

2: Location and Scale

March 14, 2022 7:14 PM

2.1 Location Model

$$x_i = \mu + u_i \quad (i = 1, \dots, n)$$

- Shift equivariant

$$\hat{\mu}(x_1 + c, \dots, x_n + c) = \hat{\mu}(x_1, \dots, x_n) + c$$

- Scale equivariant

$$\hat{\mu}(cx_1, \dots, cx_n) = c\hat{\mu}(x_1, \dots, x_n).$$

2.2 Formalizing departures from normality

- Normal mixture

$$\begin{aligned} F(t) &= P(x \leq t) = P(x \leq t|A')P(A') + P(x \leq t|A)P(A) \\ &= G(t)(1 - \epsilon) + H(t)\epsilon. \end{aligned}$$

where

$$1 - \epsilon$$

observations generated by normal model

2.3 M-estimators of location

1. Standard maximum likelihood (ML) estimator

$$\sum_{i=1}^n \left\{ \log f_0 \left(\frac{x_i - \mu}{\sigma} \right) - \log \sigma \right\}.$$

- ML-estimator of location

$$\frac{1}{n} \sum_{i=1}^n \psi_{ML} \left(\frac{x_i - \hat{\mu}_{ML}}{\sigma} \right) = 0,$$

- ML-estimator of scale verifies

$$\frac{1}{n} \sum_{i=1}^n \rho_{ML} \left(\frac{x_i - \mu}{\hat{\sigma}_{ML}} \right) = 1,$$

where

$$\psi_{ML}(u) = -f'_0(u)/f_0(u)$$

and

$$\rho_{ML}(u) = \psi_{ML}(u)u.$$

- Properties of ML-estimator: (1) 100% efficient (asymptotic variance equals the inverse of the Fisher information, the lower bound of the Cramer-Rao inequality) (2) not robust

2. General M-estimators of locations

- M-estimator of location is defined as the solution of the equation (not necessarily the score function)
can be any non-decreasing odd function)

$$\sum_{i=1}^n \psi \left(\frac{x_i - \mu}{\hat{\sigma}_0} \right) = 0.$$

- Some choices for ψ
 - Huber

$$\rho_k(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq k \\ k(|x| - \frac{k}{2}) & \text{if } |x| > k \end{cases}$$

$$\psi_k(x) = \begin{cases} x & \text{if } |x| \leq k \\ k \operatorname{sign}(x) & \text{if } |x| > k \end{cases}$$

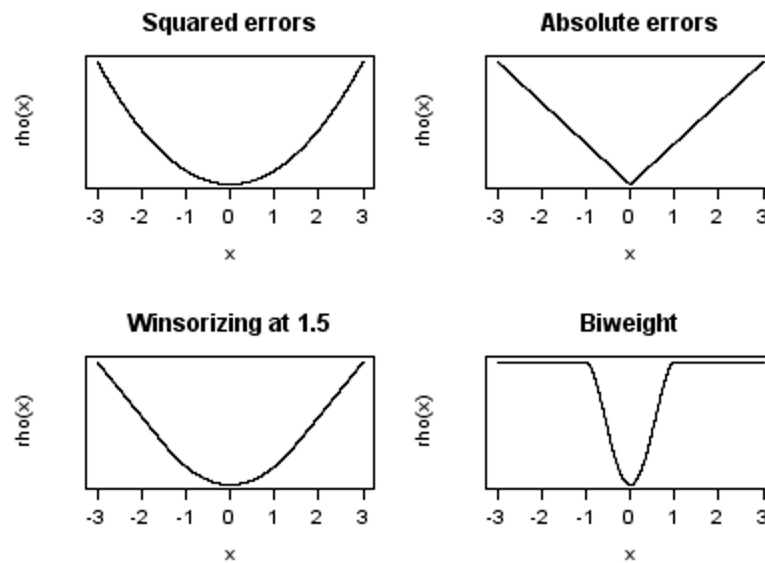
- Bisquare

$$\rho(x) = \begin{cases} 1 - [1 - (x/k)^2]^3 & \text{if } |x| \leq k \\ 1 & \text{if } |x| > k \end{cases}$$

