ANSWERS TO STAT 695R HW2

XIONGZHI CHEN

Abstract. It is more of a think sheet than mere answers.

1. Notations and Conventions

a: Unless otherwise specified, all notations presented here bear their generic meanings as implied or required by the contexts.

b: Unless otherwise stated, minimal assumptions that are needed to induce or disprove the conclusions but are not explicitly given in the homework problems are automatically proposed.

c: As usual $dx$ denotes a Lebesgue measurable change in $x \in \mathbb{R}^n$.

2. Problems and Answers

Problem 1. (Chp8, pg123, prob 2) Calculate the asymptotic efficiency of the sample mean and sample median based on a sample of size $n$ for the Laplace distribution (with density $p(x) = e^{-|x|/2}$)

Solution. Let $X$ be with the Laplace distribution. Then $\theta_0 = E(X) = 0$ and $\text{Var}(X) = 2$. For $X_i, i = 1, .., n$, an iid sample for $X$, let $S_n = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $T_n = \text{Med}(X_i; 1 \leq i \leq n)$. Then by CLT, it follows

$$\sqrt{n}(S_n - \theta_0) \sim N(\theta_0, \text{Var}(X)) = N(0, 2)$$

To find the asymptotic distribution of $\sqrt{n}(T_n - \theta_0)$, it suffices to use Lemma 5.10 and Theorem 5.23 on the text. Following the notations there in by letting

$$\Psi_n(\theta) = n^{-1}\sum_{i=1}^{n} \text{sign}(X_i - \theta)$$

and

$$\Psi(\theta) = E\text{sign}(X - \theta) = P(X > \theta) - P(X < \theta)$$

it is clear the

$$\Psi_n(\theta) \rightarrow^p \Psi(\theta) \quad (2.1)$$

by the LLN for each fixed $\theta$.

Through Lemma 5.10, the consistency of $T_n$ for $\theta_0$ is implied jointly by:

(2.1); $\Psi_n(T_n) = 0$; $\Psi_n(\theta)$ nonincreasing in $\theta$; the uniqueness of the median for $X$, i.e., $\text{Med}(X) = \theta_0 = 0$, where Med denotes median.
To establish asymptotic normality of $\sqrt{n} (T_n - \theta_0)$, some more details are needed than those provided by the text. Since

$$T_n = \arg \max_{\theta} \left( -\sum_{i=1}^{n} |X_i - \theta| \right)$$

(2.2)

it is reasonable to use

$$m_{\theta} (x) = |x - \theta| - |\theta| = \begin{cases} \theta & \text{if } x < a \\ \theta - 2x & \text{if } a \leq x < b \\ -\theta & \text{if } x \geq b \end{cases}$$

(2.3)

where $a = \min (0, \theta)$, $b = \max (0, \theta)$.

Clearly $m_{\theta} (x)$ is with Lipschitz constant 1 (in $\theta$) since it’s the translated (by $x$) Euclidean norm and

$$\dot{m}_{\theta} (x) = -\text{sign} (x - \theta), \forall x \neq \theta$$

(2.4)

as well. Further

$$P m_{\theta} = \int_{\mathbb{R}} m_{\theta} (x) dF (x)$$

(2.5)

$$= \left( \int_{\{x < a\}} \theta + \int_{[a, b]} (\theta - 2x) - \int_{\{x \geq b\}} \theta \right) dF (x)$$

$$= \theta F (a-) + \theta (F (b) - F (a)) - \theta (1 - F (b)) - 2 \int_{[a, b]} x dF (x)$$

$$= -\theta dF (a) + \theta [F (b-) + F (b)] - \theta - 2 \left[ \theta F (\theta-) \text{sign} \theta - \int_{[a, b]} F (x) dx \right]$$

where the cdf $F$ is assumed to be right-continuous.

Notice however the text assumes $\theta \geq 0$, which is unnecessary.

For the settings here with $F' (x) = p (x), x \in \mathbb{R}$, (2.5) reduces to

$$P m_{\theta} = 2 \theta F (b) - \theta - 2 \int_{[a, b]} x dF (x)$$

(2.6)

$$= 2 \theta F (b) - \theta - 2 \theta F (\theta) \text{sign} \theta + 2 \int_{[a, b]} F (x) dx$$

$$= 2 \theta F (\max (0, \theta)) - \theta - 2 \theta F (\theta) \text{sign} \theta + 2 \int_{[\min (0, \theta), \max (0, \theta)]} F (x) dx$$

and $P m_{\theta}$ is twice differentiable in $\theta$ except at $\theta = 0$ since $p' (0+) = -1$ but $p' (0-) = -1$ since $\min (0, \theta)$ and $\max (0, \theta)$ functions are obviously so).

