Bayesian Statistics
Note 3
Bayesian Inference

Keunbaik Lee

Sungkyunkwan University

Point Estimation I

Two candidates are posterior mode or posterior mean.

Definition
The generalized maximum likelihood or ML-2 estimate of θ is the value $\hat{\theta}$ which maximize $P(\theta|y)$ considered as a function of θ.

Definition
If θ is a real valued parameter with posterior distribution $P(\theta|y)$, and δ is an estimate of θ, then the **posterior mean squared error** of δ is $V_\delta(y) = E[(\theta - \delta)^2|y]$.

Point Estimation II

Note:

$$
\begin{aligned}
V_\delta(y) &= E[(\theta - \delta(y))^2 | y] \\
&= E[(\theta - \mu(y) + \mu(y) - \delta(y))^2 | y], \quad \text{where } \mu(y) = E(\theta | y) \\
&= E[(\theta - \mu(y))^2 | y] + (\mu(y) - \delta(y))^2 \\
&= v(y) + (\mu(y) - \delta(y))^2, \quad \text{where } v(y) = Var(\theta | y)
\end{aligned}
$$

Note that $\mu(y)$ minimizes $V_\delta(y)$ (over all δ) and it is the estimate with the smallest posterior mean squared error.
For this reason, it is customary to use $\mu(y)$ as the estimate of θ and report $\sqrt{v(y)}$ as the standard error.

Point Estimation III

Example 1:

$$
\begin{aligned}
& Y|\theta \sim N(\theta, \sigma^2), \quad \theta \sim N(\mu_0, \tau_0^2) \\
\Rightarrow \quad & \theta|y \sim N((1-B_0)y + B_0\mu_0, \quad \sigma^2(1-B_0)),
\end{aligned}
$$

where $B_0 = \frac{\sigma^2}{\sigma^2+\tau_0^2}$. Then, ML-2 estimate of θ is posterior mode$= (1-B_0)y + B_0\mu_0$
For example, $y = 5$ (See R code).

Point Estimation IV

Example 2:

$$
\begin{aligned}
P(y|\theta) &= \exp(\theta - y) I_{[y \geq \theta]} \\
P(\theta) &= \frac{1}{\pi(1+\theta^2)} I_{(-\infty,\infty)}(\theta) \\
\Rightarrow P(\theta|y) &= \frac{P(\theta)P(y|\theta)}{P(y)} = \frac{\exp(\theta - y) I_{[y \geq \theta]}}{P(y)\pi(1+\theta^2)} \\
P(y) &= \int_{-\infty}^{y} P(y|\theta)P(\theta)d\theta
\end{aligned}
$$

Then, ML-2 estimate of θ is $\hat{\theta} = y$.

Point Estimation V

The posterior mean is

$$E(\theta|y) = \frac{\int_{-\infty}^{y} \theta \frac{e^{\theta-y}}{1+\theta^2} d\theta}{\int_{-\infty}^{y} \frac{e^{\theta-y}}{1+\theta^2} d\theta}.$$

Need numerical integration. (See R code.)

Point Estimation VI

Example 3:
$y|\theta \sim N(\theta, \sigma^2)$, where $\sigma^2(> 0)$ is known, but $\theta \in (0,\ \infty)$ is unknown.

- The UMVUE of θ is y, which is clearly unsuitable, since y can take negative values with positive probability.
- The MLE of θ is $\max(y,\ 0)$. also not very appealing for large negative values of y.

Suppose, no specific prior knowledge is available. Then a reasonable non-informative prior is $I_{(0,\infty)}(\theta)$.

$$\Rightarrow P(\theta|y) = \frac{P(\theta)P(y|\theta)}{P(y)} = \frac{\exp(-\frac{1}{2\sigma^2}(\theta - y)^2)I_{(0,\infty)}(\theta)}{\int_0^\infty \exp(-\frac{1}{2\sigma^2}(\theta - y)^2)d\theta}$$

