STAT 200B 2019 Week10

Soyeon Ahn

1 Bayes estimator under specific loss functions

The Bayes rule $\hat{\theta}$ minimizes

$$r(\hat{\theta}|x) = \int L(\theta, \hat{\theta}(x)) f(\theta|x) d\theta$$

1.1 Sqaured error loss

$$L(\theta, a) = (\theta - a)^2$$

The Bayes estimator is the posterior mean.

Taking the derivative of

$$\int (\theta - a)^2 f(\theta|x) d\theta$$

with respect to a,

$$\int 2(\theta - \hat{\theta}) f(\theta|x) d\theta = 0$$

Then,

$$\hat{\theta} = \int \theta f(\theta|x) d\theta = E(\theta|X = x)$$

1.2 Absolute error loss

$$L(\theta, a) = |\theta - a|$$

The Bayes estimator is the posterior median.

(I follow the proof from Berger, Statistical Decision Theory and Bayesian Analysis, 2nd edition) Let m denote a median of $f(\theta|x)$, and let $a > m$ be another action.

$$L(\theta, m) - L(\theta, a) = \begin{cases} m - a & \text{if } \theta \leq m, \\ 2\theta - (m + a) & \text{if } m < \theta < a, \\ a - m & \text{if } \theta \geq m, \end{cases}$$

$$L(\theta, m) - L(\theta, a) \leq (m - a)I_{(-\infty,m)}(\theta) + (a - m)I_{(m,\infty)}(\theta).$$

Since $P(\theta \leq m|x) \geq \frac{1}{2}$ and $P(\theta > m|x) \leq \frac{1}{2}$,

$$\begin{aligned} E[L(\theta, m) - L(\theta, a)] &\leq (m - a)P(\theta \leq m|x) + (a - m)P(\theta > m|x) \\ &\leq (m - a)\frac{1}{2} + (a - m)\frac{1}{2} = 0 \end{aligned}$$

m has posterior expected loss at least small as a. Similar argument holds for $a < m$.

1.3 Zero-one loss

$$L(\theta, a) = \begin{cases} 0 & \text{if } a = \theta \\ 1 & \text{if } a \neq \theta \end{cases}$$

The Bayes estimator is the posterior mode. discrete case

$$1 - \int \mathbb{1}(\theta = a) f(\theta|x) d\theta = 1 - f(a|x)$$

continuous case, define $1 - \delta(a - \theta)$ (Dirac delta function) as loss function

$$\int (1 - \delta(a - \theta)) f(\theta|x) d\theta = 1 - f(a|x)$$

To minimize the posterior risk, take a posterior mode, modal value of $f(\theta|x)$ is most probable.

2 Solutions

2.1 Week 05

Example 1: Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, where σ^2 is known. Consider testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$, using rejection region

$$R = \{x_1, \ldots, x_n : |\bar{X}_n| > c\}$$

Find and plot $\beta(\mu)$. (HW05 4(b))

solution
$\Theta = \mathbb{R}$, $\Theta_0 = \{0\}$. The likelihood ratio statistic is

$$\lambda = 2 \log \left(\frac{\mathcal{L}_n(\hat{\mu})}{\sup_{\mu \in \Theta_0} \mathcal{L}_n(\mu)} \right) = 2 \log \left(\frac{\mathcal{L}_n(\hat{\mu})}{\mathcal{L}_n(0)} \right) = 2(\ell_n(\hat{\mu}) - \ell_n(0)),$$

where $\hat{\mu}$ is the MLE on Θ.

$$\lambda = 2(\ell_n(\hat{\mu}) - \ell_n(0)) = \frac{n(\bar{X}_n)^2}{2\sigma^2} \geq 0.$$

The rejection region is $\lambda > c$, with $c > 0$ to be determined by the level $\alpha = 0.05$. This is equivalent to $|\bar{X}_n| > d$, with $d > 0$ to be determined by $\alpha = 0.05$.

