Lecture 6: Bayesian Hypothesis Testing Continued

Merlise Clyde

September 14, 2021

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \;\; \text{vs } \mathcal{H}_1 : \theta \neq 0$

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \ \text{vs}\ \mathcal{H}_1 : \theta \neq 0$

1. Put a prior on $\theta \mid \mathcal{H}_i$,

$$\pi(\theta \mid \mathcal{H}_1) = \mathcal{N}(0, 1/\tau_0)$$

$$p(\theta \mid \mathcal{H}_0) = \delta_o(\theta)$$

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \text{ vs } \mathcal{H}_1 : \theta \neq 0$

1. Put a prior on $\theta \mid \mathcal{H}_i$,

$$\pi(\theta \mid \mathcal{H}_1) = \mathcal{N}(0, 1/\tau_0)$$

$$p(\theta \mid \mathcal{H}_0) = \delta_o(\theta)$$

2. Put a prior on the hypotheses $\pi(\mathcal{H}_0)$ and $\pi(\mathcal{H}_1)$.

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \;\; \text{vs } \mathcal{H}_1 : \theta \neq 0$

1. Put a prior on $\theta \mid \mathcal{H}_i$,

$$\pi(\theta \mid \mathcal{H}_1) = \mathcal{N}(0, 1/\tau_0)$$

$$p(\theta \mid \mathcal{H}_0) = \delta_o(\theta)$$

2. Put a prior on the hypotheses $\pi(\mathcal{H}_0)$ and $\pi(\mathcal{H}_1)$.
3. Likelihood of the hypotheses $p(y^{(n)} \mid \mathcal{H}_i)$

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \ \ \text{vs}\ \mathcal{H}_1 : \theta \neq 0$

1. Put a prior on $\theta \mid \mathcal{H}_i$,

$$\pi(\theta \mid \mathcal{H}_1) = \mathcal{N}(0, 1/\tau_0)$$

$$p(\theta \mid \mathcal{H}_0) = \delta_o(\theta)$$

2. Put a prior on the hypotheses $\pi(\mathcal{H}_0)$ and $\pi(\mathcal{H}_1)$.
3. Likelihood of the hypotheses $p(y^{(n)} \mid \mathcal{H}_i)$
4. Obtain posterior probabilities of $\mathcal{H}_0$ and $\mathcal{H}_1$ via Bayes Theorem or Bayes Factors

Hypothesis Testing Setup Recap

- univariate data $y_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
- test $\mathcal{H}_0 : \theta = 0; \;\; \text{vs} \; \mathcal{H}_1 : \theta \neq 0$

1. Put a prior on $\theta \mid \mathcal{H}_i$,

$$\pi(\theta \mid \mathcal{H}_1) = \mathcal{N}(0, 1/\tau_0)$$

$$p(\theta \mid \mathcal{H}_0) = \delta_o(\theta)$$

2. Put a prior on the hypotheses $\pi(\mathcal{H}_0)$ and $\pi(\mathcal{H}_1)$.
3. Likelihood of the hypotheses $p(y^{(n)} \mid \mathcal{H}_i)$
4. Obtain posterior probabilities of $\mathcal{H}_0$ and $\mathcal{H}_1$ via Bayes Theorem or Bayes Factors
5. Report based on loss (optional)

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.

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.

- **Rule of thumb**: $\mathcal{BF}_{10} > 10$ is strong evidence for $\mathcal{H}_1$; $\mathcal{BF}_{10} > 100$ is decisive evidence for $\mathcal{H}_1$.

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.

- **Rule of thumb**: $\mathcal{BF}_{10} > 10$ is strong evidence for $\mathcal{H}_1$; $\mathcal{BF}_{10} > 100$ is decisive evidence for $\mathcal{H}_1$.
  Not worth a bare mention $1 < \mathcal{BF}_{10} < 3$

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.

- **Rule of thumb**: $\mathcal{BF}_{10} > 10$ is strong evidence for $\mathcal{H}_1$; $\mathcal{BF}_{10} > 100$ is decisive evidence for $\mathcal{H}_1$.
  Not worth a bare mention $1 < \mathcal{BF}_{10} < 3$

- Posterior probabilities

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

- $\mathcal{O}_{01}$ prior odds of $\mathcal{H}_0$ to $\mathcal{H}_1$

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.

