Bayesian Hypothesis Testing

Yingbo Li

Clemson University

MATH 9810

Hypothesis Testing

- Traditional setting: $X_i \mid \theta \sim f(x \mid \theta)$ with $\theta \in \Theta$. Let $\{\Theta_0, \Theta_1\}$ partition of Θ. To test:

$$
\begin{aligned}
H_0 :& \theta \in \Theta_0 \\
H_1 :& \theta \in \Theta_1
\end{aligned}
$$

- Decision between H_0 and H_1 is simply based on their posterior probabilities:

$$
\begin{aligned}
\alpha_0 &= P(H_0 \mid \mathbf{X}) \\
\alpha_1 &= P(H_1 \mid \mathbf{X}) = 1 - \alpha_0
\end{aligned}
$$

- Conceptual advantages over frequentist counterpart
  - Posterior probabilities are easy to interpret
  - It does not matter which hypothesis is labeled H_0

- We approach hypothesis testing as a *model selection* problem. Hypotheses must have prior believability. Let:

$$\pi_0 = P(H_0) \quad \text{prior probability of } H_0$$
$$\pi_1 = P(H_1) = 1 - \pi_0 \quad \text{prior probability of } H_1$$

- Prior odds ratio (of H_0 to H_1) $= \frac{\pi_0}{\pi_1}$
- Posterior odds ratio (of H_0 to H_1) $= \frac{\alpha_0}{\alpha_1}$
- *Bayes factor* (in favor of H_0):

$$B_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}$$

- B_{01} is often thought of as "the odds of H_0 to H_1 provided by the data" (but can depend on prior)

$$\underset{\text{(posterior odds)}}{\frac{P(H_0|\mathbf{X})}{P(H_1|\mathbf{X})}} = \underset{\text{(prior odds)}}{\frac{P(H_0)}{P(H_1)}} \times \underset{\text{(Bayes factor)}}{B_{01}}$$

Simple vs Simple

Assume $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$

- Posterior probabilities:

$$\alpha_0 = \frac{\pi_0 f(\mathbf{X} \mid \theta_0)}{\pi_0 f(\mathbf{X} \mid \theta_0) + \pi_1 f(\mathbf{X} \mid \theta_1)} = 1 - \alpha_1$$

- Posterior odds:

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 f(\mathbf{X} \mid \theta_0)}{\pi_1 f(\mathbf{X} \mid \theta_1)}$$

- Bayes factor:

$$B_{01} = \frac{f(\mathbf{X} \mid \theta_0)}{f(\mathbf{X} \mid \theta_1)} = \text{likelihood ratio}$$

Example: Normal

$X_i \mid \theta \sim \mathsf{N}(\theta, 1), i = 1, \ldots, n$. To test:

$$H_0 : \theta = 0$$
$$H_1 : \theta = 1$$

$\bar{X}$ is sample mean, so $\bar{X} \mid H_0 \sim \mathsf{N}(0, 1/n)$ and $\bar{X} \mid H_1 \sim \mathsf{N}(1, 1/n)$

Posterior odds: with $\pi_0 = \pi_1$,

$$\frac{\alpha_0}{\alpha_1} = \exp\left\{-\frac{n}{2}(2\bar{X} - 1)\right\}$$

Since prior odds $= 1$, posterior odds $=$ Bayes factor.
If $n = 10, \bar{X} = 2$, posterior odds $= 3.2 \times 10^{-7}$

General Formulation

- Basic ingredients are
  - Prior probability π_i that hypothesis i is the true one
  - Assuming that H_i is true, a density $g_i(\theta)$ describing how θ is distributed in Θ_i: $g_0(\theta)$ and $g_1(\theta)$.
- Note that $\int_{\Theta_0} g_0(\theta)d\theta = 1$ and $\int_{\Theta_1} g_1(\theta)d\theta = 1$.
- Overall prior is

$$\pi(\theta) = \begin{cases} \pi_0 g_0(\theta) & \text{if } \theta \in \Theta_0 \\ \pi_1 g_1(\theta) & \text{if } \theta \in \Theta_1 \end{cases}$$

  or equivalently, in the mixture format:

$$\pi(\theta) = \pi_0 g_0(\theta)\mathbf{1}_{\Theta_0}(\theta) + \pi_1 g_1(\theta)\mathbf{1}_{\Theta_1}(\theta)$$

