Foundations of Statistical Inference

J. Berestycki & D. Sejdinovic

Department of Statistics
University of Oxford

MT 2019

Chapter 12: Bayesian Hypothesis Tests

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Psychokinesis example

The experiment: Schmidt, Jahn and Radin (1987) used electronic and quantum-mechanical random event generators with visual feedback. Subject with alleged ability tries to "influence" the generator.

1. Stream of particles arrive at 'quantum gate'; each goes on to either red or green light.
2. Quantum mechanics implies a 50/50 ratio.
3. Subject tries to *influence* particles to go to red.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,490,000$. Observe $x = 52,263,471$.

Question : Has the subject influenced the particles ?

$H_0 : \theta = 1/2$ vs $H_1 : \theta \neq 1/2$

P-value $P_{\theta=1/2}(X \geq x) \approx .0003$. Strong evidence of paranormal ability?!

Bayesian tests

Model : $X|\theta \sim f_\theta(\cdot)$, $\theta \in \Theta \sim \Pi$
Testing problem

$$H_0 : \theta \in \Theta_0, \quad \text{vs} \quad H_1 : \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

$0 - 1$ loss function $\delta \in \{0, 1\}$

$$L(\theta, \delta) = \begin{cases} 1 & \text{if} \quad \mathbf{1}_{\theta \in \Theta_1} \neq \delta \\ 0 & \quad \text{otherwise} \end{cases}$$

Bayesian test

$$\delta(X) = \begin{cases} 1 & \text{if} \quad \Pi(\Theta_0|X) \leq \Pi(\Theta_1|X) \\ 0 & \text{if} \quad \Pi(\Theta_0|X) > \Pi(\Theta_1|X) \end{cases}$$

[Proof : exercise]

Simple/composite hypotheses

Definition

- A hypothesis $H_j : \theta \in \Theta_j$ is called simple iff Θ_j is a singleton.
- A hypothesis $H_j : \theta \in \Theta_j$ is called composite iff Θ_j is NOT a singleton.

Psychokinesis example:
$H_0 : \theta = 1/2$ is simple, $H_1 : \theta \neq 1/2$ is composite.

Construction of priors in the case of a simple hypothesis

We cannot use a continuous prior on Θ if Π has density (wrt Lebesgue) then

$$\Pi(\Theta_0) = \int_{\{\theta_0\}} \pi(\theta) d\theta = 0$$

We construct a prior as a mixture between a prior on Θ_0 and a prior on Θ_1.

$$\Pi(d\theta) = p_0 \Pi_0(d\theta) + (1 - p_0)\Pi_1(d\theta)$$

where Π_0 is a probability distribution on Θ_0 and Π_1 is a probability distribution on Θ_1.
$p_0 = \Pi(\Theta_0)$ Then if $\Theta_1 = \{\theta \in \Theta, \theta \neq \theta_0\}$

$$\Pi = p_0 \delta_{(\theta_0)} + (1 - p_0)\Pi_1, \quad \Pi_1(\Theta_1) = 1.$$

Test in the case of a simple hypothesis

$$\Pi = p_0 \delta_{(\theta_0)} + (1-p_0)\Pi_1, \quad \Pi_1(\Theta_1) = 1$$

Posterior

$$\Pi(\{\theta_0\}|X) = \frac{p_0 f_{\theta_0}(X)}{p_0 f_{\theta_0}(X) + (1-p_0)\int_{\Theta_1} f_\theta(X)\pi_1(\theta)d\theta}.$$

Bayes test

$$\begin{aligned} \delta(X) = 1 \quad &\Leftrightarrow \quad p_0 f_{\theta_0}(X) < (1-p_0)\int_{\Theta_1} f_\theta(X)\pi_1(\theta)d\theta \\ &\Leftrightarrow \quad \underbrace{\frac{f_{\theta_0}(X)}{\int_{\Theta_1} f_\theta(X)\pi_1(\theta)d\theta}}_{\text{Bayes factor}} < \frac{1-p_0}{p_0}. \end{aligned}$$

