STAT 4010 Bayesian Learning

Tutorial 6

Spring 2023

Cheuk Hin (Andy) CHENG (Email | Homepage)

Di SU (Email | Homepage)

1 Decision Theoretic Testing

Definition 1. *(Decision Theory) Consider testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta_1 \in \Theta_1 = \Theta \backslash \Theta_0$. Then, the parameter of interest is $\psi = \mathbb{1}(\theta \in \Theta_1)$ and the decision space is $\mathscr{D} = \{0, 1\}$.*

Definition 2. *(**P-value**) The p-value is defined as the probability that more or equally extreme result is observed under H_0. Note that the probability is a function of the data and thus is a statistic. Smaller p-value indicates stronger evidence against the null.*

Definition 3. *(Type-I and II error) Let the test be $\hat{\psi}(x)$. Then,*

- ***Type-I error** is $\alpha_0 = \Pr(\hat{\psi}(x) = 1 \mid \theta \in \Theta_0)$, i.e. probability of rejecting the Null wrongly. This is also known as **size**.*
- ***Type-II error** is $\alpha_1 = \Pr(\hat{\psi}(x) = 0 \mid \theta \in \Theta_1)$, i.e. probability of failing to reject the Null.*
- ***Power** is $\Pr(\hat{\psi}(x) = 1 \mid \theta \in \Theta_1)$, i.e. probability of rejecting the Null correctly. **Power = 1-Type-II error**.*

Theorem 1.1. *Consider the weighted 0-1 loss defined as*

$$L(\theta, \hat{\psi}) = a_0 \mathbb{1}(\psi < \hat{\psi}) + a_1 \mathbb{1}(\psi > \hat{\psi}),$$

where $a_0, a_1 \geq 0$ defined in above. Then the Bayes estimator is,

$$\hat{\psi}_\pi = \mathbb{1}(\hat{p}_0 < \frac{a_1}{a_1 + a_0}) = \mathbb{1}(\hat{p}_1 > \frac{a_0}{a_1 + a_0}),$$

where $\hat{p}_j = \Pr(\theta \in \Theta_j \mid x)$.

Proof of Theorem 1.1. The posterior loss is

$$\begin{aligned} L(\pi, \widehat{\psi} \mid x) &= \mathrm{E}\{L(\theta, \widehat{\psi}) \mid x\} \\ &= a_0 \mathrm{P}(\psi = 0 \mid x) \mathbb{1}(\widehat{\psi} = 1) + a_1 \mathrm{P}(\psi = 1 \mid x) \mathbb{1}(\widehat{\psi} = 0) \\ &= \begin{cases} a_0 \mathrm{P}(\psi = 0 \mid x) & \text{if } \widehat{\psi} = 1; \\ a_1 \mathrm{P}(\psi = 1 \mid x) & \text{if } \widehat{\psi} = 0. \end{cases} \end{aligned}$$

Hence, the minimizer satisfies

$$\widehat{\psi} = 1 \quad \Leftrightarrow \quad a_0 \mathrm{P}(\psi = 0 \mid x) < a_1 \mathrm{P}(\psi = 1 \mid x) \quad \Leftrightarrow \quad \mathrm{P}(\psi = 0 \mid x) < \frac{a_1}{a_0 + a_1}.$$

Thus, the result follows.

Remark 1.1.

- $\hat{p}_0$ acts as p-value. $\alpha = a_1/(a_1 + a_0)$ acts as p-value. Just 'act as', they may not be the same.
- The procedure of constructing the test is as followed.
  1. Specify the loss.
  2. Specify the model (prior and sampling distribution).
  3. Derive the posterior loss and bayes estimator that minimizes the posterior loss.

2 Bayes Factor

Definition 4. *Consider $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$. The **Bayes factor** is defined as*

$$\begin{aligned} B_{10} =& \frac{\Pr(\theta \in \Theta_1 \mid x)}{\Pr(\theta \in \Theta_0 \mid x)} \bigg/ \frac{\Pr(\theta \in \Theta_1)}{\Pr(\theta \in \Theta_0)} \\ =& \textit{Posterior odd} \bigg/ \textit{Prior odd} \\ =& \frac{\hat{p}_1}{\hat{p}_0} \bigg/ \frac{\varrho_1}{\varrho_0}, \end{aligned}$$

where $\varrho_j = Pr(\theta \in \Theta_j), j = 0, 1$.

Theorem 2.1. *Let $\theta \sim \pi(\theta)$ and $\varrho_j = \Pr(\theta \in \Theta_j) > 0$ for $j = 0, 1$. Then,*

$$B_{10} = \frac{\kappa_1(x)}{\kappa_0(x)},$$

where

$$\kappa_j(x) = \int_{\Theta_j} f(x \mid \theta)\pi_j(\theta)\mathrm{d}\theta, \quad \pi_j(\theta) = \frac{1}{\varrho_j}\pi(\theta)\mathbb{1}(\theta \in \Theta_j).$$

Remark 2.1. We reject the null based on the magnitude of B_{10}, which serves as evidence against H_0. When $B_{10} \geq 3.2, 10, 100$, the strength of evidence is substantial, strong, and conclusive, respectively.

Example 2.1. Consider $x_{1:n} \mid \theta \overset{\text{IID}}{\sim} \text{Laplace}(\theta)$, where $\theta > 0$. The Laplace distribution has density $(2\theta)^{-1}\exp\{-|x|/\theta\}$. We are interested in testing $H_0 : \theta \in [3, 5]$ against $H_0 : \theta \in [0, 3) \cup (5, \infty)$. In addition, we collected $n = 30$ observations and obtain $A_n = \sum_{i=1}^n |x_i| = 240$.

1. Find the conjugate prior for θ. Compute the posterior.
2. Consider the 0-1 loss with $\alpha = 5\%$. Derive and compute the bayes estimator.
3. Compute the Bayes factor B_{10}. We reject the null if $B_{10} > 10$. Compare and comment the conclusion to that of the bayes estimator you derived in the last part.

4. Consider a grid of θ from $[0.5, 10]$. Using simulation, plot the power curve for the bayes estimator and the test constructed by the Bayes factor. Comment.

5. Find a weakly informative prior for θ. Compute the Bayes factor and compare the result to part 3.

Solution:

1. Let $\eta = 1/\theta$. We can rewrite the density of the sampling distribution as

$$f(x \mid \eta) = 0.5 \exp\{\eta(-|x|) + \log \eta\}.$$

By theorem 2.3 in the lecture note, the conjugate prior for η is $Ga(\alpha)/\beta$ since

$$f(\eta) \propto \exp\{\beta\eta - \alpha\eta\} = \eta^{\alpha} \exp\{\beta\eta\}.$$

Therefore, the conjugate prior for θ is $\beta/Ga(\alpha)$. The posterior can be computed as a result

$$f(\theta \mid x_{1:n}) \propto f(\theta) f(x_{1:n} \mid \theta) = \theta^{-\alpha-n-1} \exp\{-1/\theta(\beta + A_n)\} \mathbb{1}(\theta > 0).$$

Therefore, $\theta \mid x_{1:n} \sim \beta_n/Ga(\alpha_n)$, where $\alpha_n = \alpha + n$ and $\beta_n = \beta + A_n$.

2. We continue by setting $\alpha = 2$ and $\beta = 4$ (as a result $\mathsf{E}[\theta] = 4$). By theorem 1.1, the bayes estimator is $\hat{\psi}_\pi = \mathbb{1}(\hat{p}_0 < \alpha)$, where

$$\hat{p}_0 = \Pr(\theta \in \Theta_0 \mid x_{1:n}) = \texttt{pinvgamma}(5, \alpha_n, \beta_n) - \texttt{pinvgamma}(3, \alpha_n, \beta_n) = 0.0043.$$

Thus, $\hat{\psi}_\pi = \mathbb{1}(\hat{p}_0 < \alpha) = 1$ and we reject the null. Note also the Bayesian p-value in this case is $\hat{p}_0 = 0.0043$.