Point Estimation VII

Hence,

$$
\begin{aligned}
E(\theta|y) &= E(\theta - y + y|y) \\
&= y + \sigma E\left[\frac{\theta - y}{\sigma}|y\right] \\
&= y + \sigma \frac{\int_0^\infty \frac{\theta - y}{\sigma} \exp(-\frac{1}{2\sigma^2}(\theta - y)^2) d\theta}{\int_0^\infty \exp(-\frac{1}{2\sigma^2}(\theta - y)^2) d\theta} \\
&= y + \sigma \frac{\int_{-\frac{y}{\sigma}}^\infty z \exp(-\frac{1}{2}z^2) dz}{\int_{-\frac{y}{\sigma}}^\infty \exp(-\frac{1}{2}z^2) dz} \\
&= y + \sigma \frac{\sqrt{2\pi}\phi\left(\frac{y}{\sigma}\right)}{\sqrt{2\pi}\left\{1 - \Phi\left(-\frac{y}{\sigma}\right)\right\}} \; = \; y + \sigma \frac{\phi\left(\frac{y}{\sigma}\right)}{\Phi\left(\frac{y}{\sigma}\right)}
\end{aligned}
$$

Point Estimation VIII

Example 4:

$y|\theta \sim N(\theta,\ \sigma^2)$, $\theta \sim N(\mu_0,\ \tau_0^2)$ then $v(y) = \frac{\sigma^2\tau_0^2}{\sigma^2+\tau_0^2}$.

Recall the IQ example. Here, $\sigma^2 = 100,\ \ \mu_0 = 100,\ \ \tau_0^2 = 225$

Hence, $v(y) = 69.23, \quad \sqrt{v(y)} = 8.32$.

The classical estimate of θ is just $\delta(y) = y$ then

$$\begin{aligned}
V_\delta(y) &= v(y) + (\mu(y) - \delta(y))^2 \\
&= v(y) + ((1 - B_0)y + B_0\mu_0 - y)^2 \\
&= v(y) + B_0^2(y - \mu_0)^2 \\
&= 69.23 + \left(\frac{4}{13}\right)^2 (y - 100)^2
\end{aligned}$$

Point Estimation IX

where

$$B_0 = \frac{\sigma^2}{\sigma^2 + \tau_0^2}$$

For example, $y = 115,\ V_\delta(115) = 90.53,\ V_\delta^{\frac{1}{2}}(115) = 9.51.$

In this example, a Bayesian would estimate θ by $\mu(y)$ with s.e. $\sigma(1 - B_0)^{1/2} < \sigma$. This is usually (but not always) true for Bayesian estimation. Since the Bayesian is using prior as well as sample information for estimation θ, the Bayesian s.e. is typically smaller than the s.e. that a classical statistician would report for the classical estimate.

Credible Sets I

Definition

A $100(1-\alpha)\%$ credible set for θ is a subset C of Θ such that

$$1-\alpha \le P(\theta \in C|y) = \begin{cases} \int_C P(\theta|y)d\theta, & \text{continuous;} \\ \sum_C P(\theta|y), & \text{discrete.} \end{cases}$$

In choosing a credible set for θ ($100(1-\alpha)\%$ coverage probability), it is usually desirable to try to minimize its size. To do this, one should include in the set only those points with the largest posterior density. i.e. the most likely values of θ.

Credible Sets II

Definition

The $100(1-\alpha)\%$ highest posterior density (HPD) credible set for θ is the subset C of Θ such that

$$C = \{\theta \in \Theta \ : \ P(\theta|y) \geq k(\alpha)\}$$

where $k(\alpha)$ is the largest constant such that $P(C|y) \geq 1-\alpha$.

Credible Sets III

Example 1:

$$
\begin{aligned}
& y|\theta \sim N(\theta,\ \sigma^2), \quad \theta \sim N(\mu_0,\ \tau_0^2) \\
\Rightarrow \quad & \theta|y \sim N((1-B_0)y + B_0\mu_0, \quad \sigma^2(1-B_0))
\end{aligned}
$$

where $B_0 = \frac{\sigma^2}{\sigma^2+\tau_0^2}$. Thus, under the posterior distribution,

$$
\frac{\theta - ((1-B_0)y + B_0\mu_0)}{\sqrt{\sigma^2(1-B_0)}} \ \sim\ N(0,\ 1).
$$

Credible Sets IV

The $100(1-\alpha)\%$ HPD credible set for θ is given by

$$
\begin{aligned}
C &= \{\theta \ : \ P(\theta|y) \geq k(\alpha)\} \\
&= \left\{\theta \ : \ |\theta - ((1-B_0)y + B_0\mu_0)| \leq z_{\frac{\alpha}{2}}\sqrt{\sigma^2(1-B_0)}\right\} \\
&= \left\{\theta \ : \ (1-B_0)y + B_0\mu_0 - z_{\frac{\alpha}{2}}\sqrt{\sigma^2(1-B_0)}\right. \\
&\qquad \left. \leq \theta \leq (1-B_0)y + B_0\mu_0 + z_{\frac{\alpha}{2}}\sqrt{\sigma^2(1-B_0)}\right\}
\end{aligned}
$$

