Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2016

Lecture 9 : bayesian hypothesis tests and model selection

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.

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.

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 2influence2 particles to go to either red or green.

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 2influence2 particles to go to either red or green.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,900,000$. Observe $x = 52,263,000$.

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 2influence2 particles to go to either red or green.

Model: $X = \#$ red particles. $X \sim \text{Bin}(n, \theta)$. $n = 104,900,000$. Observe $x = 52,263,000$.P-value $= P_{\theta=.5}(X \geq x) \approx .0003$. Strong evidence of ability?

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.Set a prior probability $\pi(H_1) = 1 - \pi(H_0)$ and a prior $\pi(\theta)$ for θ on H_1.

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.Set a prior probability $\pi(H_1) = 1 - \pi(H_0)$ and a prior $\pi(\theta)$ for θ on H_1. Let us chose $\pi(H_1) = frac12$ and $\pi(\theta) = 1$.

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.Set a prior probability $\pi(H_1) = 1 - \pi(H_0)$ and a prior $\pi(\theta)$ for θ on H_1. Let us chose $\pi(H_1) = frac12$ and $\pi(\theta) = 1$.
We want the posterior probability

$$P(H_0|x)$$

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.Set a prior probability $\pi(H_1) = 1 - \pi(H_0)$ and a prior $\pi(\theta)$ for θ on H_1. Let us chose $\pi(H_1) = frac12$ and $\pi(\theta) = 1$.
We want the posterior probability

$$P(H_0|x) = \frac{f(x|\theta = \frac{1}{2})\pi(H_0)}{\pi(H_0)f(x|H_0) + \pi(H_1)f(x|H_1)}$$

Psychokinesis example

Bayesian analysis: Set $H_0 = \{\theta = \frac{1}{2}\}$ and $H_1 = \{\theta \neq \frac{1}{2}\}$.Set a prior probability $\pi(H_1) = 1 - \pi(H_0)$ and a prior $\pi(\theta)$ for θ on H_1. Let us chose $\pi(H_1) = frac12$ and $\pi(\theta) = 1$.
We want the posterior probability

$$P(H_0|x) = \frac{f(x|\theta = \frac{1}{2})\pi(H_0)}{\pi(H_0)f(x|H_0) + \pi(H_1)f(x|H_1)}$$

Caclculation shows

$$P(H_0|x = 52,263,000) \approx 0.94.$$

(recall p-value$\approx$ 0.0003)

Hypothesis Testing I

In Bayesian inference, hypotheses are represented by prior distributions. There is nothing special about H_0, and H_0 and H_1 need not be nested.

Let $\pi(\theta|H_0),\ \theta \in \Theta_0$ be the prior distribution of θ under hypothesis H_0. Here $\pi(\theta|H_0)$ is a pmf/pfd as $\theta|H_0$ is continuous/discrete.

Composite $H_0 : \theta \in \Theta_0$, with $\Theta_0 \subset \Theta$,

$$\pi(\theta|H_0) = \frac{\pi(\theta)}{\pi(\theta \in \Theta_0)} \mathbb{I}(\theta \in \Theta_0)$$

If Θ_0 has more than one element then H_0 is a composite hypothesis.

Simple $H_0 : \theta = \theta_0$, so that $\pi(\theta_0|H_0) = 1$. This is a simple hypothesis.

However, since any statement about the form of the prior amounts to a hypothesis about θ, we are not restricted to statements about set membership (not just simple and composite).

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$
3. Prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$
3. Prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$
4. prior densities $\pi_0(\theta)$ on Θ_0 and $\pi_1(\theta)$ on Θ_1 sometime written $\pi(\theta|H_{0,1})$.

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$
3. Prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$
4. prior densities $\pi_0(\theta)$ on Θ_0 and $\pi_1(\theta)$ on Θ_1 sometime written $\pi(\theta|H_{0,1})$.

Observe that you can think of this as a hierarchical model since

1. $X|\theta \sim f(x,\theta)$,
2. $\theta\,|H \sim \pi(\theta|H)$ where $H \in \{H_0, H_1\}$

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$
3. Prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$
4. prior densities $\pi_0(\theta)$ on Θ_0 and $\pi_1(\theta)$ on Θ_1 sometime written $\pi(\theta|H_{0,1})$.