$$\begin{aligned}
0.05 &= P_{\mu=0}(|\bar{X}_n| > d) \\
&\text{When } \mu = 0, \bar{X}_n \sim N(0, \sigma^2/n) \\
&= P_{\mu=0}\left(\sqrt{n}\bar{X}_n/\sigma > \sqrt{n}d/\sigma\right) + P_{\mu=0}\left(\sqrt{n}\bar{X}_n/\sigma < -\sqrt{n}d/\sigma\right) \\
&= 1 - \Phi(\sqrt{n}d/\sigma) + \Phi(-\sqrt{n}d/\sigma) = 2(1 - \Phi(\sqrt{n}d/\sigma)).
\end{aligned}$$

$$\Rightarrow d = \frac{\Phi^{-1}(0.975)\sigma}{\sqrt{n}} = \frac{z_{0.025}\sigma}{\sqrt{n}}.$$

The power function of the test is

$$\begin{aligned}
\beta(\mu) &= P_\mu\left(|\bar{X}_n| > d\right) \\
&= P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} > \frac{\sqrt{n}(d-\mu)}{\sigma}\right) + P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} < -\frac{\sqrt{n}(d-\mu)}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{\sqrt{n}(d-\mu)}{\sigma}\right) + \Phi\left(-\frac{\sqrt{n}(d-\mu)}{\sigma}\right) \\
&= 1 - \Phi\left(\Phi^{-1}(0.975) - \frac{\sqrt{n}\mu}{\sigma}\right) + \Phi\left(-\Phi^{-1}(0.975) - \frac{\sqrt{n}\mu}{\sigma}\right) \\
&= 1 - \Phi\left(z_{0.025} - \frac{\sqrt{n}\mu}{\sigma}\right) + \Phi\left(-z_{0.025} - \frac{\sqrt{n}\mu}{\sigma}\right).
\end{aligned}$$

Example 2: Let $X \sim Bin(5, p)$. Consider testing $H_0 : p \leq 1/2$ versus $H_1 : p > 1/2$. Consider two different rejection regions:

$$\begin{aligned}
R_1 &= \{x : x = 5\} \\
R_2 &= \{x : x \geq 3\}
\end{aligned}$$

Plot and compare the corresponding power functions $\beta_1(p)$ and $\beta_2(p)$.

solution

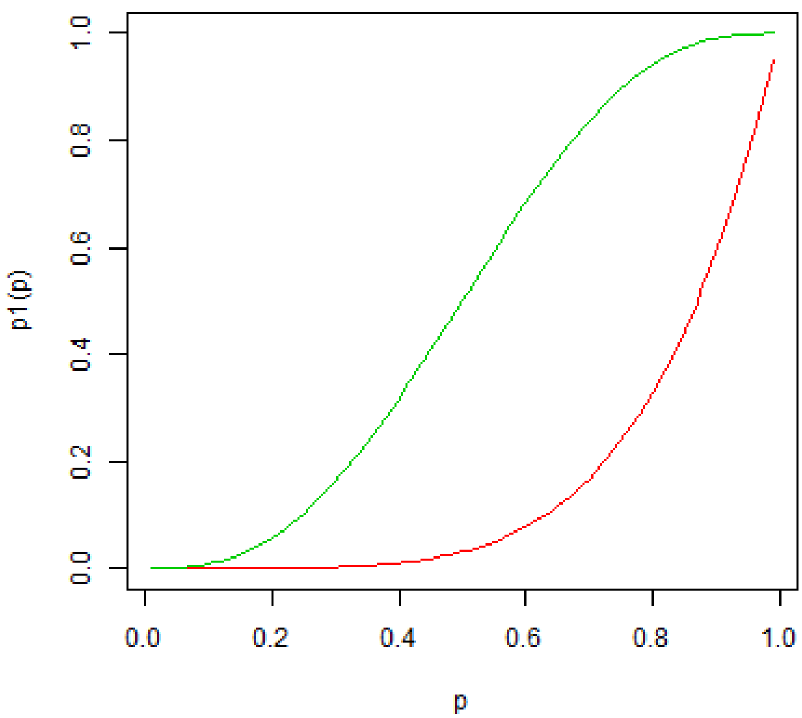
- $R_1 = \{x : x = 5\}$ We observe all success;

