BIOS 7731 HW 6

Tim Vigers

10/21/2020

BD 3.5.11

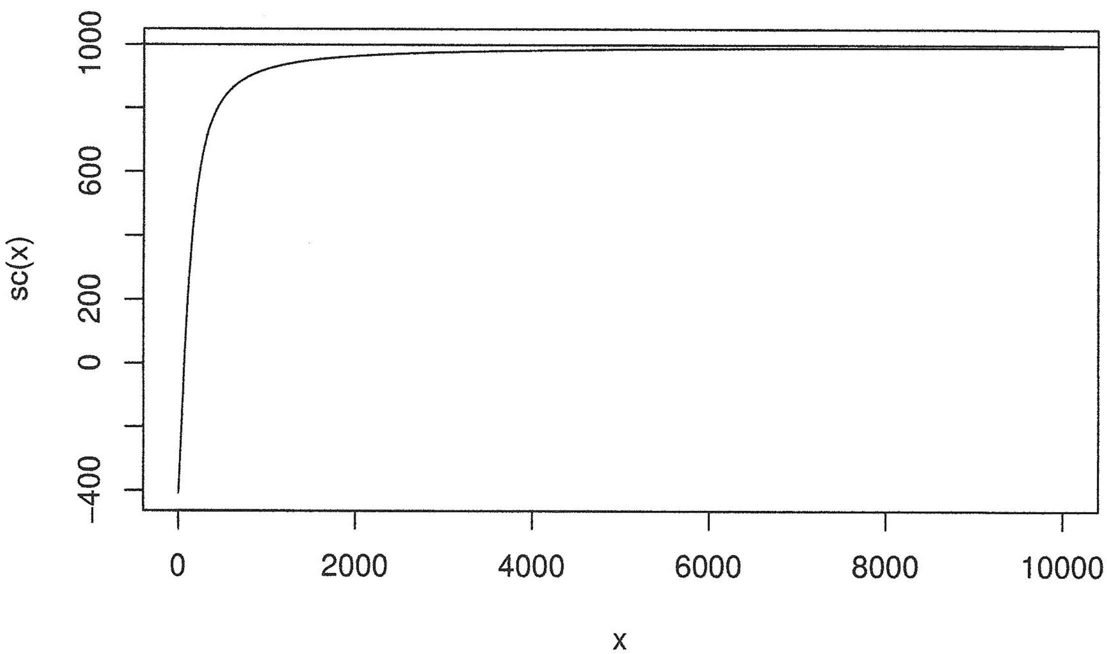
If we set $\mu_0 = 0$ and the ideal sample mean of $x_1, ..., x_{n-1}$, $\bar{X}_{n-1} = 0$, then the sensitivity curve of $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n-1}}}$ simplifies to:

$$sc(x) = n(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sqrt{\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}}} - 0) = n[\frac{\sqrt{n}(\bar{X})}{\sqrt{\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}}}]$$

a)

From this we can see that the limit of $sc(x)$ as $|x| \to \infty$ is 1, assuming n is fixed. When the observation x is added to the ideal sample with sample mean 0, the new sample mean is pushed away from 0 (with the direction depending on the sign of x). As x gets extremely large, the function approaches $n\frac{\bar{X}}{\sqrt{\bar{X}^2}} = n$ due to the Law of Large Numbers. In order to check this, I wrote some quick R code:

```
set.seed(1017)
# Make n-1 sample with mean 0 (or close enough)
xn_1 <- rnorm(999,0,5)
# N
n <- length(xn_1)+1
# Values of x going toward infinity
xs <- 1:10000
# SC function
sc <- lapply(xs, function(x){
  xn <- c(xn_1,x)
  stat <- n*sqrt(n)*mean(xn)/sd(xn)
  stat
})
# Plot
plot(xs,unlist(sc),type = "l",xlab = "x",ylab = "sc(x)")
abline(n,0)
```



b)

It's a little more obvious to see the limit of $sc(x)$ as $n \to \infty$ with x fixed. The function can be rearranged to $[\frac{n\sqrt{n}\sqrt{n-1}(\bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}]$. With x fixed this is increasing in n, so the limit as n approaches ∞ does not exist.