Observe that you can think of this as a hierarchical model since

1. $X|\theta \sim f(x,\theta)$,
2. $\theta\,|H \sim \pi(\theta|H)$ where $H \in \{H_0, H_1\}$ (this is the ψ parameter from previous lecture),

Hypothesis Testing I

Framework for Bayesian hypothesis testing

1. $X|\theta \sim f(x;\theta)$
2. $H_0 := \{\theta \in \Theta_0\}$, $H_1 := \{\theta \in \Theta_1\}$
3. Prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$
4. prior densities $\pi_0(\theta)$ on Θ_0 and $\pi_1(\theta)$ on Θ_1 sometime written $\pi(\theta|H_{0,1})$.

Observe that you can think of this as a hierarchical model since

1. $X|\theta \sim f(x,\theta)$,
2. $\theta\,|H \sim \pi(\theta|H)$ where $H \in \{H_0, H_1\}$ (this is the ψ parameter from previous lecture),
3. H has (hyper)-prior $p(H)$ (think of a parameter $H \in \{H_0, H_1\}$ with likelihood $P(x|H)$, prior $p(H)$ and posterior $P(H|x)$).

Hypothesis Testing I

Definition (Marginal likelihood)

The Marginal likelihood of x under $H_i, i = 0, 1$ is

$$\begin{aligned} p(x|H_i) &= \int f(x|\theta, H_i)\pi(\theta|H_i)d\theta \\ &= \int f(x;\theta)\pi(\theta|H_i)d\theta \end{aligned}$$

Hypothesis Testing I

Definition (Marginal likelihood)

The Marginal likelihood of x under $H_i, i = 0, 1$ is

$$\begin{aligned} p(x|H_i) &= \int f(x|\theta, H_i)\pi(\theta|H_i)d\theta \\ &= \int f(x;\theta)\pi(\theta|H_i)d\theta \end{aligned}$$

$$\begin{aligned} \text{Continuous case} \quad P(x|H_0) &= \int_{\Theta_0} L(\theta; x)\pi(\theta|H_0)d\theta, \\ \text{Discrete case} \quad P(x|H_0) &= \sum_{\theta\in\Theta_0} L(\theta; x)\pi(\theta|H_0), \\ \text{Simple hyp. case} \quad P(x|H_0) &= L(\theta_0; x). \end{aligned}$$

Example 1

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 $H_0 : \theta \leq 0.2$ against $H_1 : \theta > 0.2$ and that the prior is $\theta \sim \text{Beta}(2, 5)$ so that

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

Now if $\pi(H_0) = \pi(\theta \in \Theta_0)$ then $\pi(H_0) = \int_0^{0.2} 30\theta(1-\theta)^4 d\theta$ so that $\pi(H_0) \simeq 0.345$ and $\pi(H_1) \simeq 1 - 0.345$ so

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

and

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

Example 1 (cont)

In the quality inspection program suppose n components are selected for independent testing. The number X that fail is $X \sim \text{Binomial}(n, \theta)$. Recall $H_0 : \theta \leq 0.2$ with $\theta \sim \text{Beta}(2, 5)$ in the prior.

The marginal likelihood for H_0 is

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

For one batch of size $n = 5$, $X = 0$ is observed. Recall that $\pi(H_0) \simeq 0.345$. Then

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

Similarly, for $H_1 : \theta > 0.2$

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

Notice that

1. $$P(x|H_0) = \mathbb{E}(L(\vartheta; x)|H_0),$$
   that is, the marginal likelihood is the average likelihood given the prior $\pi(\theta|H_0)$
2. the marginal likelihood is the normalizing constant we often leave off when we write the posterior

$$\pi(\theta|x, H_0) = \frac{L(\theta; x)\pi(x|H_0)}{P(x|H_0)},$$

$$\text{posterior} = \frac{\text{likelihood} \quad \times \quad \text{prior}}{\text{marginal likelihood}},$$

Prior and Posterior

We have a posterior probability for H_0 itself. This is actually where we started with Bayesian inference. In the simple case where we have two hypotheses H_0, H_1, exactly one of which is true,

Posteriors

$$P(H_0 \mid x) = \frac{P(H_0)P(x \mid H_0)}{P(x)} = 1 - P(H_1|x),$$

where

$$P(x) = P(H_0)P(x \mid H_0) + P(H_1)P(x \mid H_1)$$

When we estimate the value of a discrete parameter $H \in \{H_0, H_1\}$, we are making a Bayesian hypothesis test.