More specifically, $P m_{\theta}$

$$= \begin{cases} 2 \int_{[0, \theta]} F (x) dx - \theta & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ 2 \theta F (0) - \theta + 2 \theta F (\theta) + 2 \int_{[0, \theta]} F (x) dx & \text{if } x \geq b \end{cases}$$

(2.7)
and

\[ \hat{P}_{m\theta} = \begin{cases} 
  2F(\theta) - 1 & \text{if } \theta > 0 \\
  2F(0) - 1 = 0 & \text{if } \theta = 0 \\
  2F(0) - 1 + 2F(\theta) + 2\theta p(\theta) - 2F(\theta) & \text{if } \theta > 0
\end{cases} \quad (2.8) \]

and

\[ \hat{P}_{m\theta} = \begin{cases} 
  2p(\theta) & \text{if } \theta > 0 \\
  2p(0) = 1 & \text{if } \theta = 0 \\
  2p(\theta) + 2\theta p'(\theta) & \text{if } \theta > 0
\end{cases} \quad (2.9) \]

where \( F'(x) = p(x) \) and a fact from analysis has been used \(^1\).

All the above shows an admissible expansion

\[ P_{m\theta} = P_{m\theta_0} + \hat{P}_{m\theta_0}(\theta - \theta_0) + \frac{1}{2} \hat{P}_{m\theta_0}(\theta - \theta_0)^2 + o\left(|\theta - \theta_0|^2\right) \quad (2.10) \]

which combined with Theorem 5.23 yields

\[ \sqrt{n} (T_n - \theta_0) \sim N \left( 0, \left( \hat{P}_{m\theta_0} \right)^{-2} \right) = N(0,1) \quad (2.11) \]

Hence by Section 8.2 on the text, the relative efficiency is

\[ \rho(T_n, S_n) = \frac{2}{1} = 2 \]

i.e., the median is more efficient. \( \blacksquare \)

**Problem 2.** (Chp24, pg 356, prob1) Show, informally, that under sufficient regularity conditions

\[ \text{MISE}_f(\hat{f}) \sim \frac{1}{nh} \int K^2(y) dy + \frac{1}{4h^4} \int f''(x)^2 dx \left( \int y^2 K(y) dy \right)^2 \]

What does this imply for an optimal choice of the bandwidth?

**Solution.** I think the optimal bandwidth does not change. The definition of \( \text{MISE}_f(\hat{f}) \) suggests

\[ \text{MISE}_f(\hat{f}) = \int_R E_f \left( \hat{f}(x) - f(x) \right)^2 dx \]

\[ = \int_R \text{Var}_f \left[ \hat{f}(x) \right] dx + \int_R [E_f \left( \hat{f}(x) \right) - f(x)]^2 dx \]

where without loss of generality (WLOG), all integrals are extended to the corresponding domains in which the integrands are zero outside their supports.

\(^1\)It is the fact that, if a derivative defined on a punctured disk is continuous at the center, then the its prime is actually differentiable at this center. That is, the center in this case is a removable singularity.
For the first part, it is easy to see

\[
\int_{\mathbb{R}} \text{Var}_f \left[ \hat{f}(x) \right] \, dx = \int_{\mathbb{R}} \frac{1}{n} \text{Var}_f \frac{1}{h} K \left( \frac{x - X_1}{h} \right) \, dx \\
\leq \int_{\mathbb{R}} \frac{1}{nh^2} E_f K^2 \left( \frac{x - X_1}{h} \right) \, dx = \int_{\mathbb{R}} \frac{1}{nh} \int_{\mathbb{R}} K^2(t) f(x - ht) \, dt \, dx \\
= \frac{1}{nh} \int_{\mathbb{R}} f(x - ht) \, dx \left( \int_{\mathbb{R}} K^2(t) \, dt \right) \\
= \frac{1}{nh} \int_{\mathbb{R}} f(x) \, dx \left( \int_{\mathbb{R}} K^2(t) \, dt \right) \leq \frac{1}{nh} \int_{\mathbb{R}} K^2(t) \, dt
\]

by Tonelli’s theorem, the translation invariance of Lebesgue measure.