So, the t-ratio is robust as a function of x, but not n.

1. Suppose $X_1, \ldots X_n$ are iid Poisson (θ) with $\theta \sim$ Gamma $(1, \lambda)$.

a) To find the Bayes rule for the loss function $l(\theta, a) = \theta^p(\theta - a)^2$ (where p is a fixed constant) we can use the weighted loss approach from BD problem 3.2.5 b) with $1/w(\theta) = \theta^p$. Also, we know that the Gamma distribution is a conjugate prior for the Poisson, so:

$$p(\theta|X) \sim \text{Gamma}(\Sigma x_i + 1, n + \lambda).$$

So, $E[l(\theta, a)|X]$ can be written:

$$\frac{\int_\theta \theta^p(\theta-a)^2 \frac{(n+\lambda)^{\Sigma x_i + 1}}{\Gamma(\Sigma x_i + 1)} \theta^{\Sigma x_i} e^{-\theta(n+\lambda)} \, d\theta}{\int_\theta \frac{(n+\lambda)^{\Sigma x_i + 1}}{\Gamma(\Sigma x_i + 1)} \theta^{\Sigma x_i} e^{-\theta(n+\lambda)} \, d\theta}$$

Multiplying top and bottom by the normalizing constant c (again see BD problems 1.4.24 and 3.2.5 for details) gives us the squared loss function $(\theta - a)^2$ with posterior density Gamma $(\Sigma x_i + p + 1, n + \lambda)$. Because we now have squared loss, the Bayes rule is the mean of the posterior:

$$\boxed{\delta^*(x) = \frac{\Sigma x_i + p + 1}{n + \lambda}}$$

b) In order for the Bayes rule to be minimax, risk $R(\theta, \delta^*)$ must be constant. In this case,

$$R(\theta, \delta^*) = E[\ell(\theta,a)] = E[\theta^p(\theta-a)^2] = \theta^p E[(\theta-a)^2]$$

Which simplifies to $\theta^p * MSE[\delta^*]$ because the function in the expectation is squared loss. So we can rewrite the risk as:

$$\theta^p\left(Var(\delta^*) + Bias(\delta^*)^2\right) = \theta^p\left[Var\left(\frac{\sum x_i + p+1}{n+\lambda}\right) + \left(\theta - E\left[\frac{\sum x_i + p+1}{n+\lambda}\right]\right)^2\right]$$

$$= \theta^p\left[\frac{n\theta}{(n+\lambda)^2} + \left(\theta - \frac{n\theta + p+1}{n+\lambda}\right)^2\right]$$

Next we need to find values of n, λ, and p that make this function independent of θ. I spent hours on this part and the only values I could find we $n=0$ and $p=0$ with $\lambda > 0$. This is out of the range of p, therefore δ^* is not minimax. Of course it's possible that there are values ~~of~~ of n, p, and λ that will make risk constant, since I didn't prove that there aren't any. However, after staring at this problem for hours I'm fairly certain δ^* isn't minimax.

2. Consider estimation of regression slopes $\theta_1, \ldots, \theta_p$ for p observations of $(X_1, Y_1), \ldots, (X_p, Y_p)$ modeled as independent with $X_i \sim N(0,1)$ and $Y_i | X_i \sim N(\theta_i x, 1)$.

a) Following a Bayesian approach, let θ_i be iid from $N(0, \tau^2)$. Find the Bayes estimate of θ_i assuming squared loss.

