Lecture 15: The James-Stein Estimator

I once met Stein! :)

- MLE great bc gives, in general, nearly unbiased estimates of nearly minimum variance & it is automatic.
- However, MLE can be dangerous to use in high-dimensional modern applications (it's OK for low-dim generally speaking)
- James-Stein made this evident even in the context of just a few unknown pars.
- Begins the story of shrinkage estimation where deliberate biases are introduced (hurts the individual parameter estimates, but it's beneficial in the overall).
- As we quickly covered in Ch 5, if wish to estimate one-dimensional μ from one observation x, using Bayesian reasoning:

Normal Prior, Normal Likelihood

If $\mu \sim N(M, A)$ and $x \mid \mu \sim N(\mu, 1)$

Then $\mu \mid X \sim N(M + B(X - M), B)$ where $\boxed{B = \frac{A}{(A+1)}}$

Normal Posterior

$$\left.\Rightarrow \hat{\mu}^{Bayes} = M + B(X - M)\right\}$$

Effect of X on our belief of μ is shrunk towards M, the prior mean.

$$\hat{\mu}^{MLE} = X$$

It can be shown that

Expected squared error $= E\{(\hat{\mu}^{Bayes} - \mu)^2\} = B$

True par.

while

$$E[(\hat{\mu}^{MLE} - \mu)^2] = 1$$

$\Rightarrow$ Then $\hat{\underline{\mu}}^{Bayes}$ has smaller ESE (risk)!

In the same way

$$\hat{\underline{\mu}}^{Bayes} = \underline{M} + B \cdot (\underline{x} - \underline{M})$$

$$\underline{M} = [M, M, M, \cdots, M]$$

Further, the total squared error of $\hat{\underline{\mu}}^{Bayes}$ is

$$E\{\|\hat{\underline{\mu}}^{Bayes} - \underline{\mu}\|^2\} = E\left\{\sum_{i=1}^{N} (\hat{\mu}_i^{Bayes} - \mu_i)^2\right\} = N \cdot B$$

$$E\{\|\hat{\underline{\mu}}^{MLE} - \underline{\mu}\|^2\} = N$$

This sounds cool if we only knew M and B.
However, since we don't know them, we...

Estimate them!

$\hat{M} = \bar{x}$ (unbiased estimate of M).

$$\hat{B} = 1 - \frac{(N-3)}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$$\boxed{\hat{\underline{\mu}}^{JS} = \hat{\underline{M}} + \hat{B}(\underline{x} - \hat{\underline{M}})}$$

James - Stein Estimator !

$\hat{\underline{M}} = [\hat{M}, \hat{M}, \ldots, \hat{M}]$

Further, it can be shown that

$$E\{\|\hat{\underline{\mu}}^{JS} - \underline{\mu}\|^2\} = NB + 3(1-B)$$

James - Stein Theorem

If $\boxed{x_i | \mu_i \sim N(\mu_i, 1)}$

independently for $i = 1, 2, \ldots, N$ with $N \geq 4$

Then, $E\{\|\hat{\underline{\mu}}^{JS} - \underline{\mu}\|^2\} < N = E\{\|\hat{\underline{\mu}}^{MLE} - \underline{\mu}\|^2\}$

Boom !

for all choices of $\underline{\mu} \in \mathbb{R}^N$

The Baseball Players Example

Like, for 18 different players

- 18 observations of batting averages

over the first 90 at-bats of a season, (1970)

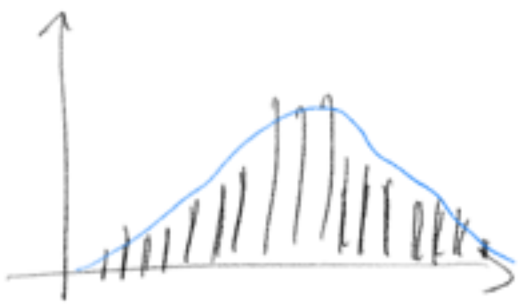
. we would like to estimate the true value of each player's batting average.

Batting avg. over the whole season.

what might this refer to?

Then, it is quite reasonable (& standard) to assume

$$\underbrace{p_i}_{\substack{\text{Proportion of}\\ \text{hits for player } i\\ \text{(Sample Batting Avg)}}} \sim \text{Bin}(90, \underbrace{\pi_i}_{\substack{\text{True proportion of}\\ \text{hits for player } i\\ \text{(True batting avg)}}}) / 90$$



$$p_i \overset{a}{\sim} N(\pi_i, \sigma_0^2)$$

$$\sigma_0^2 = \frac{\cancel{90} \cdot \bar{p}(1-\bar{p})}{90^{\cancel{2}}} = \frac{\bar{p}(1-\bar{p})}{90}$$

$V(a \cdot X) = a^2 \cdot V(X)$

$$X_i = p_i / \sigma_0$$

Plug in into the $\hat{\mu}^{JS}$ estimator, get

estimate, & translate back $\hat{p}^{JS} = \sigma_0 \underbrace{\hat{\mu}_i^{JS}}$

The one we computed using the X_i's

Turns out that doing this, renders equation

$$\hat{p}_i^{JS} = \bar{p} + \left[1 - \frac{(N-3)\sigma_0^2}{\Sigma(p_i - \bar{p})^2} \right] (p_i - \bar{p})$$