03/03/22

- Find :

14	15	16	17	18
M	T	W	Th	F

- HW#3 - Solution
- HW#4

$H_0: \theta = \theta_0$

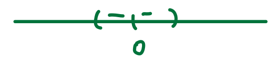
† Testing a point null hypothesis

- For continuous $\pi(\theta)$, $P^\pi(\theta = \theta_0) = 0$ & $P^\pi(\theta = \theta_0 \mid x) = 0$

- $H_0 : \theta = \theta_0$ will virtually never be the case that one seriously entertains the possibility that $\theta = \theta_0$ *exactly*.

  $\Rightarrow$ $H_0 : \theta = \theta_0$ *unrealistic!*

$\theta=0$



- More reasonable that

  ⋆⋆ Define $H_0 : \theta \in \Theta_0 = (\theta_0 - \epsilon, \theta_0 + \epsilon)$

  ⋆⋆ Choose some constant $\epsilon > 0$ such that all θ in Θ_0 can be considered "indistinguishable" from θ_0.

  ⋆⋆ *indistinguishable?* For any $\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$, the observed likelihood function $f(x \mid \theta)$ is approximately constant.

† Testing a point null hypothesis (contd)

- Consider $H_0 : \theta = \theta_0$ vs $H_0 : \theta \neq \theta_0$.
- Consider the following prior

$$\theta \begin{cases} = \theta_0 & \text{with probability } \rho_0, \\ \sim g_1(\theta) & \text{with probability } \rho_1 = 1 - \rho_0, \end{cases}$$

where probability distribution $g_1(\theta)$ gives probability zero to the event $\theta = \theta_0$.

⇒ We rewrite

$$\pi(\theta) = \rho_0 \delta_{\theta_0} + (1 - \rho_0) g_1(\theta),$$

where δ_{θ_0} is the Dirac mass at θ_0.

$$= \frac{f(x \mid \theta_0)\,\pi(\theta = \theta_0)}{m(x)}$$

- Let's find $\pi(\theta = \theta_0 \mid x)$.

★★ *Step 1* The marginal distribution for X is

$$\pi(\theta) = \rho_0 \delta(\theta_0) + (1-\rho_0)\cdot g_1(\theta)$$

$$\begin{aligned} m(x) &= \int f(x \mid \theta)\pi(\theta)d\theta \\ &= \rho_0 f(x \mid \theta_0) + (1-\rho_0)\underbrace{\int f(x \mid \theta)g_1(\theta)d\theta}_{= m_1(x)} \\ &= \rho_0 f(x \mid \theta_0) + (1-\rho_0)m_1(x). \end{aligned}$$

- Let's find $\pi(\theta = \theta_0 \mid x)$ (contd).

  ⋆⋆ *Step 2* The posterior probability of $\theta = \theta_0$ is

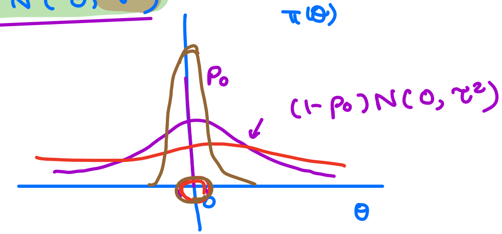
$$
\begin{aligned}
\pi(\theta = \theta_0 \mid x) &= \frac{\rho_0 f(x \mid \theta_0)}{m(x)} \\
&= \frac{\rho_0 f(x \mid \theta_0)}{\rho_0 f(x \mid \theta_0) + (1 - \rho_0) m_1(x)} \\
&= \left\{ 1 + \frac{1 - \rho_0}{\rho_0} \frac{m_1(x)}{f(x \mid \theta_0)} \right\}^{-1}.
\end{aligned}
$$

- **Example 5.2.8** (Example 5.2.4 continued) Consider the test of $H_0 : \theta = 0$. It seems reasonable to choose $\overset{g_1}{\cancel{\pi_1}}$ as $\underline{N(\mu, \tau^2)}$ and $\mu = 0$, if no additional information is available. Find the posterior probability, $\underline{\pi(\theta = 0 \mid x)}$.