- **Rule of thumb**: $\mathcal{BF}_{10} > 10$ is strong evidence for $\mathcal{H}_1$; $\mathcal{BF}_{10} > 100$ is decisive evidence for $\mathcal{H}_1$.
  Not worth a bare mention $1 < \mathcal{BF}_{10} < 3$

- Posterior probabilities

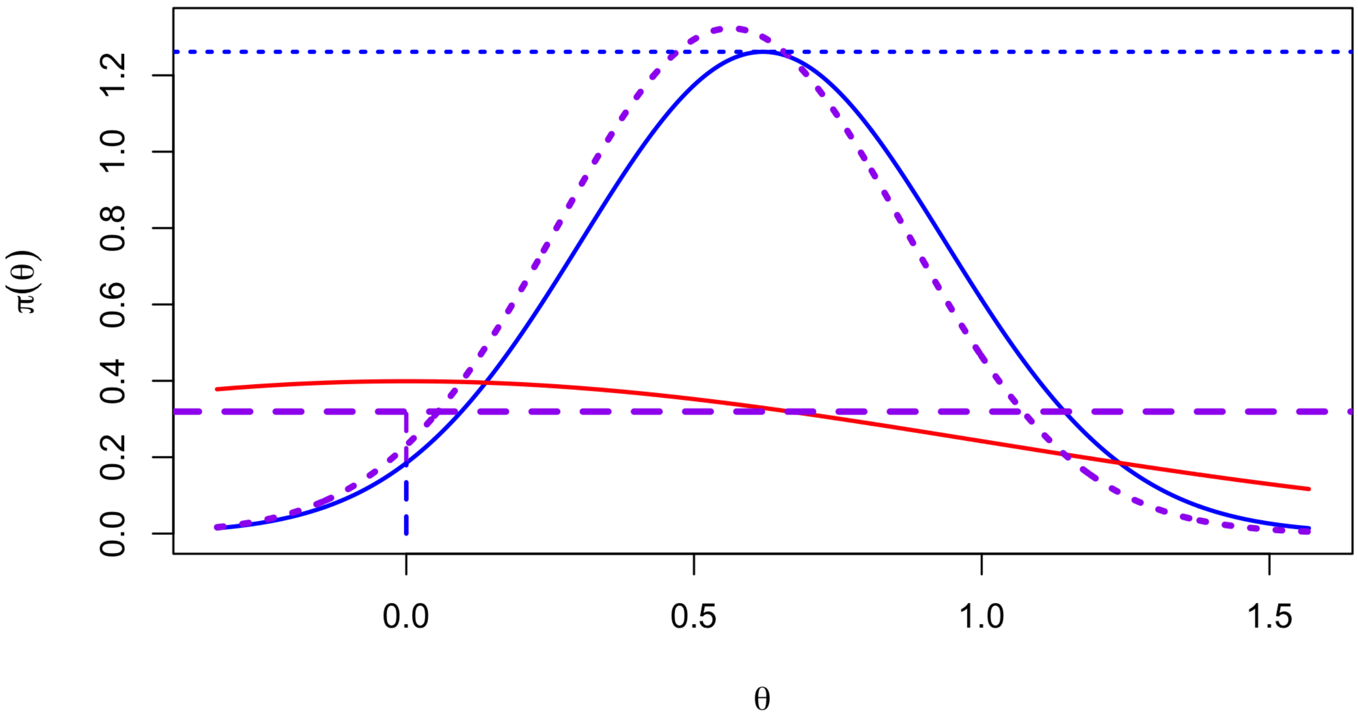
$$\pi(\mathcal{H}_1 \mid Y) = \frac{1}{\frac{\pi(\mathcal{H}_0)}{\pi(\mathcal{H}_1)} \frac{p(y^{(n)}|\mathcal{H}_0)}{p(y^{(n)}|\mathcal{H}_1)} + 1} = \frac{1}{\mathcal{O}_{01}\mathcal{BF}_{01} + 1}$$

