STA 365: Applied Bayesian Statistics

Boris Babic
Assistant Professor, University of Toronto

Week 6B: Hypothesis Testing

Hypothesis Testing: Overview

- Recall that everything on the Bayesian approach goes through the posterior distribution.
- HT is not exception. We'll want to identify hypotheses, derive their posterior distributions, and then consider some test statistics.
- But on the Bayesian approach, the rationale for doing this is less obvious. We already have the full posterior.
- The epistemology underlying Bayesianism (confirmation of hypotheses by evidence) is very different from the epistemology underlying classical NHST (falsification of hypotheses by evidence).
- However, you may want to perform hypothesis tests for various reasons and as such we look at how to do so.

Classical Hypothesis Testing

- Set up: $X_1, ..., X_n = \mathbf{X} \in \mathbf{Z} \overset{\text{iid}}{\sim} f(\mathbf{x}|\theta),\ \theta \in \Theta \subset \mathbb{R}$.

  $H : \theta \in \Theta_0, \Theta_0 \subset \Theta$

  $K : \theta \in \Theta_1, \Theta_1 \subset \Theta$

  $\Theta_0 \cap \Theta_1 = \emptyset$

  $\Theta_0 \cup \Theta_1 = \Theta$ is not necessary.
- Goal: To decide if $\theta \in \Theta_0$ or $\theta \in \Theta_1$
- Define:

  $H_0 : \theta \in \Theta_0 \to$ Null

  $H_1 : \theta \in \Theta_1 \to$ Alternate

Classical Hypothesis Testing

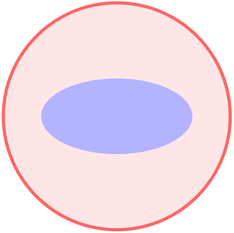
Error Probabilities

	H_0 True	H_1 True
H_0	✓	Type II error
H_1	Type I error	✓

$\mathrm{P}(\text{Type I}) = \alpha$
$\mathrm{P}(\text{Type II}) = \beta$

Classical Hypothesis Testing

- Test (NP Framework): To choose a decision function for testing H_0 vs. H_1 is to partition Z into A and R, where $A \cap R = \emptyset$ and $A \cup R = Z$.



$$\alpha = \Pr\{\mathbf{X} \in R | \theta \in \Theta_0\}$$

$$\beta = \Pr\{\mathbf{X} \in A | \theta \in \Theta_1\}$$

Classical Hypothesis Testing

- Steps in applying the NP framework:
- Choose α close to zero (e.g., $\alpha = 0.01$).
- Restrict attention to tests $T = \{\delta(\mathbf{X}) : \Pr(\mathbf{X} \in R | \theta \in \Theta_0) \leq \alpha\}$.
- among all tests in T choose $\arg\min_\delta \Pr(\mathbf{X} \in A | \theta \in \Theta_1)$.

Classical Hypothesis Testing

- Simple vs simple:

$$H_0 : \theta = \theta_0 \in \Theta$$

$$H_1 : \theta = \theta_1 \in \Theta$$

$$\Theta = \{\theta_0, \theta_1\}, \theta_1 > \theta_0$$

- NP Lemma: A post powerful level α test (i.e, that minimizes β), is of the form

$$R = \{\mathbf{X} \in Z : LR \leq k\}$$

where

$$LR = \frac{f(\mathbf{x}|H_0)}{f(\mathbf{x}|H_1)} = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)} = \frac{1}{\Lambda}$$

Classical Hypothesis Testing

Example: Normal Distribution

Suppose $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$ where σ^2 is known. Consider testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ where $\theta_1 > \theta_0$.

Classical Hypothesis Testing

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} \left[\frac{1}{2\pi\sigma^2} \exp\left\{ - \frac{(x_i - \theta)^2}{2\sigma^2} \right\} \right]$$

$$\frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} = \frac{\exp\left\{ -\sum_{i=1}^{n} \frac{(x_i-\theta_1)^2}{2\sigma^2} \right\}}{\exp\left\{ -\sum_{i=1}^{n} \frac{(x_i-\theta_0)^2}{2\sigma^2} \right\}}$$

$$= \exp\left\{ - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\sigma^2} + \sum_{i=1}^{n} \frac{(x_i - \theta_0)^2}{2\sigma^2} \right\}$$

$$= \exp\left\{ \frac{\sum_{i=1}^{n}(x_1 - \theta_0)^2 - \sum_{i=1}^{n}(x_1 - \theta_1)^2}{2\sigma^2} \right\}$$

$$= \exp\left\{ \frac{n(\theta_0^2 - \theta_1^2) + 2\sum_{i=1}^{n} x_i(\theta_1 - \theta_0)}{2\sigma^2} \right\}$$

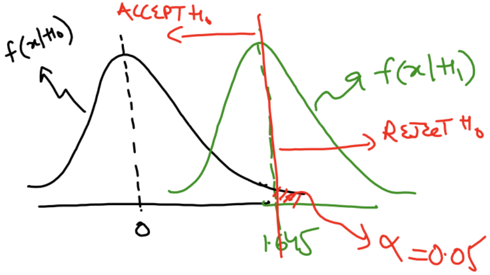
$$\rightarrow \frac{n(\theta_0^2 - \theta_1^2) + 2\sum_{i=1}^{n} x_i(\theta_1 - \theta_0)}{2\sigma^2} > \log k$$

$$\rightarrow \sum_{i=1}^{n} x_i > k^* \rightarrow \alpha = \Pr\left(\sum_{i=1}^{n} X_i > k^* | \theta_0 \right)$$

$$\rightarrow \frac{k^*/n - \theta_0}{\sigma/\sqrt{n}} = z_\alpha \rightarrow k^* = n(\theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}})$$

Classical Hypothesis Testing

- $H_0 : \mu = 0$ vs. $H_1 : \mu = 1.645$, $X \sim \mathrm{N}(\mu, 1)$, $n = 1$.
- $\alpha = 0.05$. $R = \{LR \leq k\} = \{X \in Z : x \geq c\}$
- $c = 1.645$



Classical Hypothesis Testing

p-value approach:

- Do not choose α a priori.
- Instead, observe a sample, compute a p-value, and then let a decision maker decide based on personal choice of α.

Classical Hypothesis Testing

Example

- Suppose that you toss a coin 12 times, obtaining 9 heads and three tails. If X_i corresponds to an outcome of the coin toss, then $Y = \sum X_i \overset{\text{iid}}{\sim} \text{Bin}(n, \theta)$.
- Let $H_0 : \theta = 0.5$. That is, you want to test if the coin is fair.
- Compute the p-value.

Classical Hypothesis Testing

Example

Let $Y = \sum X_i$. Then

$$Y \sim \text{Binomial}(n, \theta)$$

$$\begin{aligned} P(Y \geq 9 | \theta = 0.5, n = 12) &= \sum_{y=9}^{12} \binom{12}{y_i} \left(\frac{1}{2}\right)^{y_i} \left(\frac{1}{2}\right)^{12-y_i} \\ &= \left[\binom{12}{3} + \binom{12}{2} + \binom{12}{1} + \binom{12}{0}\right] \left(\frac{1}{2}\right)^{12} \\ &= \frac{299}{4096} \approx 0.07 \end{aligned}$$

The result is not statistically significant. Thus we cannot reject the hypothesis that the coin is fair.

Classical Hypothesis Testing

- Suppose the experiment was performed by your RA.
- You now learn there was a mix up.
- You thought you told the RA to toss the coin 12 times.
- But your RA actually tossed it until observing 3 tails.
- As it happens, it took 12 tosses to get 3 tails.
- Should this matter? The evidence is what it is isn't it?
- But now n is random and Y is fixed.
- Thus, a result more extreme than $(9,3)$ is no longer $(10,2)$, $(11,1)$, and $(12,0)$. Rather, it is $(10,3)$, $(11,3)$, $(12,3)$, and so on.
- We have to re-calculate the p-value.

