

Recitation: M Estimation

M estimation is process of estimating a parameter for a distribution using some kind of minimisation

M-estimation

$\text{d} X_1, \dots, X_n \text{ iid } P_\mu = f(x-\mu)$
↳ pdf "centered" at 0

Choose a fcn ρ
e.g. $\rho(x) = |x|$
 $\rho(x) = x^2$

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

1) Calculate asympt var (avar)
of sample med.

2) Compare avar for
o Mean & Med under Laplace(μ)
o Mean, Med, Huber est. under
Cauchy(μ)

- trying to find the mu that is the centre of the distribution
- try to find a b that minimises the distance between mu and mu_hat
- rho are different loss functions
- if one estimator has a lower asymptotic variance than another then it is deviating from our value less

Calculate asymptotic variance for sample median

1) Assume f cts, n is odd

Sample med m_n

$$\circ \sqrt{n}(m_n - \mu) \xrightarrow{(d)} N(0, \sigma^2)$$

(df:

$$P(\sqrt{n}(m_n - \mu) \leq a) = P\left(m_n \leq \frac{a}{\sqrt{n}} + \mu\right)$$

$$= P\left(\#\left(X_i \geq \frac{a}{\sqrt{n}} + \mu\right) \geq \frac{n+1}{2}\right)$$

$$= P\left(\bar{Y}_n \geq \frac{n+1}{2}\right), Y_i = \mathbb{1}(X_i \geq \frac{a}{\sqrt{n}} + \mu)$$

- the number of values greater than the median has to be at least half of the values
- Y_i is now Bernoulli, but it is dependent on n

$$= P\left(\bar{Y}_n \geq \frac{n+1}{2}\right), (Y_i = \mathbb{1}(X_i \geq \frac{a}{\sqrt{n}} + \mu))$$

$\hookrightarrow \sim \text{Bern}(p_n)$

$$P\left(\frac{\bar{Y}_n - p_n \cdot n}{\sqrt{n(1-p_n)p_n}} \geq \frac{\frac{n+1}{2} - p_n \cdot n}{\sqrt{n(1-p_n)p_n}}\right)$$

- p_n is the median (and parameter of the Bernoulli)
- as $n \rightarrow \infty$ p_n goes to 1/2

$$= P\left(Z \geq \frac{\frac{n+1}{2} - p_n \cdot n}{\sqrt{n(1-p_n)p_n}}\right)$$
$$Z \sim N(0, 1)$$

Continuing calculation

1) Assume f cts, n is odd, $F(\mu) = \frac{1}{2}$

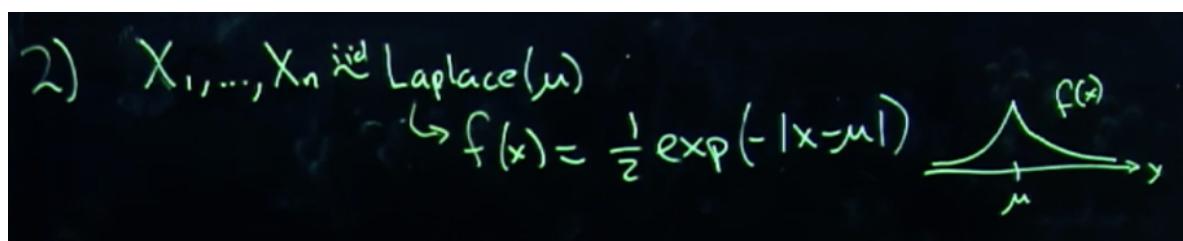
- Sample med m_n
- $P(\sqrt{n}(m_n - \mu) \leq a) \rightarrow P(Z \geq \lim_{n \rightarrow \infty} \frac{\frac{n+1}{2} - \mu n}{\sqrt{n(1-p_n)p_n}})$
- $\frac{\frac{n+1}{2} - \mu n}{\sqrt{n(1-p_n)p_n}} = \frac{\frac{1}{2} - p_n n}{\sqrt{n(1-p_n)p_n}}$ $(Z \sim N(0,1))$
- $p_n = P(Y_i = 1) = P(X_i \leq \frac{a}{\sqrt{n}} + \mu) \rightarrow \frac{1}{2}$
- $= \frac{\frac{1}{2} - p_n n}{\sqrt{n(1-p_n)p_n}} + \frac{\frac{1}{2}}{\sqrt{n(1-p_n)p_n}} \xrightarrow[n \rightarrow \infty]{\rightarrow 0}$
- $= \frac{\frac{1}{2} - p_n}{\frac{\sqrt{(1-p_n)p_n}}{\sqrt{n}}} = \frac{F(\mu) - F(\frac{a}{\sqrt{n}} + \mu)}{\frac{a}{\sqrt{n}}} \cdot \frac{a}{\sqrt{(1-p_n)p_n}}$
- $= -F'(\mu) \cdot 2a = -2af(\mu)$
- o (i) $\rightarrow P(Z \geq -2af(\mu)) = P(-Z \leq 2af(\mu))$