First we need to find the posterior distribution for θ_i given x_i and y_i:

$$p(\theta_i | x_i, y_i) \propto p(x_i)\, p(y_i | x_i)\, \pi(\theta)$$

which is proportional to:

$$\exp\left(\frac{-x_i^2}{2}\right) \exp\left(\frac{-(y_i - \theta_i x_i)^2}{2}\right) \exp\left(-\frac{\theta_i^2}{2\tau^2}\right) =$$

$$\exp\left(\frac{-\left(x_i^2 + (y_i - \theta_i x_i)^2\right)}{2}\right) \exp\left(\frac{-\theta_i^2}{2\tau^2}\right)$$

Combining these gives us:

$$\exp\left(\frac{-\left(\tau^2\left(x_i^2 + (y_i - \theta_i x_i)^2\right) + \theta_i^2\right)}{2\tau^2}\right)$$

This can be simplified using the same complete the square approach as a simple case of normal with normal prior (see C&B exercise 7.22), but using $(1 + x_i \tau^2)$ in place of $(1 + \tau^2)$. Thus, using C&B example 7.2.16 we see that

$$p(\theta_i \mid X, Y) \sim N\left(\frac{x_i y_i \tau^2}{1 + x_i^2 \tau^2}, \frac{\tau^2}{1 + x_i^2 \tau^2}\right)$$

Because we are assuming the squared loss function, δ^* is the expected vale of $p(\theta_i \mid X, Y)$:

$$E[p(\theta_i \mid X, Y)] = \frac{X_i Y_i \tau^2}{1 + X_i^2 \tau^2}$$

b) To find $E[Y_i^2]$, we can use the double expectation theorem:

$$E[Y_i^2] = E[E[Y_i^2 \mid X_i, \theta_i]]$$

The inner expectation can be written:

$$E[Y_i^2 \mid X_i, \theta_i] = Var[Y_i \mid X_i, \theta_i] + E[Y_i \mid X_i, \theta_i]^2$$

$$= 1 + (\theta_i x_i)^2$$

So $E[Y_i^2] = E[1 + (\theta_i x_i)^2] = 1 + E[(\theta_i x_i)^2]$
$= 1 + E[\theta_i^2] E[X_i^2] = \boxed{1 + \tau^2(1)}$

This is the expected value for a single Y_i, so a estimator using all the observations would be:

$$\frac{\sum_{i=1}^{P} Y_i^2}{P} - 1 = \hat{\tau}^2$$

c) The empirical Bayes estimator of θ_i is simply the Bayes estimator from a) but substituting the known τ^2 with $\hat{\tau}^2$ above:

$$\hat{\theta}_i = \frac{X_i Y_i \hat{\tau}^2}{1 + X_i^2 \hat{\tau}^2}$$

3. BD 3.4.2: Suppose that $l(\theta, a)$ is convex and $\delta^*(X) = E[\delta(X) \mid T(X)]$. First, define $R(\theta, \delta(X))$ and $R(\theta, \delta^*(X))$:

$R(\theta, \delta(X)) = E[l(\theta, \delta(X))] = E[E[l(\theta, \delta(X)) \mid T(X)]]$ by the double expectation theorem

$R(\theta, \delta^*(X)) = E[l(\theta, \delta^*(X))] = E[l(\theta, E[\delta(X) \mid T(X)])]$

Because the loss function is convex, $E[l(\theta, \delta(X)) \mid T(X)]$ must be $\geq l(\theta, E[\delta(X) \mid T(X)])$ by Jensen's inequality (where $g(x)$ is the loss function).

3.4.3. Let $X \sim p(x, \theta)$ and assume regularity conditions hold for calculating Fisher's information number. Set $\eta = h(\theta)$ and $q(x, \eta) = p(x, h^{-1}(\eta))$.

a) Calculate the Information number $I_q(\eta)$:

$$I_q(\eta) = E\left[\left(\partial/\partial\eta \ \log q(x,\eta)\right)^2\right]$$ usin