$X \sim N(\theta, \sigma^2)$

$H_0 \longrightarrow \quad \theta = 0 \quad$ w/p $\; p_0$

$H_1 \longrightarrow \quad \theta \sim N(\underbrace{\mu}_{=0}, \tau^2) \quad$ w/p $\; p_1 = 1 - p_0$



$$\pi(\theta = 0 \mid x) = \left\{ 1 + \frac{1-p_0}{p_0} \, \frac{m_1(x)}{f(x \mid \theta_0)} \right\}^{-1}$$

$m_1(x) = N(0, \sigma^2 + \tau^2)$

$f(x \mid \theta = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}$

$$= \left\{ 1 + \frac{(1-p_0)}{p_0} \cdot \frac{\frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \, e^{-x^2/2(\sigma^2+\tau^2)}}{\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-x^2/2\sigma^2}} \right\}^{-1}$$

$$= \left\{ 1 + \frac{1-p_0}{p_0} \cdot \sqrt{\frac{\sigma^2}{\tau^2+\sigma^2}} \cdot \exp\left(\frac{\tau^2 x^2}{2(\sigma^2+\tau^2)} \right) \right\}^{-1}$$

- **Example 5.2.8** (contd)

⋆⋆ $\rho_0 = 1/2$ and $\tau = \sigma$

Table 5.2.2. *Posterior probabilities of $\theta = 0$ for different values of $z = x/\sigma$ and for $\tau = \sigma$.*

α $\quad$ $\alpha = 0.025$

z	0	0.68	1.28	1.96
$\pi(\theta = 0 \mid z)$	0.586	0.557	0.484	0.351

⋆⋆ $\rho_0 = 1/2$ and $\tau^2 = 10\sigma^2$ (more diffuse prior)

Table 5.2.3. *Posterior probabilities of $\theta = 0$ for $\tau^2 = 10\sigma^2$ and $z = x/\sigma$.*

z	0	0.68	1.28	1.96
$\pi(\theta = 0 \mid x)$	0.768	0.729	0.612	0.366

- **(JB p151)** Let's change the example a bit. Now we have $x_i \mid \theta \overset{iid}{\sim} N(\theta, \sigma^2)$, $i = 1, \ldots, n$ and let $\sigma = \tau$. Show the posterior probability on the null hypothesis is shown below to be given by

$$\pi(\theta = 0 \mid \bar{x}) = \frac{1}{1 + \frac{1}{\sqrt{n+1}} \exp\left\{ \frac{g^2}{2(1+1/n)} \right\}},$$

where $g = \frac{\sqrt{n}|\bar{x}|}{\sigma}$.

$H_0: \theta = 0$

$\bar{x} \mid \theta \sim N(\theta, \sigma^2/n)$

⋆⋆ Find the p-value.

$$\text{p-value} = \Pr(\underbrace{\bar{x} \geq |\bar{x}|}_{=z}) \quad \text{observed value}$$

$$= \Pr\left(\frac{\bar{x} - \theta}{\sigma/\sqrt{n}} \geq \frac{|\bar{x} - 0|}{\sigma/\sqrt{n}} \right)$$

$$= \Pr(z \geq |g|)$$

- Values of the Posterior Probabilities $P(H_0 \mid x)$.

p-value	g	$n = 1$	$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 1000$
0.100	1.645	0.42	0.44	0.47	0.56	0.65	0.72	0.89
0.050	1.960	0.35	0.33	0.37	0.42	0.52	0.60	0.82
0.010	2.576	0.21	0.13	0.14	0.16	0.22	0.27	0.53
0.001	3.291	0.086	0.026	0.024	0.026	0.034	0.045	0.124

- Observe when $g = 1.96$ for $n = 50$

⋆⋆ The frequentist researcher could reject H_0 at $p = 0.05$

⋆⋆ The Bayesian hypothesis tester, the evidence against the null hypothesis is weaker (little or no evidence against H_0).

