# STA 360/602L: Module 6.4

## Bayesian hypothesis testing

Dr. Olanrewaju Michael Akande

# Bayesian hypothesis testing

- How to do Bayesian hypothesis testing for a simple model?
- Suppose we have univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\mu, 1)$ and wish to test $\mathcal{H}_0 : \mu = 0; \;\; \text{vs.} \mathcal{H}_1 : \mu \neq 0$ under the Bayesian paradigm.
- Informal approach:
  1. Put a prior on $\mu$, $\pi(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$.
  2. Compute posterior $\mu | Y = (y_1, \ldots, y_n) \sim \mathcal{N}(\mu_n, \sigma_n^2)$ for updated parameters $\mu_n$ and $\sigma_n^2$.
  3. Compute a 95% CI based on the posterior.
  4. Reject $\mathcal{H}_0$ if interval does not contain zero.

STA 602L

# Bayesian hypothesis testing

- Formal approach:

  1. Put a prior on the actual hypotheses/models, that is, on $\pi(\mathcal{H}_0) = \Pr(\mathcal{H}_0 = \text{True})$ and $\pi(\mathcal{H}_1) = \Pr(\mathcal{H}_1 = \text{True})$.

     For example, set $\pi(\mathcal{H}_0) = 0.5$ and $\pi(\mathcal{H}_1) = 0.5$, if apriori, we believe the two hypotheses are equally likely.

  2. Put a prior on the parameters in each model.

     In our simple normal model, the only unknown parameter is $\mu$, so for example, our prior can once again be $\pi(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$.

  3. Compute marginal posterior probabilities for each hypothesis, that is, $\pi(\mathcal{H}_0|Y)$ and $\pi(\mathcal{H}_1|Y)$. Can start with the joint posterior between each hypothesis and the parameter, then integrate out the parameter.

  4. Conclude based on the magnitude of $\pi(\mathcal{H}_1|Y)$ relative to $\pi(\mathcal{H}_0|Y)$.

STA 602L

# Bayesian hypothesis testing

- Using Bayes theorem,

$$\pi(\mathcal{H}_1|Y) = \frac{p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)}{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)},$$

  where $p(Y|\mathcal{H}_0)$ and $p(Y|\mathcal{H}_1)$ are the marginal likelihoods for the data under the null and alternative hypotheses respectively.

- If for example we set $\pi(\mathcal{H}_0) = 0.5$ and $\pi(\mathcal{H}_1) = 0.5$ apriori, then

$$\begin{aligned}\pi(\mathcal{H}_1|Y) &= \frac{0.5p(Y|\mathcal{H}_1)}{0.5p(Y|\mathcal{H}_0) + 0.5p(Y|\mathcal{H}_1)} \\ &= \frac{p(Y|\mathcal{H}_1)}{p(Y|\mathcal{H}_0) + p(Y|\mathcal{H}_1)} = \frac{1}{\frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)} + 1}.\end{aligned}$$

- The ratio $\frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)}$ is known as the Bayes factor in favor of $\mathcal{H}_0$, and often written as $\mathcal{BF}_{01}$. Similarly, we can compute $\mathcal{BF}_{10}$.

STA 602L

# Bayes factors

- Bayes factor: is a ratio of marginal likelihoods and it provides a weight of evidence in the data in favor of one model over another.

- It is often used as an alternative to the frequentist p-value.

- **Rule of thumb**: $\mathcal{BF}_{01} > 10$ is strong evidence for $\mathcal{H}_0$; $\mathcal{BF}_{01} > 100$ is decisive evidence for $\mathcal{H}_0$.

- Notice that for our example,

$$\pi(\mathcal{H}_1|Y) = \frac{1}{\frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)} + 1} = \frac{1}{\mathcal{BF}_{01} + 1}$$

  the higher the value of $\mathcal{BF}_{01}$, that is, the weight of evidence in the data in favor of $\mathcal{H}_0$, the lower the marginal posterior probability that $\mathcal{H}_1$ is true.

- That is, here, as $\mathcal{BF}_{01} \uparrow$, $\pi(\mathcal{H}_1|Y) \downarrow$.

STA 602L

# Bayes factors

- Let's look at another way to think of Bayes factors. First, recall that

$$\pi(\mathcal{H}_1|Y) = \frac{p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)}{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)},$$

so that

$$\begin{aligned}
\frac{\pi(\mathcal{H}_0|Y)}{\pi(\mathcal{H}_1|Y)} &= \frac{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0)}{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)} \div \frac{p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)}{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)} \\
&= \frac{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0)}{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)} \times \frac{p(Y|\mathcal{H}_0)\pi(\mathcal{H}_0) + p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)}{p(Y|\mathcal{H}_1)\pi(\mathcal{H}_1)} \\
\therefore \underbrace{\frac{\pi(\mathcal{H}_0|Y)}{\pi(\mathcal{H}_1|Y)}}_{\text{posterior odds}} &= \underbrace{\frac{\pi(\mathcal{H}_0)}{\pi(\mathcal{H}_1)}}_{\text{prior odds}} \times \underbrace{\frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)}}_{\text{Bayes factor } \mathcal{BF}_{01}}
\end{aligned}$$

- Therefore, the Bayes factor can be thought of as the factor by which our prior odds change (towards the posterior odds) in the light of the data.

- In linear regression, **BIC** approximates the $\mathcal{BF}$ comparing a model to the null model.