- Posterior odds

$$\frac{\alpha_0}{\alpha_1} = \frac{\int_{\Theta_0} p_0(\theta \mid \mathbf{X}) d\theta}{\int_{\Theta_1} p_1(\theta \mid \mathbf{X}) d\theta} = \frac{\pi_0 \int_{\Theta_0} f(\mathbf{X} \mid \theta) g_0(\theta) d\theta}{\pi_1 \int_{\Theta_1} f(\mathbf{X} \mid \theta) g_1(\theta) d\theta}$$

- Bayes factor

$$B_{01} = \frac{\int_{\Theta_0} f(\mathbf{X} \mid \theta) g_0(\theta) d\theta}{\int_{\Theta_1} f(\mathbf{X} \mid \theta) g_1(\theta) d\theta} = \frac{m_0(\mathbf{X})}{m_1(\mathbf{X})}$$

  the ratio of "weighted" likelihoods (contrast with likelihood ratio).

- *Marginal likelihood* $m_i(\mathbf{X})$ is predictive under H_i evaluated at observed $\mathbf{X}$.
- B_{01} depends on the prior g_0, g_1, but often sensibly robust

Decision as to whether accept H_0 or accept H_1 (reject H_0)

- Based on the posterior odds. By default, H_0 accepted if $\alpha_0 > \alpha_1$ but often decisions are not reported
- Alternatively, report Bayes factor B_{01}, either because
  - is to be combined with personal prior odds
  - the 'default' $\pi_0 = \pi_1$ is used

As a *decision problem*, decide between $\begin{cases} a_0 & \text{accept } H_0 \\ a_1 & \text{accept } H_1 \end{cases}$

With a 0-1 loss function $L(\theta, a_i) = \begin{cases} 0 & \text{if } \theta \in \Theta_i \\ 1 & \text{if } \theta \in \Theta_j, j \leq i \end{cases}$

Optimal decision minimizes expected posterior loss

$$E_{\pi(\theta|X)} L(\theta, a_1) = \int L(\theta, a_1) \pi(\theta \mid \mathbf{X}) d\theta = P(\Theta_0 \mid \mathbf{X})$$

Therefore, prefer a_0 to a_1 iff $P(\Theta_0 \mid \mathbf{X}) > P(\Theta_1 \mid \mathbf{X})$.

An Alternative Way of Specifying the Prior $\pi(\theta)$

The encompassing prior approach: sometimes instead of separately assessing π_0, π_1, g_0, g_1 and then deriving the overall $\pi(\theta)$, it is possible to start with an overall, conventional $\pi(\theta)$ and deduce π_0, π_1, g_0, g_1 from π:

$$\pi_0 = \int_{\Theta_0} \pi(\theta) d\theta \quad \text{and} \quad \pi_1 = \int_{\Theta_1} \pi(\theta) d\theta$$

$$g_0(\theta) = \frac{1}{\pi_0} \pi(\theta) \mathbf{1}_{\Theta_0}(\theta) \quad \text{and} \quad g_1(\theta) = \frac{1}{\pi_1} \pi(\theta) \mathbf{1}_{\Theta_1}(\theta)$$

This is a conveniently easy approach, but to be sensible:

- π_0 and π_1 must make sense
- g_0 and g_1 must make sense as distributions under H_i
- With this formulation, two statisticians can not obviously agree on the g_i's and disagree on the π_i's nor vice versa (has to be done through the overall π).

Example: Intelligence Testing

- $X \mid \theta \sim \mathsf{N}(\theta, 100)$, overall $\theta \sim \mathsf{N}(100, 225)$, $n = 100, \bar{X} = 115$
- To test "below average" versus "above average"

$$H_0 : \theta \leq 100 \quad \text{vs} \quad H_1 : \theta > 100$$

- Recall, if $\theta \sim \mathsf{N}(m_0, v_0^2)$, and σ^2 known, posterior is $\mathsf{N}(m_1, v_1^2)$ with

$$m_1 = \frac{\sigma^2/n}{v_0^2 + \sigma^2/n} m_0 + \frac{v_0^2}{v_0^2 + \sigma^2/n} \bar{X}, \quad v_1^2 = \left[1/v_0^2 + n/\sigma^2\right]^{-1}$$

- Posterior $\theta \mid \mathbf{X} \sim \mathsf{N}(110.39, 69.23)$
- Posterior probabilities:

$$\alpha_0 = P(\theta \leq 100 \mid \mathbf{X}) = 0.106, \quad \alpha_1 = P(\theta > 100 \mid \mathbf{X}) = 0.894$$

- Posterior odds: $\frac{\alpha_0}{\alpha_1} = \frac{1}{8.44}$

- *Induced* prior probabilities of hypotheses

$$\pi_0 = P(\theta \leq 100) = 1/2 = \pi_1$$

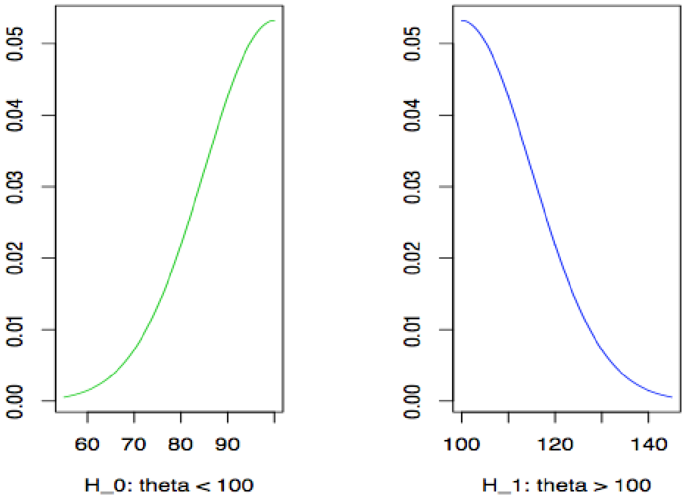
  A usual *default choice*, giving prior odds $= 1$

- *Induced* densities under each hypothesis:

$$g_0(\theta) = 2\mathrm{N}(\theta; 100, 225)\mathbf{1}_{-\infty,100}(\theta)$$
$$g_1(\theta) = 2\mathrm{N}(\theta; 100, 225)\mathbf{1}_{100,\infty}(\theta)$$

  maybe not too bad

- The beauty of the "conventional overall prior" approach is that these derivations are not formally needed (except for checking that the implied priors are sensible)


0.00
0.01
0.02
0.03
0.04
0.05
60
70
80
90
H_0: theta < 100
100
120
140
H_1: theta > 100

Important Caution: Improper g_i's

Let's in general denote $g_i(\theta) = c_i g_i^*(\theta)$. Bayes factor is:

$$B_{01} = \frac{c_0 \int_{\Theta_0} f(\mathbf{X} \mid \theta) g_0^*(\theta) d\theta}{c_1 \int_{\Theta_1} f(\mathbf{X} \mid \theta) g_1^*(\theta) d\theta}$$

For improper g_i's (and a "formal" definition of B_{01}), the c_i's are arbitrary, and hence B_{01} is also arbitrary (as well as the "formal" posterior odds ratio).

In some scenarios, they can be used

Of course, prior odds can not be defined (not worrisome)

One-Sided Testing

With this we refer to situations where:

- $\Theta \in \mathbb{R}$, and Θ_1 is to one side of Θ_0,
- similar with more than 2 hypotheses

Testing is easy and direct, does not pose any special problems. It has some "nice" peculiarities:

- The "alternative way" of specifying $\pi(\theta)$ (encompassing $\pi(\theta)$) is often used
- Non-informative, improper g_i are used sometimes; taking $c_0 = c_1$: cancel in the definition of the BF

Example: Normal, Objective Prior

- $X \mid \theta \sim \mathsf{N}(\theta, \sigma^2)$, overall $\pi(\theta) \propto$ constant and $\theta \mid X \sim \mathsf{N}(X, \sigma^2)$
- One-sided testing:

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

- Posterior probabilities:

$$\alpha_0 = P(\theta \leq \theta_0 \mid X) = \Phi\left(\frac{\theta_0 - X}{\sigma}\right) = 1 - \alpha_1$$

- Posterior odds: α_0/α_1. Equivalent to (conventionally) taking $g_0 = g_1 =$ same constant, and $\pi_0 = \pi_1 = 1/2$.

- Bayes factor can also be (formally) defined:

$$B_{01} = \frac{\int_{\Theta_0} f(\mathbf{X} \mid \theta) g_0(\theta) d\theta}{\int_{\Theta_1} f(\mathbf{X} \mid \theta) g_1(\theta) d\theta} = \frac{\text{const} \int_{\Theta_0} f(\mathbf{X} \mid \theta) d\theta}{\text{const} \int_{\Theta_1} f(\mathbf{X} \mid \theta) d\theta}$$

and (somehow arbitrarily) assuming *the same constant*, the Bayes factor is defined. In this case:

$$B_{01} = \frac{\alpha_0}{\alpha_1} = \text{posterior odds ratio}$$

Point Null Hypothesis

For $\Theta \in \mathbb{R}$, to test

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

(one-sided versions are dealt with similarly)