Psychokinesis example cont'd

Recall

$$H_0 : \left\{\theta = \frac{1}{2}\right\} \quad \text{vs} \quad H_1 = \left\{\theta \neq \frac{1}{2}\right\}.$$

Let us chose $p_0 = \Pi(H_0) = \frac{1}{2}$ and $\Pi_1 = \mathcal{U}(0,1)$. Note that $\Pi_1(\{1/2\}) = 0$

Posterior probability of hypothesis H_0:

$$\Pi(H_0|x) = \Pi(\{1/2\}|x) = \frac{p_0 f(x|\theta = \frac{1}{2})}{p_0 f(x|\theta = \frac{1}{2}) + (1-p_0)\int_0^1 f_\theta(x)d\theta}$$

$$= \frac{\binom{n}{x}2^{-n}}{\binom{n}{x}2^{-n} + \frac{1}{n+1}}.$$

Direct calculation gives a very different conclusion from the one based on the p-value (recall $p \approx 0.0003$):

$$\Pi(H_0|x = 52,263,471) \approx 0.92.$$

Psychokinesis example cont'd

Recall

$$H_0 : \left\{\theta = \frac{1}{2}\right\} \quad \text{vs} \quad H_1 = \left\{\theta \neq \frac{1}{2}\right\}.$$

Let us chose $p_0 = \Pi(H_0) = \frac{1}{2}$ and $\Pi_1 = \mathcal{U}(0,1)$. Note that $\Pi_1(\{1/2\}) = 0$

Posterior probability of hypothesis H_0:

$$\Pi(H_0|x) = \Pi(\{1/2\}|x) = \frac{p_0 f(x|\theta = \frac{1}{2})}{p_0 f(x|\theta = \frac{1}{2}) + (1-p_0)\int_0^1 f_\theta(x)d\theta}$$

$$= \frac{\binom{n}{x}2^{-n}}{\binom{n}{x}2^{-n} + \frac{1}{n+1}}.$$

Direct calculation gives a very different conclusion from the one based on the p-value (recall $p \approx 0.0003$):

$$\Pi(H_0|x = 52,263,471) \approx 0.92.$$

Point composite hypothesis

Example $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$.

$$H_0 : \mu = 0, \quad \text{vs} \quad H_1 : \mu \neq 0 \quad \Theta_0 = \{0\} \times \mathbb{R}_{+*}$$

Same approach as for simple hypothesis But

- Π_0 is defined as $\delta_{(0)} \otimes \Pi_\sigma$ where Π_σ is the prior distribution on σ with density π_σ
- Π_1 has density (wrt Lebesgue) on $\mathbb{R} \times \mathbb{R}_{+*}$, e.g.

$$\pi_1(\mu, \sigma^2) = \frac{\varphi\left(\frac{\mu-\mu_0}{\sigma\tau}\right)}{\sigma\tau} \times (\sigma^2)^{-a-1} e^{-b/\sigma^2} \frac{b^a}{\Gamma(a)}$$

i.e . hierarchical prior : under H_1

$$\mu|\sigma \sim \mathcal{N}(\mu_0, \sigma^2\tau^2), \quad \sigma^2 \sim \text{IGamma}(a, b)$$

Posterior calculations

$$\Pi(\Theta_0|x) = \frac{p_0 m_0(x)}{p_0 m_0(x) + (1-p_0) m_1(x)}$$

- $$m_0(x) = \int_0^\infty f(x|\mu = 0, \sigma^2)\pi_\sigma(\sigma) d\sigma$$

  marginal likelihood under H_0

- $$m_1(x) = \int_{\mathbb{R}} \int_0^\infty f(x|\mu, \sigma^2)\pi(\mu|\sigma^2,)\pi_\sigma(\sigma) d\sigma d\mu$$

  marginal likelihood under H_1

Posterior calculations

Generally speaking one can write the posterior probability of Θ_0 as

$$\Pi(\Theta_0|X) = \frac{p_0 m_0(X)}{p_0 m_0(X) + (1-p_0) m_1(X)}$$

where

- $$m_0(X) = \int_{\Theta_0} f_\theta(X)\Pi_0(d\theta)$$

  marginal likelihood under H_0

- $$m_1(X) = \int_{\Theta_1} f_\theta(X)\Pi_1(d\theta)$$

  marginal likelihood under H_1

Bayes factor

Definition (Bayes factor)

