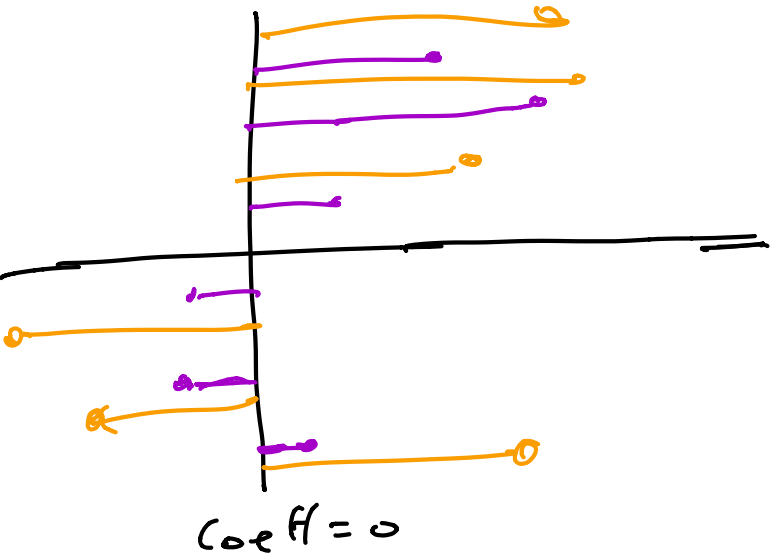
- Previously, we covered ridge regression, which is a popular shrinkage estimator.



→ Ridge is shrinking the coefficients towards a common value (zero).

- Ridge is not the only shrinkage estimator. Today we'll cover the first shrinkage estimator that came to be:

I once met Stein! ☺!!

The James-Stein estimator

- Suppose we wish to estimate a single parameter μ from observation x in Bayesian context:

$$\mu \sim N(M, A) \quad \text{and} \quad x|\mu \sim N(\mu, 1)$$

Normal Prior. Normal $(\mu, 1)$ likelihood

- In this case $\mu|x$ has posterior distribution

$$\mu \mid x \sim N\left(M + B(x - M),\ B\right)$$

where $B = \dfrac{A}{A+1}$

Example:

→ If this was MLE and $x = 10$

$$\Rightarrow \hat{\mu}^{MLE} = \frac{10}{1} = 10$$

- However, in this context, say we had $M = 9$,

then $\hat{\mu}^{Bayes} = M + \left(\dfrac{A}{A+1}\right)\cdot(x - M)$ ← Changes

$= B$; $= B$

Always the same.

$$= 9 + \left(\frac{A}{A+1}\right)(10 - 9)$$

$$0 \le \frac{A}{A+1} \le 1$$

- Let's now compare the overall expected error of either estimator:

$$E\left\{(\hat{\mu}^{MLE} - \mu)^2\right\} = 1$$ ←

If we had $x_1 = 10 \rightarrow \hat{\mu}_1^{MLE} = 10$

$x_2 = 10.5 \rightarrow \hat{\mu}_2^{MLE} = 10.5$

$\vdots$

Var = 1

$$E\left\{(\hat{\mu}^{Bayes} - \mu)^2\right\} = B^2$$

$$V(\hat{\mu}^{Bayes}) = V(M + B(x - M))$$
$$= V(Bx - BM)$$
$$= B^2 V(x) = B^2$$

This variance is smaller than that of MLE.

Now, say ... if we have N independent observations,

$$\underline{\mu} = [\mu_1, \mu_2, \ldots, \mu_N]^T$$

and

$$\underline{x} = [x_1, x_2, \ldots, x_N]^T$$

The Data Generation Process (DGP) of each x_i is not the same

Different μ_i's for each observation.

$$\mu_i \sim N(M, A) \quad \text{and} \quad x_i | \mu_i \sim N(\mu_i, 1)$$

we're assuming each μ_i has the same prior.

x_i's have the same variance around their different means.

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

And $\hat{\mu}^{MLE} = \underline{x}$

Thus, $E\left\{ \| \hat{\mu}^{MLE} - \mu \|^2 \right\} = E\left\{ \sum_{i=1}^{N} (\hat{\mu}^{MLE} - \mu)^2 \right\}$

$= N$

$$E\left\{ \| \hat{\mu}^{Bayes} - \mu \|^2 \right\} = N B^2$$

Again, the overall error of $\hat{\mu}^{Bayes}$ is smaller. vs. $\hat{\mu}^{MLE}$

Recall

$$V(\hat{\theta}) = E\left[(\hat{\theta} - E(\hat{\theta}))^2 \right]$$

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$\text{Bias}^2 = \left[E(\hat{\theta}) - \theta \right]^2$$

If true θ was 10.1, then

$\text{Bias}(\hat{\theta}) = -0.1$

$\text{Bias}^2(\hat{\theta}) = 0.01$

Sample 1

$\hat{\theta}_1 = 10.5$

Sample 2

$\hat{\theta}_2 = 10$

⋮

Sample N

$\hat{\theta}_N = 9.5$

$E(\hat{\theta}) = 10$

$$V(\hat{\theta}) = \frac{(10.5 - 10)^2 + (10 - 10)^2 + \cdots + (9.5 - 10)^2}{n - 1}$$

$V(\hat{\theta}) + \text{Bias}^2(\hat{\theta})$ to check how estimator performs

Overall Error: $\frac{(10.5 - 10.1)^2 + (10 - 10.1)^2 + \ldots + (9.5 - 10.1)^2}{N}$

($\hat{\theta}$: 10.5, 10, ...; True θ: 10.1)

→ If we know M and A, then we're in great shape with $\hat{\mu}_{Bayes}$. However, we usually don't have prior information. what to do then?

- Estimate M ! $\hat{M} = \frac{\sum_{i=1}^{n} x_i}{N}$

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

Then, $\hat{\mu}_i^{JS} = \hat{M} + \hat{B}(x_i - \hat{M})$ for $i = 1, \ldots, N$.

James-Stein Estimator

It can be shown that

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

James-Stein Theorem

Assumption

normally. Different means. Same variance

Assumption

- Suppose that $X_i \mid \mu_i \sim N(\mu_i, 1)$ independently for $i = 1, 2, \ldots, N$ with $N \geq 4$.

(Regardless of the prior),

Statement

Then, $E\left(\|\hat{\mu}^{JS} - \mu\|^2\right) < N = E\left(\|\hat{\mu}^{MLE} - \mu\|^2\right)$

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

Overall error of JS estimator

Overall error of MLE.

Baseball Players

- Same season of MLB
- Each of them batted a # of hits over a number of at-bats.
- If we divide $\frac{\#\text{ of hits}}{\#\text{ of at bats}}$ = batting avg.

→ Lets get the JS for these batting avgs and compare it to the MLE.

Example: player 1 : $\frac{31}{90} = .345 = \hat{\mu}_1^{MLE}$

player 18 : $\frac{13}{90} = .145 = \hat{\mu}_{18}^{MLE}$

Long-run batting avg. of the player

probabilities, p_i's

$p_i \sim \frac{Bi(90, p_i)}{90}$

$V(p_i) = \frac{1}{n^2} \cdot n\,p(1-p) = \frac{p(1-p)}{n}$

These are like the x_i's

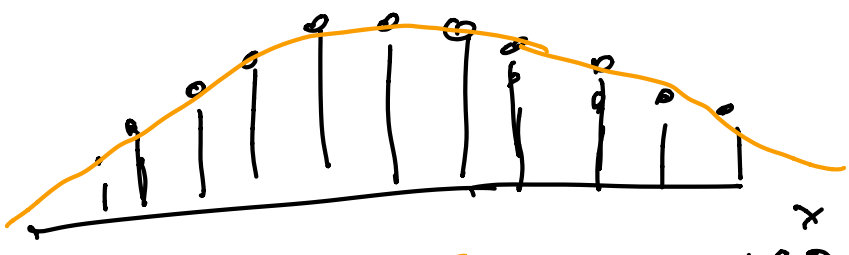
These are like the μ_i's

Issue: this is Binomial! not normal !!.

Good News is that we can approximate the Binomial w/ a Normal.

$\Rightarrow p_i \overset{approx}{\sim} N(p_i, \sigma_0^2)$

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



$z = \frac{p_i - p_i}{\sigma_0}$, $z \sim N(0,1)$

$x \sim N(p_i, 1)$ where $\bar{p} = \frac{\sum_{i=1}^{n} p_i}{n}$

Now, let $x_i = \frac{p_i}{\sigma_0}$ why? $\Rightarrow x_i \overset{approx}{\sim} N(p_i, 1)$

and apply JS estimator

$$\hat{\mu}^{JS} = \hat{\mu} + \hat{B}(x_i - \hat{\mu})$$