Stat 563
5-30-19

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

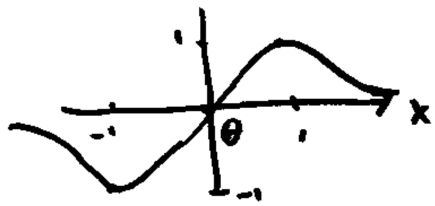
$$= \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1+(x_i-\theta)^2}$$

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

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

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

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



4

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

Summary:

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

(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

7

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

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

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

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

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

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 $\frac{p}{2}\%$ of 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)

$$= P\left(\underbrace{\frac{\sum_{i=1}^{n} Y_i - np_n}{\sqrt{np_n q_n}}}_{\xrightarrow{D}\ N(0,1)} \ge \frac{\frac{n+1}{2} - np_n}{\sqrt{np_n q_n}}\right)$$

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

This will be finished on Tuesday.

Also, the Final Exam will be handed out on Tuesday.

Due Monday 6/10 5pm.