The Bayes factor of H_0 over H_1 is given by

$$B_{0/1}(X) = \frac{m_0(X)}{m_1(X)}.$$

The Bayes test associated to the $0-1$ loss function verifies

$$\delta(X) = 1 \quad \Leftrightarrow B_{0/1}(X) < \frac{1-p_0}{p_0}$$

- $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

$$B_{0/1} = \frac{f_{\theta_0}(X)}{\int_{\Theta_1} f_\theta(X)\pi(\theta)d\theta}$$

Interpreting Bayes Factors

Adrian Raftery gives this table (values are approximate, and adapted from a table due to Jeffreys) interpreting B.

'$P(H_0 \mid x)$'	B	$2\log(B)$	evidence for H_0
< 0.5	$<$ 1	$<$ 0	negative (supports H_1)
0.5 to 0.75	1 to 3	0 to 2	barely worth mentioning
0.75 to 0.92	3 to 12	2 to 5	positive
0.92 to 0.99	12 to 150	5 to 10	strong
> 0.99	$>$ 150	$>$ 10	very strong

$2\log(B)$ sometimes reported because it is on the same scale as the familiar deviance and likelihood ratio test statistic.

In Psychokinesis example $B = 12$, corresponding to positive-to-strong evidence in favour of H_0 (no paranormal ability).

Interpreting Bayes Factors

Adrian Raftery gives this table (values are approximate, and adapted from a table due to Jeffreys) interpreting B.

'$P(H_0 \mid x)$'	B	$2\log(B)$	evidence for H_0
< 0.5	$<$ 1	$<$ 0	negative (supports H_1)
0.5 to 0.75	1 to 3	0 to 2	barely worth mentioning
0.75 to 0.92	3 to 12	2 to 5	positive
0.92 to 0.99	12 to 150	5 to 10	strong
> 0.99	$>$ 150	$>$ 10	very strong

$2\log(B)$ sometimes reported because it is on the same scale as the familiar deviance and likelihood ratio test statistic.
In Psychokinesis example $B = 12$, corresponding to positive-to-strong evidence in favour of H_0 (no paranormal ability).

The Bayes factor is the ratio of marginal likelihoods

$$\text{Continuous case} \quad m_j(X) = P(x|H_j) = \int_{\Theta_j} L(\theta; x)\pi(\theta|H_j)d\theta,$$

$$\text{Discrete case} \quad m_j(X) = P(x|H_j) = \sum_{\theta \in \Theta_j} L(\theta; x)\pi(\theta|H_j),$$

$$\text{Simple hyp. case} \quad m_j(X) = P(x|H_j) = L(\theta_0; x).$$

The Bayes factor is the ratio of marginal likelihoods

$$
\begin{aligned}
\text{Continuous case} \quad m_j(X) = P(x|H_j) &= \int_{\Theta_j} L(\theta; x)\pi(\theta|H_j)d\theta, \\
\text{Discrete case} \quad m_j(X) = P(x|H_j) &= \sum_{\theta \in \Theta_j} L(\theta; x)\pi(\theta|H_j), \\
\text{Simple hyp. case} \quad m_j(X) = P(x|H_j) &= L(\theta_0; x).
\end{aligned}
$$

Example

In a quality inspection program components are selected at random from a batch and tested. Let θ denote the failure probability. Suppose that we want to test for $H_0 : \theta \le 0.2$ against $H_1 : \theta > 0.2$.

$$\pi(\theta) = 30\theta(1-\theta)^4,\ 0 < \theta < 1.$$

And this implies $p_0 = p(H_0) = \pi(\theta \in \Theta_0)$ then $p(H_0) = \int_0^{0.2} 30\theta(1-\theta)^4 d\theta$ so that $p(H_0) \simeq 0.345$ and $p(H_1) \simeq 1 - 0.345$.