3. Let $\varrho_0 = \Pr(\theta \in \Theta_0) = \texttt{pinvgamma}(5, \alpha, \beta) - \texttt{pinvgamma}(3, \alpha, \beta) = 0.1937$. Since $\Theta_1 = \Theta \setminus \Theta_0$, we have

$$B_{10} = \frac{\Pr(\theta \in \Theta_1 \mid x)}{\Pr(\theta \in \Theta_0 \mid x)} \bigg/ \frac{\Pr(\theta \in \Theta_1)}{\Pr(\theta \in \Theta_0)} = \frac{1 - \hat{p}_0}{\hat{p}_0} \bigg/ \frac{1 - \varrho_0}{\varrho_0} = 55.1068.$$

Therefore, we have strong evidence against the null and decide to reject the null. The conclusion is in line with the last part.

4. Let ψ_1 and ψ_2 be the Bayes estimator and the test based on B_{10} respectively. The test ψ_2 rejects the null when $B_{10} > 10$. The power curve is a curve that takes power on the y-axis and θ on the x-axis. The following procedure is used to obtain the power curve.

   (a) Create a grid of θ.

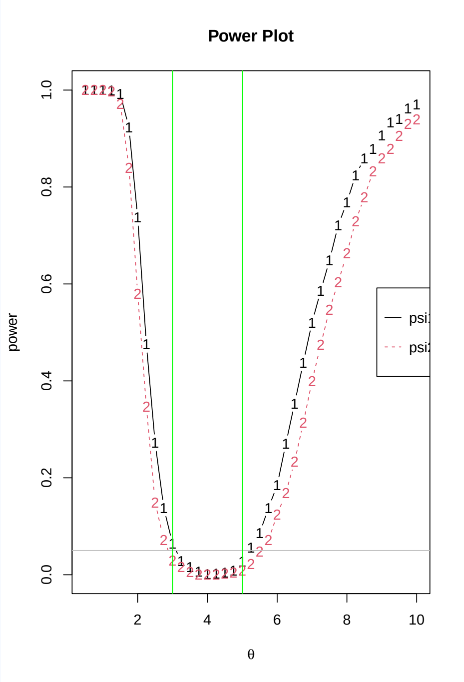
   (b) Fix a θ from the grid. Simulate $x_{1:n} \mid \theta$. Compute $\psi_1^{(j)}$ and $\psi_2^{(j)}$ where j denotes the j-th iteration.

   (c) Repeat step 2 for $j = 1, \dots, \texttt{nRep}$. Estimate the power by Monte Carlo

$$\widehat{\Pr}(\psi_p = 1 \mid \theta \in \Theta_1) = \frac{\sum_j \mathbb{1}(\psi_p^{(j)} = 1)}{\texttt{nRep}}.$$

(d) Repeat step (b) and (c) for all the θ in the grid.