$$\beta_1(\theta) = P_\theta(X \in R_1) = P(X = 5) = p^5$$

- $R_1 = \{x : x \geq 3\}$ =

$$\beta_2(\theta) = P_\theta(X \in R_2) = P(X = 3, 4, \text{ or }, 5) = \binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p)^1 + \binom{5}{5}p^5$$

```
p <- seq(0.01, 0.99, 0.01)
p1 <- function(p){  1-pbinom(4, 5, p)}
p2 <- function(p){ 1-pbinom(2, 5, p)  }
plot(p, p1(p), ylim=c(0,1), type="l", col=2)
lines(p, p2(p), col=3)
```



Continuation of Example 2: Consider a rejection region of the form $R = \{x : x \geq c\}$.

- What values of c do we need to consider?
- For each of these, find the size of the corresponding test.
- What c should we choose if we want a probability of Type I error of no more than 10%?

solution Since possible x values are $\{0, 1, 2, 3, 4, 5\}$, we consider c from $\{0, 1, 2, 3, 4, 5\}$. The size is $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \leq \frac{1}{2}} P_\theta(X \geq c) = P_{\theta = \frac{1}{2}}(X \geq c)$

```
> 1-pbinom(0:4, 5, 0.5) # 1 - P(X<=c)
[1] 0.96875 0.81250 0.50000 0.18750 0.03125
```

if we want a probability of Type I error of no more than 10%, c should be 5.

Examples

- Consider again $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where σ^2 is known. Show that the size α Wald test for $H_0 : \mu = 0$ produces a rejection region as in Example 1 above. (Actually the size is exactly α in this case).
- Now suppose that σ^2 is unknown. Construct a size α Wald test for $H_0 : \mu = 0$.
- Suppose that $X_1 \sim Bin(m, p_1)$ and $X_2 \sim Bin(n, p_2)$. Construct a size α Wald test for $H_0 : p_1 = p_2$.
- Let $F(u, v)$ be the joint distribution of two r.v. U and V. Let $\theta = T(F) = \rho(U, V)$, where ρ denotes the correlation. Describe how to construct a size α Wald test for $H_0 : \rho = 0$ using the plug-in estimator and the bootstrap.

solution

- $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$.

  $$\hat{\mu}_n = \frac{\sum X_i}{n}$$

  is an estimator such that $(\hat{\mu}_n)/\widehat{se}(\hat{\mu}_n) \xrightarrow{D} N(0,1)$. The size α Wald test rejects H_0 when $T > z_{\alpha/2}$, where

  $$T = \left| \frac{\hat{\mu}_n}{\widehat{se}(\hat{\mu}_n)} \right| = \left| \frac{\bar{X}_n}{\sigma/\sqrt{n}} \right|$$

  .

  Thus, $\left|\bar{X}_n\right| \geq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ is a size α Wald test.
- A size α Wald test for unknown variance is $\left|\bar{X}_n\right| \geq z_{\alpha/2} \frac{s}{\sqrt{n}}$, where s the sample variance.
- The Wald statistic is

  $$W = \frac{\hat{p}_1 - \hat{p}_2}{\widehat{se}(\hat{p}_1 - \hat{p}_2)} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1(1-\hat{p}_1)/m + \hat{p}_2(1-\hat{p}_2)/n}}.$$

  where $\hat{p}_1 = \frac{X_1}{m}$ and $\hat{p}_2 = \frac{X_2}{n}$. The rejection region is $|W| > z_{\alpha/2}$.
- Use the sample correlation coefficient and estimate the standard error using the bootstrap.

