Bayesian Hypothesis Testing in ExpAn

Shan Huang

May 22, 2017

This documentation explains how Bayes factor is implemented in ExpAn.
This is not a formal paper. The intention is to share knowledge between developers internally.

1 Problem

Given samples **x** from treatment group, samples **y** from control group, we want to know whether there is a statistically significant difference between the mean of two variants.

2 Model setup

We assume that

$$\begin{aligned} x &\sim \mathcal{N}(\mu + \alpha, \sigma^2) \\ y &\sim \mathcal{N}(\mu, \sigma^2) \end{aligned} \tag{2.1}$$

δ represents how many units of std is the mean difference:

$$\delta = \frac{\alpha}{\sigma} \tag{2.2}$$

We place non-informative priors on the data model:

$$\begin{aligned} \mu &\sim Cauchy(0, 1) \\ \sigma &\sim Gamma(2, 2) \end{aligned} \tag{2.3}$$

3 Bayesian hypothesis testing

We are interested in the "standardised" difference δ. Therefore, we need to make hypothesis on δ and compare the hypothesis.
Our null hypothesis of δ is always a fixed value 0, which means we believe there is no difference at all. On the other hand, our alternative hypothesis H_1 is a probability distribution, such that we can update the posterior $p(H_1|D)$, or $p_1(D)$, after seeing the data. The prior distribution of δ is a vague cauchy.

$$\begin{aligned} &H_0 : \delta = 0 \\ &H_1 : \delta \sim Cauchy(0,1) \end{aligned} \tag{3.1}$$

After seeing the data, we update the posterior distribution $p(H_1|D)$ represented by $p_1(\delta|D)$. We compute the 95% highest density interval of the posterior $p_1(\delta|D)$ as credible interval of δ. We can then use the credible interval of δ to compute credible interval of α. Finally, credible interval of α can be used to decide whether the result is statistically significant (whether 0 is outside the interval).

4 Early stopping

4.1 Stop by Bayes precision

We can use the width of credible interval as the stop criteria, because the smaller the interval is, the more certain we are about the posterior. If we are certain about our posterior, then we can stop. ;)
From best practice, we set the threshold of width to 0.08.

4.2 Stop by Bayes factor

Bayes factor simply compares the ratio of likelihood:

$$BF_{01} = \frac{p(D|H_0)}{p(D|H_1)} \tag{4.1}$$

BF_{01} smaller than 1/3 can be interpreted as support for the null hypothesis (no difference), higher than 3 can be interpreted as support for the alternative hypothesis (significant difference). Values between 1/3 and 3 are inconclusive.
The threshold $BF_{01} > 3$ or $BF_{01} < \frac{1}{3}$ are heuristic values from best practice. The definition of strong evidence is subjective, one can also use a threshold of 10 or more.

4.3 Stop by credible interval

Note that we can also stop and/or analyse siginificance by credible interval as described in section 3. A comparison of different stopping criteria and significance analysis will be summarised in another documentation.
(Not implemented yet.)

5 Savage-Dickey density ratio

Since the likelihood in Bayes factor involves an intractable integral over δ, we use Savage-Dickey density ratio to compute Bayes factor in implementation.
For simplicity, we use subscripts 0 and 1 to denote the probability densities under hypothesis H_0 and H_1, respectively:

$$BF_{01} = \frac{p_0(D)}{p_1(D)} = \frac{p_1(\delta = 0|D)}{p_1(\delta = 0)} \tag{5.1}$$

Proof:
Assume the model parameter falls into two categories $\theta = (\phi, \psi)$.
ϕ are the parameters of interest for the hypothesis testing. In our model setup, our ϕ is δ. We further assume under null hypothesis ϕ is set to a special fixed value $\phi = \phi_0$, whereas the alternative hypothesis places a distribution over ϕ — the same setup in section 3.
ψ are so-called nuisance parameters. We don't care about the value ψ in the experiment. We introduce ψ here only for the reason of deriving the formula of Savage-Dickey density ratio.

Now assume that $p(\psi|\phi)$ does not depend on the model. Then if ϕ is continuous at ϕ_0,

$$\lim_{\phi \to \phi_0} p_1(\psi|\phi) = p_0(\psi|\phi_0) \tag{5.2}$$

By setting $\phi = \phi_0$ for H_1, and $\phi = \phi_0$ by definition for H_0, we get a lemma:

$$p_1(\psi|\phi = \phi_0) = p_0(\psi) \tag{5.3}$$

The marginal likelihood over ψ under H_0 is given by

$$p_0(D) = \int p_0(D|\psi)\, p_0(\psi)\, \mathrm{d}\psi. \tag{5.4}$$

Using the continuity condition and our lemma, this can be rewritten as

$$p_0(D) = \int p_1(D|\psi, \phi = \phi_0)\, p_1(\psi|\phi = \phi_0)\, \mathrm{d}\psi. \tag{5.5}$$

Note that the right-hand side of the equation above is a marginalization over ψ, thus ψ disappears!

$$p_0(D) = p_1(D|\phi = \phi_0) \tag{5.6}$$

Finally, we apply Bayes' rule to get

$$p_0(D) = \frac{p_1(\phi = \phi_0|D)\, p_1(D)}{p_1(\phi = \phi_0)} \tag{5.7}$$

Substituting $p_0(D)$ in the equation of Bayes factor

$$BF_{01} = \frac{p(D|H_0)}{p(D|H_1)} = \frac{p_0(D)}{p_1(D)} = \frac{p_1(\phi = \phi_0|D)}{p_1(\phi = \phi_0)} \tag{5.8}$$

■