- Bayesian and frequentist answers are radically different
- Testing $H_0 : \theta = \theta_0$ is often an approximation to testing $H_0 : \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$
- The "alternative" approach of assessing a conventional, continuous overall $\pi(\theta)$ can not be used since then $\pi_0 = \alpha_0 = 0$
- Prior distribution:
  - Give to θ_0 probability π_0 (maybe the mass of the real null $(\theta_0 - \epsilon, \theta_0 + \epsilon)$ with an overall $\pi(\theta)$).
  - Give to $\theta \neq \theta_0$ density $\pi_1 g_1(\theta)$, where $\pi_1 = 1 - \pi_0$ and $\int g_1(\theta) d\theta = 1$.

- analysis proceeds as usual, taking into account that prior $\pi(\theta)$ has discrete and continuous parts
- Overall predictive distribution of $\mathbf{X}$

$$m(\mathbf{X}) = \pi_0 f(\mathbf{X} \mid \theta_0) + \pi_1 \underbrace{\int_{\theta \neq \theta_0} f(\mathbf{X} \mid \theta) g_1(\theta) d\theta}_{m_1(\mathbf{X})}$$

- Posterior probabilities:

$$\alpha_0 = 1 - \alpha_1 = p(\theta_0 \mid \mathbf{X}) = \frac{\pi_0 f(\mathbf{X} \mid \theta_0)}{m(\mathbf{X})}$$

- Posterior odds ratio:

$$\frac{\alpha_0}{\alpha_1} = \frac{p(\theta_0 \mid \mathbf{X})}{1 - p(\theta_0 \mid \mathbf{X})} = \frac{\pi_0}{\pi_1} \frac{f(\mathbf{X} \mid \theta_0)}{m_1(\mathbf{X})}$$

- Bayes factor for H_0 versus H_1 is

$$B_{01} = \frac{f(\mathbf{X} \mid \theta_0)}{m_1(\mathbf{X})} = \frac{\text{likelihood of observed data under } H_0}{\text{“average” likelihood of observed data under } H_1}$$

  Of course, $B_{10} = 1/B_{01}$.

  Reporting the Bayes factor: an “objective” alternative to choosing $P(H_0) = P(H_1) = 1/2$.

- *Important*: no “cancellation” can occur $\Longrightarrow$ $g_1(\theta)$ *proper.*
- Posterior odds ratio and posterior probabilities are naturally expressed in terms of the Bayes factor:

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0}{\pi_1} B_{01}, \quad \alpha_0 = \left[1 + \frac{\pi_1}{\pi_0}\frac{1}{B_{01}}\right]^{-1}$$

Normal Example

- $X_i \mid \theta \overset{iid}{\sim} \mathsf{N}(\theta, \sigma^2)$, σ^2 is known. To test:

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

- Likelihood:

$$f(\mathbf{X} \mid \theta) \propto \mathsf{N}(\bar{X} \mid \theta, \sigma^2/n)$$

($\bar{X}$ is sufficient statistic under both hypotheses)

- Prior on $H_1 : g_1(\theta) = \mathsf{N}(\theta; \theta_0, v_0^2)$. Taking prior mean $m_0 = \theta_0$ is usual, sensible choice
- Marginal likelihood under H_1: $m_1(\mathbf{X}) = \mathsf{N}(\bar{X}; \theta_0, v_0^2 + \sigma^2/2)$
- Posterior probability:

$$\alpha_0 = \left[1 + \frac{\pi_1}{\pi_0} \frac{\exp\left\{\frac{1}{2} z^2 [1 + \sigma^2/(nv_0^2)]^{-1}\right\}}{(1 + nv_0^2/\sigma^2)^{1/2}} \right]^{-1}$$

where $z = \frac{|\bar{X} - \theta_0|}{\sigma/\sqrt{n}}$ is the frequentist test statistic for this problem.