⋆⋆ (but keep in mind) p-values are not a posterior probability of a hypothesis! For more, read JB 4.3.3.

- Testing with a noninformative prior

Example 5.2.9 Consider $x \sim \mathrm{N}(\theta, 1)$ and test $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$. For the diffuse distribution, $\pi(\theta) = 1$, find $P(\Theta_0 \mid x)$.

$$\Rightarrow \quad \pi(\theta \mid x) = N(x, 1)$$

$$\Rightarrow \quad Pr(H_0 \mid x) = Pr(\theta \leq 0 \mid x)$$

$$= \Phi(-x)$$

the same as the p-value under the classical procedure.

- **Example 5.2.8** (contd, Section 5.2.5) Assume $x \sim \mathrm{N}(\theta, 1)$. Consider the test of $H_0 : \theta = 0$ to test against $H_1 : \theta \neq 0$. To express vague prior information, assume the improper prior $\pi(\theta) = c$ on $\{\theta \neq 0\}$.

$$\pi(\theta = 0 \mid x) = \frac{p_0\, f(x \mid 0)}{m(x)} = \frac{p_0 \cdot f(x \mid 0)}{p_0 \cdot f(x \mid 0) + (1 - p_0) \cdot \underbrace{\int_{\mathbb{R}} c \cdot f(x \mid \theta)\, d\theta}_{= c}}$$

$$= \frac{\cancel{p_0} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{\cancel{p_0} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + (1 - \cancel{p_0})\, c} \quad \leftarrow \text{assuming } p_0 = \tfrac{1}{2}$$

$$= \frac{1}{1 + \sqrt{2\pi} \cdot c \cdot e^{x^2/2}}$$

- **Example 5.2.8** (contd)

★★ $\pi(\theta = \theta_0 \mid x)$ for the Jeffreys prior $\pi(\theta) = 1$

Table 5.2.5. *Posterior probabilities of $\theta = 0$ for the Jeffreys prior $\pi(\theta) = 1$.*

x	0.0	1.0	1.65	1.96	2.58
$\pi(\theta = 0 \mid x)$	0.285	0.195	0.089	0.055	0.014

★★ $\pi(\theta = \theta_0 \mid x)$ for the Jeffreys prior $\pi(\theta) = 10$

Table 5.2.6. *Posterior probabilities of $\theta = 0$ for the Jeffreys prior $\pi(\theta) = 10$.*

x	0.0	1.0	1.65	1.96	2.58
$\pi(\theta = 0 \mid x)$	0.0384	0.0236	0.0101	0.00581	0.00143

• **Example 5.2.8**(contd) Another illustration of the delicate issue of improper priors in testing setting.

→ $\theta \sim N(\mu, \tau^2) = g_1$

as $\tau^2 \to \infty$, $\pi(\theta) \propto c$

$$\begin{aligned}\pi(\theta = \theta_0 \mid x) &= \left\{1 + \frac{\rho_1}{\rho_0}\frac{m_1(x)}{f(x \mid \theta_0)}\right\}^{-1} \\ &= \left\{1 + \frac{\rho_1}{\rho_0}\sqrt{\frac{\sigma^2}{\sigma^2+\tau^2}}\exp\left(\frac{\tau^2 x^2}{2\sigma^2(\sigma^2+\tau^2)}\right)\right\}^{-1}.\end{aligned}$$

⋆⋆ For every x, as $\tau^2 \to \infty$, $\pi(\theta = \theta_0 \mid x) \to 1$.