Prior and Posterior

We have a posterior probability for H_0 itself. This is actually where we started with Bayesian inference. In the simple case where we have two hypotheses H_0, H_1, exactly one of which is true,

Posteriors

$$P(H_0 \mid x) = \frac{P(H_0)P(x \mid H_0)}{P(x)} = 1 - P(H_1|x),$$

where

$$P(x) = P(H_0)P(x \mid H_0) + P(H_1)P(x \mid H_1)$$

When we estimate the value of a discrete parameter $H \in \{H_0, H_1\}$, we are making a Bayesian hypothesis test. Recall that in the psychokinesis example we computed such a posterior $P(H_0|x) \approx .94$.

Example 1 (cont)

$X \sim \text{Binomial}(5, \theta)$ with $\theta \sim \text{Beta}(2, 5)$ in the prior and $H_0 : \theta \leq 0.2$ and $H_1 : \theta > 0.2$. The posterior probability for H_0 given we observe $X = 0$ is

$$
\begin{aligned}
P(H_0|x) &= \frac{P(x|H_0)P(H_0)}{P(x)} \\
P(H_0) &= \frac{P(\theta \in \Theta_0)}{(P(\theta \in \Theta_0) + P(\theta \in \Theta_1))} = \pi(H_0) \\
P(x|H_0)\pi(H_0) &\simeq 0.185 \\
P(x|H_1)\pi(H_1) &\simeq 0.088 \\
P(x) &\simeq P(x|H_0)P(H_0) + P(x|H_1)P(H_1) \\
&\simeq 0.273 \\
P(H_0|x) &\simeq 0.185/0.273 \\
&= 0.678 \\
P(H_1|x) &\simeq 0.322
\end{aligned}
$$

Hypothesis Testing II, Bayes factors

Suppose we have two hypotheses H_0, H_1, exactly one of which is true. Data x.

The Prior Odds

$$Q = \frac{P[H_0]}{P[H_1]}$$

These are prior *odds* since $P[H_1] = 1 - P[H_0]$. Here H_0 *is* Q *times more probable than* H_1*, given the prior model*.

The Posterior Odds

$$Q^* = \frac{P[H_0 \mid x]}{P[H_1 \mid x]}$$

are the posterior odds, so that H_0 *is* Q^* *times more probable than* H_1*, given the data and prior model*.

Hypothesis Testing II, Bayes factors

The posterior odds for H_0 against H_1 can be written

$$Q^* = \frac{P[H_0]}{P[H_1]} \times \frac{P(x \mid H_0)}{P(x \mid H_1)} = Q \times B$$

where Q is the prior odds and

$$B = \frac{P(x \mid H_0)}{P(x \mid H_1)}$$

is the Bayes Factor.

The Bayes Factor is a criterion for model comparison since H_0 *is* B *times more probable than* H_1*, given the data and a prior model which puts equal probability on* H_0 *and* H_1. The Bayes factor tells us how the data shifts the strength of belief (measured as a probability) in H_0 relative to H_1.

Example 1 (cont)

$X \sim \text{Binomial}(5, \theta)$ with $\theta \sim \text{Beta}(2, 5)$ in the prior and $H_0 : \theta \leq 0.2$ and $H_1 : \theta > 0.2$.

The prior odds are

$$\begin{aligned} Q &= P(H_0)/P(H_1) \\ &\simeq 0.345/(1 - 0.345) \simeq 0.527 \end{aligned}$$

The posterior odds are

$$\begin{aligned} Q^* &= P(H_0|x)/P(H_1|x) \\ &\simeq 0.678/(1 - 0.678) \simeq 2.1 \end{aligned}$$

The Bayes factor comparing H_0 and H_1 is

$$\begin{aligned} B &= \frac{P(x|H_0)}{P(x|H_1)} \\ &\simeq 0.536/0.134 = 4 \end{aligned}$$