- $\mathcal{O}_{01}$ prior odds of $\mathcal{H}_0$ to $\mathcal{H}_1$

Alternative expression for Bayes Factor

$$\mathcal{BF}_{10} = \frac{p(y^{(n)} \mid \mathcal{H}_1)}{p(y^{(n)} \mid \mathcal{H}_0)} = \frac{\pi_\theta(0 \mid \mathcal{H}_1)}{\pi_\theta(0 \mid y^{(n)}, \mathcal{H}_1)}$$

Marginal Likelihoods & Evidence



$\mathcal{BF}_{10}$ = 1.73 Posterior Probability of $\mathcal{H}_0$ = 0.3665 versus p-value of 0.05

Decisions

- Selection 0-1 loss;
  - if $\pi(\mathcal{H}_1 \mid y^{(n)}) > .5$ choose $\mathcal{H}_1$,
  - otherwise $\mathcal{H}_0$

Decisions

- Selection 0-1 loss;
    - if $\pi(\mathcal{H}_1 \mid y^{(n)}) > .5$ choose $\mathcal{H}_1$,
    - otherwise $\mathcal{H}_0$
- Estimation of θ under squared error loss
- report $\hat{\theta}$ that minimizes Bayes expected loss

$$\mathsf{E}_{\theta \mid y^{(n)}} \left[(\theta - \hat{\theta})^2 \right]$$

Decisions

- Selection 0-1 loss;
  - if $\pi(\mathcal{H}_1 \mid y^{(n)}) > .5$ choose $\mathcal{H}_1$,
  - otherwise $\mathcal{H}_0$
- Estimation of θ under squared error loss
- report $\hat{\theta}$ that minimizes Bayes expected loss

$$\mathsf{E}_{\theta \mid y^{(n)}} \left[(\theta - \hat{\theta})^2 \right]$$

- Bayes optimal estimator under squared error is the posterior mean $\mathsf{E}[\theta \mid y^{(n)}]$

Decisions

- Selection 0-1 loss;
  - if $\pi(\mathcal{H}_1 \mid y^{(n)}) > .5$ choose $\mathcal{H}_1$,
  - otherwise $\mathcal{H}_0$
- Estimation of θ under squared error loss
- report $\hat{\theta}$ that minimizes Bayes expected loss

$$\mathsf{E}_{\theta \mid y^{(n)}}\left[(\theta - \hat{\theta})^2\right]$$

- Bayes optimal estimator under squared error is the posterior mean $\mathsf{E}[\theta \mid y^{(n)}]$
- no $\mathcal{H}_i$!

Decisions

- Selection 0-1 loss;
  - if $\pi(\mathcal{H}_1 \mid y^{(n)}) > .5$ choose $\mathcal{H}_1$,
  - otherwise $\mathcal{H}_0$
- Estimation of θ under squared error loss
- report $\hat{\theta}$ that minimizes Bayes expected loss

$$\mathsf{E}_{\theta \mid y^{(n)}} \left[(\theta - \hat{\theta})^2 \right]$$

- Bayes optimal estimator under squared error is the posterior mean $\mathsf{E}[\theta \mid y^{(n)}]$
- no $\mathcal{H}_i$!
- marginal posterior distribution of θ

Averaging over Hypotheses

Prior on θ is a mixture model:

$$p(\theta) = \pi_0 \delta_0(\theta) + (1 - \pi_0)\pi(\theta \mid \mathcal{H}_1)$$

Averaging over Hypotheses

Prior on θ is a mixture model:

$$p(\theta) = \pi_0 \delta_0(\theta) + (1 - \pi_0)\pi(\theta \mid \mathcal{H}_1)$$

