Recitation Note: M-estimation

Tyler Maunu
maunut@mit.edu

September 3, 2019

This note will cover some topics in M-estimation.

The Huber Function

In class you learned about M-estimation, which tries to estimate a parameter by minimizing some loss function. In the univariate case with data X_i, this procedure takes the form

$$\hat{\mu} = \operatorname{argmin}_{b \in \mathbb{R}} \sum_{i=1}^{n} \rho(x_i - m),$$

for some choice of function ρ. Some examples include $\rho(x) = x^2$, which yields the sample mean; $\rho(x) = |x|$, which yields the sample median, and

$$\rho(x) = \begin{cases} |x - m|, & |x - m| \geq \delta \\ \frac{(x-m)^2}{2\delta} + \frac{\delta}{2}, & |x - m| < \delta, \end{cases}$$

for some parameter δ, which is called the Huber loss function.

Questions

1. What M-estimator corresponds to the maximum likelihood estimator for a $\mathsf{Laplace}(\mu, 1)$ distribution?
2. What are the asymptotic variance of the sample mean and the MLE for this $\mathsf{Laplace}(\mu, 1)$ distribution? How do things change if we instead consider a $\mathsf{Laplace}(\mu, b)$ distribution?
3. Compare the asymptotic variance for the sample mean and sample median when $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathsf{Cauchy}(\mu, 1)$.
4. Calculate the asymptotic variance for the Huber estimator for general symmetric functions, and apply this to the case of $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathsf{Cauchy}(\mu, 1)$.

Solutions

1. For the Laplace distribution, the likelihood of $X_1, \ldots, X_n$ is

$$L(X_1, \ldots, X_n; \mu) = \prod_{i=1}^{n} \frac{1}{2} \exp(-|x_i - \mu|). \tag{1}$$

The log-likelihood is

$$\ell(X_1, \ldots, X_n; \mu) = -n \log(2) - \sum_{i=1}^{n} |x_i - \mu|. \tag{2}$$

Notice that maximizing ℓ is equivalent to minimizing $-\ell$, and so

$$\hat{\mu}_{MLE} = \min_{\mu} \sum_{i=1}^{n} |x_i - \mu|. \tag{3}$$

This is an M-estimator that exactly corresponds to the sample median.

2. First, we derive the asymptotic variance of the sample mean. This is just

$$\text{avar}(\overline{X_n}) = \text{Var}(X_1) = 2. \tag{4}$$

Here, we assume we know the variance of the $\mathsf{Laplace}(\mu, 1)$ distribution. If you do now remember this variance, then it can be obtained using integration by parts twice on the Laplace$(0, 1)$ distribution (notice that $\mathsf{Laplace}(\mu, 1)$ and $\mathsf{Laplace}(0, 1)$ must have the same variance):

$$\begin{aligned}
\text{Var}(X) = \mathbb{E}X^2 &= \int_{-\infty}^{\infty} \frac{x^2}{2} \exp(-|x|) dx \\
&= 2 \int_{0}^{\infty} \frac{x^2}{2} \exp(-x) dx \\
&= 2 \left(-\frac{x^2}{2} \exp(-x) \Big|_0^{\infty} - \int_0^{\infty} x \exp(-x) dx \right) \\
&= 2 \left(- \int_0^{\infty} x \exp(-x) dx \right) \\
&= 2 \left(x \exp(-x) \Big|_0^{\infty} + \int_0^{\infty} \exp(-x) dx \right) \\
&= 2 \left(- \exp(-x) \Big|_0^{\infty} \right) \\
&= 2.
\end{aligned} \tag{5}$$

On the other hand, we will derive the asymptotic variance of the sample median. Let m_n be the sample median and suppose that we observe data $X_1, \ldots, X_n \sim F$, where F is a continuous distribution, and let μ be the population median $F(\mu) = 1/2$ For the following, assume that n is odd, so that the sample median is the unique point $m_n = X_{((n+1)/2)}$. Then,

