The Bayes Factor

- The **Bayes Factor** provides a way to formally compare two **competing models**, say M_1 and M_2.
- It is similar to testing a "full model" vs. "reduced model" (with, e.g., a likelihood ratio test) in classical statistics.
- However, with the **Bayes Factor**, one model **does not have to be nested** within the other.
- Given a data set $\mathbf{x}$, we compare models

$$M_1 : f_1(\mathbf{x}|\boldsymbol{\theta}_1) \text{ and } M_2 : f_2(\mathbf{x}|\boldsymbol{\theta}_2)$$

- We may specify prior distributions $p_1(\boldsymbol{\theta}_1)$ and $p_2(\boldsymbol{\theta}_2)$ that lead to prior probabilities for each model $p(M_1)$ and $p(M_2)$.

The Bayes Factor

By Bayes' Law, the **posterior odds** in favor of Model 1 versus Model 2 is:

$$
\begin{aligned}
\frac{\pi(M_1|\mathbf{x})}{\pi(M_2|\mathbf{x})} &= \frac{\int_{\Theta_1} \frac{p(M_1) f_1(\mathbf{x}|\boldsymbol{\theta}_1) p_1(\theta_1)\,\mathrm{d}\boldsymbol{\theta}_1}{p(\mathbf{x})}}{\int_{\Theta_2} \frac{p(M_2) f_2(\mathbf{x}|\boldsymbol{\theta}_2) p_2(\theta_2)\,\mathrm{d}\boldsymbol{\theta}_2}{p(\mathbf{x})}} \\
&= \frac{p(M_1)}{p(M_2)} \cdot \frac{\int_{\Theta_1} f_1(\mathbf{x}|\boldsymbol{\theta}_1) p_1(\theta_1)\,\mathrm{d}\boldsymbol{\theta}_1}{\int_{\Theta_2} f_2(\mathbf{x}|\boldsymbol{\theta}_2) p_2(\theta_2)\,\mathrm{d}\boldsymbol{\theta}_2} \\
&= [\text{prior odds}] \times [\text{Bayes Factor } B(\mathbf{x})]
\end{aligned}
$$

The Bayes Factor

Rearranging, the Bayes Factor is:

$$\begin{aligned} B(\mathbf{x}) &= \frac{\pi(M_1|\mathbf{x})}{\pi(M_2|\mathbf{x})} \times \frac{p(M_2)}{p(M_1)} \\ &= \frac{\pi(M_1|\mathbf{x})/\pi(M_2|\mathbf{x})}{p(M_1)/p(M_2)} \end{aligned}$$

(the ratio of the posterior odds for M_1 to the prior odds for M_1).

The Bayes Factor

- **Note**: If the prior model probabilities are equal, i.e., $p(M_1) = p(M_2)$, then the Bayes Factor equals the posterior odds for M_1.
- **Note**: If $p(M_1) = p(M_2)$ and the parameter spaces Θ_1 and Θ_2 are the same, then the Bayes Factor reduces to a **likelihood ratio**.

  Note that:

$$B(\mathbf{x}) = \frac{\pi(M_1|\mathbf{x})}{\pi(M_2|\mathbf{x})} \times \frac{p(M_2)}{p(M_1)} = \frac{\frac{\pi(M_1,\mathbf{X})}{p(\mathbf{X})p(M_1)}}{\frac{\pi(M_2,\mathbf{X})}{p(\mathbf{X})p(M_2)}}$$

$$= \frac{\frac{\pi(M_1,\mathbf{X})}{p(M_1)}}{\frac{\pi(M_2,\mathbf{X})}{p(M_2)}} = \frac{\pi(\mathbf{x}|M_1)}{\pi(\mathbf{x}|M_2)}$$

The Bayes Factor

- Clearly a Bayes Factor much greater than 1 supports Model 1 over Model 2.
- Jeffreys proposed the following rules, if Model 1 represents a null model:

Result	**Conclusion**
$B(\mathbf{x}) \geq 1 \rightarrow$	Model 1 supported
$0.316 \leq B(\mathbf{x}) < 1 \rightarrow$	Minimal evidence against Model 1
	(Note $0.316 = 10^{-1/2}$)
$0.1 \leq B(\mathbf{x}) < 0.316 \rightarrow$	Substantial evidence against Model 1
$0.01 \leq B(\mathbf{x}) < 0.1 \rightarrow$	Strong evidence against Model 1
$B(\mathbf{x}) < 0.01 \rightarrow$	Decisive evidence against Model 1

- Clearly these labels are fairly arbitrary.