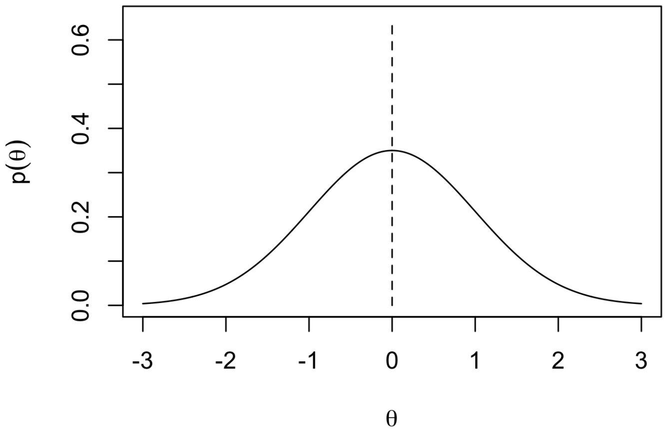
- Dirac delta - degenerate distribution at 0

Averaging over Hypotheses

Prior on θ is a mixture model:

$$p(\theta) = \pi_0 \delta_0(\theta) + (1 - \pi_0)\pi(\theta \mid \mathcal{H}_1)$$

- Dirac delta - degenerate distribution at 0
- "spike & slab" prior

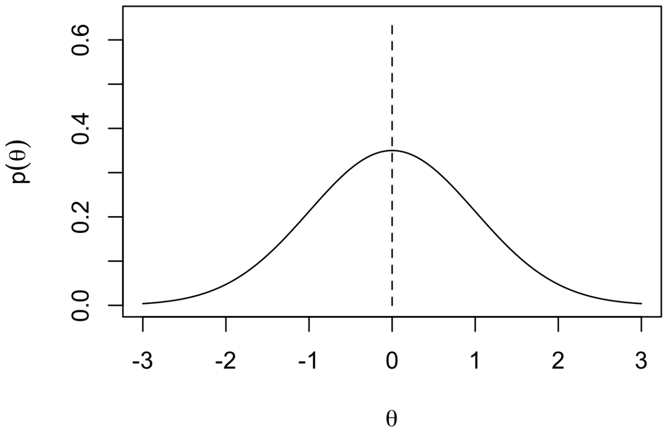


Averaging over Hypotheses

Prior on θ is a mixture model:

$$p(\theta) = \pi_0 \delta_0(\theta) + (1 - \pi_0)\pi(\theta \mid \mathcal{H}_1)$$

- Dirac delta - degenerate distribution at 0
- "spike & slab" prior



- how to sample from prior?

Posterior under Spike & Slab Prior

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_0, y^{(n)}) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

Posterior under Spike & Slab Prior

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_0, y^{(n)}) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\delta_0(\theta) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

- posterior also has a spike & slab

Posterior under Spike & Slab Prior

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_0, y^{(n)}) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\delta_0(\theta) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

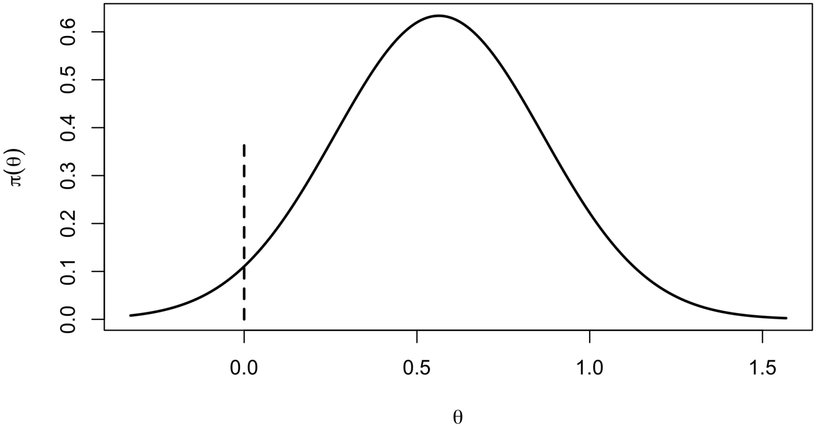
- posterior also has a spike & slab
- mixture weights are updated

Posterior under Spike & Slab Prior

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_0, y^{(n)}) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

$$\pi(\theta \mid y^{(n)}) = \Pr(\mathcal{H}_0 \mid y^{(n)})\delta_0(\theta) + \Pr(\mathcal{H}_1 \mid y^{(n)})\pi(\theta \mid \mathcal{H}_1, y^{(n)})$$