Credible Sets V

eg) In the IQ example, if a child scores 115 in a given test,

- the 95% HPD credible set for θ (the true score) is

$$[110.39-(1.96)(69.23)^{\frac{1}{2}},\ 110.39+(1.96)(69.23)^{\frac{1}{2}}] = [94.08,\ 126.70]$$

- Since a random test score $y \sim N(\theta,\ 100)$, the classical 95% CI for θ is

$$[115-(1.96)(10),\ 115+(1.96)(10)] = [95.40,\ 134.60]$$

Note

If one assumes the improper prior $P(\theta) = 1$, the $100(1-\alpha)\%$ HPD credible set matches perfectly the $100(1-\alpha)\%$ classical CI for θ.

Credible Sets VI

Example 2:

Let $y_1, \ \cdots, \ y_n$ be iid Cauchy with common pdf

$$f(y|\theta) = \frac{1}{\pi(1+(y-\theta)^2)} I_{(-\infty,\infty)}(y).$$

Assume $\theta \in \Theta = (0, \ \infty)$
noninformative prior for θ: $P(\theta) = I_{(0, \ \infty)}(\theta)$

$$\Rightarrow \quad P(\theta|y_1, \ \cdots, \ y_n) = \frac{\prod_{j=1}^{n} \left\{ \frac{1}{1+(\theta-y_i)^2} \right\} I_{(0, \ \infty)}(\theta)}{\int_0^\infty \prod_{j=1}^{n} \left\{ \frac{1}{1+(\theta-y_i)^2} \right\} d\theta}.$$

Credible Sets VII

Although not analytically tractable, finding a $100(1-\alpha)\%$ HPD credible set on a computer is not too difficult.

eg) If $n = 5,\ y = (4.0,\ 5.5,\ 7.5,\ 4.5,\ 3.0)^T$, the resulting 95% HPD credible set is $[3.10,\ 6.06]$.
In contrast, it is not clear at all how to construct a good classical CI.

Credible Sets VIII

The general idea behind numerical calculation of the HPD credible set in a situation where $P(\theta|y)$ is continuous in θ is to set up a program along the following lines:

1. Create a subroutine, which for a given k finds all solutions to the equation $P(\theta|y) = k$. The set $C(k) = \{\theta; P(\theta|y) > k\}$ can typically be easily constrcted from these solutions. For instance, if Θ is an infinite interval in R, and only two solutions $\theta_1(k)$ and $\theta_2(k)$ are found, then $C(k) = [\theta_1(k), \theta_2(k)]$
2. Create a subroutine which calculates

$$P(C(k)) = \int_{C(k)} P(\theta|y) d\theta$$

Credible Sets IX

3. Numerically solve the equation

$$P(C(k)) = 1 - \alpha$$

calling on the above two subroutines as k varies.

Credible Sets X

- Often a useful approximation to a HPD credible set can be achieved through the use of normal approximation to posteriors.
- It can be shown that for large sample sizes, the posterior distribution will be approximately normal.
- Even for small samples, a normal likelihood function will usually yield a roughly normal posterior so that the approximation can be useful even in small samples.

Credible Sets XI

- $$\theta|y \sim N(\mu(y),\ v(y))$$

  where $\mu(y) = E(\theta|y),\ v(y) = E[(\theta - \mu(y))^2|y]$.
  :approximate $100(1-\alpha)\%$ HPD credible set is

  $$C = [\mu(y) \pm z_{\frac{\alpha}{2}}\sqrt{v(y)}].$$

See R code

Credible Sets XII

Example 2: (Continued)
The posterior density is clearly normal, and with 5 Cauchy observations, one might imagine that normal approximation would be somewhat inaccurate.
The posterior distribution is, however, unimodal, and the normal approximation seems to be excellent out to the 2.5% and 97.5% tails.