Bayes factor is

$$B_{01} = \frac{(1 + nv_0^2/\sigma^2)^{1/2}}{\exp\left\{\frac{1}{2}z^2[1 + \sigma^2/(nv_0^2)]^{-1}\right\}}$$

Common default options (not optimal, but usually sensible)

- $\pi_0 = \pi_1 = 1/2$
- $g_1(\theta)$ has to be *proper*. "Convenient" objective choice is $g_1(\theta) = \mathsf{N}(\theta \mid \theta_0, \sigma^2)$

and

$$B_{01} = \sqrt{1+n}\exp\left\{-\frac{n}{2(1+n)}z^2\right\} = \frac{\alpha_0}{\alpha_1}, \quad \alpha_0 = \left(1 + \frac{1}{B_{01}}\right)^{-1}$$

Comparing α_0 with Classical p-value for Various n

z	p-value	$n=5$	$n=20$	$n=100$	α_0
1.645	0.1	0.44	0.56	0.72	0.4121
1.960	0.05	0.33	0.42	0.60	0.3221
2.576	0.01	0.13	0.16	0.27	0.1334

The conflict Bayesian-frequentist reports is evident. The last column is the smallest α_0 can be among all normal priors with mean θ_0.

Is the conflict due to prior choice?

- The normal choice for g_1 is usually not crucial
- The choices $m_0 = \theta_0$ and $\pi_0 = \pi_1 = 1/2$ are standard.
- *The choice for v_0^2 is important*. $v_0^2 = \sigma^2$ is based on Jeffreys proposal

Multiple Hypothesis Testing

The previous analysis generalizes in obvious ways to multiple testing problems: compute posterior probability of each hypothesis.

Example: Intelligence testing (cont.)

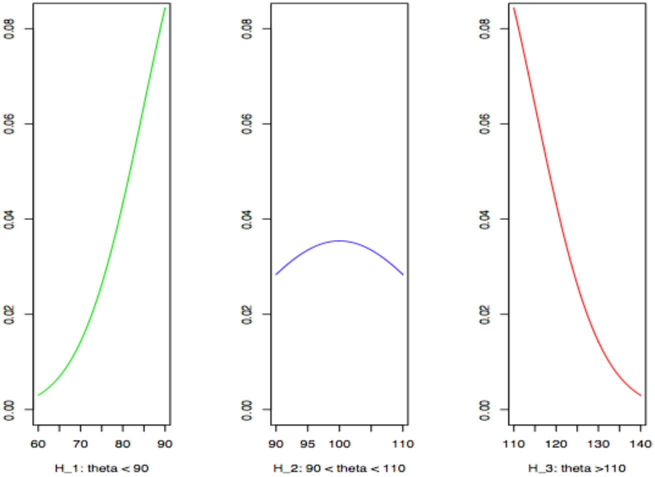
- We had $\theta \mid \bar{X} = 115 \sim \mathsf{N}(110.39, 69.23)$
- To test "below average" versus "average" versus "above average"

$$
\begin{aligned}
H_1 :& \theta < 90 \\
H_2 :& 90 \leq \theta \leq 110 \\
H_3 :& \theta > 110
\end{aligned}
$$

- Posterior probabilities:

$$
\begin{aligned}
\alpha_1 &= P(\theta < 90 \mid \mathbf{X}) \\
\alpha_2 &= P(90 \leq \theta \leq 110 \mid \mathbf{X}) \\
\alpha_3 &= P(\theta > 110 \mid \mathbf{X})
\end{aligned}
$$

Note: this is done with the "encompassing" prior, that is, taking an overall N$(100, 225)$ as prior for θ, and hence $\pi_1 = 0.2525, \pi_2 = 0.495$, and $\pi_3 = 0.2525$ which seems sensible. The g_i's seem fine too.



Automatic Occam's Razor

- Attributed to thirteen-century Franciscan monk William of Ockham (Occam in latin)

  "It is vain to do with more what can be done with fewer"

- Preferring the simpler of two hypothesis to the more complex when both agree with data is an old principle in science
- Regard H_0 as *simpler* than H_1 if it makes *sharper predictions* about what data will be observed
- Complex hypothesis have extra adjustable parameters that allow them to accommodate a larger set of potential observations than can simple ones
  - "coin is fair" vs. "coin has unknown bias θ"
  - "relationship is $s = a + ut + gt^2$" vs
    "relationship is $s = a + ut + gt^2 + ct^3$"