- posterior also has a spike & slab
- mixture weights are updated
- updated "slab" hyperparameters



Prior

An important issue with hypothesis testing and using Spike & Slab prior is choice of hyperparameters (τ_0 in this case)

Prior

An important issue with hypothesis testing and using Spike & Slab prior is choice of hyperparameters (τ_0 in this case)

- Bayes Factor and posterior probabilities of $\mathcal{H}_i$ depend on τ_0 through $p(y^{(n)} \mid \mathcal{H}_1)$

Prior

An important issue with hypothesis testing and using Spike & Slab prior is choice of hyperparameters (τ_0 in this case)

- Bayes Factor and posterior probabilities of $\mathcal{H}_i$ depend on τ_0 through $p(y^{(n)} \mid \mathcal{H}_1)$

1. What is impact of τ_0 on $\mathcal{BF}_{01}$?

Prior

An important issue with hypothesis testing and using Spike & Slab prior is choice of hyperparameters (τ_0 in this case)

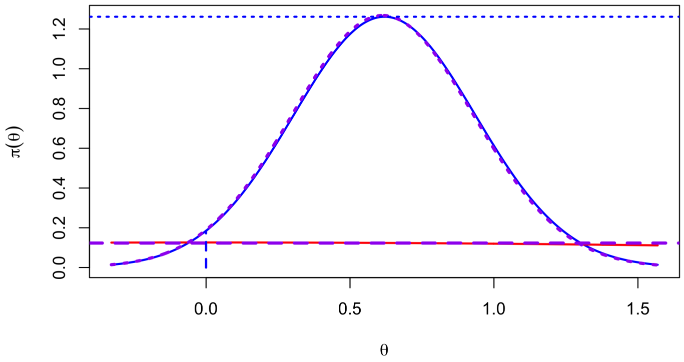
- Bayes Factor and posterior probabilities of $\mathcal{H}_i$ depend on τ_0 through $p(y^{(n)} \mid \mathcal{H}_1)$

1. What is impact of τ_0 on $\mathcal{BF}_{01}$?
2. How do we choose τ_0?

Question 1.

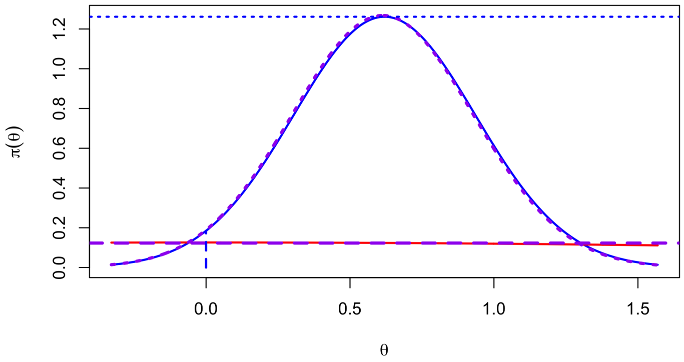
$$\mathcal{BF}_{01} = \frac{\pi(0 \mid \mathcal{H}_1, y^{(n)})}{\pi(0 \mid \mathcal{H}_1)}$$

Precision



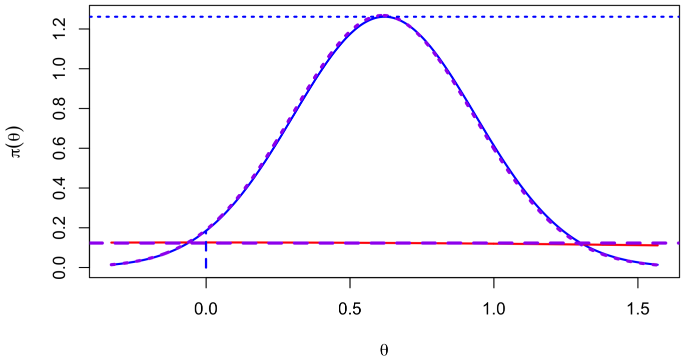
- $\tau_0 = 1/10$

Precision



- $\tau_0 = 1/10$
- Bayes Factor for $\mathcal{H}_0$ to $\mathcal{H}_1$ is 1.5

Precision



- $\tau_0 = 1/10$
- Bayes Factor for $\mathcal{H}_0$ to $\mathcal{H}_1$ is 1.5
- Posterior Probability of $\mathcal{H}_0$ = 0.6001

What about even more vague priors?