Credible Sets XIII

Numerical calculation:

$$\mu(y) = 4.55, \quad v(y) = 0.562$$

Actual and Approximate Posterior Percentiles

α	2.5	25	50	75	97.5
α th percentile of $P(\theta\|y)$	3.10	4.07	4.52	5.00	6.06
α th percentile of $N(\mu(y),\ v(y))$	3.08	4.05	4.55	5.06	6.02

Credible Sets XIV

- The approximate 95% HPD credible set is $C = [3.08,\ 6.02]$, which is very close to the actual 95% HPD credible set $[3.10,\ 6.06]$. This approximated C has actual posterior probability 0.948 which is extremly close to 0.95.
- The approximate 90% HPD credible set is $[3.32,\ 5.78]$ and has actual posterior probability of 0.906.

For multivariate θ, the definition of a HPD credible set remains the same.

Bayesian hypothesis Testing I

- Classical approach bases accept/reject decision on $p - value = p\{T(Y)$ more "extreme" than $T(y_{obs})|\theta, H_0\}$, where "extremeness" is in the direction of H_1.
- Several *troubles* with this approach:
  - hypotheses must be nested
  - p-value can only offer evidence against the null
  - p-value is not the "probability that H_0 is true" (but is often erroneously interpreted this way)
  - As a result of the dependence on "more extreme" $T(Y)$ values, two experiments with different designs but identical likelihoods could result in different p-values, violating the Likelihood Principle!

Bayesian hypothesis Testing II

The posterior probabilities : $\alpha_0 = P(H_0|y)$ and $\alpha_1 = P(H_1|y)$
The prior probabilities : π_0, π_1

Definition

- The ratio $\frac{\alpha_0}{\alpha_1}$ is called the posterior odds of H_0 and H_1.
- The ratio $\frac{\pi_0}{\pi_1}$ is called the prior odds.
- The quantity $B = \dfrac{\text{posterior odds}}{\text{prior odds}} = \frac{\frac{\alpha_0}{\alpha_1}}{\frac{\pi_0}{\pi_1}} = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0}$ is called the **Bayes factor** in favor of H_0.

Bayesian hypothesis Testing III

The hypothesis are simple, i.e. $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$:

$$
\begin{aligned}
\alpha_0 &= P(\theta_0|y) = \frac{\pi_0 P(y|\theta_0)}{\pi_0 P(y|\theta_0) + \pi_1 P(y|\theta_1)} \\
\alpha_1 &= P(\theta_1|y) = \frac{\pi_1 P(y|\theta_1)}{\pi_0 P(y|\theta_0) + \pi_1 P(y|\theta_1)} \\
\Rightarrow \qquad & \frac{\alpha_0}{\alpha_1} = \frac{\pi_0 P(y|\theta_0)}{\pi_1 P(y|\theta_1)} \\
& B = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = \frac{P(y|\theta_0)}{P(y|\theta_1)},
\end{aligned}
$$

B is just the most powerful test of H_0 to H_1.

Bayesian hypothesis Testing IV

In general, B will depend on the prior input. Suppose the prior as

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

where g_0 and g_1 are (proper) densities which describe how the prior mass is spread out over the two hypotheses, and $\pi_0 + \pi_1 = 1$.

$$\begin{aligned} B &= \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = \frac{\int_{\Theta_0} P(y|\theta) g_0(\theta) d\theta}{\int_{\Theta_1} P(y|\theta) g_1(\theta) d\theta} \\ &= \frac{\text{Marginal pdf of } y \text{ under } H_0}{\text{Marginal pdf of } y \text{ under } H_1} \end{aligned}$$

Bayesian hypothesis Testing V

B is the ratio of weighted by g_0 and g_1 likelihoods of $Ⓗ_0$ to$Ⓗ_1$.

The Bayes factor evaluate the evidence *in favor* of the null hypothesis, which is familiar to classical significance tests. Suggested interpretation of the Bayes factor is provided by Kass and Raftery (1995); see the following table.