Classical Hypothesis Testing

Second Attempt

$$
\begin{aligned}
P(N \geq 12 | \theta = 0.5, r = 3) &= \sum_{n=12}^{\infty} \binom{n_i - 1}{r - 1} .5^r .5^{n_i - r} \\
&= \sum_{n=12}^{\infty} \binom{n_i - 1}{2} .5^{n_i} \\
&= 1 - \sum_{n=1}^{11} \binom{n_i - 1}{2} .5^{n_i} \\
&\approx 0.03
\end{aligned}
$$

- Now the result *is* statistically significant!
- What if the RA stopped tossing the coin so that they can get a coffee?

  Sometimes there are ethical reasons to stop collecting data
  (HIV antiretroviral drug example)

Bayesian Approach

- Priors:

$$\pi(H_0), \pi(H_1) = 1 - \pi(H_0)$$

- Posteriors:

$$\pi(H_0|y) = \frac{\pi(H_0)f(y|H_0)}{\pi(H_0)f(y|H_0) + \pi(H_1)f(y|H_1)}$$

$$\pi(H_1|y) = \frac{\pi(H_1)f(y|H_1)}{\pi(H_1)f(y|H_1) + \pi(H_0)f(y|H_0)}$$

- Prior odds:

$$\Omega' = \frac{\pi(H_0)}{\pi(H_1)}$$

- Posterior odds:

$$\Omega'' = \frac{\pi(H_0|y)}{\pi(H_1|y)} = \frac{\pi(H_0)f(y|H_0)}{\pi(H_1)f(y|H_1)} \to \Omega'' = \Omega' \times LR$$

More on Losses

Bayesian Decision Theory

	H_0 True	H_1 True
Decide H_0	0	L_{II}
Decide H_1	L_I	0

$$\mathrm{E}\left[\text{Decide } H_0\right] = 0 \times \pi(H_0|y) + L_{II} \times \pi(H_1|y)$$

$$\mathrm{E}\left[\text{Decide } H_1\right] = L_I \times \pi(H_0|y) + 0 \times \pi(H_1|y)$$

$$\text{Decide } H_1 \text{ if} L_I \pi(H_0|y) < L_{II} \pi(H_1|y)$$

Odds and Likelihood Ratios

Decide H_1 if

$$\Omega'' < \frac{L_{II}}{L_I}$$

$$\Omega' \times LR < \frac{L_{II}}{L_I}$$

$$LR < \frac{L_{II}/L_I}{\Omega'}$$

The last line suggests that only relative severity of each type of error is what matters.

Example

Example

Let $X \sim \mathrm{N}(\mu, 1)$, $n = 1$, $H_0 : \mu = 0, H_1 : \mu = 1.645$
If $L_{II}/L_I = 1/2$ and $\Omega' = 1$, express the decision rule.

If $X = 1$, what will you do?

Odds and Likelihood Ratios

Solution

Decide H_1 if $LR < \frac{1}{2}$

$$\begin{aligned}
LR &= \frac{f(x|H_0)}{f(x|H_1)} \\
&= \frac{\exp\left[-\frac{x^2}{2}\right]}{\exp\left[-\frac{(x-1.645)^2}{2}\right]} \\
&= \exp\left[-\frac{x^2}{2} + \frac{(x-1.645)^2}{2}\right] \\
&= \exp\left[\frac{(x-1.645)^2}{2} - \frac{x^2}{2}\right] \\
&= \exp\left[\frac{(x-1.645)^2 - x^2}{2}\right] \\
&= \exp\left[\frac{x^2 + 1.645^2 - 2x1.645 - x^2}{2}\right] \\
&= \exp\left[\frac{1.645(1.645-2x)}{2}\right]
\end{aligned}$$

If $x = 1$, $LR \approx 0.7 \rightarrow$ decide H_0.

