SURE
Stein’s Unbiased Risk Estimate

Bradley Efron
Stanford University

A Symposium in Probability and Statistics
In honor of Charles Stein
on his 90th Birthday
March 22, 2010
Mallows’ Cp

- **Observe** \( y \sim (\mu, \sigma^2 I) \)
- **Estimate** \( \hat{\mu} = My \)
- **Future Vector** \( y^0 \sim (\mu, \sigma^2 I) \)
- **Prediction Error** \( \text{Err} = ||y^0 - \hat{\mu}||^2 \)
- **Apparent Error** \( \text{err} = ||y - \hat{\mu}||^2 \) usually too optimistic
- **Mallows’ Cp**
  \[
  \widehat{\text{Err}} = \text{err} + 2\sigma^2 \text{trace}(M) \]
  “degrees of freedom”
- \( E\{\widehat{\text{Err}}\} = E\{\text{Err}\} \)
Stein’s Unbiased Risk Estimate (1981)

- \( \hat{\mu} = m(y) \), not necessarily linear but smoothly differentiable

- Covariance Penalty
  
  \[
  E\{\text{Err}\} = E \{ \text{err} + 2 \sum_{i=1}^{n} \text{cov} (\hat{\mu}_i, y_i) \} \leftarrow \text{“cov penalty”}
  \]

- Normal Case
  
  \( y \sim N_n(\mu, \sigma^2 I) \): \( \text{cov} (\hat{\mu}_i, y_i) = \sigma^2 E \{ \partial \hat{\mu}_i / \partial y_i \} \)

SURE

\[
\hat{\text{Err}} = \text{err} + 2\sigma^2 \sum_{i=1}^{n} \frac{\partial \hat{\mu}_i}{\partial y_i} \leftarrow \text{“df”}
\]
From the Kidney Laboratory

- **Observe** \((x_i, y_i)\) \(i = 1, 2, \ldots, n = 157\)
  \[
  \begin{cases}
    x_i = \text{age} \\
    y_i = \text{kidney score}
  \end{cases}
  \]

- **Fitted Curve** \(\hat{\mu} = \text{lowess}(x, y, \text{window} = 1/3)\)

- **err** = \(||y - \hat{\mu}||^2 = 495\)

- \(\hat{\sigma}^2 = 495/156 = 3.17\)

- How well would \(\hat{\mu}\) predict the next 157 scores?
kidney function vs age for 157 healthy volunteers; fitted curve from lowess(f=1/3); residual sum of squares 495

\[ x = \text{age} \]
\[ y = \text{total kidney function} \]

estimated standard deviation 3.17
Applying SURE

- **Numerically Evaluate** \( \frac{\partial \hat{\mu}_i}{\partial y_i} \) using \( y_i \pm \epsilon \):
- \( 2\hat{\sigma}^2 \sum_{i=1}^{157} \frac{\partial \hat{\mu}_i}{\partial y_i} = 43.4 \)
- \( \hat{\text{Err}} = 495 + 43.4 = 538.4 \)
- \( \hat{df} = 6.85 \)
Circles are componentwise SURE covariance penalties, sum=43.4; line shows parametric bootstrap estimates, sum=42.3; both add 9%
Parametric Bootstrap

• \( y \sim N_N(\mu, \sigma^2 I) \)

• \( \text{Boot } y^* \sim N_N(\hat{\mu}, \hat{\sigma}^2 I) \)

• \( \hat{\mu}^* = m(y^*) \), \( "m" = \text{lowess(age,tot^*,1/3)} \)

• \( \text{Estimate } \text{cov}(\hat{\mu}_i, y_i) \) from boot covariance of \( (\hat{\mu}_i^*, y_i^*) \)

• Gave \( 2 \sum \text{cov}_i = 42.3, \ \hat{df} = 6.67 \)
The “Steinian” (Efron 2004, JASA)

- \( y \sim \text{Bern}(\mu) \) gives \( \hat{\pi} = m(y) \)
- Prediction error = \#\{on different sides of 1/2\}
- Define \( y_{i1} = y \) with \( i \)th entry 1 and \( y_{i0} = y \) with \( i \)th entry 0, giving \( \hat{\pi}_{i1} \) and \( \hat{\pi}_{i0} \)
- \[ \widehat{\text{Err}} = \text{err} + 2 \sum_i \hat{\pi}_i (1 - \hat{\pi}_i) \cdot D_i \]
  with \( D_i \) equal 2 or 0 as \( \hat{\pi}_{i1}(i), \hat{\pi}_{i0}(i) \) on different or same sides of 1/2.
So in SuperBowl Notation:

Happy XCth Birthday to Super Charles!