Stat 563
6-2-26

Laplace example

$$f(x\mid\theta) = \frac{1}{2} e^{-|x-\theta|} \qquad -\infty < x < \infty \qquad (1)$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i-\theta|}$$

$$= \frac{1}{2^n} e^{-\sum_{i=1}^{n} |x_i-\theta|}$$

$$\ell(\theta) = -n\ln 2 - \sum_{i=1}^{n} |x_i-\theta|$$

$$\ell'(\theta) = -\sum_{i=1}^{n} \begin{cases} -1 & x_i > \theta \\ 0 & x_i = \theta \\ 1 & x_i < \theta \end{cases} = \sum_{i=1}^{n} \begin{cases} 1 & x_i > \theta \\ 0 & x_i = 0 \\ -1 & x_i < \theta \end{cases}$$

$$\ell'(\theta) = \sum_{i=1}^{n} \operatorname{sgn}(x_i-\theta) \qquad (2)$$

$\psi(x-\theta) = \operatorname{sgn}(x-\theta)$ This is bounded between ± 1,

so the MLE of θ will be robust

Cauchy example

$$f(x\mid\theta) = \frac{1}{\pi\left(1+(x-\theta)^2\right)} \qquad -\infty < x < \infty$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\pi\left(1+(x_i-\theta)^2\right)}$$

3

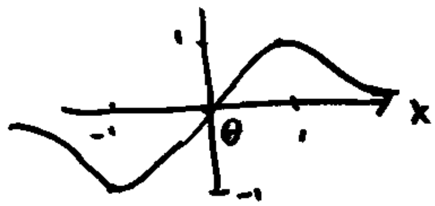
$$= \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1+(x_i-\theta)^2}$$

$$\ell(\theta) = -n \ln \pi + \sum_{i=1}^{n} -\ln\left(1+(x_i-\theta)^2\right)$$

$$\ell'(\theta) = -\sum_{i=1}^{n} \frac{1}{1+(x_i-\theta)^2} 2(x_i-\theta)(-1)$$

$$= \sum_{i=1}^{n} \frac{2(x_i-\theta)}{1+(x_i-\theta)^2}$$

$$\psi(x-\theta) = \frac{2(x-\theta)}{1+(x-\theta)^2}$$



4

This is bounded between ± 1, so the MLE will be robust (even though there is no closed form)

Summary:

Normal	Laplace	Cauchy
$\psi(x-\theta) = \frac{x-\theta}{\sigma^2}$	$\psi(x-\theta) = \mathrm{sgn}(x-\theta)$	$\psi(x-\theta) = \frac{2(x-\theta)}{1+(x-\theta)^2}$
$w(x-\theta) = \frac{1}{\sigma^2}$	$w(x-\theta) = \frac{\mathrm{sgn}(x-\theta)}{x-\theta}$	$w(x-\theta) = \frac{2}{1+(x-\theta)^2}$

(5)

The Huber influence function is

$$\psi(x-\theta) = \begin{cases} -k & x-\theta < -k \\ x-\theta & -k \le x-\theta \le k \\ k & x-\theta > k \end{cases}$$

This is bounded, so solving

$$\sum_{i=1}^{n} \psi(x_i - \theta) = 0 \quad \text{for } \theta$$

will give you an M-estimator that is robust.

(6)

Breakdown

Start with a sample of size n

- compute an estimator $\hat{\theta}$.

Find the smallest proportion $\frac{m}{n}$ of the sample values such that $|\hat{\theta}^* - \hat{\theta}|$ can be made arbitrarily large by corrupting m data values + computing $\hat{\theta}^*$.

This proportion is called the finite breakdown point

The limit as $n \to \infty$ is called the breakdown point. (7)

<u>Example</u>: $\bar{x}$

Replace $x_1, \dots, x_n$ with $x_1, \dots, x_{n-1}, x_n^*$

$$|\hat{\theta}^* - \hat{\theta}| = \left| \cancel{\frac{1}{n}\sum_{i=1}^{n-1} x_i} + \frac{1}{n} x_n^* - \left(\cancel{\frac{1}{n}\sum_{i=1}^{n-1} x_i} + \frac{1}{n} x_n\right)\right|$$

$$= \frac{1}{n} | x_n^* - x_n |$$

Finite breakdown point is $\frac{1}{n}$

Breakdown point is 0

(8)

<u>Example</u>: Trimmed mean $\bar{x}_p$

Trim $p\%$ of values ($\frac{p}{2}\%$ from each end) + average the remaining values

We would have to corrupt $\frac{p}{2}\%$ $\cdot n$ plus 1 values in order to corrupt $\bar{x}_p$.