Example

Thus

$$\pi(\theta|H_0) = \frac{30\theta(1-\theta)^4}{p(H_0)},\ 0 < \theta \leq 0.2$$

and

$$\pi(\theta|H_1) = \frac{30\theta(1-\theta)^4}{p(H_1)},\ 0.2 < \theta < 1$$

Example (cont)

In the quality inspection program suppose n components are selected for independent testing. The number X that fail is $X \sim \text{Binomial}(n, \theta)$.

The marginal likelihood for H_0 is

$$
\begin{aligned}
m_0(x) &= \int_{\Theta_0} L(\theta; x)\pi(\theta|H_0)d\theta \\
&= \binom{5}{x} \int_0^{0.2} \theta^x (1-\theta)^{n-x} \frac{30\theta(1-\theta)^4}{\pi(H_0)} d\theta
\end{aligned}
$$

For one batch of size $n = 5$, $X = 0$ is observed. Recall that $pH_0) \simeq 0.345$. Then

$$
\begin{aligned}
m_0(x) &= \binom{5}{0} \int_0^{0.2} \frac{30\theta(1-\theta)^9}{p(H_0)} d\theta \\
&\simeq 0.185/0.345 = 0.536.
\end{aligned}
$$

Similarly, for $H_1 : \theta \sim \text{Beta}(2,5)|\theta > 0.2$

$$\begin{aligned} m_1(x) &= \binom{5}{0} \int_{0.2}^{1} \frac{30\theta(1-\theta)^9}{p(H_1)} d\theta \\ &\simeq 0.134. \end{aligned}$$

$$\begin{aligned} \Pi(H_0|x) &= \frac{P(x|H_0)p_0}{m(x)} \\ p_0 &= \frac{P(\theta \in \Theta_0)}{(P(\theta \in \Theta_0) + P(\theta \in \Theta_1))} = \pi(H_0) \\ m_0(x)p_0 &\simeq 0.185 \\ m_1(x)(1-p_0) &\simeq 0.088 \\ m(x) &\simeq m_0(x)p_0 + m_1(x)(1-p_0) \simeq 0.273 \\ \Pi(H_0|x) &\simeq 0.185/0.273 = 0.678 \quad \Pi(H_1|x) \simeq 0.322 \end{aligned}$$

$$B = \frac{m_0(x)}{m_1(x)} \simeq 0.536/0.134 = 4$$

Example

$X_1, \ldots, X_n$ are iid $N(\theta, \sigma^2)$, with σ^2 known.
$H_0 : \theta = 0,\ H_1 : \theta | H_1 \sim N(\mu, \tau^2)$. Bayes factor is m_0/m_1, where

$$\begin{aligned} m_0 &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum x_i^2\right) \\ m_1 &= (2\pi\sigma^2)^{-n/2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\sum (x_i - \theta)^2\right) \\ &\qquad \times (2\pi\tau^2)^{-1/2} \exp\left(-\frac{(\theta-\mu)^2}{2\tau^2}\right) d\theta. \end{aligned}$$

Completing the square in m_1 and integrating $d\theta$,

$$\begin{aligned} m_1 &= (2\pi\sigma^2)^{-n/2} \left(\frac{\sigma^2}{n\tau^2 + \sigma^2}\right)^{1/2} \\ &\qquad \times \exp\left[-\frac{1}{2}\left\{\frac{n}{n\tau^2+\sigma^2}(\bar{x} - \mu)^2 + \frac{1}{\sigma^2}\sum (x_i - \bar{x})^2\right\}\right] \end{aligned}$$

So

$$B = \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{n\bar{x}^2}{\sigma^2} - \frac{n}{n\tau^2 + \sigma^2}(\bar{x} - \mu)^2\right\}\right]$$

Defining $t = \sqrt{n}\bar{x}/\sigma, \eta = -\mu/\tau, \rho = \sigma/(\tau\sqrt{n})$, this can be written as

$$B = \left(1 + \frac{1}{\rho^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{(t - \rho\eta)^2}{1 + \rho^2} - \eta^2\right\}\right]$$

This example illustrates a problem choosing the prior. If we take a diffuse prior under H_1: i.e. $\tau \to +\infty$, then $B \to \infty$ whatever x, giving overwelming support for H_0.

