DRAFT 0.3

Glen Cowan
RHUL Physics
15 January, 2012

Asymptotic distribution of the ratio of two profile likelihoods

1 Introduction

In Ref. [1], distributions for test statistics based on likelihood ratios were given, using approximations that become exact in the limit of a large data sample. The main results were given for test statistics based on a ratio of the profile likelihood (defined below) to the maximized likelihood. Another type of likelihood ratio, namely, a ratio of two profile likelihoods corresponding to two hypotheses has also been widely used in HEP searches, especially at the Tevatron. A special case of such a likelihood ratio was also discussed in Sec. 3.8 of [1], but it was restricted to the case where the distributions of the corresponding likelihood-ratio statistic (defined below) was approximately Gaussian. The main purpose of the present note is to extend the result of [1] to cases where the Gaussian approximation for the test statistic is not valid. The more general result is needed when the standard deviation of the estimator of the signal strength depends significantly on the hypothesized signal strength.

Consider a model containing a rate parameter μ as well as nuisance parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, described by a likelihood function $L(\mu, \boldsymbol{\theta})$. The tests of hypothesized parameter value μ were based on the ratio of the profile to maximized likelihoods:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \ . \tag{1}$$

Here $\hat{\mu}$, and $\hat{\boldsymbol{\theta}}$ are the values of the parameters that maximize the likelihood and $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ denotes the values of the $\boldsymbol{\theta}$ that maximize the likelihood for a given value of μ. That is, the numerator of (1) is the profile likelihood and the denominator is the maximized likelihood.

In this note, we will assume one measures a continuous variable x that follows a Gaussian distribution with mean μ and standard deviation σ_μ. This is equivalent to the assumption that the asymptotic approximations of Wald and Wilks hold, as described in Ref. [1]. There, the Maximum Likelihood (ML) estimator for μ, $\hat{\mu}$, played the role of the measured quantity x. We assume that there are two specific hypothesis of interest:

$$H_0 \quad : \quad x \sim \text{Gauss}(\mu_0, \sigma_0) \ , \tag{2}$$

$$H_1 \quad : \quad x \sim \text{Gauss}(\mu_1, \sigma_1) \ . \tag{3}$$

In a search for a new process, H_0 typically represents the hypothesis that the data sample consists entirely of events from known (background) processes, and and H_1 is the hypothesis that includes a sought after signal process.

Often, for example, the measurement is based on counting a number of events n that are Poisson distributed with a mean value $\mu s + b$, where s and b represent the expected number of events from signal and background processes, respectively, and μ is a strength parameter defined such that $\mu = 0$ gives the background-only hypothesis H_0 and $\mu = 1$ corresponds to the nominal alternative H_1. In this case the ML estimator for μ is

$$\hat{\mu} = \frac{n - b}{s} \ . \tag{4}$$

Since n is Poisson distributed with mean $\mu s + b$, the variance of n is $V[n] = \mu s + b$ and therefore the standard deviation of $\hat{\mu}$ is

$$\sigma_\mu = \frac{\sqrt{\mu s + b}}{s} \ . \tag{5}$$

Here we will assume that specification of the hypotheses H_0 and H_1 provides known values for the two mean values μ_0 and μ_1 as well as the standard deviations σ_0 and σ_1. Here for illustrative purposes we will take the parameter μ under the two hypotheses H_0 and H_1 to be $\mu_0 = 0$ and $\mu_1 = 1$, with the corresponding standard deviations

$$\sigma_0 = \frac{\sqrt{b}}{s} \ , \tag{6}$$

$$\sigma_1 = \frac{\sqrt{s + b}}{s} \ . \tag{7}$$

Although this rule for the standard deviations is derived from the assumption of Poisson distributed data, we nevertheless consider here that the data value x follows a Gaussian distribution, as would be the case for $x = \hat{\mu}$ given a sufficiently large value for the mean of n.

In more complicated problems with nuisance parameters, one would nevertheless have a ML estimator $\hat{\mu}$ for a corresponding signal rate parameter μ, and this will have a certain standard deviation σ_μ. To apply the procedure here to such analyses, one must obtain the standard deviation σ_μ of $\hat{\mu}$ (i.e., what we call here x) by using, for example, the matrix of second derivatives of the log-likelihood function. The value of σ_μ will depend in general on the assumed values of both μ and of any nuisance parameters.

2 Definition of the test statistic q and distribution for $\sigma_0 = \sigma_1$

We assume that the measurement can be represented by a single Gaussian distributed variable x with mean μ and standard deviation σ_μ, so that the likelihood function is

$$L(\mu, \sigma_\mu) = \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{-(x-\mu)^2/2\sigma_\mu^2} \ . \tag{8}$$

The two hypotheses H_0 and H_1 defined above thus correspond to two pairs of values for μ and σ_μ, namely, (μ_0, σ_0) and (μ_1, σ_1).