Finite breakdown is $\dfrac{\frac{p}{2}\% \cdot n + 1}{n} = \frac{p}{2}\% + \frac{1}{n}$

Breakdown is $\frac{p}{2}\%$

(9)

Example : $\hat{\theta}$ = sample median , n is even

Finite breakdown is $\frac{n/2}{n} = \frac{1}{2}$

Breakdown $= \frac{1}{2}$

If n is odd,

finite breakdown is $\frac{\frac{n+1}{2}}{n} = \frac{1}{2} + \frac{1}{2n}$

breakdown $= \frac{1}{2}$

(10)

Let F be a continuous cdf

$X_1, \dots, X_n \sim$ iid F Assume n is odd

Let M be the population median,

i.e. $P(X_i \le M) = \frac{1}{2}$

Let m_n be the sample median,

i.e. $m_n = X_{(\frac{n+1}{2})}$

Let $Y_i = \begin{cases} 1 & X_i \le M + \frac{a}{\sqrt{n}} \\ 0 & \text{otherwise} \end{cases}$ for some a

(11)

Then $Y_i \sim \text{Bernoulli}\left(p_n = \underbrace{P\left(X_i \le M + \frac{a}{\sqrt{n}}\right)}_{F\left(M + \frac{a}{\sqrt{n}}\right)}\right)$

So $\sum_{i=1}^{n} Y_i \sim \text{Bino}(n, p_n)$

$X_{\left(\frac{n+1}{2}\right)} = m_n \le M + \frac{a}{\sqrt{n}}$ iff $\sum_{i=1}^{n} Y_i \ge \frac{n+1}{2}$

$$P\left(\sqrt{n}(m_n - M) \le a\right) = P\left(m_n \le M + \frac{a}{\sqrt{n}}\right)$$
$$= P\left(\sum_{i=1}^{n} Y_i \ge \frac{n+1}{2}\right)$$

(12)

$$= P\left(\underbrace{\frac{\sum_{i=1}^{n} Y_i - np_n}{\sqrt{np_nq_n}}}_{\downarrow \mathcal{D}} \ge \frac{\frac{n+1}{2} - np_n}{\sqrt{np_nq_n}}\right)$$

$N(0,1)$ by Central Limit Theorem

13

$$\lim_{n\to\infty} \frac{\frac{n+1}{2} - np_n}{\sqrt{np_nq_n}} = \lim_{n\to\infty} \frac{n\left(\frac{1}{2} - p_n\right) + \frac{1}{2}}{\sqrt{np_nq_n}}$$

$$= \lim_{n\to\infty} \frac{n\left(\frac{1}{2} - p_n\right)}{\sqrt{np_nq_n}} + 0$$

$$= \lim_{n\to\infty} \frac{\sqrt{n}\left(\frac{1}{2} - p_n\right)}{\sqrt{p_nq_n}} = \lim_{n\to\infty} \frac{1}{\sqrt{p_nq_n}} \cdot \lim_{n\to\infty} \sqrt{n}\left(\tfrac{1}{2} - p_n\right)$$

$$= 2 \lim_{n\to\infty} \frac{\frac{1}{2} - p_n}{1/\sqrt{n}} = 2 \lim_{n\to\infty} \frac{-\frac{dp_n}{dn}}{-\frac{1}{2}n^{-3/2}}$$

since $p_n \to \frac{1}{2}$

14

what is $\frac{dp_n}{dn}$?

$$p_n = P\left(X_i \le M + \frac{a}{\sqrt{n}}\right) = F\left(M + \frac{a}{\sqrt{n}}\right)$$

$$\frac{dp_n}{dn} = f\left(M + \frac{a}{\sqrt{n}}\right) \cdot a\left(-\tfrac{1}{2}\right)n^{-3/2}$$

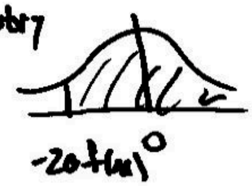
$$2 \lim_{n\to\infty} \cancel{2n^{3/2}}\, f\left(M + \frac{a}{\sqrt{n}}\right) a \cancel{\left(-\tfrac{1}{2}\right)n^{-3/2}} = -2f(M)a$$

(15)

So $\lim\limits_{n\to\infty} P\left[\sqrt{n}\,(m_n - M) \le a\right]$

$= P\left[Z \ge -2af(M)\right]$ where $Z \sim N(0,1)$

$= P\left[Z \le 2af(M)\right]$ by symmetry



Let $z = 2af(M)$

$a = \dfrac{z}{2f(M)}$

$$\lim_{n\to\infty} P\left[\sqrt{n}\,(m_n - M) \le \frac{z}{2f(M)}\right] = P\left[Z \le z\right] = \Phi(z)$$

(16)

$$\lim_{n\to\infty} P\left[\frac{m_n - M}{\frac{1}{2\sqrt{n}f(M)}} \le z\right] = \Phi(z)$$

That is, $\dfrac{m_n - M}{\left(\frac{1}{2\sqrt{n}f(M)}\right)} \xrightarrow{\mathcal{D}} N(0,1)$

For large n, $m_n \approx N\left(M, \dfrac{1}{4nf^2(M)}\right)$

The sample median is $\begin{cases} \text{asymptotically normal} \\ \quad \text{"} \qquad \text{unbiased} \\ \text{consistent} \end{cases}$

(17)

Consider the specific case of the normal distribution. We have $\mu = M$

$$f(M) = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

For large n, $m_n \simeq N\left(\mu, \dfrac{1}{4n\frac{1}{\sigma^2 2\pi}}\right)$

$$\simeq N\left(\mu, \frac{\pi\sigma^2}{2n}\right)$$

Compare this to $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

(18)

(approximate)
relative efficiency of m_n to $\bar{x}$

$$= \frac{V(\bar{x})}{V(m_n)} = \frac{\sigma^2/n}{\frac{\pi}{2}\frac{\sigma^2}{n}} = \frac{2}{\pi}$$

The asymptotic rel. eff. of m_n to $\bar{x}$ is $\frac{2}{\pi} \approx .64$

C.I. for μ based on $\bar{x}$:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

C.I. for μ based on m_n:

$$m_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{\pi}{2}}$$