Standard One-Tailed Test

$H_0 : \theta \leq \theta_0$
$H_1 : \theta > \theta_0$
Priors:

$$\pi(H_0) = \int_{-\infty}^{\theta_0} f(\theta) d\theta$$

$$\pi(H_1) = \int_{\theta_0}^{\infty} f(\theta) d\theta$$

Posteriors:

$$\pi(H_0|y) = \int_{-\infty}^{\theta_0} f(\theta|y) d\theta$$

$$\pi(H_1|y) = \int_{\theta_0}^{\infty} f(\theta|y) d\theta$$

Bayes Factor

$$\Omega'' = \Omega' \times LR \rightarrow LR = \frac{\Omega''}{\Omega'} = BF$$

Example

Example

- Let $X \sim \mathrm{N}(\mu, 1)$, $n = 1$, $L_{II}/L_I = 1/2$.
- Let $\mu \sim \mathrm{N}(\mu_0 = 0, \tau_0 = 3)$.
- Consider $H_0 : \mu \leq 0$ vs $H_1 : \mu > 0$.
- Suppose $x = 1$.
- What should you decide?

Example

Solution

Note that $\sigma_0^2 = 1/3$ and $\tau = 1$. And $\mu|x \sim \mathrm{N}(\mu_1, \sigma^2/\tau_1)$, where

$$\mu_1 = \frac{n\tau + \tau_0\mu_0}{n + \tau_0} = \frac{1+0}{4} = 1/4$$

$$\tau_1 = 4$$

$$\Omega' = \frac{\pi(H_0)}{\pi(H_1)} = \frac{\pi(\mu \leq 0)}{\pi(\mu > 0)} = \frac{\Pr\left(Z \leq \frac{0-0}{1/\sqrt{1/3}}\right)}{\Pr\left(Z > \frac{0-0}{1/\sqrt{1/3}}\right)} = 1$$

$$\Omega'' = \frac{\pi(H_0|x)}{\pi(H_1|x)} = \frac{\pi(\mu \leq 0|x)}{\pi(\mu > 0|x)} = \frac{\Pr\left(Z \leq \frac{0-0.25}{0.5}\right)}{\Pr\left(Z > \frac{0-0.25}{0.5}\right)} = \frac{0.31}{0.69} = 0.45$$

Since $L_{II}/L_I = 1/2$ and $\Omega'' \leq 1/2$, decide H_1.

Two Tailed Test

Example

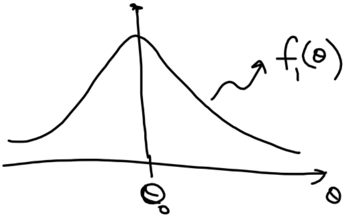
Suppose $X_1, ...X_n \overset{\text{iid}}{\sim} f(x|\theta)$. And consider

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

Priors

- Priors:

$$\pi(H_0) = \pi_0 \text{ and } \pi(H_1) = \pi_1 \sim f_1(\theta)$$



Two Tailed Test

Remaining Ingredients

- Likelihoods:

$$f(x|H_0) = f(x|\theta_0)$$

$$f(x|H_1) = \int_{-\infty}^{\infty} f(x|\theta) f_1(\theta) d\theta = f_1(x) \rightarrow \text{PPD under } H_1$$

- Full predictive:

$$f(x) = \pi_0 f(x|\theta_0) + \pi_1 f_1(x)$$

- Posteriors:

$$\begin{aligned}
\pi(H_0|x) &= \frac{\pi(H_0) f(x|H_0)}{\pi(H_0) f(x|H_0) + [1 - \pi(H_0)] f(x|H_1)} \\
&= \frac{\pi_0 f(x|\theta_0)}{\pi_0 f(x|\theta_0) + \pi_1 f_1(x)} \\
&= \frac{\pi_0 f(x|\theta_0)}{f(x)}
\end{aligned}$$

$$\pi(H_1|x) = \frac{\pi_1 f_1(x)}{f(x)}$$

Two Tailed Test

Odds Ratios

$$\begin{aligned}\Omega'' &= \Omega' \times LR \\ &= \frac{\pi_0}{\pi_1}\frac{f(x|\theta_0)}{f_1(x)} \\ &= \frac{\pi_0}{\pi_1} \times BF\end{aligned}$$