Bayesian hypothesis Testing VI

Table: Bayes factor Calibration

Bayes factor	Strenght of evidence in favor of H_0
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 32	Strong
32 to 100	Very strong
> 100	Decisive
	Strenght of evidence against H_0
$1/3.2$ to 1	Not worth more than a bare mention
$1/10$ to $1/3.2$	Substantial
$1/32$ to $1/10$	Strong
$1/100$ to $1/32$	Very strong
$< 1/100$	Decisive

Bayesian hypothesis Testing VII

Using Bayes factors, we can

- Evaluate the evidence in favor of the null hypothesis
- Compare two or more non-nested models
- Draw inferences without ignoring model uncertainty
- Determine which set of explanatory variables gives better predictive results

Bayesian hypothesis Testing VIII

Example 1: Diagnostic testing (Spiegelhalter et al., 2004)

- A new HIV test is claimed to have 95% sensitivity and 98% specificity, and is used in population with an HIV prevalence of $1/1000$. Tabulate expected status of 100,000 individuals in that population who are tested:

	HIV−	HIV+	Marginal
Test−	97,902	5	97,907
Test+	1,998	95	2,093
Marginal	99,900	100	100,000

  Thus of the 2,093 who have test positive, only 95 are truely HIV positive, giving a "predictive value positive" of $95/2093 = 4.5\%$.

Bayesian hypothesis Testing IX

- Bayesian language: Let H_0 be hypothesis that individual is truly HIV positive. Let y be observation that an individual tests positive. The disease prevalence is the prior probability $p(H_0) = 0.001$. We are interested in the probability that someone who tests positive is truly HIV positive=posterior probability $p(H_0|y)$. Let H_1 be the hypothesis of truly HIV negative. Then, 95% sensitivity means $p(y|H_0) = 0.95$, and 98% specificity means $p(y|H_1) = 0.02$.
  - Prior odds$= 1/999$; Posterior odds$= 95/1998$; Bayes factor$= 0.95/0.02$.
  - These odds correspond to $p(H_0|y) = 95/(95 + 1998) = 0.045$.

Bayesian hypothesis Testing X

Example 2: Consumer preference data
Suppose 16 taste testers compare two types of ground beef patty (one stored in a deep freeze, the other in a less expensive freezer). The chain is interested in whether storage in the higher-quality freezer translates into a "substantial improvement in taste."

Experiment: In a test kitchen, the patties are defrosted and prepared by a single chef/statistician, who randomizes the order in which the patties are served in double-blind fashion.

Result: 13 of the 16 testers state a preference for the more expensive patty.

Bayesian hypothesis Testing XI

Let

$$\theta = \text{probability consumers prefer more expensive patty.}$$

$$Y_i = \begin{cases} 1, & \text{if tester } i \text{ prefers more expensive patty;} \\ 0, & \text{otherwise.} \end{cases}$$

Assuming independent testers and constant θ, then if $X = \sum_{i=1}^{16} Y_i$, we have $X|\theta \sim \text{Bin}(16, \theta)$,

$$f(x|\theta) = \binom{16}{x} \theta^x (1-\theta)^{16-x}.$$

Bayesian hypothesis Testing XII

The beta distribution offers a conjugate family, since

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

The posterior is then Beta$(x+\alpha, 16-x+\beta)$.

Table: Posterior summaries

Prior distribution	Posterior quantile .025	.500	.975	$P(\theta > .6\|x)$
Beta(.5,.5)	.579	.806	.944	.964
Beta(1,1)	.566	.788	.932	.954
Beta(2,2)	.544	.758	.909	.930

Bayesian hypothesis Testing XIII

Suppose we define "*substantial* improvement in taste" as $\theta \geq 0.6$. Then under the uniform prior, the Bayes factor in favor of $M_1 : \theta \geq 0.6$ over $M_2 : \theta < 0.6$ is

$$BF_{21} = \frac{0.954/0.046}{0.4/0.6} = 31.1$$

or fairly strong evidence in favor of a substantial improvement in taste.

Bayesian hypothesis Testing XIV

Example 3: The IQ Example
The child taking the IQ test is to be classified as having below average IQ(< 100) or above average IQ(> 100).
Formally, we test $H_0 : \theta \leq 100$ vs $H_1 : \theta > 100$

$$y|\theta \sim N(\theta,\ 100),\ \ \theta \sim N(100,\ 225)\ \Rightarrow\ \theta|y = 115 \sim N(110.39,\ 69.23)$$

$\Rightarrow\ \alpha_0 = P(\theta \leq 100|y) = 0.106$ and $\alpha_1 = P(\theta > 100|y) = 0.894$
Hence, $\frac{\alpha_0}{\alpha_1} = 0.1185$.
Since the prior is $N(100,\ 225)$, $\pi_0 = P(\theta \leq 100) = \frac{1}{2}$,
$\pi_1 = P(\theta > 100) = \frac{1}{2}$
Hence, $B = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0} = 0.1185$.
See R code.

