# Bayesian Hypothesis Testing in ExpAn

Shan Huang

April 28, 2017

This documentation explains how Bayes factor is implemented in ExpAn.
It is not a formal paper. The intention of the documentation is to share knowledge between developers internally, and I want to keep it short.

## 1 Problem

Given samples **x** from treatment group, samples **y** from control group[1], we want to know whether there is a statistically significant difference.

## 2 Model setup

We assume that

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

The "z-scored" or "standardised" difference we are interested in is:

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

Before seeing any data, we place the following non-informative priors on the parameters:

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

[1] Control is the group for baseline metric. So $y$ is the baseline. We add a random variable on $y$ to get $x$. This is somewhat confusing to me, but it is the terminology we are using throughout ExpAn.

## 3 Bayesian hypothesis testing

We are interested in the "standardised" difference $\delta$. Therefore, we need to make hypothesis on $\delta$ and compare the hypothesis.
Note that our null hypothesis of $\delta$ 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 in other notation $p_1(D)$, after seeing the data.

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

After seeing the data, we compute the posterior distribution $p(H_1|D)$ represented by $p_1(\delta|D)$. This is done via MCMC sampling from `StanModel.sampling`, which is pretty slow[2] and might have some potential improvements.
After getting the posterior, we compute the .95 posterior density interval and show it in our result object in column `pctile` and `value`. This interval is computed via another MCMC sampling, here we implemented this method ourselves in `early_stopping.HDI_from_MCMC`.

## 4 Early stopping

### 4.1 Bayes precision

We use the width of .95 posterior interval as the stop criteria. The smaller the density interval is, the more certain we are about the posterior. If we are certain about our posterior, then we can stop. ;)
For some reason, we set the threshold of width to 0.08.

### 4.2 Bayes factor

Bayes factor simply compares the ratio of likelihood:

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

If this value is big or small enough (which means the difference between two hypothesis is pretty large), we make the decision of early stopping. For some reason, we set the threshold to a heuristic value: $BF_{01} > 3$ or $BF_{01} < \frac{1}{3}$.
Note that our analysis result is always the .95 credible interval as described in section 3. We only use Bayes factor for the decision of early stopping.

[2] Is there a better way? Maybe there exist an analytical solution.

## 4.3 Savage-Dickey density ratio

Since the likelihood in Bayes factor involves an intractable integral over $\delta$, 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:

$$BF_{01} = \frac{p_0(D)}{p_1(D)} = \frac{p_1(\delta = 0|D)}{p_1(\delta = 0)} \tag{4.2}$$

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

$$\lim_{\phi \to \phi_0} p_1(\psi|\phi) = p_0(\psi|\phi_0) \tag{4.3}$$

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

$$p_1(\psi|\phi = \phi_0) = p_0(\psi) \tag{4.4}$$

The marginal likelihood over $\psi$ under $H_0$ is given by

$$p_0(D) = \int p_0(D|\psi)\, p_0(\psi)\, \mathrm{d}\psi. \tag{4.5}$$

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

$$p_0(D) = \int p_1(D|\psi, \phi = \phi_0)\, p_1(\psi|\phi = \phi_0)\, \mathrm{d}\psi. \tag{4.6}$$

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

$$p_0(D) = p_1(D|\phi = \phi_0) \tag{4.7}$$

Finally, we apply Bayes' rule to get

$$p_0(D) = \frac{p_1(\phi = \phi_0|D)\, p_1(D)}{p_1(\phi = \phi_0)} \tag{4.8}$$

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

$$BF_{01} = \frac{p(D|H_0)}{p(D|H_1)} = \frac{p_0(D)}{p_1(D)} = \frac{p_1(\phi = \phi_0|D)}{p_1(\phi = \phi_0)} \tag{4.9}$$

■