2: Location and Scale

March 14, 2022 7:14 PM

2.1 Location Model

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

- Shift equivariant

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

- Scale equivariant

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

2.2 Formalizng departures from normality

- Normal mixture

$$F(t) = \mathrm{P}(x \leq t) = \mathrm{P}(x \leq t|A')P(A') + \mathrm{P}(x \leq t|A)P(A)$$

$$= G(t)(1 - \varepsilon) + H(t)\varepsilon.$$

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

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

   - ML-estimator of scale verifies

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

     where

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

     and

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

   - 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

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

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

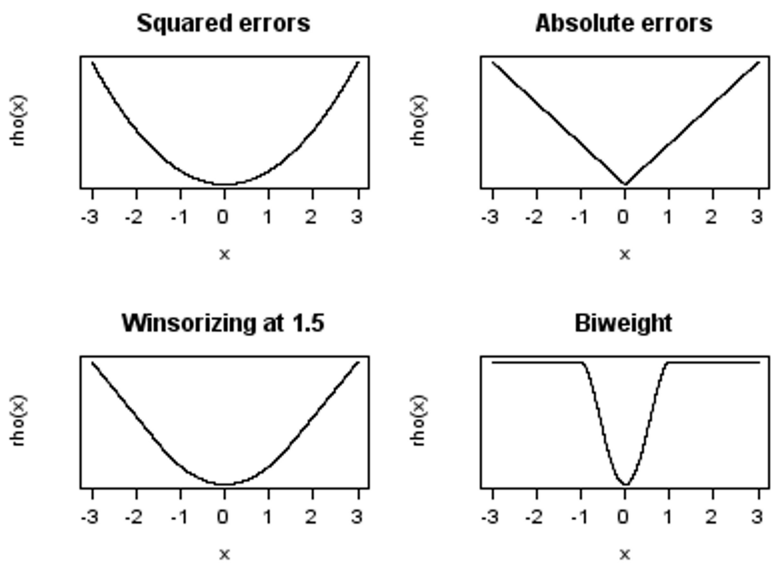
- Bisquare

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

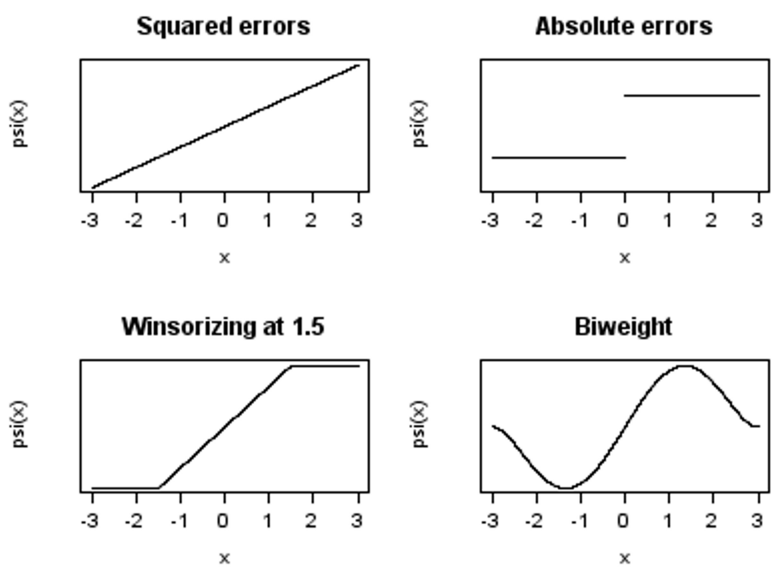
$$\rho'(x) = 6\psi(x)/k^2$$

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

- 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

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

- • Example: 2, 4, 5, 10, 200
  Mean = 44.2 Median = 5
  20% trimmed mean = (4 + 5 + 10) / 3 = 6.33
  20% Windsorized 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

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

shift and scale equivariant

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

- Mean absolute deviation (MD)

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

$$\mathrm{MD}(x) = \mathrm{E}|x - \mathrm{E}x|$$

- Median absolute deviation

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

- Interquartile range

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

2.7 M-estimators of location with unknown dispersion

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

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

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

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

  $x_i = \mu + \sigma u_i$

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

  The solutions are

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

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

where

$$\psi(x) = -\rho_0', \ \rho_{\text{scale}}(x) = x\psi(x), \ \delta = 1.$$

2.8 Numerical Computation of M-estimators

- Algorithm to compute robust location M-estimator

  1. **1.** Compute $\widehat{\sigma} = \text{MADN}(x)$ and $\mu_0 = \text{Med}(\mathbf{x})$.
  2. **2.** For $k = 0, 1, 2, \ldots,$ compute the weights (2.77) and then $\widehat{\mu}_{k+1}$ in (2.78).
  3. **3.** Stop when $|\widehat{\mu}_{k+1} - \widehat{\mu}_k| < \varepsilon\widehat{\sigma}$.

- Algorithm to compute robust scale M-estimator

  1. **1.** For $k = 0, 1, 2, \ldots,$ compute the weights (2.79) and then $\widehat{\sigma}_{k+1}$ in (2.80).
  2. **2.** Stop when $|\widehat{\sigma}_{k+1}/\widehat{\sigma}_k - 1| < \varepsilon$.

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
  - use robust *t* statistic