Following a slightly different notation for Sobolev spaces, we define

\[
\mathcal{S}_{m,p} (N^{M}) = \left\{ g \in C^\alpha (N^{M}) : \| \partial^\alpha g \|_p < \infty \text{ for } \forall |\alpha| \leq m \right\}
\]

where \( \alpha \) is a multi-index, \( m \in \mathbb{N}, 0 < p \leq \infty \) and measurability is automatically assumed. Now for any \( f \in \mathcal{S}_{2,2} (\mathbb{R}^2) \), the Taylor expansion of \( f \) at \( x \) with Lagrange remainder

\[
f(x - ht) - f(x) = -ht f'(x) + \frac{h^2 t^2}{2} f''(x - \alpha ht)
\]

for some \( \alpha \in (0, 1) \) implies

\[
\text{Bias}_f \hat{f}(x) = E_f \left( \hat{f}(x) \right) - f(x) = \int_{\mathbb{R}} K(t) (f(x - ht) - f(x)) \, dt \\
= \int_{\mathbb{R}} K(t) \left( -ht f'(x) + \frac{h^2 t^2}{2} f''(x - \alpha ht) \right) \, dt \\
= \int_{\mathbb{R}} \frac{h^2 t^2}{2} K(t) f''(x - \alpha ht) \, dt = \int_0^1 ds \int_{\mathbb{R}} \frac{h^2 t^2}{2} K(t) f''(x - \alpha ht) \, dt \\
= \frac{h^2}{2} \int_{\mathbb{R} \times [0,1]} t \times t f''(x - \alpha ht) (K(t) \, dt \otimes ds)
\]

By Cauchy-Schwartz inequality,

\[
\left| \int_{\mathbb{R} \times [0,1]} t \times t f''(x - \alpha ht) (K(t) \, dt \otimes ds) \right|^2 \\
\leq \left( \int_{\mathbb{R} \times [0,1]} t^2 (K(t) \, dt \otimes ds) \right) \left( \int_{\mathbb{R} \times [0,1]} t^2 f''(x - \alpha ht)^2 (K(t) \, dt \otimes ds) \right) \\
\leq \left( \int y^2 K(y) \, dy \right) \int_{\mathbb{R}} t^2 f''(x - \alpha ht)^2 K(t) \, dt
\]
and thus
\[ \int_{\mathbb{R}} \left( \text{Bias}_f \hat{f}(x) \right)^2 \, dx \]
\[ \leq \frac{h^4}{4} \left( \int_{\mathbb{R}} t^2 K(t) \, dt \right) \int_{\mathbb{R}} dx \int_{\mathbb{R}} t^2 f''(x - \alpha h t)^2 K(t) \, dt \]
\[ = \frac{h^4}{4} \left( \int_{\mathbb{R}} t^2 K(t) \, dt \right) \left( \int_{\mathbb{R}} f''(x - \alpha h t)^2 \, dx \right) \]
\[ = \frac{h^4}{4} \left( \int_{\mathbb{R}} f''(x)^2 \, dx \right) \left( \int_{\mathbb{R}} y^2 K(y) \, dy \right)^2 \]
by Tonelli’s theorem and translation invariance of Lebesgue measure.

Consequently
\[ MISE_f \left( \hat{f} \right) \leq \frac{1}{nh} \int_{\mathbb{R}} K^2(t) \, dt + \frac{h^4}{4} \left( \int_{\mathbb{R}} t^2 K(t) \, dt \right) \left( \int_{\mathbb{R}} f''(x)^2 \, dx \right) \]
\[ = \frac{C}{nh} + \frac{h^4}{4} B^2 A \leq \max \{ C, B^2 A/4 \} \left( \frac{1}{nh} + h^4 \right) \]
and choosing \( h \sim n^{-1/5} \) yields
\[ \lim_{n \to \infty} \sup_{h \sim n^{-1/5}} \left( \frac{MISE_f \left( \hat{f} \right)}{\left( \frac{1}{nh} + h^4 \right)} \right) = \lim_{n \to \infty} \sup_{h \sim n^{-1/5}} \frac{MISE_f \left( \hat{f} \right)}{n^{-4/5}} \leq \max \{ C, B^2 A/4 \} \]
as desired, which gives the optimal rate \( n^{-2m/(2m+1)} \) with \( m = 2 \) in this case.