The Bayes Factor

- In the case when there are only **two possible models**, M_1 and M_2, then given the Bayes Factor $B(\mathbf{x})$, we can calculate the posterior probability of Model 1 as:

$$
\begin{aligned}
P(M_1|\mathbf{x}) = 1 - P(M_2|\mathbf{x}) &= 1 - \frac{P(\mathbf{x}|M_2)P(M_2)}{P(\mathbf{x})} \\
&= 1 - \frac{P(\mathbf{x}|M_1)}{B(\mathbf{x})}\frac{P(M_2)}{P(\mathbf{x})} \\
\Rightarrow P(M_1|\mathbf{x}) &= 1 - \left\{\frac{1}{B(\mathbf{x})}\frac{P(M_2)}{P(M_1)}\right\}P(M_1|\mathbf{x}) \\
\Rightarrow 1 &= \left[1 + \left\{\frac{1}{B(\mathbf{x})}\frac{P(M_2)}{P(M_1)}\right\}\right]P(M_1|\mathbf{x}) \\
\Rightarrow P(M_1|\mathbf{x}) &= \frac{1}{1 + \left\{\frac{1}{B(\mathbf{x})}\frac{P(M_2)}{P(M_1)}\right\}}
\end{aligned}
$$

Example: Comparing Two Means

Example 2(a): Comparing Two Means (Bayes Factor Approach)

- **Data**: Blood pressure reduction was measured for 11 patients who took calcium supplements and for 10 patients who took a placebo.
- We model the data with normal distributions having common variance:

$$\text{Calcium data}: X_{1j} \overset{\text{iid}}{\sim} N(\mu_1, \sigma^2), \quad j = 1, \ldots, 11$$

$$\text{Placebo data}: X_{2j} \overset{\text{iid}}{\sim} N(\mu_2, \sigma^2), \quad j = 1, \ldots, 10$$

Consider the two-sided test for whether the mean BP reduction differs for the two groups:

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$

Example: Comparing Two Means

- We will place a prior on the difference of standardized means

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma}$$

with specified prior mean μ_Δ and variance σ_Δ^2.

- Consider the classical two-sample t-statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}/\sqrt{n^*}},$$

where $n^* = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}$.

Example: Comparing Two Means

- H_0 and H_a define two specific models for the distribution of T.
- Under H_0, $T \sim$ (central) t with $(n_1 + n_2 - 2)$ degrees of freedom.
- Under H_a, $T \sim$ noncentral t.
- With this prior, the Bayes Factor for H_0 over H_a is:

$$B(\mathbf{x}) = \frac{t_{n_1+n_2-2}(t^*, 0, 1)}{t_{n_1+n_2-2}(t^*, \mu_\Delta \sqrt{n^*}, 1 + n^* \sigma_\Delta^2)}$$

  where t^* is the observed t-statistic.
- See `R` example to get $B(\mathbf{x})$ and $P[H_0|\mathbf{x}]$.

Example: Comparing Two Means

Example 2(a): Comparing Two Means (Gibbs Sampling Approach)

- Same data set, but suppose our interest is in testing whether the calcium yields a **better** BP reduction than the placebo:

$$H_0 : \mu_1 \leq \mu_2 \text{ vs. } H_a : \mu_1 > \mu_2$$

- We set up the sampling model:

$$X_{1j} = \mu + \tau + \epsilon_{1j}, j = 1, \ldots, 11$$
$$X_{2j} = \mu - \tau + \epsilon_{2j}, j = 1, \ldots, 10$$

  where $\epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

- Thus $\mu_1 = \mu + \tau$ and $\mu_2 = \mu - \tau$.

Example: Comparing Two Means

We can assume independent priors for μ, τ, and σ^2:

$$\mu \sim N(\mu_\mu, \sigma_\mu^2)$$
$$\tau \sim N(\mu_\tau, \sigma_\tau^2)$$
$$\sigma^2 \sim IG(\nu_1/2, \nu_1\nu_2/2)$$

Then it can be shown that the full conditional distributions are:

$$\mu | \mathbf{x}_1, \mathbf{x}_2, \tau, \sigma^2 \sim \text{Normal}$$
$$\tau | \mathbf{x}_1, \mathbf{x}_2, \mu, \sigma^2 \sim \text{Normal}$$
$$\sigma^2 | \mathbf{x}_1, \mathbf{x}_2, \mu, \tau \sim IG$$

where the appropriate parameters are given in the R code.

Example: Comparing Two Means

- `R` **example**: Gibbs Sampler can obtain approximate posterior distributions for μ and (especially of interest) for τ.
- Note $P[\mu_1 > \mu_2|\mathbf{x}] = P[\tau > 0|\mathbf{x}]$.
- We can also find the **posterior predictive** probability $P[X_1 > X_2]$.