2.2 Week 06

Examples: Let

$$T(X) = \frac{\sup_{\theta\in\Theta} \mathcal{L}_n(\theta)}{\sup_{\theta\in\Theta_0} \mathcal{L}_n(\theta)}$$

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Find $T(X)$ and find a simplified expression for the form of the rejection region. Use it to find the size α LRT.

solution (HW05 4(b))

2.3 Week 09

Examples

Example:

1. Find the Jeffreys prior for λ when $X_1, \ldots, X_n \overset{iid}{\sim} Pois(\lambda)$.
2. Is the Jeffreys prior proper?
3. Find the implied prior distribution for $\phi = \log\lambda$.
4. Show the prior in 3 is the same as the Jeffreys prior for ϕ.

solution

1. $f(x;\lambda) \sim \lambda^x \exp^{-\lambda}$
$$f(\lambda) \propto \frac{1}{\sqrt{\lambda}} \sim \lambda^{\frac{1}{2}-1}$$
2. The Jeffreys prior is improper
3. $\phi = \log\lambda = g(\lambda)$,
$$\begin{aligned} I(\lambda) &\propto -E\left[\frac{\partial^2 \log f(x;\lambda)}{\partial\lambda^2}\right] \\ &= E\left[\frac{\partial^2 \log f(x;\lambda = g^{-1}(\phi))}{\partial\phi^2}\left|\frac{\partial\phi}{\partial\lambda}\right|^2\right] \\ &= I(\phi)\left|\frac{1}{\lambda}\right|^2 \end{aligned}$$

$$\sqrt{I(\phi)} = \sqrt{I(\lambda)} \times \lambda = \sqrt{\lambda} = \sqrt{\exp\phi}$$

4. Show the prior in 3 is the same as the Jeffreys prior for ϕ.

$$f_\phi(\phi) = f_\lambda(g^{-1}(\phi)) \left| \frac{dg^{-1}(\phi)}{d\phi} \right| = \frac{1}{\sqrt{\exp(\phi)}} \exp(\phi) = \sqrt{\exp(\phi)}$$

Computing the Bayes Factor:

$$\frac{P(H_i|x^n)}{P(H_j|x^n)} = \frac{f(x^n|H_i)}{f(x^n|H_j)} \times \frac{P(H_i)}{P(H_j)}$$

which is Posterior odds = Prior odds $\times$ Bayes factor. $p = P(H_1)$ and $p^* = P(H_1|Data)$,

$$\text{(LHS)} = \frac{p^*}{1-p^*}$$

$$\text{(RHS)} = BF_{10} \times \frac{p}{1-p}$$

thus

$$p^* = \frac{\frac{p}{1-p} BF_{10}}{1 + \frac{p}{1-p} BF_{10}}$$

The posterior probability is related to the prior probability of H_1. If $BF_{10} = 1$, the posterior probability is the same as the prior probability. If the Bayes Factor (BF_{10}) is greater than 1, the posterior probability of H_1 will be increased. BF is a summary of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another.

(from Bayes factors. Kass, Robert E; Raftery, Adrian E. Journal of the American Statistical Association; 1995)

- From Jeffreys' Bayesian viewpoint, the purpose of hypothesis testing is to evaluate the evidence in favor of a scientific theory.
- Bayes factors offer a way of evaluating evidence in favor of a null hypothesis.
- Bayes factors provide a way of incorporating external information into the evaluation of evidence about a hypothesis.
- Bayes factors are very general and do not require alternative models to be nested.

Example: Suppose $X \sim N(\theta, 1)$ and we are estimating θ under squared error loss. Consider $\hat{\theta}_c(x) = cx$.

- Calculate the risk in terms of c and θ.

- Calculate the risk when $c = 1$.
- Show that $\hat{\theta}_c$ is inadmissible when $c > 1$.
- Make a plot comparing the risk when $c = 1/2$ and $c = 1$.

solution

- $E_\theta[(\theta - cX + c\theta - c\theta)^2] = E_\theta[(\theta(1-c) - c(X-\theta))^2] = \theta^2(1-c)^2 + c^2$.
- When $c = 1$, the risk is 1
- When $c > 1$, cX is dominated by X, which has risk c^2.
- Make a plot comparing the risk when $c = 1/2$ and $c = 1$.

```
theta1 <- seq(-2,2,by=0.1)
plot(theta1, (theta1^2 + 1)*0.25, type="l",ylab="Risk",ylim=c(0,6))
abline(h=1, col=2)
lines(theta1, (theta1^2 + 4), col=3)
legend(-1.5,3, c("c=0.5","c=1","c=2"),col=1:3,lwd=1)
```