We will show that

$$P(\sqrt{n}(m_n - \mu) \leq a) \to N(0, \sigma^2), \tag{6}$$

for some asymptotic variance σ^2. To this end, notice that

$$\{m_n \leq \mu + a/\sqrt{n}\} = \left\{\#(X_i \leq \mu + a/\sqrt{n}) \geq \frac{n+1}{2}\right\} = \left\{\overline{Y_n} \geq \frac{n+1}{2}\right\}, \tag{7}$$

where the random variables Y_i are defined by $Y_i = \mathbb{1}(X_i \leq \mu + a/\sqrt{n})$. Then we have

$$\begin{aligned} P(\sqrt{n}(m_n - \mu) \leq a) &= P\left(\overline{Y_n} \geq \frac{n+1}{2}\right) \\ &= P\left(\overline{Y_n} - p_n n \geq \frac{n+1}{2} - p_n n\right) \\ &= P\left(\frac{\overline{Y_n} - p_n n}{\sqrt{np_n(1-p_n)}} \geq \frac{\frac{n+1}{2} - p_n n}{\sqrt{np_n(1-p_n)}}\right), \end{aligned} \tag{8}$$

where $p_n = F(\mu + a/\sqrt{n})$ is the probability that $Y_i = 1$ (notice that Y_i is nothing more than a $\mathsf{Bernoulli}(p_n)$ random variable).

Let $p = F(\mu) = 1/2$. Notice that

$$\frac{\overline{Y_n} - p_n n}{\sqrt{np_n(1-p_n)}} - \frac{\overline{Y_n} - pn}{\sqrt{np(1-p)}} \xrightarrow{p} 0, \tag{9}$$

which then yields

$$\frac{\overline{Y_n} - p_n n}{\sqrt{np_n(1-p_n)}} \xrightarrow{d} N(0,1), \tag{10}$$

Some manipulation yields that

$$P(m_n \leq \mu + a/\sqrt{n}) = P\left(Z \geq \frac{\frac{n+1}{2} - p_n n}{\sqrt{np_n(1-p_n))}}\right). \tag{11}$$

Finally, we notice that

$$\begin{aligned} \frac{\frac{n+1}{2} - p_n n}{\sqrt{np_n(1-p_n))}} &= \frac{\frac{n}{2} + \frac{1}{2} - F(\mu + a/\sqrt{n})n}{\sqrt{np_n(1-p_n))}} \\ &= \frac{\frac{n}{2} - F(\mu + a/\sqrt{n})n}{\sqrt{np_n(1-p_n))}} + \frac{\frac{1}{2}}{\sqrt{np_n(1-p_n))}} \\ &= \frac{F(\mu) - F(\mu + a/\sqrt{n})}{a/\sqrt{n}} \cdot \frac{a}{\sqrt{p_n(1-p_n)}} + \frac{\frac{1}{2}}{\sqrt{np_n(1-p_n))}} \\ &\xrightarrow{d} -2aF'(\mu). \end{aligned} \tag{12}$$

Therefore,

$$\sqrt{n}(m_n - \mu) \xrightarrow{d} N\left(0, \frac{1}{4f(\mu)^2}\right). \tag{13}$$

In the case of a $\mathsf{Laplace}(\mu, 1)$ distribution, $f(\mu) = 1/2$. Thus, the asymptotic variance of the median is 1, which is smaller than the asymptotic variance of the sample mean.

When we instead have $\mathsf{Laplace}(\mu, b)$, the asymptotic variance of the sample mean is $2b^2$ and the asymptotic variance of the sample median is b^2.

3. For $X \sim \mathsf{Cauchy}(\mu, 1)$, we have

$$f_X(x) = \frac{1}{\pi(1 + (x-\mu)^2}, \tag{14}$$

and

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x-\mu), F_X^{-1}(t) = \tan\left(\pi\left(t - \frac{1}{2}\right)\right) + \mu. \tag{15}$$

Notice that these are both continuous on $\mathbb{R}$.