According to the Neyman-Pearson Lemma (see, e.g., Ref. [2]), when defining a test of a given size α of H_0, to obtain the highest power with respect to the alternative H_1 the optimal test statistic is given by the likelihood ratio $L(\mu_1, \sigma_1)/L(\mu_0, \sigma_0)$. Equivalently we can use a monotonic function of this and therefore we define the statistic

$$q = -2 \ln \frac{L(\mu_1, \sigma_1)}{L(\mu_0, \sigma_0)} = \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{(x - \mu_0)^2}{\sigma_0^2} + 2 \ln \frac{\sigma_1}{\sigma_0} \ . \tag{9}$$

In the special case where the two standard deviations are the same, i.e., $\sigma_0 = \sigma_1 \equiv \sigma$, Eq. (9) becomes

$$q = \frac{\mu_1^2 - \mu_0^2 - 2(\mu_1 - \mu_0)x}{\sigma^2} \ . \tag{10}$$

The statistic q is thus a linear function of x and therefore also follows a Gaussian distribution with expectation value and variance given by

$$E[q|\mu, \sigma_\mu] = \frac{\mu_1^2 - \mu_0^2 - 2(\mu_1 - \mu_0)\mu}{\sigma^2} \ , \tag{11}$$

$$V[q|\mu, \sigma_\mu] = \frac{4(\mu_1 - \mu_0)^2 \sigma_\mu^2}{\sigma^4} \ . \tag{12}$$

These results were given in Eqs. (73) and (74) of Ref. [1] for the special cases of $\sigma_\mu = \sigma$, $\mu_0 = 0$ and $\mu_1 = \mu$. The purpose of the present note is to extend this result to the case where σ_0 and σ_1 are not equal, in which case the resulting distribution for q is no longer Gaussian.

3 Distribution of q for $\sigma_0 \neq \sigma_1$

In this section we treat the case of $\sigma_0 \neq \sigma_1$. Suppose $\sigma_1 > \sigma_0$, as is the case if their values are assigned according to Eqs. (6) and (7). The dependence of q on x given by Eq. (9) is shown in Fig. 1. It is a parabola with a maximum (or minimum, if $\sigma_0 > \sigma_1$) that occurs at

$$x_{\rm m} = \frac{\mu_0 \sigma_1^2 - \mu_1 \sigma_0^2}{\sigma_1^2 - \sigma_0^2} \ , \tag{13}$$

and at this point, q takes on the value

$$q_{\rm m} = \frac{(\mu_1 - \mu_0)^2}{\sigma_1^2 - \sigma_0^2} + 2 \ln \frac{\sigma_1}{\sigma_0} \ . \tag{14}$$

To find the distribution of q, we first solve for x in terms of q. The quadratic relation (9) has two solutions, x_+ and x_-, given by

$$x_\pm = x_{\rm m} \pm \sqrt{\frac{2(q - q_{\rm m})}{q''}} \tag{15}$$

where

$$q'' = \frac{2}{\sigma_1^2} - \frac{2}{\sigma_0^2} \tag{16}$$

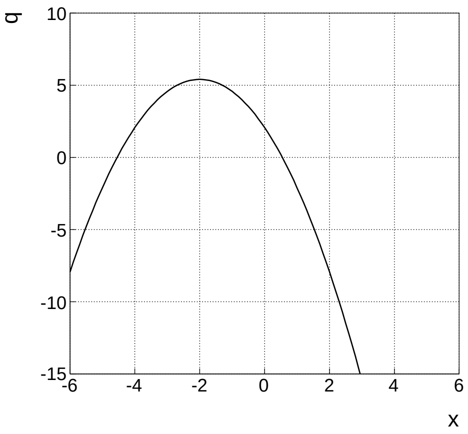

Figure 1: The test statistic q as a function of the measured variable x. The curve has been computed using $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0 = 0.632$ and $\sigma_1 = 0.775$, which result from Eqs. (6) and (7) using $s = 5$ and $b = 10$.

is the second derivative of q with respect to x. As x is Gaussian distributed with mean μ and standard deviation σ_μ, the distribution of q is given by

$$f(q|\mu,\sigma_\mu) = \left|\frac{dx_+}{dq}\right| \frac{1}{\sigma_\mu}\varphi\left(\frac{x_+(q)-\mu}{\sigma_\mu}\right) + \left|\frac{dx_-}{dq}\right| \frac{1}{\sigma_\mu}\varphi\left(\frac{x_-(q)-\mu}{\sigma_\mu}\right) , \tag{17}$$

where the function φ is the standard Gaussian pdf (zero mean and unit variance), x_+ and x_- are given by Eq. (15), and the derivatives of x with respect to q for the two solutions x_+ and x_- are

$$\frac{dx_\pm}{dq} = \pm\frac{1}{\sqrt{2(q-q_{\rm m})q''}} = \pm\frac{\sigma_0\sigma_1}{2}\frac{1}{\sqrt{(\sigma_0^2-\sigma_1^2)(q-q_{\rm m})}} . \tag{18}$$