For $\mathbf{X} = (X_1, \ldots, X_n)$, suppose we have two hypothesis H_i: for $i = 0, 1$, $\boldsymbol{\theta}_i = (\theta_{i,1}, \ldots, \theta_{i,p_i})$, and log-likelihood $l(\boldsymbol{\theta}_i) = \log L(\boldsymbol{\theta}_i) = \log f_i(\mathbf{X} \mid \boldsymbol{\theta}_i)$. Under Laplace approximation, marginal likelihood

$$
\begin{aligned}
m(\mathbf{X} \mid H_i) &= \int \pi(\boldsymbol{\theta}_i \mid H_i) L(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \\
&\approx \int \pi(\hat{\boldsymbol{\theta}}_i \mid H_i) \exp\left\{ l(\hat{\boldsymbol{\theta}}_i) - \frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)' \ddot{l}(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \right\} d\boldsymbol{\theta}_i \\
&\approx L(\hat{\boldsymbol{\theta}}_i) \times \pi(\hat{\boldsymbol{\theta}}_i \mid H_i) \int \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)' n I_i(\hat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \right\} d\boldsymbol{\theta}_i \\
&= L(\hat{\boldsymbol{\theta}}_i) \times \pi(\hat{\boldsymbol{\theta}}_i \mid H_i) \, (2\pi/n)^{\frac{p_i}{2}} \, |I_i(\hat{\boldsymbol{\theta}}_i)|^{-\frac{1}{2}} \\
&= \text{maximum likelihood} \times \text{Occam factor}
\end{aligned}
$$

$$
B_{01} \approx \frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}}_1)} \times \frac{\pi(\hat{\boldsymbol{\theta}}_0 \mid H_0)}{\pi(\hat{\boldsymbol{\theta}}_1 \mid H_1)} \cdot \left(\frac{n}{2\pi}\right)^{\frac{p_1 - p_0}{2}} \cdot \frac{|I_i(\hat{\boldsymbol{\theta}}_0)|^{-\frac{1}{2}}}{|I_i(\hat{\boldsymbol{\theta}}_1)|^{-\frac{1}{2}}}
$$

Normal Illustration

- $X_i \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \sigma^2)$ with σ^2 known.
- To test:

$$H_1 : \mu = \mu_0, \quad \text{versus} \quad H_2 : \mu \neq \mu_0$$

- Prior on $H_2 : \mu \sim \mathsf{Unif}(m_0, m_1)$ with m_0 and m_1 chosen based on genuine prior information
- Marginal likelihood under H_1:

$$m_1(\mathbf{X}) = \prod_{i=1}^{n} \mathsf{N}(X_i \mid \mu_0, \sigma^2)$$

- Marginal likelihood under H_2:

$$m_2(\mathbf{X}) = \frac{(\sigma\sqrt{2\pi})^{1-n}}{\sqrt{n}} \frac{\Phi(\frac{m_1 - \bar{X}}{\sigma/\sqrt{n}}) - \Phi(\frac{m_0 - \bar{X}}{\sigma/\sqrt{n}})}{m_1 - m_0} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\}$$

- Bayes factor:

$$B_{12} \approx \frac{m_1 - m_0}{\sigma/\sqrt{n}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2\sigma^2}(\mu_0 - \bar{X})^2\right\}$$

  assuming m_0, m_1 "far" from $\bar{X}$ (in terms of $\sigma/\sqrt{n}$).
- If we observe $\bar{X} = \mu_0$, B_{12} increase: favors the simple H_1.
- But B_{12} favors H_1 for some $\bar{X} \neq \mu_0$ even though H_2 with best-fit $\mu = \bar{X}$ fits data (slightly) better: H_2 is being penalized for being more complex.
- The likelihood ratio (ratio of best-fit likelihoods) is

$$R_{12} = \exp\left\{-\frac{n}{2\sigma^2}(\mu_0 - \bar{X})^2\right\}$$

  So the Bayes factor is $B_{12} = R_{12} \times S_{12}$.
  R_{12} always favors the complex model. S_{12} is the "simplicity factor" or "Occam factor".
  Natural quantification of Occams Razor: prefer the simpler model unless the more complicated model gives a much better fit

The Jeffreys-Lindley and Barlett "Paradoxes"

In the normal testing scenario of testing $H_0 : \theta = \theta_0$ with a normal $\mathsf{N}(\theta_0, v_0^2)$ prior on the alternative hypothesis H_1,

$$B_{01} = \frac{(1 + nv_0^2/\sigma^2)^{1/2}}{\exp\left\{\frac{1}{2}z^2[1 + \sigma^2/(nv_0^2)]^{-1}\right\}}$$

- *Jeffreys-Lindley paradox:* for large n,

  $$B_{01} \approx \sqrt{n}\frac{v_0}{\sigma}\exp\{-\frac{1}{2}z^2\}$$

  so that a classical test can strongly reject the null (large z) and a Bayesian analysis strongly support it.
- *Bartlett paradox* (sometimes also called Lindley paradox): as $v_0^2 \to \infty$, then $B_{01} \to \infty$ so that proper priors in testing can not be "arbitrarily flat".