Explicitly, from the beginning,

$$
\begin{aligned}
B &= \frac{\int_{\Theta_0} L(x;\theta)\pi(\theta|H_0)d\theta}{\int_{\Theta_1} L(x;\theta)\pi(\theta|H_1)d\theta} \\
&= \frac{\int_{\Theta_0} L(x;\theta)\pi(\theta)d\theta}{\int_{\Theta_1} L(x;\theta)\pi(\theta)d\theta} \times \frac{\pi(H_1)}{\pi(H_0)} \\
&= \frac{\binom{5}{0}\int_0^{0.2} 30\theta(1-\theta)^9 d\theta}{\binom{5}{0}\int_{0.2}^{1} 30\theta(1-\theta)^9 d\theta} \frac{\int_{0.2}^{1} 30\theta(1-\theta)^4 d\theta}{\int_0^{0.2} 30\theta(1-\theta)^4 d\theta} \\
&= 6619897/1654272 \simeq 4.002 \qquad \text{(Maple)}.
\end{aligned}
$$

This is 'positive' evidence for $\theta \leq 0.2$. Notice that the Bayes factor is 'more positive' than the posterior odds, as the prior odds were weighted against H_0.

Interpreting Bayes Factors

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

'$P(H_0\|x)$'	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

I added the leftmost column (posterior for prior odds equal one). We sometimes report $2\log(B)$ because it is on the same scale as the familiar deviance and likelihood ratio test statistic.

Simple-Simple hypothesis

If both hypotheses are simple $H_0 : \theta = \theta_0;\ H_1 : \theta = \theta_1$, with priors $P(H_0)$ and $P(H_1)$ for the two hypotheses, the posterior probability for H_0 is

$$\begin{aligned} P(H_0|x) &= \frac{P(x|H_0)P(H_0)}{P(x)} \\ &= \frac{L(\theta_0; x)P(H_0)}{L(\theta_0; x)P(H_0) + L(\theta_1; x)P(H_1)}, \end{aligned}$$

since $P(x|H_0)$ is just $L(\theta_0; x)$. The Bayes factor is then just likelihood ratio

$$B = \frac{L(\theta_0; x)}{L(\theta_1; x)}.$$

Simple-Composite hypothesis

If one hypothesis is simple and the other composite, for example, $H_0 : \theta = \theta_0; \quad H_1 : \pi(\theta|H_1), \ \theta \in \Theta$, with priors $P(H_0)$ and $P(H_1)$ for the two hypotheses, the Bayes factor is

$$B = \frac{L(x;\theta_0)}{\int_\Theta L(x;\theta)\pi(\theta|H_1)d\theta}$$

The denominator is just $\int_\Theta L(x;\theta)\pi(\theta)d\theta$ when $\pi(\theta|H_1)$ is a pdf.

Exercise Show that the posterior probability for H_0 is

$$P(H_0|x) = \frac{L(\theta_0;x)P(H_0)}{P(H_0)L(\theta_0;x) + P(H_1)\int_\Theta L(x;\theta)\pi(\theta)d\theta}$$

when $\pi(\theta|H_1)$ is a pdf, in this simple-composite comparison.

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 P_0/P_1, where

$$\begin{aligned} P_0 &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum x_i^2\right) \\ P_1 &= (2\pi\sigma^2)^{-n/2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\sum (x_i - \theta)^2\right) \\ &\qquad \times (2\pi\tau^2)^{-1/2} \exp\left(-\frac{(\theta - \mu)^2}{2\tau^2}\right) d\theta. \end{aligned}$$

Completing the square in P_1 and integrating $d\theta$,

$$\begin{aligned} P_1 &= (2\pi\sigma^2)^{-n/2} \left(\frac{\sigma^2}{n\tau^2 + \sigma^2}\right)^{1/2} \\ &\qquad \times \exp\left[-\frac{1}{2}\left\{\frac{n}{n\tau^2 + \sigma^2}(\bar{x} - \mu)^2 + \frac{1}{\sigma^2}\sum (x_i - \bar{x})^2\right\}\right] \end{aligned}$$

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, for ρ so that $\rho \to 0$, then $B \to \infty$, giving overwelming support for H_0.

This is an instance of Lindley's paradox. The point here is that B compares the *models* $\theta = \theta_0$ and $\theta \sim \pi(\cdot|H_1)$, not the *sets* θ_0 against $\theta \setminus \{\theta_0\}$. If the $\pi(\theta|H_1)$-prior becomes very diffuse then the *average* likelihood (ie $P(H_1|x)$, the marginal likelihood, which is the denominator of B) goes to zero, while $P(H_0|x) = L(\theta_0; x)$ is fixed.

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

- $X \sim f_i(x; \theta_i)$ where θ_i unknown parameter.
- Prior for θ_i is $\pi_i(\theta)$.
- Prior probability $P(M_i)$ $(= 1/k$ in the uniform prior case)

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \dots, M_k$. Under model M_i;

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

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

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

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/P(x|M_i)$.