Putting together the ingredients and writing the Gaussian terms explicitly gives

$$\begin{aligned} f(q|\mu,\sigma_\mu) &= \frac{1}{\sqrt{2(q-q_{\rm m})q''}}\frac{1}{\sqrt{2\pi}\sigma_\mu}\left\{\exp\left[-\frac{(x_{\rm m}+\sqrt{(2/q'')(q-q_{\rm m})}-\mu)^2}{2\sigma_\mu^2}\right]\right. \\ &+ \left.\exp\left[-\frac{(x_{\rm m}-\sqrt{(2/q'')(q-q_{\rm m})}-\mu)^2}{2\sigma_\mu^2}\right]\right\} \end{aligned} \tag{19}$$

for $q < q_{\rm m}$ and zero otherwise. Distributions of q according to Eq. (19) are shown in Figs. 2 for $\mu_0 = 0$, $\mu_1 = 1$, and with values of σ_0 and σ_1 computed using Eqs. (6) and (7) based on the values of s and b shown.

4 Cumulative distribution of q

To compute the p-value of a hypothesized value of μ, one requires the cumulative distribution

$$F(q|\mu,\sigma_\mu) = \int_{-\infty}^{q} f(q'|\mu,\sigma_\mu)\,dq' \tag{20}$$

Because of the parabolic shape of the function $q(x)$, the probability to find q less than a given value is equal to one minus the probability to find x between the corresponding values of x_- and x_+, i.e.,

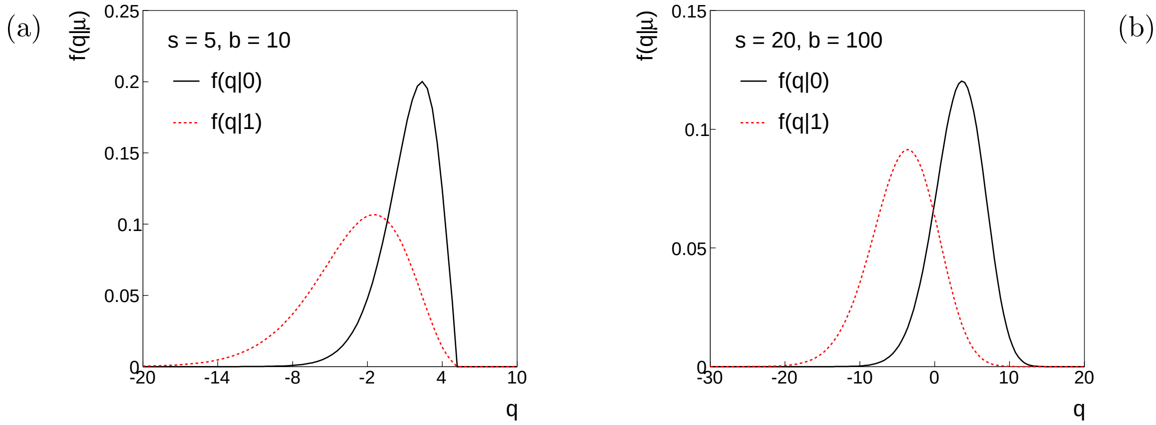

Figure 2: Distributions of q given values of $\mu = 0$ and $\mu = 1$ using values of σ_0 and σ_1 based on the values of s and b shown. These give $\sigma_0 = 0.632$ and $\sigma_1 = 0.775$ for (a) and $\sigma_0 = 0.5$ and $\sigma_1 = 0.548$ for (b).

$$F(q|\mu, \sigma_\mu) = 1 - P(x_- < x < x_+) = 1 - \Phi\left(\frac{x_+ - \mu}{\sigma_\mu}\right) + \Phi\left(\frac{x_- - \mu}{\sigma_\mu}\right) , \tag{21}$$

where x_+ and x_- are given by Eq. (15) and Φ is the cumulative distribution of the standard Gaussian.

If one has $\mu_1 > \mu_0$, as used in the examples here, then the p-value of H_0 (i.e., μ_0, σ_0) is the probability, assuming H_0, to find q less than or equal to what one observed, i.e.,

$$p_0 = F(q|\mu_0, \sigma_0) \; . \tag{22}$$

In a similar way, the p-value of the H_1 hypothesis is then the probability, assuming H_1, to find q greater than or equal to the value observed,

$$p_1 = 1 - F(q|\mu_1, \sigma_1) \; . \tag{23}$$

References

[1] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C 71 (2011) 1554.

[2] A. Stuart, J.K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2A: *Classical Inference and the Linear Model* 6th Ed., Oxford Univ. Press (1999), and earlier editions by Kendall and Stuart.