Using the result of the previous problem, the asymptotic variance of the sample median is

$$\text{avar}(\tilde{X}_n) = \frac{1/4}{f_X(F_X^{-1}(1/2))^2} = \frac{1/4}{f_X(\mu)^2} = \frac{\pi^2}{4}. \tag{16}$$

On the other hand,

$$\text{avar}(\overline{X_n}) = n\text{Var}(\overline{X_n}) = \text{Var}(X_i) = \infty. \tag{17}$$

4. Finally, we look at the Huber loss function for continuous, symmetric distributions.

Let $X \sim F$ be a random variable following a continuous cdf F (with continuous pdf $f = F'$), and suppose we observe n independent copies of $X - \mu$, $X_1 - \mu, \ldots, X_n - \mu$. Suppose also that f is symmetric about 0, and thus $f(x-\mu)$ is symmetric about μ . Consider the following form of the Huber loss:

$$\rho(x) = \begin{cases} k|x| - \frac{1}{2}k^2, & |x| > k \\ \frac{(x)^2}{2}, & |x| \leq k, \end{cases} \tag{18}$$

Then, ρ is continuous and differentiable everywhere, with derivative

$$\rho'(x) = \begin{cases} k\text{sign}(x), & |x| > k \\ x, & |x| \leq k, \end{cases} \tag{19}$$

Let $\psi(x) = \rho'(x)$.

Notice that the Huber estimator has the correct mean:

$$\begin{aligned} \mathbb{E}\psi(X-\mu) = &\int_{\mu-k}^{\mu+k} (x-\mu)f(x-\mu)dx \\ &+ \int_{-\infty}^{\mu-k} k\text{sign}(x-\mu)f(x-\mu)dx + \int_{\mu+k}^{\infty} k\text{sign}(x-\mu)f(x-\mu)dx \\ = &\int_{-k}^{k} uf(u)du + \int_{-\infty}^{-k} k\text{sign}(u)f(u)du + \int_{k}^{\infty} k\text{sign}(u)f(u)du \\ = &\ 0. \end{aligned} \tag{20}$$

This means that the Huber function has the correct asymptotic mean: if $X - \mu$ has pdf f, then μ is the stationary point of $\mathbb{E}\rho(X-b)$ (as a function of b).

The Huber estimator for the sample is the solution of

$$\hat{\mu} = \text{argmin}_{b\in\mathbb{R}} \sum_{i=1}^{n} \rho(X_i - b). \tag{21}$$

This can be found by solving

$$\sum_{i=1}^{n} \psi(X_i - b) = 0. \tag{22}$$

We wish to find the asymptotic variance of this estimator.

To this end, we can do a Taylor expansion of $\sum_{i=1}^{n} \psi(X_i - b)$ at $b = \mu$ to find

$$0 = \sum_{i=1}^{n} \psi(X_i - \hat{\mu}) = \sum_{i=1}^{n} \psi(X_i - \mu) + (\hat{\mu} - \mu) \sum_{i=1}^{n} \psi'(X_i - \mu) + \text{ higher order terms.} \tag{23}$$

Rearranging,

$$\begin{aligned} \sqrt{n}(\hat{\mu} - \mu) &= \sqrt{n} \frac{\sum_{i=1}^{n} \psi(X_i - \mu)}{-\sum_{i=1}^{n} \psi'(X_i - \mu)} + \ldots \\ &= \frac{-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i - \mu)}{\frac{1}{n} \sum_{i=1}^{n} \psi'(X_i - \mu)} + \ldots. \end{aligned} \tag{24}$$

Where the higher order terms go to zero. By the central limit theorem,

$$\sqrt{n} \left(-\frac{1}{n} \sum_{i=1}^{n} \psi(X_i - \mu) - 0 \right) \xrightarrow{d} N(0, \mathbb{E}_\mu \psi^2(X - \mu)), \tag{25}$$

and by the law of large numbers

$$\frac{1}{n} \sum_{i=1}^{n} \psi'(X_i - \mu) \xrightarrow{p} \mathbb{E}_\mu \psi'(X - \mu) \tag{26}$$