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \dots, M_k$. Under model M_i;

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

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/P(x|M_i)$.
2. Bayes factor of M_j to M_i is $B_{ji} = P(x|M_j)/P(x|M_i)$.

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

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

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/P(x|M_i)$.
2. Bayes factor of M_j to M_i is $B_{ji} = P(x|M_j)/P(x|M_i)$.
3. Posterior

$$P(M_i|x) = \frac{P(M_i)P(x|M_i)}{\sum_j P(M_j)P(x|M_j)}$$

Model selection

Framework for Bayesian Model selection

Models (or hypothesis) for data x: $M_1, \ldots, M_k$. Under model M_i;

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

1. Posterior density $\pi_i(\theta_i|x) = f_i(x|\theta_i)\pi_i(\theta_i)/P(x|M_i)$.
2. Bayes factor of M_j to M_i is $B_{ji} = P(x|M_j)/P(x|M_i)$.
3. Posterior

$$P(M_i|x) = \frac{P(M_i)P(x|M_i)}{\sum_j P(M_j)P(x|M_j)} = \left[\sum_j \frac{P(M_j)}{P(M_i)} B_{ji}\right]^{-1}.$$

Model selection: Example

Suppose tat $X_1, \ldots, X_n$ are i.i.d. with density

$$f(x|\mu, \sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$$

and several models are considered:

M_1 g is $N(0, 1)$

M_2 g is uniform $(0, 1)$

M_3 g is Cauchy $(0, 1)$

M_4 g is left-exponential ($\propto e^{x-\mu}, x \leq \mu$)

M_5 g is right-exponential ($\propto e^{-(x-\mu)}, x \geq \mu$)

Model selection: Example

Suppose tat $X_1, \dots, X_n$ are i.i.d. with density

$$f(x|\mu, \sigma) = \frac{1}{\sigma}\, g\left(\frac{x-\mu}{\sigma}\right)$$

and several models are considered:

M_1 g is $N(0,1)$

M_2 g is uniform $(0,1)$

M_3 g is Cauchy $(0,1)$

M_4 g is left-exponential ($\propto e^{x-\mu}, x \le \mu$)

M_5 g is right-exponential ($\propto e^{-(x-\mu)}, x \ge \mu$)

Choose $P(M_i) = 1/5, i = 1, \dots, 5$ and $\pi_i(\theta, \sigma) = \frac{1}{\sigma}$.

Model selection: Example (location-scale)

Marginal $P(x|M_i)$ can be calculated in close form for these distributions.

- M_1 Normal: $\frac{\Gamma((n-1)/2)}{(2\pi)^{(n-1)/2}\sqrt{n}(\sum(x_i-\bar{x})^2)^{(n-1)/2}}$
- M_2 Uniform $(0,1)$: $\frac{1}{n(n-1)(x_{(n)}-x_{(1)})^{n-1}}$
- M_3 Cauchy $(0,1)$: Given in Spiegelhalter (1985).
- M_4 Left-exponential: $\frac{(n-2)!}{n^n(x_{(n)}-\bar{x})^{n-1}}$
- M_5 Right-exponential: $\frac{(n-2)!}{n^n(\bar{x}-x_{(1)})^{n-1}}$

Model selection: Example (location-scale)

Consider four data sets

- Darwin's data ($n = 15$),
- Cavendish's data ($n = 29$),
- Stigler's data ($n = 20$),
- Randomly generated Cauchy sampple ($n = 14$).

Model selection: Example (location-scale)

Consider four data sets

- Darwin's data ($n = 15$),
- Cavendish's data ($n = 29$),
- Stigler's data ($n = 20$),
- Randomly generated Cauchy sampple ($n = 14$).

Objective posterior probability $P(M_i|x)$

	Normal	Unif	Cauchy	L. exp.	R. exp.
Darwin	.390	.056	.430	.124	.0001
Cavendish	.986	.010	.004	$4 \cdot 10^{-8}$	.0006
Stigler	$7 \cdot 10^{-8}$	$4 \cdot 10^{-5}$	.994	.006	$2 \cdot 10^{-13}$
Cauchy	$5 \cdot 10^{-13}$	$9 \cdot 10^{-12}$	.999	$7 \cdot 10^{-18}$	.0001