Hence for \( f \in S_{m,2} (\mathbb{R}^2) \), the optimal choice of bandwidth is always \( h \sim n^{-1/(2m+1)} \) as proved by the text and relies generically only on the order \( m \) of differentiability of \( f \). \( \blacksquare \)

**Problem 3.** Let \( X_i \) be iid uniform on \([0, 1]\) with \( n \geq 2 \). Denote \( X_{(1)} \) and \( X_{(n)} \) as the min and max of \( X_i \): (a) find the joint for \( (X_{(1)}, X_{(n)}) \), (b) find \( E \big( X_{(1)} | X_{(n)} \big) \), (c) find the distribution of \( X_{(n)} - X_{(1)} \)

**Solution.** Let \( Y_i = X_{(i)} \) and let \( F \) be the cdf of \( X_1 \). Given \( x \geq y \), then
\[ G(x,y) = P(Y_n < y, Y_1 < x) = P(Y_n < y) \]
\[ = P \left( \bigcap_{i=1}^n \{ X_i < y \} \right) = \prod_{i=1}^n F(y) = [F(y)]^n \]
Further, for \( x < y \), then
\[ G(x,y) = P(Y_n < y, Y_1 < x) = P(Y_n < y) - P(Y_1 \geq x, Y_n < y) \]
\[ = [F(y)]^n - P \left( \bigcap_{i=1}^n \{ x \leq X_i < y \} \right) = [F(y)]^n - [F(y) - F(x)]^n \]
(a) Now since \( F(t) = t \) if \( t \in [0, 1] \); = 0 if \( t < 0 \); = 1 if \( t > 1 \), it is clear that
\[
g(x, y) = \frac{\partial^2 G(x, y)}{\partial x \partial y} = n (n - 1) [F(y) - F(x)]^{n-2} F'(x) F'(y)
\]
\[
= n (n - 1) [y - x]^{n-2}
\]
for \( 0 \leq x < y \leq 1 \) and \( g(x, y) = 0 \) otherwise.

(b) So the marginal
\[
h(X(1) = x|X(n) = y) = \frac{n (n - 1) [y - x]^{n-2}}{\int_0^y n (n - 1) [y - x]^{n-2} \, dx} = \frac{(n - 1) (y - x)^{n-2}}{y^{n-1}}
\]
and
\[
E(X(1)|X(n) = y) = \int_0^y x \frac{(n - 1) (y - x)^{n-2}}{y^{n-1}} \, dx
\]
\[
= \frac{n - 1}{y^{n-1} n (n - 1)} = \frac{y^n}{n}
\]
by using
\[
\int x (y - x)^{n-2} \, dx = \frac{-(y - x)^{n-1}}{n - 1} + \int \frac{(y - x)^{n-1} \, dx}{n - 1}
\]
\[
= \frac{-(y - x)^{n-1}}{n - 1} - \frac{(y - x)^n}{(n - 1) n}
\]
(c) Let \( R_n = X(n) - X(1) \). From
\[
P(Y_n - Y_1 < z) = \int \int_{\{y-x<z\}} g(x, y) \, dxdy = \int \int_{\{0<y-x<0<x, y<1\}} g(x, y) \, dxdy
\]
\[
= n (n - 1) \left( \int_0^{1-z} x \int_x^{x+z} \, dx + \int_{1-z}^1 \int_x^1 \, dx \right) (y - x)^{n-2} \, dy
\]
\[
= n (n - 1) \left( \int_0^{1-z} \frac{(y - x)^{n-1}}{n - 1} \left[ \frac{y = x + z}{y = x} \right] \, dx + \int_{1-z}^1 \frac{(y - x)^{n-1}}{n - 1} \left[ \frac{y = 1}{y = x} \right] \, dx \right)
\]
\[
= n \left( \int_0^{1-z} z^{n-1} \, dx + \int_{1-z}^1 (1 - x)^{n-1} \, dx \right) = n (1 - z) z^{n-1} + z^n
\]
for \( z \in [0, 1] \), it follows that \( R_n \) has density
\[
l(z) = \frac{d}{dz} P(Y_n - Y_1 < z) = n (n - 1) (1 - z) z^{n-2}
\]
for \( z \in [0, 1] \) and \( l(z) = 0 \) otherwise. ■