nucl. Instrum. Meth. A **595** (2008) 480. DOI:10.1016/j.nima.2008.07.086, arXiv:physics/0702156.

[12] Text in this section is based on that from a draft note by Mingshui Chen and Andrey Korytov.

[13] The reference for notation of the ROOT implementation are in the source code, `http://root.cern.ch/root/html/src/TMath.cxx.html`. They use the Engineering Statistics Handbook: NIST/SEMATECH e-Handbook of Statistical Methods, `http://www.itl.nist.gov/div898/handbook/`. For lognormal the online page is `http://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm`, and for Gamma it is `http://www.itl.nist.gov/div898/handbook/eda/section3/eda366b.htm`.

[14] D.E. Alexandreas et. al., "Point source search techniques in ultra high energy gamma ray astronomy" Nucl. Instr. Meth. A **328** (1993) 570.

[15] Luc Demortier, Supriya Jain, Harrison B. Prosper, "Reference priors for high energy physics," arXiv:1002.1111v1 [stat.AP]