Glen Cowan
RHUL Physics
30 April, 2012

Bayes Factors for Discovery

The fundamental quantity one should use in the Bayesian framework to quantify the significance of a discovery is the posterior probability of the signal model. This depends, however, on the prior probability for all possible hypotheses, including that of the signal, on which different people may have widely differing views.

Alternatively, one can one can summarize an experimental result by use of a Bayes factor, B_{10}, which quantifies the degree to which one of two hypotheses, H_0 or H_1, is preferred by the data. This requires no overall prior probabilities for H_0 or H_1, but priors must be given for all of the internal parameters of the two models.

For a pair of hypotheses H_0 and H_1 the Bayes factor is defined as the posterior odds divided by the prior odds,

$$B_{10} = \frac{P(H_1|x)}{P(H_0|x)} \frac{\pi_0}{\pi_1} = \frac{P(x|H_1)}{P(x|H_0)} \; . \tag{1}$$

Here x refers to the data and π_i $(i = 0, 1)$ are the prior probabilities. That is, B_{10} is the same as the posterior odds if one were to assume equal prior probabilities, and it is thus an indicator of which model is preferred by the data. The second equality in (1) follows from Bayes' theorem, and therefore the Bayes factor is also equal to the ratio of likelihoods.

If a model contains any internal parameters, then to obtain the likelihood these must be characterized by a meaningful prior pdf and marginalized, i.e.,

$$P(x|H_i) = \int P(x|H_i, \theta_i)\pi_i(\theta_i) \, d\theta_i \; , \tag{2}$$

where θ_i are the internal parameters for model H_i $(i = 0, 1)$ and $\pi_i(\theta_i)$ is the corresponding prior pdf. It is important to note that in this case the prior pdf cannot be improper, as this would only be defined up to an arbitrary constant and the Bayes factor would not be well defined. Furthermore, if an improper prior is made proper by imposing a cut-off, then the Bayes factor will retain a dependence on this cut-off. Thus all internal parameters of the models must be characterized by meaningful, proper priors.

When using the Bayes factor to quantify significance of a Higgs signal, the hypothesis H_0 always refers to the background-only model, and its internal parameters include the full set of nuisance parameters $\boldsymbol{\theta}$ (of course the Higgs mass does not appear in H_0).

For the alternative hypothesis, H_1, one can choose specific values of the strength parameter μ and/or the Higgs mass $m_{\rm H}$ and calculate the Bayes factor as a function of these. In this case the look-elsewhere effect is not taken into account, and so one may find a large value of B_{10} for some values $m_{\rm H}$ if a large mass range is searched. To account for the look-elsewhere effect, one should specify a prior for the Higgs mass and integrate over $m_{\rm H}$ when calculating the marginal likelihood.

An important difficulty in computing Bayes factors is related to numerical challenges in computing the required marginal likelihoods given by Eq. (2). Here this has been done using nested sampling [1] as implemented in the `MultiNest` package [2]. In this algorithm one defines

$$X(\lambda) = \int_{L(\theta)>\lambda} \pi(\theta)\, d\theta \, , \tag{3}$$

so that the desired integral can be written

$$\int L(\theta)\pi(\theta)\, d\theta = \int_0^1 \lambda(X)\, dX \, . \tag{4}$$

Here $\lambda(X)$ is the inverse of Eq. (3), and in this way the marginal likelihood is reduced to a one-dimensional integral.

1 Bayes factors for the Poisson problem

Consider a measured number of events n that follows a Poisson distribution with mean $s+b$. Suppose b is known, and we want to distinguish between two hypotheses:

$$H_0 \; : \; s = 0 \, ,$$
$$H_1 \; : \; s > 0 \, .$$

The likelihood for H_0 is

$$L(n|H_0) = \frac{b^n}{n!} e^{-b} \, , \tag{5}$$

and for H_1 it is

$$L(n|s, H_1) = \frac{(s+b)^n}{n!} e^{-(s+b)} \, , \tag{6}$$

Suppose that the overall prior probabilities for the two hypotheses are $P(H_0)$ and $P(H_1) = 1 - P(H_0)$. Furthermore suppose that the prior probability for s under assumption of H_1 is

$$\pi(s|H_1) = \frac{1}{s_{\max}} \tag{7}$$

for $0 < s \leq s_{\max}$ and zero otherwise.