Therefore,

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{p} \frac{-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i - \mu)}{\frac{1}{n} \sum_{i=1}^{n} \psi'(X_i - \mu)} \xrightarrow{d} N \left(0, \frac{\mathbb{E}_\mu \psi^2(X - \mu)}{[\mathbb{E}_\mu \psi'(X - \mu)]^2} \right) \tag{27}$$

For the Huber function, we have

$$\mathbb{E}_\mu \psi'(X - \mu) = \int_{\mu-k}^{\mu+k} f(x - \mu) dx = P(|X| \leq k), \tag{28}$$

and

$$\begin{aligned} \mathbb{E}_\mu \psi^2(X - \mu) &= \int_{\mu-k}^{\mu+k} x^2 f(x - \mu) dx + k^2 \int_{-\infty}^{\mu-k} f(x - \mu) dx + k^2 \int_{\mu+k}^{\infty} f(x - \mu) dx \\ &= \int_{\mu-k}^{\mu+k} x^2 f(x - \mu) dx + 2k^2 P(X < -k). \end{aligned} \tag{29}$$

For the Cauchy distribution, we have

$$\mathbb{E}\psi'(X-\mu) = \int_{-k}^{k} f(x)dx = \int_{-k}^{k} \frac{1}{\pi(1+x^2)}dx = \frac{1}{\pi}\left(\arctan(k) - \arctan(-k)\right). \tag{30}$$

$$\begin{aligned}\mathbb{E}\psi^2(X-\mu) &= \int_{\mu-k}^{\mu+k} (x-\mu)^2 f(x-\mu)dx + \int_{-\infty}^{\mu-k} k^2 f(x-\mu)dx + \int_{\mu+k}^{\infty} k^2 f(x-\mu)dx \\ &= \int_{-k}^{k} x^2 \frac{1}{\pi(1+x^2)}dx + \int_{-\infty}^{-k} k^2 \frac{1}{\pi(1+x^2)}dx + \int_{k}^{\infty} k^2 \frac{1}{\pi(1+x^2)}dx.\end{aligned} \tag{31}$$

For the first integral, using the substitution $\tan u = x$, we have

$$\begin{aligned}\int x^2 \frac{1}{\pi(1+x^2)}dx &= \int \frac{1}{\pi}\tan^2(u)\cos^2(u)\sec^2(u)du \\ &= \frac{1}{\pi}\int \tan^2(u)du \\ &= \frac{1}{\pi}\int \sec^2(u) - 1du \\ &= \frac{1}{\pi}(\tan(u) - u) \\ &= \frac{1}{\pi}(x - \arctan(x)).\end{aligned} \tag{32}$$

We thus have

$$\begin{aligned}\mathbb{E}\psi^2(X-\mu) &= \frac{k^2}{\pi}(\arctan(-k) - \arctan(-\infty)) + \frac{k^2}{\pi}(\arctan(\infty) - \arctan(k)) \\ &+ \frac{1}{\pi}(k - \arctan(k) + k + \arctan(-k)) \\ &= \frac{k^2}{2} + \frac{2k}{\pi} - \frac{2}{\pi}\arctan(k).\end{aligned} \tag{33}$$

Putting this all together, the asymptotic variance of the Huber estimator is

$$\frac{\mathbb{E}\psi^2(X-\mu)}{[\mathbb{E}\psi'(X-\mu)]^2} = \frac{\frac{2k}{\pi} - \frac{2k^2}{\pi}\arctan(k)}{[\frac{2}{\pi}\arctan(k)]^2} = \frac{\pi}{2}\frac{k - k^2\arctan(k)}{\arctan(k)^2}. \tag{34}$$

As $k \to 0$, this converges to 1.