(e) Plot the powers vs θ.




```r
library(invgamma)
a = 2
b = 4
n = 30
An = 240
an = a+n
bn = b+ An
p0 = pinvgamma(5,an,bn) -pinvgamma(3,an,bn)
p0
[1] 0.004341378

##Bayes factor
q0 = pinvgamma(5,a,b) -pinvgamma(3,a,b)
q0
[1] 0.1937321
BF = ((1-p0)/p0)/((1-q0)/q0)
BF
[1] 55.1068

##Power curve
get_psi1 <- function(An,n,cri = 0.05,a = 2, b = 4){
  an = a+n
  bn = b+ An
  p0 = pinvgamma(5,an,bn) -pinvgamma(3,an,bn)
  return(list(p0 = p0, psi =p0< cri))
}

get_psi2 <- function(An,n,cri = 10,a = 2, b = 4){
  an = a+n
  bn = b+ An
  p0 = pinvgamma(5,an,bn) -pinvgamma(3,an,bn)
```

```
  q0 = pinvgamma(5,a,b) -pinvgamma(3,a,b)
  BF = ((1-p0)/p0)/((1-q0)/q0)
  return(list(BF = BF, psi =BF>cri))
}

theta_grid = seq(0.5,10,by = 0.25)
mgrid = length(theta_grid)
nRep = 2^10
power = array(NA,c(mgrid,2))
for (i in 1:mgrid) {
  theta = theta_grid[i]
  out = array(NA,c(nRep,2))
  for (j in 1:nRep) {
    set.seed(j)
    ##By representation if L~Laplace(theta), E1, E2~iid Exp(1)
    ##L = theta(E1 - E2)
    x = theta*(rexp(n,1) - rexp(n,1))
    An_sim = sum(abs(x))
    out[j,1] = get_psi1(An_sim,n)$psi
    out[j,2] = get_psi2(An_sim,n)$psi
  }
  power[i,] = apply(out,2,mean)
}
matplot(theta_grid,power,type = 'b',pch=c('1','2'),main = "Power Plot",
    col = 1:2,lty = 1:2,xlab = expression(theta))
legend('right',c('psi1','psi2'),col = 1:2,lty = 1:2)
abline(h = 0.05,col = 'grey')
abline(v = c(3,5),col = "green")
```

5. It is easy to see that θ is a scale parameter. Therefore, an invariant prior is $f(\theta) \propto 1/\theta$. Since it is impropoer, we regularize it by considering $f(\theta) \propto 1/\theta \mathbb{1}(\theta \in [l,u])$ where $l = 0.001$ and $u = 999$. It can be shown that $f(\theta) = c/\theta$ where $c = 1/\ln(u/l)$. As a result, we have

$$\begin{aligned}
\varrho_0 = \Pr(\theta \in \Theta_0) = \int_3^5 c/\theta \mathrm{d}\theta = \frac{\ln(5/3)}{\ln(u/l)};\\
\pi_0(\theta) &= \frac{1}{\varrho_0}\pi(\theta)\mathbb{1}(\theta \in \Theta_0);\\
\kappa_0(x) &= \int_{\Theta_0} f(x_{1:n} \mid \theta)\pi_0(\theta)\mathrm{d}\theta\\
&= \int_3^5 \frac{1}{2^n\theta^n}\exp\{-A_n/\theta\}\left(\frac{1}{\varrho_0}\right)\left(\frac{c}{\theta}\right)\mathrm{d}\theta\\
&= \frac{c}{2^n\varrho_0}\frac{\Gamma(n)}{A_n^n}\int_3^5 \frac{A_n^n}{\Gamma(n)}\theta^{-n-1}exp\{-A_n/\theta\}\mathrm{d}\theta\\
&= \frac{c\Gamma(n)}{2^n\varrho_0 A_n^n}\left[\mathtt{pinvgamma}(5,n,A_n) - \mathtt{pinvgamma}(3,n,A_n)\right]
\end{aligned}$$

Similarly, we also have

$$\begin{aligned}
\kappa_1(x) = \frac{c\Gamma(n)}{2^n(1-\varrho_0)A_n^n}&[\{\mathtt{pinvgamma}(3,n,A_n) - \mathtt{pinvgamma}(l,n,A_n)\}\\
&+\{\mathtt{pinvgamma}(u,n,A_n) - \mathtt{pinvgamma}(5,n,A_n)\}]
\end{aligned}$$

By theorem 2.1, we have

$$BF_{10} = \frac{\kappa_1(x)}{\kappa_0(x)} = 17.58956.$$

Therefore, we have strong evidence to reject the null.
As a remark, note that $\varrho_0/(1-\varrho_0) \to 0$ as $l \to 0$ and $u \to \infty$. Therefore, $BF_{10} \to 0$. If we consider the invariant prior, we always do not reject the null.

```
####part 5
##note Gamma(n), A_n and 2^n can be ignored as they cancel out each other
l = 0.001
u = 999
q0 = log(5/3)/log(u/l)
kappa1 = (pinvgamma(3,n,An) - pinvgamma(l,n,An) + pinvgamma(u,n,An) -
    pinvgamma(5,n,An))/(1-q0)
kappa0 = (pinvgamma(5,n,An) -pinvgamma(3,n,An))/q0
BF = kappa1/kappa0
BF
[1] 17.58956
```

2.1 Well-defined Bayes Factor

When testing simple hypotheses, if the underlying random variable is continuous, $\varrho_j = 0$ and definition 4 is not well-defined. This motivated the following modification.