# Bayes factors

- While Bayes factors can be appealing, calculating them can be computationally demanding.
- Why have we been "mildly obsessed" with MCMC sampling? To avoid computing any **marginal likelihoods**! Well, guess what? Bayes factors are ratios of marginal likelihoods, taking us back to the problem we always try to avoid.
- Of course this isn't all *"doom and gloom"*, there are various ways (once again!) of getting around computing those likelihoods analytically.
- Unfortunately, we will not have time to cover them in this course.

STA 602L

# Bayes factors

- As a teaser, one approach is to flip the relationship on the previous slide:

$$\underbrace{\frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)}}_{\text{Bayes factor } \mathcal{BF}_{01}} = \underbrace{\frac{\pi(\mathcal{H}_0|Y)}{\pi(\mathcal{H}_1|Y)}}_{\text{posterior odds}} \times \underbrace{\frac{\pi(\mathcal{H}_1)}{\pi(\mathcal{H}_0)}}_{\text{prior odds}},$$

  which is easy to compute as long as we can use posterior samples to compute/approximate the posterior odds.

- Bayes factors can work well when the underlying model is discrete but do not work well for models that are inherently continuous.

- For more discussions on this, see Chapter 7.4 of Bayesian Data Analysis (Third Edition).

- Even in the discrete case, Bayes factors are not perfect, as we see in the following example.

STA 602L

# Hypothesis testing example

- Suppose we have univariate data $y_1, \ldots, y_n | \theta \sim \text{Bernoulli}(\theta)$.
- Also, suppose we wish to test $\mathcal{H}_0 : \theta = 0.5 \text{ vs. } \mathcal{H}_1 : \theta \neq 0.5$, using the Bayes factor.
- First, we need to put priors on the two hypotheses. Again, if apriori we believe the two hypotheses are equally likely, then we can set

$$\pi(\mathcal{H}_0) = \text{Pr}(\mathcal{H}_0 = \text{True}) = 0.5; \ \ \pi(\mathcal{H}_1) = \text{Pr}(\mathcal{H}_1 = \text{True}) = 0.5.$$

- Next, we need to put priors on the parameters in each model.
  - When $\mathcal{H}_0$ is true, we have that $\theta = 0.5$ and so there's no need for a prior on $\theta$.
  - When $\mathcal{H}_1$ is true, we can set a conjugate prior for $\theta$, that is, $\text{Beta}(a, b)$.

STA 602L

# Hypothesis testing example

- To compute the Bayes factor, we need to compute $p(Y|\mathcal{H}_0)$ and $p(Y|\mathcal{H}_1)$.
- For each, we need to start with the joint distribution of the data and parameter, given each hypothesis, then integrate out the parameter.
- For $p(Y|\mathcal{H}_0)$, we have

$$\begin{aligned}
p(Y|\mathcal{H}_0) &= \int_0^1 p(Y,\theta|\mathcal{H}_0)\mathrm{d}\theta \\
&= \int_0^1 p(Y|\mathcal{H}_0,\theta)\cdot\pi(\theta|\mathcal{H}_0)\mathrm{d}\theta \\
&= \int_0^1 p(Y|\theta=0.5)\cdot 1\ \mathrm{d}\theta \\
&= \int_0^1 0.5^{\sum_{i=1}^n y_i}(1-0.5)^{n-\sum_{i=1}^n y_i}\cdot 1\ \mathrm{d}\theta \\
&= 0.5^n\int_0^1 1\ \mathrm{d}\theta \\
&= 0.5^n
\end{aligned}$$

# Hypothesis testing example

- For $p(Y|\mathcal{H}_1)$, we have

$$\begin{aligned} p(Y|\mathcal{H}_1) &= \int_0^1 p(Y|\mathcal{H}_1,\theta)\cdot\pi(\theta|\mathcal{H}_1)\mathrm{d}\theta \\ &= \int_0^1 \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i}\cdot\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathrm{d}\theta \\ &= \frac{1}{B(a,b)}\int_0^1 \theta^{a+\sum_{i=1}^n y_i-1}(1-\theta)^{b+n-\sum_{i=1}^n y_i-1}\mathrm{d}\theta \\ &= \frac{B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)}{B(a,b)} \end{aligned}$$

- Bayes factor in favor of $\mathcal{H}_0$, $\mathcal{BF}_{01}$, is therefore

$$\mathcal{BF}_{01} = \frac{p(Y|\mathcal{H}_0)}{p(Y|\mathcal{H}_1)} = \frac{0.5^n B(a,b)}{B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)}.$$

- Also,

$$\pi(\mathcal{H}_1|Y) = \frac{1}{\mathcal{BF}_{01}+1} = \frac{1}{\frac{0.5^n B(a,b)}{B(a+\sum_{i=1}^n y_i, b+n-\sum_{i=1}^n y_i)}+1}.$$

STA 602L

# Hypothesis testing example

- Suppose the true value of $\theta = 0.6$. Suppose that in $n = 20$ trials, we observe $13$ successes, that is, $\sum_{i=1}^{n} y_i = 13$.
- If we assume a $\text{Beta}(a = 1, b = 1)$ prior on $\theta$, then $\mathcal{BF}_{01}$ is

```
0.5^20*beta(1,1)/beta(1+13,1+7)
```

```
## [1] 1.552505
```

- On the other hand, $\mathcal{BF}_{10} \approx 0.64$. So that even though based on the data, our estimate of $\theta$ is $\hat{\theta} = \frac{13}{20} = 0.65$, we still have stronger evidence in favor of $\mathcal{H}_0$ over $\mathcal{H}_1$, which is interesting!
- There are a few contributing factors, including the sample size, our choice of prior, and how far $\hat{\theta}$ is from the true $\theta$.
- You will explore this in more detail on the homework.

# What's next?

Move on to the readings for the next module!