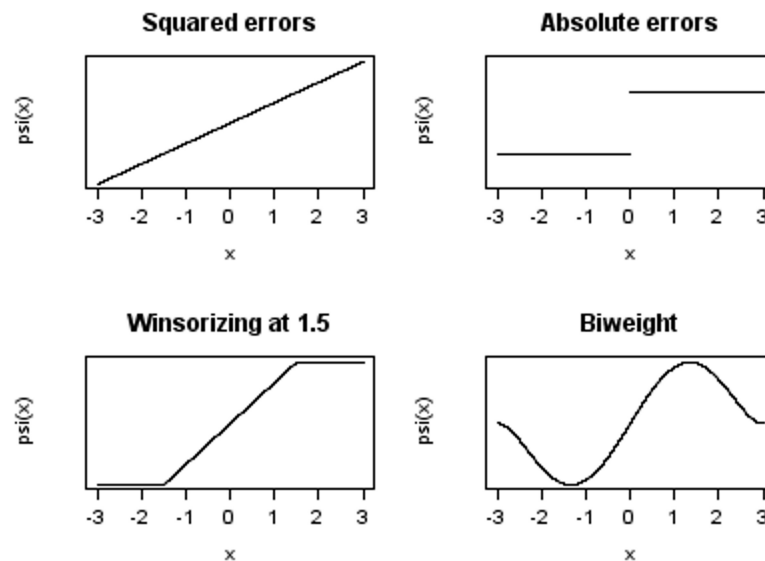
$$\rho'(x) = 6\psi(x)/k^2$$

$$\psi(x) = x \left[1 - \left(\frac{x}{k} \right)^2 \right]^2 I(|x| \leq k).$$

- Errors using different rho functions



psi functions



- Redescending M-estimators vs Monotone M-estimators:

- Redescending: for symmetric heavy tailed distributions, it is better to use "redescending" ψ that tend to zero at infinity, implies that for large x-values, the ρ
- Monotone: ψ that

2.4 Trimmed and Winsorized means

- α -trimmed mean: Trim a proportion of α from both ends of the data set and then take the mean.

$$\bar{x}_\alpha = \frac{1}{n-2m} \sum_{i=m+1}^{n-m} x_{(i)},$$

- α -Winsorized mean: Replace a proportion of α from both ends of the data set by the next closest observation and then take the mean

$$\tilde{x}_\alpha = \frac{1}{n} \left(mx_{(m)} + mx_{(n-m+1)} + \sum_{i=m+1}^{n-m} x_{(i)} \right),$$

- Example: 2, 4, 5, 10, 200

Mean = 44.2 Median = 5

- 20% trimmed mean = $(4 + 5 + 10) / 3 = 6.33$
20% Winsorized mean = $(4 + 4 + 5 + 10 + 10) / 5 = 6.6$

2.5 M-estimator of scale

- Delta = M-estimator of scale

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{x_i}{\hat{\sigma}} \right) = \delta,$$

- Equivariant

$$\hat{\sigma}(c\mathbf{x}) = |c| \hat{\sigma}(\mathbf{x})$$

2.6 Dispersion estimators

- Traditional standard deviation

$$SD(\mathbf{x}) = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}.$$

shift and scale equivariant

$$SD(\mathbf{x} + c) = SD(\mathbf{x}), \quad SD(c\mathbf{x}) = |c| SD(\mathbf{x}).$$

- Mean absolute deviation (MD)

$$MD(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

$$MD(x) = E|x - Ex|$$

- Median absolute deviation

$$MAD(\mathbf{x}) = \text{Med}(|\mathbf{x} - \text{Med}(\mathbf{x})|)$$

- Interquartile range

$$IQR(x) = x_{(n-m+1)} - x_{(m)}$$

2.7 M-estimators of location with unknown dispersion

- Previous estimation of dispersion: to obtain scale equivariant M-estimators of location, use

$$\hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n \rho \left(\frac{x_i - \mu}{\hat{\sigma}} \right),$$

$$\hat{\mu} = \arg \min_{\mu} \sum_{i=1}^n \rho \left(\frac{x_i - \mu}{\hat{\sigma}} \right),$$

- Simultaneous M-estimators of location and dispersion: two unknown parameters

$$x_i = \mu + \sigma u_i$$

$$(\hat{\mu}, \hat{\sigma}) = \arg \max_{\mu, \sigma} \frac{1}{\sigma^n} \prod_{i=1}^n f_0 \left(\frac{x_i - \mu}{\sigma} \right)$$

The solutions are

$$\sum_{i=1}^n \psi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^n \rho_{\text{scale}} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = \delta,$$

where

$$\psi(x) = -\rho'_0, \quad \rho_{\text{scale}}(x) = x\psi(x), \quad \delta = 1.$$

2.8 Numerical Computation of M-estimators

- Algorithm to compute robust location M-estimator
 1. Compute $\hat{\sigma} = \text{MADN}(x)$ and $\mu_0 = \text{Med}(x)$.
 2. For $k = 0, 1, 2, \dots$, compute the weights (2.77) and then $\hat{\mu}_{k+1}$ in (2.78).
 3. Stop when $|\hat{\mu}_{k+1} - \hat{\mu}_k| < \varepsilon \hat{\sigma}$.
- Algorithm to compute robust scale M-estimator
 1. For $k = 0, 1, 2, \dots$, compute the weights (2.79) and then $\hat{\sigma}_{k+1}$ in (2.80).
 2. Stop when $|\hat{\sigma}_{k+1}/\hat{\sigma}_k - 1| < \varepsilon$.

2.9 Robust Confidence Intervals

- Confidence Intervals: Robust confidence intervals that are not much influenced by outliers can be obtained by imitating the form of the classical Student t confidence interval, but replacing the average and SD by robust location and dispersion estimators
- Tests: we conclude that if the data are symmetric but heavy tailed, the intervals will be longer than necessary, with the consequence that the actual Type 1 error rate may be much smaller than α , but the Type 2 error rate may be too large; that is, the test will have low power
 - o use robust t statistic