This is an instance of Lindley's paradox.

Bayes factor & Bayes tests (under $0-1$ loss) are not defined when the priors are improper

So

$$B = \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{n\bar{x}^2}{\sigma^2} - \frac{n}{n\tau^2 + \sigma^2}(\bar{x} - \mu)^2\right\}\right]$$

Defining $t = \sqrt{n}\bar{x}/\sigma, \eta = -\mu/\tau, \rho = \sigma/(\tau\sqrt{n})$, this can be written as

$$B = \left(1 + \frac{1}{\rho^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{(t - \rho\eta)^2}{1 + \rho^2} - \eta^2\right\}\right]$$

This example illustrates a problem choosing the prior. If we take a diffuse prior under H_1: i.e. $\tau \to +\infty$, then $B \to \infty$ whatever x, giving overwelming support for H_0.
This is an instance of Lindley's paradox.
Bayes factor & Bayes tests (under $0-1$ loss) are not defined when the priors are improper

So

$$B = \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{n\bar{x}^2}{\sigma^2} - \frac{n}{n\tau^2+\sigma^2}(\bar{x}-\mu)^2\right\}\right]$$

Defining $t = \sqrt{n}\bar{x}/\sigma, \eta = -\mu/\tau, \rho = \sigma/(\tau\sqrt{n})$, this can be written as

$$B = \left(1 + \frac{1}{\rho^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{(t-\rho\eta)^2}{1+\rho^2} - \eta^2\right\}\right]$$

This example illustrates a problem choosing the prior. If we take a diffuse prior under H_1: i.e. $\tau \to +\infty$, then $B \to \infty$ whatever x, giving overwelming support for H_0.
This is an instance of Lindley's paradox.
Bayes factor & Bayes tests (under $0-1$ loss) are not defined when the priors are improper

So

$$B = \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{n\bar{x}^2}{\sigma^2} - \frac{n}{n\tau^2 + \sigma^2}(\bar{x} - \mu)^2\right\}\right]$$

Defining $t = \sqrt{n}\bar{x}/\sigma, \eta = -\mu/\tau, \rho = \sigma/(\tau\sqrt{n})$, this can be written as

$$B = \left(1 + \frac{1}{\rho^2}\right)^{1/2} \exp\left[-\frac{1}{2}\left\{\frac{(t - \rho\eta)^2}{1 + \rho^2} - \eta^2\right\}\right]$$

This example illustrates a problem choosing the prior. If we take a diffuse prior under H_1: i.e. $\tau \to +\infty$, then $B \to \infty$ whatever x, giving overwelming support for H_0.
This is an instance of Lindley's paradox.
Bayes factor & Bayes tests (under $0 - 1$ loss) are not defined when the priors are improper

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i) d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i) m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j) m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji} \right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \dots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $\mathfrak{M}_1, \dots, \mathfrak{M}_k$. Under model $\mathfrak{M}_i$;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(\mathfrak{M}_i)$ ($= 1/k$ in the uniform prior case)
- Marginal density of X is $m_i(x) = m(x|\mathfrak{M}_i) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i$.

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/m(x|\mathfrak{M}_i)$.
2. Bayes factor of $\mathfrak{M}_j$ to $\mathfrak{M}_i$ is $B_{ji} = m(x|\mathfrak{M}_j)/m(x|\mathfrak{M}_i)$.
3. Posterior

$$\Pi(\mathfrak{M}_i|x) = \frac{(\mathfrak{M}_i)m(x|\mathfrak{M}_i)}{\sum_j \Pi(\mathfrak{M}_j)m(x|\mathfrak{M}_j)} = \left[\sum_j \frac{\Pi(\mathfrak{M}_j)}{\Pi(\mathfrak{M}_i)} B_{ji}\right]^{-1}.$$