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!
- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Known as **Bartlett's Paradox** - the paradox is that a seemingly non-informative prior for θ is very informative about $\mathcal{H}$!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Known as **Bartlett's Paradox** - the paradox is that a seemingly non-informative prior for θ is very informative about $\mathcal{H}$!

- General problem with nested sequence of models. If we choose vague priors on the additional parameter in the larger model we will be favoring the smaller models under consideration!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Known as **Bartlett's Paradox** - the paradox is that a seemingly non-informative prior for θ is very informative about $\mathcal{H}$!

- General problem with nested sequence of models. If we choose vague priors on the additional parameter in the larger model we will be favoring the smaller models under consideration!

Bottom Line Don't use vague priors!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Known as **Bartlett's Paradox** - the paradox is that a seemingly non-informative prior for θ is very informative about $\mathcal{H}$!

- General problem with nested sequence of models. If we choose vague priors on the additional parameter in the larger model we will be favoring the smaller models under consideration!

Bottom Line Don't use vague priors!

Vague Priors & Hypothesis Testing

- As $\tau_0 \to 0$ the $\mathcal{BF}_{01} \to \infty$ and $\Pr(\mathcal{H}_0 \mid y^{(n)}) \to 1$!

- As we use a less & less informative prior under $\mathcal{H}_1$ we obtain more & more evidence for $\mathcal{H}_0$ over $\mathcal{H}_1$!

Known as **Bartlett's Paradox** - the paradox is that a seemingly non-informative prior for θ is very informative about $\mathcal{H}$!

- General problem with nested sequence of models. If we choose vague priors on the additional parameter in the larger model we will be favoring the smaller models under consideration!

Bottom Line Don't use vague priors!

What then?

Objective Bayes

- Conventional Priors

Objective Bayes

- Conventional Priors
- Simplest is Unit Information Prior (UIP)

Objective Bayes

- Conventional Priors
- Simplest is Unit Information Prior (UIP)
- center prior at MLE ($\bar{y}$) but choose prior precision to be the equivalent of a sample size of 1

Objective Bayes

- Conventional Priors
- Simplest is Unit Information Prior (UIP)
- center prior at MLE ($\bar{y}$) but choose prior precision to be the equivalent of a sample size of 1
- center prior at 0, but choose prior precision to be the equivalent of a sample size of 1 (UIP)

Objective Bayes

- Conventional Priors
- Simplest is Unit Information Prior (UIP)
- center prior at MLE ($\bar{y}$) but choose prior precision to be the equivalent of a sample size of 1
- center prior at 0, but choose prior precision to be the equivalent of a sample size of 1 (UIP)

Default UIP

$$\theta \mid \mathcal{H}_1 \sim \mathsf{N}(0, 1)$$

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

- BIC is derived in more general settings by taking a Laplace approximation to the marginal likelihood and making some simplifying assumptions

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

- BIC is derived in more general settings by taking a Laplace approximation to the marginal likelihood and making some simplifying assumptions
- BIC chooses model with highest marginal likelihood

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

- BIC is derived in more general settings by taking a Laplace approximation to the marginal likelihood and making some simplifying assumptions
- BIC chooses model with highest marginal likelihood
- BIC has a well known tendency to choose/favor simpler models due to UIP containing very little information

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

- BIC is derived in more general settings by taking a Laplace approximation to the marginal likelihood and making some simplifying assumptions
- BIC chooses model with highest marginal likelihood
- BIC has a well known tendency to choose/favor simpler models due to UIP containing very little information
- consistent for model selection i.e. $\Pr(\mathcal{H}_i \mid y^{(n)})$ goes to 1 for the true model as $n \to \infty$

UIP & BIC

Note: UIP is the basis for the Bayes Information Criterion (BIC)

- BIC is derived in more general settings by taking a Laplace approximation to the marginal likelihood and making some simplifying assumptions
- BIC chooses model with highest marginal likelihood
- BIC has a well known tendency to choose/favor simpler models due to UIP containing very little information
- consistent for model selection i.e. $\Pr(\mathcal{H}_i \mid y^{(n)})$ goes to 1 for the true model as $n \to \infty$
- Is a fixed τ_0 consistent as $n \to \infty$?