Bayesian hypothesis Testing XV

Example 3-1: Two-Sided Tests

- Two-sided tests about θ have the form:

$$H_0 : \theta = c \quad vs \quad H_1 : \theta \neq c$$

  for some constant c.

- We cannot test this using a continuous prior on θ, because that would result in a prior probability $P(\theta \in \Theta_0) = 0$ and thus a posterior probability $P(\theta \in \Theta_0 \mid y) = 0$ for **any** data set y.

Bayesian hypothesis Testing XVI

- We could place a prior probability mass on the point $\theta = c$. but many Bayesians are uncomfortable with this since the value of this point mass impossible to judge and is likely to greatly affect the posterior.
- **One solution**: Pick a small value $\epsilon > 0$ such that if θ is within ϵ of c, it is considered "practically indistinguishable" from c.
- Then let $\Theta_0 = (c - \epsilon, c + \epsilon)$ and find the posterior probability that $\theta \in \Theta_0$.
- In IQ example, $H_0 : \theta = 100$ vs $H_1 : \theta \neq 100$. Letting $\epsilon = 0.01$, then $\Theta_0 = (99.99, 100.01)$ and

$$p(\theta \in \Theta_0|y) = \int_{99.01}^{100.01} p(\theta|y)d\theta = \int_{-1.250}^{-1.248} \phi(z)dz = 0.00037.$$

Bayesian hypothesis Testing XVII

- **Another solution** (mimicking classical approach): Derive a $100(1-\alpha)\%$ (two-sided) HPD credible interval for θ. Reject $H_0 : \theta = c$ "at level α" if and only if c falls outside this credible interval.
- **Note**: Bayesian **decision theory** attempts to specify the **cost** of a wrong decision to conclude H_0 or H_1 through a **loss function**.
- We might evaluate the **Bayes risk** of some decision rule, i.e., its expected loss with respect to the posterior distribution of θ.

Bayesian hypothesis Testing XVIII

Bayes Factors and Model Uncertainty (Adrian Raftery, Biometrika, 1996, 251-268)

- Select the model with the largest posterior probability, for $k = 1, 2$,

  $$p(M_k|y) = p(y|M_k)p(M_k)/p(y),$$

  where $P(y|M_k) = \int P(y|\theta_k, M_k)P(\theta_k|M_k)d\theta_k$ and $P(\theta_k|M_k)$ is the prior density under M_k.
- The Bayes Factor B_{12} for Model M_1 against M_2 is given by

  $$B_{12} = \frac{p(M_1|y)/p(M_2|y)}{p(M_1)/p(M_2)} = \frac{P(y|M_1)}{P(y|M_2)}.$$

  i.e., the likelihood ratio if both hypotheses are simple.
- Problem: If $p(\theta_k|M_k)$ is improper, then $p(y|M_k)$ necessarily is as well $=>$ BF is not well-defined.

Example 4: Comparing Two Means

Comparing two means using 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:

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

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

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

$$H_0 : \mu_1 = \mu_2 \;\; vs \;\; H_1 : \mu_1 \neq \mu_2.$$

Bayesian hypothesis Testing XX

- 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^2_Δ.

- Consider the classical two-sample $t-$statistic

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

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

- H_0 and H_1 define two specific models for the distribution of T.

Bayesian hypothesis Testing XXI

- Under H_0, $T \sim$ (central) $t-$ distribution with $n_1 + n_2 - 2$ degrees of freedom.
- Under H_1, $T \sim$ (noncentral) $t-$ distribution with $n_1 + n_2 - 2$ degrees of freedom.
- With this prior, the Bayes factor for H_0 over H_1 is:

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

  where t^* is the observed $t-$ statistic.
- See R code to get BF and $p(H_0|x)$.

Bayesian hypothesis Testing XXII

Example 5:
Under M_1,

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon.$$

Under M_0,

$$y = X_1\beta_1 + \epsilon.$$

Let $\theta_1 = (\beta_1^T, \beta_2^T)^T$ and $\theta_0 = \beta_1$.

Bayesian hypothesis Testing XXIII

Example 6:
Under M_1,

$$P(y|\theta_1) = \frac{1}{\pi(1 + (y - \theta_1)^2)}.$$

Under M_0,

$$P(y|\theta_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - \theta_0)^2}.$$