Definition 5. *(Modification of prior and BF). Let the prior of θ be defined in two steps.*

(1) Let the prior probabilities of H_j be $\varrho_j = \mathrm{P}(\theta \in \Theta_j)$ for $j = 0, 1$ such that $\varrho_1 + \varrho_0 = 1$ and $\varrho_0, \varrho_1 > 0$.

(2) Let the prior of θ under $H_j : \theta \in \Theta_j$ be $\theta \sim \pi_j(\theta)$.

So, the (overall) implied prior of θ is

$$\pi(\theta) = \varrho_0\pi_0(\theta) + \varrho_1\pi_1(\theta). \tag{2.1}$$

Then the Bayes factor is given by

$$B_{10} = \frac{\pi_1(x)}{\pi_0(x)} \quad \text{where} \quad \pi_j(x) = \int_{\Theta_j} f(x \mid \theta)\pi_j(\theta)\mathrm{d}\theta, \quad j = 0, 1. \tag{2.2}$$

Remark 2.2. In equation 2.2, the "between-group" prior belief ϱ_0, ϱ_1 are eliminated. However, it still depends on the "within-group" prior belief which is reflected by $\pi_j(\theta), j = 0, 1$.

2.2 Relationship with Decision Theoretic Testing

Example 2.2. Let $H_0 : \theta \in \Theta_0$, and $H_1 : \theta \in \Theta_1$. The Bayes factor is a one-to-one transformation of the posterior probability $\hat{p}_0$. And the conclusion derived from Bayes factor is equivalent to that from posterior probability.

(1) $\widehat{p}_0 = \left(1 + \frac{\varrho_1}{\varrho_0} B_{10}\right)^{-1}$,

(2) If $\Theta_1 = \Theta \backslash \Theta_0$, then

$$\text{Reject } H_0 \Leftrightarrow \underbrace{\widehat{p}_0 < \frac{a_1}{a_0 + a_1}}_{\text{The Bayes test in (4.4)}} \quad \Leftrightarrow \quad \underbrace{B_{01} < \frac{a_1 \varrho_1}{a_0 \varrho_0}}_{\text{The BF test}}. \tag{2.3}$$

Example 2.3. Assume $\varrho_0 = 0.99, \varrho_1 = 0.01$ and $\hat{p}_0 = 0.9, \hat{p}_1 = 0.1$. Then we have $\hat{p}_0 > 1/2 > \hat{p}_1$. However,

$$B_{10} = \frac{0.1/0.9}{0.01/0.99} = 11,$$

suggesting a strong evidence against H_0.

Takeaway: The probability $\hat{p}_0$ represents the "exact" posterior belief on H_0. Bayes factor represents the "change" in the belief on H_0 after collecting data. Therefore, BF alleviates the prior preference.

3 Continuous Decision Space $\mathscr{D} = [0, 1]$

Theorem 3.1. *Consider $D = [0, 1]$. The Bayes estimators $\widehat{\psi}_0, \widehat{\psi}_1, \widehat{\psi}_2$ under L^0, L^1, L^2 losses are given by*

$$\widehat{\psi}_0 = \widehat{\psi}_1 = \mathbb{1}\left\{\mathrm{P}\left(\theta \in \Theta_0 \mid x\right) < 1/2\right\} \qquad \textit{and} \qquad \widehat{\psi}_2 = \mathrm{P}\left(\theta \in \Theta_1 \mid x\right).$$