Other Options

- Place a prior on τ_0

Other Options

- Place a prior on τ_0

$$\tau_0 \sim \textsf{Gamma}(1/2, 1/2)$$

$$p(\tau_0 \mid \mathcal{H}_1) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \tau_0^{1/2 - 1} \exp\left(-\tau_0/2\right)$$

Other Options

- Place a prior on τ_0

$$\tau_0 \sim \textsf{Gamma}(1/2, 1/2)$$

$$p(\tau_0 \mid \mathcal{H}_1) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \tau_0^{1/2-1} \exp\left(-\tau_0/2\right)$$

- If $\theta \mid \tau_0, \mathcal{H}_1 \sim \textsf{N}(0, 1/\tau_0)$, then $\theta_0 \mid \mathcal{H}_1$ has a $\textsf{Cauchy}(0, 1)$ distribution! Recommended by Jeffreys (1961)

Other Options

- Place a prior on τ_0

$$\tau_0 \sim \text{Gamma}(1/2, 1/2)$$

$$p(\tau_0 \mid \mathcal{H}_1) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \tau_0^{1/2-1} \exp\left(-\tau_0/2\right)$$

- If $\theta \mid \tau_0, \mathcal{H}_1 \sim \mathsf{N}(0, 1/\tau_0)$, then $\theta_0 \mid \mathcal{H}_1$ has a $\mathsf{Cauchy}(0, 1)$ distribution! Recommended by Jeffreys (1961)

- no closed form expressions for marginal likelihood!

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$
- use part of the data $y(l)$ to update an improper prior on θ to get a proper posterior $\pi(\theta \mid \mathcal{H}_i, y(l))$

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$
- use part of the data $y(l)$ to update an improper prior on θ to get a proper posterior $\pi(\theta \mid \mathcal{H}_i, y(l))$
- use $\pi(\theta \mid y(l), \mathcal{H}_i)$ to obtain the posterior for θ based on the rest of the training data

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$
- use part of the data $y(l)$ to update an improper prior on θ to get a proper posterior $\pi(\theta \mid \mathcal{H}_i, y(l))$
- use $\pi(\theta \mid y(l), \mathcal{H}_i)$ to obtain the posterior for θ based on the rest of the training data
- Calculate a Bayes Factor (avoids arbitrary normalizing constants!)

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$
- use part of the data $y(l)$ to update an improper prior on θ to get a proper posterior $\pi(\theta \mid \mathcal{H}_i, y(l))$
- use $\pi(\theta \mid y(l), \mathcal{H}_i)$ to obtain the posterior for θ based on the rest of the training data
- Calculate a Bayes Factor (avoids arbitrary normalizing constants!)
- Choice of training sample $y(l)$?
- Berger & Pericchi (1996) propose "averaging" over training samples **intrinsic Bayes Factors**

Intrinsic Bayes Factors & Priors (Berger & Pericchi)

- Can't use improper priors under $\mathcal{H}_1$
- use part of the data $y(l)$ to update an improper prior on θ to get a proper posterior $\pi(\theta \mid \mathcal{H}_i, y(l))$
- use $\pi(\theta \mid y(l), \mathcal{H}_i)$ to obtain the posterior for θ based on the rest of the training data
- Calculate a Bayes Factor (avoids arbitrary normalizing constants!)
- Choice of training sample $y(l)$?
- Berger & Pericchi (1996) propose "averaging" over training samples **intrinsic Bayes Factors**
- **intrinsic prior** on θ that leads to the IBF