Two Tailed Test

Sufficient Statistics

Suppose $y = g(x)$ is a SS so that $f(x|\theta) = f(y|\theta)f(x|y)$.
Then,

$$\pi(H_0|x) = \pi(H_0|y)$$

$$\pi(H_1|x) = \pi(H_1|y)$$

And,

$$\Omega'' = \frac{\pi_0}{\pi_1}\frac{f(y|\theta_0)}{f_1(y)}$$

Two Tailed Test

Example

Let $X_1, ...X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma_0^2)$ where σ_0^2 is known.
$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
$\pi_0 = \pi_1 = 1/2$
$Y = \overline{X} \sim \mathrm{N}(\mu, \frac{\sigma_0^2}{n})$
Prior on μ under H_1:

$$f_1(\mu) \sim \mathrm{N}(\mu_0, \frac{\sigma^2}{n'})$$

where $\mu_0 = m'$.

Two Tailed Test

Example

Posteriors:

$$f(y|h_0) = f(y|\mu_0)$$

$$f(y|h_1) = \int_{-\infty}^{\infty} f(y|\mu) f_1(\mu) d\mu = f_1(y)$$

$$f_1(y) = f_1(\overline{x}) \sim \mathrm{N}(\mu_0, \frac{\sigma_0^2}{n} + \frac{\sigma_0^2}{n'})$$

Two Tailed Test

Example

Posterior Odds:

$$\Omega'' = \Omega' \times LR$$

$$= \frac{\pi_0}{\pi_1} \frac{\frac{\sqrt{n}}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(\overline{x}-\mu_0)^2\right]}{\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_0^2}{n}+\frac{\sigma_0^2}{n'}}} \exp\left[-\frac{1}{2}\frac{(\overline{x}-\mu_0)^2}{\frac{\sigma_0^2}{n}+\frac{\sigma_0^2}{n'}}\right]}$$

Bayes Factor:

$$BF = \frac{f(y|H_0)}{f_1(y)}$$

$$= \sqrt{\frac{n+n'}{n'}} \exp\left[\left\{-\frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2/n}\right\}\left\{1-\frac{n'}{n+n'}\right\}\right]$$

$$= \sqrt{\frac{n+n'}{n'}} \exp\left[-\frac{1}{2}z^2\left\{1-\frac{n'}{n+n'}\right\}\right]$$

Two Tailed Test

Example: Special Case

Suppose $n' = 1$, $n = 15$, $\pi_0 = \pi_1 = 1/2$ and $z = 1.96$
Then,

$$BF = \sqrt{16/1}\exp[-1/2z^2(1 - 1/16)] = 0.66$$

$$\Omega'' = \frac{\pi_0''}{1 - \pi_0''} = \frac{\pi_0}{\pi_1}BF \rightarrow \pi_0'' = \frac{BF}{1 + BF} = 0.4$$

This is the posterior probability of the null hypothesis. Compare this to p-value of 0.05!

Two Tailed Test

Extreme Result

$$BF = \sqrt{\frac{n+n'}{n'}} \exp[-1/2z^2(1 - \frac{n'}{n+n'})]$$

As $n \to \infty$, $BF \to \infty$

Lindley's Paradox

We can find a sample size and a sample result such that both of the following two conditions are satisfied:
(1) A classical two tailed test reveals an arbitrarily small p-value, and
(2) The LR/BF (and hence posterior odds ratio) is arbitrarily large.

A bound on BF

In general,

$$f_1(\overline{x}) = f_1(y) = \int_{-\infty}^{\infty} f(\overline{x}|\theta) f_1(\theta) d\theta \leq f(\overline{x}|\hat{\theta}_{MLE})$$

where $f(\overline{x}|\hat{\theta}_{MLE}) = \sup_\theta f(\overline{x}|\theta)$.

Normal Example

Let $\hat{\theta} = \overline{x}$. So $f_1(x) \leq \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_0}$

$$BF = \frac{f(\overline{x}|\mu_0)}{f_1(x)} \geq \exp -1/2z^2$$

Normal Example

z	p-value	Lower Bound on B	Bound on π_0''
1.645	0.1	0.258	0.205
1.96	0.05	0.146	0.128
2.58	0.01	0.036	0.035
3.29	0.001	0.004	0.004