⋆⋆ Compare to $\pi(\theta = \theta_0 \mid x)$ with the improper prior $\pi(\theta) = 1$ on $\{\theta \neq 0\}$,

$$\pi(\theta = 0 \mid x) = \frac{1}{1 + \sqrt{2\pi}\exp(x^2/2)}$$

⋆⋆ i.e., limiting arguments are not valid in the testing settings and prevent an alternative derivation of noninformative answers.

$H_0 \in p_0$ $\quad \theta \sim g_0$

$H_1 \in (1-p_0)$ $\quad \theta \sim g_1$

† Testing with noninformative priors. (contd)

- In many (**not all**) one-sided testing situations (& estimation situations), vague prior information tends to result in posterior probabilities that are similar to p-values.

- Improper priors should not be used at all in tests – DeGroot, 1973

- A testing problem cannot be treated in a coherent way if no prior information is available.

- Read CR 5.2.5 very interesting things regarding testing with improper priors.

- How about Bayes factor when the prior is improper?

  ⋆⋆ Intrinsic Bayes factor and fractional Bayes factor.

  Both use some part of data to make the improper prior proper ("propertize" the improper prior) and proceed the posterior inference **as if** it were a regular proper prior for the remainder of the sample.

  ⋆⋆ CR 5.2.6 Pseudo-Bayes Factor.

STAT 206B
Chapter 7: Data Augmentation and Model Choice

Winter 2022

† Data Augmentation

- Data augmentation = adding auxiliary variables.

$$p_y(y \mid \theta) = \int p_{y|V}(y \mid \theta, V) p_V(V \mid \theta) dV$$

⋆⋆ Y is the variable of interest, but $p_y(y \mid \theta)$ not easy to sample from.

⋆⋆ V's are auxiliary variables that cannot be directly observed.

⋆⋆ $p_{y|V}(y \mid \theta, V)$ and $p_V(V \mid \theta)$ are easy to sample from.

- Gibbs sampler computations can often be simplified or convergence accelerated by data augmentation.

$X \sim N(\theta, V)$

$V \sim f$

- **Example 1** Scale mixtures of normal distributions

Suppose $p(y)$ is a t-distribution with d.f ν, location parameter μ and scale parameter σ^2,

$$p(y) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\sigma^2\pi}\Gamma(\nu/2)} \left\{1 + \frac{(y-\mu)^2}{\nu\sigma^2}\right\}^{-\frac{\nu+1}{2}}$$

⋆⋆ ~~We may directly simulate y from the marginal distribution~~.

⋆⋆ Alternatively, we utilize the hierarchical structure,

$$p(y) = \int_{\mathbb{R}^+} p_{y|V}(y \mid \mu, V) p_V(V \mid \sigma^2) dV,$$

where $p_{y|V}(y \mid V) = \mathrm{N}(\mu, V)$ and $p_V(V \mid \sigma^2) = \mathrm{IG}(\nu/2, \sigma^2\nu/2) = \text{Inv-}\chi^2(\nu, \sigma^2)$.

- **Example 1** (contd) Consider the following model;

fixed

$$
\begin{aligned}
y_i \mid \nu, \mu, \sigma &\overset{iid}{\sim} t(\nu, \mu, \sigma^2), i = 1, \ldots, n, \\
\pi(\mu, \sigma^2) &\propto 1/\sigma^2,
\end{aligned}
$$

where degrees of freedom ν is fixed.

⋆⋆ The joint posterior is

$$
p(\mu, \sigma^2 \mid y_1, \ldots, y_n) \propto \underbrace{\frac{1}{\sigma^2}}_{\pi(\theta)} \prod_{i=1}^{n} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma} \left\{1 + \frac{1}{\nu}\left(\frac{y_i - \mu}{\sigma}\right)^2\right\}^{-(\nu+1)/2}.
$$

$$
\sigma^{-n-2} \cdot \prod \left\{1 + \frac{1}{\nu} \cdots \left(\frac{y_i - \mu}{\sigma}\right)^2 \right\}^{-(\nu+1)/2}
$$