- showing that the initial Probability (i, above) converges in probability to :

$$= P\left(\frac{Z}{2af(\mu)} \leq a\right) \sim N\left(0, \frac{1}{4f(\mu)^2}\right)$$

- this gives a general formula for the asymptotic variance

Compare asymptotic variances for Laplace (mean and median)



- graph of probability density function

$$\sqrt{n}(\bar{x} - \mu) \xrightarrow{d} N(0, \text{Var}(x_i))$$

$$\text{avar}(\bar{x}) = \text{var}(x_i) = \text{var}(x'_i) \quad , \quad x'_i \sim \text{Laplace}(\mu)$$

$$= E(x'_i)^2 = 2$$

- asymptotic variance of the sample mean for a Laplace = 2

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

$$\text{avar}(m_n) = \frac{1}{4f(\mu)^2} = \frac{1}{4} \cdot \frac{1}{\left[\frac{1}{2} \exp(-|x-\mu|)\right]^2}$$

$$= 1$$

- asymptotic variance of the sample median for a Laplace (from previous section) = 1
- this means the median is a better estimator than the mean

Compare asymptotic variances for Cauchy (mean and median)

2) $X_1, \dots, X_n \sim \text{Cauchy}(\mu)$ if

pdf: $f(x) = \frac{1}{\pi(1 + (x - \mu)^2)}$

cdf: $F(x) = \int_{-\infty}^x f(t) dt$

$$= \int_{-\infty}^x \frac{1}{\pi(1 + (t - \mu)^2)} dt$$

$$= \frac{\arctan(t - \mu)}{\pi} \Big|_{-\infty}^x = \frac{1}{2} + \frac{\arctan(x - \mu)}{\pi}$$

Note: $F(\mu) = \frac{1}{2}$

Let's use \bar{X} to estimate μ

$$E\bar{X} = E\bar{X}_i = \int_{-\infty}^{\infty} x \cdot \frac{1}{\pi(1+x^2)} dx \text{ diverges}$$

$$\text{Var}(\bar{X}) = \frac{1}{n} \text{Var}(X_i) = \text{diverges}$$

- has no variance or mean, they are both infinity so diverge
- same is true for asymptotic variance

so try m_n instead (sample median)

$\text{avar}(m_n) = \frac{1}{4f(\mu)^2} = \frac{1}{4} \cdot \frac{1}{\left[\frac{1}{\pi}\right]^2} = \frac{\pi^2}{4}$

- the median is a better estimator than the mean for Cauchy

Huber estimator for Cauchy

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

- Huber switches to a quadratic between $-\delta$ and δ to make it continuously differentiable
-

$$\rho'(x) = \begin{cases} x & |x| < \delta \\ \delta \text{sgn}(x) & |x| \geq \delta \end{cases}$$

$$\rho''(x) = \begin{cases} 1 & |x| < \delta \\ 0 & |x| \geq \delta \end{cases}$$

Huber est: $\hat{\mu}_{\text{huber}} = \underset{b \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \rho(x_i - b)$

$\sqrt{n}(\hat{\mu}_{\text{huber}} - \mu) \xrightarrow{(d)} N(0, \frac{\text{Var}(\rho'(x))}{[\mathbb{E} \rho''(x)]^2})$