The posterior probability density for s given n under assumption of H_1 is from Bayes' theorem

$$p(s|n, H_1) = \frac{L(n|s, H_1)\pi(s|H_1)}{\int L(n|s, H_1)\pi(s|H_1)\, ds} \tag{8}$$

$$= \frac{(s+b)^n e^{-(s+b)}}{\int_0^{s_{\max}} (s+b)^n e^{-(s+b)}\, ds} \, . \tag{9}$$

In the limit where $s_{\max}$ to infinity, this probability goes to

$$p(s|n, H_1) = \frac{(s+b)^n e^{-(s+b)}}{\Gamma(n+1) - \gamma(n+1, b)} \,, \tag{10}$$

where

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} \, dt \tag{11}$$

is the lower incomplete gamma function and $\Gamma(a) = \gamma(a, \infty)$ is the usual Euler gamma function. That is, the posterior probability $p(s|n, H_1)$ approaches a limiting form that is independent of $s_{\max}$.

In addition to $p(s|n, H_1)$, however, we would also like to know the posterior probabilities of the two hypotheses, $P(H_0|n)$ and $P(H_1|n)$. Applying again Bayes' theorem, these are found to be

$$P(H_0|n) = \frac{L(n|H_0)P(H_0)}{P(n)} \,, \tag{12}$$

$$P(H_1|n) = \frac{\int_0^{s_{\max}} L(n|s, H_1)P(H_1)\pi(s|H_1) \, ds}{P(n)} \,, \tag{13}$$

where $P(n)$ is the probability for n summed over all hypotheses,

$$P(n) = L(n|H_0)P(H_0) + \int_0^{s_{\max}} L(n|s, H_1)P(H_1)\pi(s|H_1) \, ds \,. \tag{14}$$

For H_1 we integrate over s to find the marginal likelihood,

$$m_1 = \int L(n|s, H_1)\pi(s|H_1) \, ds \tag{15}$$

$$= \frac{1}{n! s_{\max}} \int_0^{s_{\max}} (s+b)^n e^{-(s+b)} \, ds \tag{16}$$

$$= \frac{1}{n! s_{\max}} \left(\gamma(n+1, s_{\max} + b) - \gamma(n+1, b) \right) \,. \tag{17}$$

The hypothesis H_0 has no nuisance parameters so its marginal likelihood is simply $m_0 = L(n|H_0)$. The desired posterior probabilities are

$$P(H_0|n) = \frac{m_0 P(H_0)}{m_0 P(H_0) + m_1 P(H_1)} \,, \tag{18}$$

$$P(H_1|n) = \frac{m_1 P(H_1)}{m_0 P(H_0) + m_1 P(H_1)} \,, \tag{19}$$

and the Bayes factor B_{10} is

$$B_{10} = \frac{P(H_0|n)/P(H_0)}{P(H_1|n)/P(H_1)} = \frac{m_1}{m_0} \tag{20}$$

$$= \frac{1}{s_{\max}} \frac{\gamma(n+1, s_{\max}+b) - \gamma(n+1, b)}{b^n e^{-b}} \,. \tag{21}$$

Although the conditional posterior probability $p(s|n, H_1)$ can be normalized to unity and thus decouples from $s_{\max}$ in the limit $s_{\max} \to \infty$, the probabilities for H_0 and H_1, and thus also the Bayes factor, retain a dependence on $s_{\max}$, such that in the limit where $s_{\max} \to \infty$, then $P(H_0|n) \to 1$, $P(H_1|n) \to 0$, and thus $B_{10} \to 0$. The Bayes factor is shown in Fig. 1 for $b = 2$, $n = 8$ as a function of $s_{\max}$.

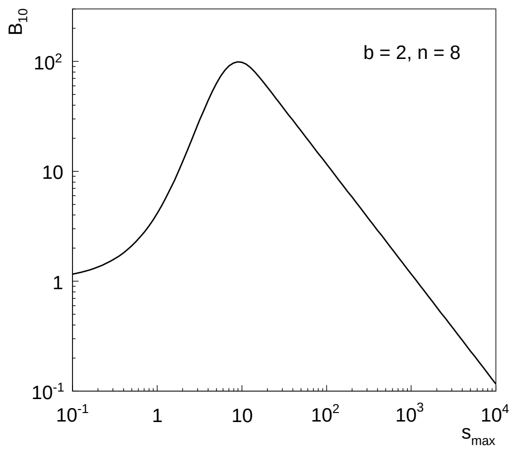

Figure 1: The Bayes factor B_{10} versus the cut-off $s_{\max}$ for $b = 2$, $n = 8$ (see text).

For $s_{\max} = 0$, by construction one has $B_{10} = 1$. As $s_{\max}$ is increased, the data initially favour the signal model H_1. If $s_{\max}$ becomes very large, however, then the increased volume of the parameter space of H_1 penalizes its probability, and the Bayes factor B_{01} tends toward zero.

References

[1] J. Skilling, Bayesian Analysis (2006) 1, Number 4, pp. 833–860.

[2] F. Feroz, M.P. Hobson and M. Bridges, Mon. Not. Roy. Astron. Soc., 398, 4, 1601-1614 (2009); arXiv:0809.3437.