- **Example 1** (contd)

  ⋆⋆ Then the full conditionals are

$$\Rightarrow \quad p(\sigma^2 \mid \mu, y_1, \ldots, y_n) \propto (\sigma^2)^{-1-n/2} \prod_{i=1}^{n} \left\{ 1 + \frac{1}{\nu} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right\}^{-(\nu+1)/2}$$

$$\Rightarrow \quad p(\mu \mid \sigma^2, y_1, \ldots, y_n) \propto \prod_{i=1}^{n} \left\{ 1 + \frac{1}{\nu} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right\}^{-(\nu+1)/2}$$

$\Rightarrow$ Not convenient.

- **Example 1** (contd) We rewrite the model using the normal-scale mixture representation of a t-distribution;

$$
\begin{aligned}
y_i \mid \mu, V_i &\overset{indep}{\sim} \mathrm{N}(\mu, V_i),\ i = 1, \ldots, n,\\
V_i \mid \sigma^2 &\overset{iid}{\sim} \text{Inv-}\chi^2(\nu, \sigma^2),\\
\pi(\mu, \sigma^2) &\propto 1/\sigma^2,
\end{aligned}
$$

where ν is fixed.

$\{\mu, \sigma^2; V_1, \ldots, V_n\}$

⋆⋆ The joint posterior is

$$
\begin{aligned}
p(\mu, \sigma^2, V_i \mid y_1, \ldots, y_n) \propto\ & \overbrace{\frac{1}{\sigma^2}}^{\pi(\mu,\sigma^2)} \prod_{i=1}^{n} \overbrace{\frac{1}{\sqrt{2\pi V_i}} \exp\left\{-\frac{(y_i-\mu)^2}{2V_i}\right\}}^{= f(y_i \mid \mu, V_i)}\\
& \times \prod_{i=1}^{n} \underbrace{\frac{(\nu\sigma^2/2)^{\nu/2}}{\Gamma(\nu/2)} V_i^{-\nu/2} \exp\left(-\frac{\nu\sigma^2}{2V_i}\right)}_{p(V_i \mid \sigma^2)}.
\end{aligned}
$$

$$\pi(\mu \mid \text{——}) \propto \exp\left\{ - \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2 V_i} \right\}$$

$$\propto \exp\left[-\frac{1}{2} \left\{ \left(\sum_{i=1}^{n} \frac{1}{V_i} \right) \mu^2 - 2 \sum_{i=1}^{n} \frac{y_i}{V_i} \cdot \mu \right\} \right]$$

$$\Rightarrow \pi(\mu \mid V_i, y) = N\left(\left(\sum_{i=1}^{n} \frac{1}{V_i} \right)^{-1} \left(\sum_{i=1}^{n} \frac{y_i}{V_i} \right), \left(\sum_{i=1}^{n} \frac{1}{V_i} \right)^{-1} \right)$$

$$\pi(\sigma^2 \mid \text{——}) \propto (\sigma^2)^{-1 + n\nu/2} \exp\left(- \sigma^2 \cdot \sum_{i=1}^{n} \frac{\nu}{2V_i} \right)$$

$$\Rightarrow \pi(\sigma^2 \mid \text{——}) = Ga\left(\frac{n\nu}{2}, \sum_{i=1}^{n} \frac{\nu}{2V_i} \right)$$

$$\pi(V_i \mid \text{——}) \propto (V_i)^{-1/2 - \nu/2} \exp\left(- \frac{(y_i - \mu)^2}{2V_i} - \frac{\nu \sigma^2}{2 V_i} \right)$$

$$\Rightarrow \pi(V_i \mid \text{——}) = IG\left(\left(\frac{\nu}{2} + \frac{1}{2} \right), \frac{1}{2} \left((y_i - \mu)^2 + \nu \cdot \sigma^2 \right) \right)$$

$i = 1, \ldots, n$