$X \sim f(x), \quad X_i \sim f(x - \mu)$

- calculate these quantities to get the asymptotic variance

$$\begin{aligned}
 \bullet \mathbb{E} \rho''(x) &= \int_{-\delta}^{\delta} 1 \cdot f(x) dx = \int_{-\delta}^{\delta} \frac{1}{\pi(1+x^2)} dx \\
 &= \frac{1}{\pi} (\arctan \delta - \arctan(-\delta)) \\
 &= \frac{2}{\pi} \arctan \delta
 \end{aligned}$$

$$\begin{aligned}
 \bullet \text{Var}(\rho'(x)) &= \mathbb{E}(\rho'(x))^2 = \int_{-\infty}^{-\delta} \int_{-\infty}^{\delta} x^2 \cdot \frac{1}{\pi(1+x^2)} dx \\
 &\quad + \int_{-\delta}^{\delta} \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\pi(1+x^2)} dx + \int_{\delta}^{\infty} \int_{-\infty}^{\delta} x^2 \cdot \frac{1}{\pi(1+x^2)} dx
 \end{aligned}$$

- because $E[\rho'(x)]$ is 0

-

calculating integrals

$$\begin{aligned}
 \bullet \mathbb{E} \rho''(x) &= \frac{2}{\pi} \arctan \delta, \quad x \sim f(x) \\
 \bullet \mathbb{E}(\rho'(x))^2 &= \int_{-\infty}^{-\delta} \int_{-\infty}^{\delta} x^2 \frac{1}{\pi(1+x^2)} dx + \int_{-\delta}^{\delta} \int_{-\infty}^{\infty} x^2 \frac{1}{\pi(1+x^2)} dx + \int_{\delta}^{\infty} \int_{-\infty}^{-\delta} x^2 \frac{1}{\pi(1+x^2)} dx \\
 &= \frac{\delta^2}{\pi} \left(\arctan(-\delta) - \left(-\frac{\pi}{2} \right) \right) + \frac{\delta^2}{\pi} \left(\frac{\pi}{2} - \arctan(\delta) \right) \\
 &\quad + \frac{2\delta}{\pi} - \frac{2 \arctan \delta}{\pi}
 \end{aligned}$$

- calculation of 3rd term below, subbing in trigonometric identities for integrating

$$\begin{aligned}
 & \cdot \int \frac{x^2}{\pi(1+x^2)} dx \quad | + \tan^2 x = \sec^2 x \\
 & \quad \tan u = x \rightarrow \sec^2 u du = dx \\
 & \approx \int \frac{\tan^2 u \sec^2 u du}{\pi(\sec^3 u)} = \frac{1}{\pi} \int \sec u - 1 du \\
 & = \frac{1}{\pi} (\tan u - u) \\
 & \left. \int_{-\delta}^{\delta} \frac{x^2}{\pi(1+x^2)} dx = \frac{1}{\pi} (x - \arctan x) \right|_{-\delta}^{\delta} \\
 & = \frac{1}{\pi} (\delta - \arctan \delta - (-\delta - \arctan(-\delta))) \\
 & = \frac{2\delta}{\pi} - \frac{2 \arctan \delta}{\pi}
 \end{aligned}$$

Asymptotic variance of Huber estimator for Cauchy distribution

$$\begin{aligned}
 & \cdot \text{avar}(\hat{\mu}_{\text{Huber}}) = \frac{\mathbb{E}(\rho'(x)^2)}{[\mathbb{E}\rho''(x)]^2} \\
 & = \frac{\frac{2\delta}{\pi} - \frac{2 \arctan \delta}{\pi} + \delta^2 - 2 \frac{\delta^2}{\pi} \arctan(\delta)}{\left[\frac{2}{\pi} \arctan(\delta) \right]^2}
 \end{aligned}$$

$$\cdot \text{avar}(\hat{\mu}_{\text{Huber}}) \xrightarrow{\delta \rightarrow 0} \frac{\pi^2}{4} = \text{avar}(m_n)$$

- for delta $\rightarrow 0$, the asymptotic variance converges to median

$$\cdot \delta \rightarrow \infty, \text{avar}(\hat{\mu}_{\text{Huber}}) \text{ diverges}$$

- for delta $\rightarrow \infty$, the asymptotic variance diverges