Using the change of variable $\eta = h(\theta)$ with $\partial/\partial\eta = \partial/\partial\theta \ \ d\theta/d\eta = \partial/\partial\theta \ \frac{1}{h'(h^{-1}(\eta))}$, rewrite this as:

$$I_q(\eta) = E\left[\left(\partial/\partial\theta \ \log p(x,\theta) \ \frac{1}{h'(h^{-1}(\eta))}\right)^2\right]$$

This is the same as:

$$\left(\frac{1}{h'(h^{-1}(\eta))}\right)^2 E\left[\left(\partial/\partial\theta \ \log p(x,\theta)\right)^2\right] = \frac{I_p(\theta)}{h'(h^{-1}(\eta))^2} \quad \text{with } h^{-1}(\eta) = \theta.$$

The denominator can be pulled out of the expectation because it is constant w.r.t. x.

b) To show the equivariance of the information bound, first write B in terms of q and η:

$$B_q(\eta) = \frac{(\psi'(\eta))^2}{I_q(\eta)}$$

Per BD 3.4.13 we can write $\psi'(\eta)$ as:

$$\int T(x)\left(\partial/\partial\eta \log q(x,\eta)\right) q(x,\eta)\, dx$$

Using the exact same change of variables as above, we can again rewrite this as:

$$\psi'(\eta) = \int T(x)\left(\partial/\partial\theta \log p(x,\theta)\right)\frac{1}{h'(h^{-1}(\eta))} p(x,\theta)\, dx$$

This is equivalent to:

$$\psi'(\eta) = \frac{1}{h'(h^{-1}(\eta))} \psi'(\theta).$$

So, from this and the result of a), we see that:

$$\frac{\psi'(\eta)^2}{I_q(\eta)} = \psi'(\eta)^2 * \frac{h'(h^{-1}(\eta))^2}{I_p(\theta)} = \frac{\psi'(\theta)^2}{h'(h^{-1}(\eta))^2} * \frac{h'(h^{-1}(\eta))^2}{I_p(\theta)}$$

Thus, $\dfrac{\psi'(\eta)^2}{I_q(\eta)} = \dfrac{\psi'(\theta)^2}{I_p(\theta)}$.

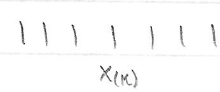
3.5.1. The sample median $\hat{x}$ is defined as:

$$\hat{x} = \begin{cases} X_{(K+1)} & \text{if } n = 2K+1 \quad (\text{odd}) \\ \frac{1}{2}\left(X_{(K)} + X_{(K+1)}\right) & \text{if } n = 2K \quad (\text{even}) \end{cases}$$

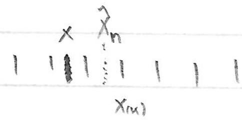
To calculate the sensitivity curve of the median for the even case, we first set the median for n-1 observations to 0 without loss of generality (WLOG). For the even case, $\hat{x}_{n-1} = X_{(K)} = 0$. As a result, the sensitivity curve $SC(x) = n\left[\hat{x}_n - \hat{x}_{n-1}\right] = n\,\hat{x}_n$. $\hat{x}_n$ depends on the value of the x being added back in, so the sensitivity curve is defined as:

$$SC(x) = \begin{cases} n\left(\dfrac{X_{(K-1)} + X_{(K)}}{2}\right) = \dfrac{n\,X_{(K-1)}}{2} & \text{if } x < X_{(K-1)} \\[2ex] n\left(\dfrac{X_{(K)} + x}{2}\right) = \dfrac{n\,x}{2} & \text{if } X_{(K-1)} \le x \le X_{(K+1)} \\[2ex] n\left(\dfrac{X_{(K)} + X_{(K+1)}}{2}\right) = \dfrac{n\,X_{(K+1)}}{2} & \text{if } x > X_{(K+1)} \end{cases}$$

I found it helpful to first draw the n-1 sample like so:



Then it becomes clear what the median is for $x < X_{(K-1)}$:



This can be repeated for the different ranges of x.

A rough sketch of $SC(x)$ looks similar to the plot for the odd case, but centered differently: