Math stats

Efron (2010) exercises

1.1

With the given distribution

$$\mu \sim \mathcal{N}(0, A) \rightarrow f(\mu) = \frac{1}{\sqrt{2\pi A}} \exp\{-\frac{\mu^2}{2A}\}$$
$$z|\mu \sim \mathcal{N}(0, 1) \rightarrow f(z|\mu) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(z-\mu)^2}{2}\}$$

So, we have

$$\begin{aligned} f(\mu) \cdot f(z|\mu) &= \frac{1}{2\pi\sqrt{A}} \exp\{-\frac{1}{2}(\frac{A+1}{A}\mu^2 - 2\mu z + z^2)\} \\ &= \frac{1}{2\pi\sqrt{A}} \exp\{-\frac{1}{2B}(\mu^2 - 2B\mu z + Bz^2)\}; B = \frac{A}{A+1} \\ &= \frac{1}{2\pi\sqrt{A}} \exp\{-\frac{1}{2}(\frac{\mu - Bz}{\sqrt{B}})^2\} \cdot \exp\{-\frac{1}{2}(1-B)z^2\} \end{aligned}$$

Then, the integral

$$\begin{aligned} f(z) &= \int f(\mu) f(z|\mu) d\mu \\ &= \frac{1}{2\pi\sqrt{A}} \exp\{-\frac{1}{2}(1-B)z^2\} \cdot \sqrt{B} \int_{-\infty}^{+\infty} \exp\{-\frac{1}{2}(\frac{\mu - Bz}{\sqrt{B}})^2\} d(\frac{\mu - Bz}{\sqrt{B}}) \\ &= \frac{\sqrt{B}}{\sqrt{2\pi A}} \exp\{-\frac{1}{2}(1-B)z^2\} \\ &= \frac{1}{\sqrt{2\pi(A+1)}} \exp\{-\frac{z^2}{2(A+1)}\} \\ \rightarrow z &\sim \mathcal{N}(0, A+1) \end{aligned}$$

Finally, we get

$$\begin{aligned} f(\mu|z) &= \frac{f(\mu) \cdot f(z|\mu)}{f(z)} \\ &= \frac{1}{\sqrt{2\pi B}} \exp\{-\frac{1}{2}(\frac{\mu - Bz}{\sqrt{B}})^2\} \end{aligned}$$

that is,

$$\mu|z \sim \mathcal{N}(Bz, B), B = A/(A+1)$$

1.2

(i) (1.17)
We have Bayes estimator

$$\hat{\boldsymbol{\mu}}^{(Bayes)} = B\boldsymbol{z}; B = \frac{A}{A+1}$$

Then, the risk function (total square error loss)

$$\begin{aligned} R^{(Bayes)}(\boldsymbol{\mu}) &= E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2\} \\ &= E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||B\boldsymbol{z} - B\boldsymbol{\mu} + B\boldsymbol{\mu} - \boldsymbol{\mu}||^2\} \\ &= B^2 E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\boldsymbol{z} - \boldsymbol{\mu}||^2\} + (B-1)^2 E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\boldsymbol{\mu}||^2\} \end{aligned}$$

With the vector $\boldsymbol{z}|\boldsymbol{\mu} \sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{I})$,

$$E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\boldsymbol{z} - \boldsymbol{\mu}||^2\} = ||\boldsymbol{I}_{N\times N}||^2 = N$$

So, we have finally got

$$R^{(Bayes)}(\boldsymbol{\mu}) = B^2 N + (1 - B)^2 ||\boldsymbol{\mu}||^2$$

(ii) (1.18)
With the vector $\boldsymbol{\mu} \sim \mathcal{N}_N(\boldsymbol{0}, A\boldsymbol{I})$,

$$E_{\boldsymbol{\mu}}\{||\boldsymbol{\mu}||^2\} = ||A\boldsymbol{I}_{N\times N}||^2 = A \cdot N$$

Using (1.17), the overall Bayes risk

$$\begin{aligned} R^{(Bayes)} &= E_{\boldsymbol{\mu}}\{R^{(Bayes)}(\boldsymbol{\mu})\} \\ &= B^2 N + (1 - B)^2 E_{\boldsymbol{\mu}}\{||\boldsymbol{\mu}||^2\}; B = \frac{A}{A+1} \\ &= \frac{NA^2}{(A+1)^2} + \frac{1}{(A+1)^2} NA \\ &= \frac{NA}{A+1} \end{aligned}$$

1.4

(a) (1.31)
We have (1.30), that is

$$E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2\} = E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\boldsymbol{z} - \hat{\boldsymbol{\mu}}||^2\} - N + 2\sum_{i=1}^{N} E_{\boldsymbol{z}|\boldsymbol{\mu}}\{\frac{\partial \hat{\mu}_i}{\partial z_i}\}$$

Then, with

$$\hat{\boldsymbol{\mu}}^{(JS)} = (1 - \frac{N-2}{S})\boldsymbol{z}$$
$$S = ||\boldsymbol{z}||^2 = \sum_{i=1}^{N} z_i^2$$

We get

$$\begin{aligned} E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\hat{\boldsymbol{\mu}}^{(JS)} - \boldsymbol{\mu}||^2\} &= E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\frac{N-2}{S}\boldsymbol{z}||^2\} - N + 2\sum_{i=1}^{N} E_{\boldsymbol{z}|\boldsymbol{\mu}}\{(1 - \frac{N-2}{S}) + \frac{N-2}{S^2} 2z_i^2\} \\ &= E_{\boldsymbol{z}|\boldsymbol{\mu}}\{(\frac{N-2}{S})^2 ||\boldsymbol{z}||^2\} - N + 2E_{\boldsymbol{z}|\boldsymbol{\mu}}\{N(1 - \frac{N-2}{S}) + \frac{2(N-2)}{S^2}\sum_{i=1}^{N} z_i^2\} \\ &= E_{\boldsymbol{z}|\boldsymbol{\mu}}\{\frac{(N-2)^2}{S}\} - N + 2N - 2E_{\boldsymbol{z}|\boldsymbol{\mu}}\{\frac{N(N-2)}{S} - \frac{2(N-2)}{S}\} \\ &= N - E_{\boldsymbol{z}|\boldsymbol{\mu}}\{\frac{(N-2)^2}{S}\} \end{aligned}$$

(b) (1.24)
We have (1.31), that is shown above; and with marginal $\boldsymbol{z} \sim \mathcal{N}_N(\boldsymbol{0}, (A+1)\boldsymbol{I})$,

$$E_{\boldsymbol{z}}\{\frac{N-2}{S}\} = \frac{1}{A+1}$$

So we can derive that

$$\begin{aligned}
R^{(JS)} &= E_{\boldsymbol{\mu}}\{R^{(JS)}(\boldsymbol{\mu})\} \\
&= E_{\boldsymbol{\mu}}\{E_{\boldsymbol{z}|\boldsymbol{\mu}}\{||\hat{\boldsymbol{\mu}}^{(JS)} - \boldsymbol{\mu}||^2\}\} \\
&= N - E_{\boldsymbol{\mu}}\{E_{\boldsymbol{z}|\boldsymbol{\mu}}\{\frac{(N-2)^2}{S}\}\} \\
&= N - (N-2)E_{\boldsymbol{z}}\{\frac{N-2}{S}\} \\
&= N - \frac{N-2}{A+1} \\
&= N\frac{A}{A+1} + \frac{2}{A+1}
\end{aligned}$$

1.5

If we assume in Table 1.2, for $i = 1, ..., n$, $\mu_i \sim \mathcal{N}(0, A)$, and we also have $\boldsymbol{z}|\boldsymbol{\mu} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{I})$, then we can use all the conclusions from the hierachical Normal-Normal model.
So based on the derivations in the book, the simulated total square error of the James-Stein estimator should be close to the theoretical value, i.e. Bayes risk of JS estimator.

$$TSE \sim R^{(JS)} = (nA+2)/(A+1)$$

Here, we can estimate

$$\hat{A} = \frac{1}{n-2}\sum_{i=1}^{n=10} \mu_i^2$$

So, we get $\hat{A} = 3.1125$ and $TSE = 8.0547$, which is close to the 8.13 value in the table.

Simulation

In each run of the simulations, $N = 1000$ runs, we set the true $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n), n = 10$ values to be equal to those in Efron's Table 1.2, and generate random samples $z_i \sim \mathcal{N}(\mu_i, 1)$, i.e., we have N copies of $\boldsymbol{z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{I})$.

The MLE estimates in each run are just

$$\hat{\boldsymbol{\mu}}_i^{(MLE)} = z_i$$

and the James-Stein estimators in each run are

$$\hat{\boldsymbol{\mu}}_i^{(JS)} = \bar{z} + (1 - \frac{(n-3)\sigma_0^2}{S}) \cdot (z_i - \bar{z})$$
$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i, \quad \sigma_0^2 = 1, \quad S = \sum_{i=1}^{n}(z_i - \bar{z})^2$$

As for the mean squared errors and total squared errors, we have

$$MSE_i^{(MLE)} = \frac{1}{N}\sum_{N}(\hat{\mu}_{i,runs}^{(MLE)} - \mu_i)^2;$$
$$TSE^{(MLE)} = \sum_{i=1}^{n} MSE_i^{(MLE)}$$

And,

$$MSE_i^{(JS)} = \frac{1}{N}\sum_{N}(\hat{\mu}_{i,runs}^{(JS)} - \mu_i)^2;$$
$$TSE^{(JS)} = \sum_{i=1}^{n} MSE_i^{(JS)}$$

So, Efron's Table 1.2 is reproduced as below.

	μ_i	$MSE_i^{(MLE)}$	$MSE_i^{(JS)}$
1	-0.81	1.01	0.91
2	-0.39	1.00	0.75
3	-0.39	0.95	0.73
4	-0.08	1.02	0.69
5	0.69	0.95	0.57
6	0.71	0.99	0.59
7	1.28	0.98	0.59
8	1.32	0.97	0.59
9	1.89	1.04	0.75
10	4.00	1.02	1.76
TSE		9.92	7.95

For MSE's, we expect to see one decimal place of agreement between my table and Efron's Table 1.2. In real simulation, we get one decimal place of agreement on the MLE estimates, but zero on the JS estimates. The one decimal place of agreement on the JS estimates have some deviations due to simulation randomness.

Shrinking radon

In this dataset, we are interested in the observed radon levels from Minnesota basements. Following the data cleaning guidelines in the assignment description, we get a total of 511 observations across 17 counties in MN, each with at least 10 observations. The dataset is then further splited into two sets: a training set with 5 randomly chosen observations from each county, and a test set with the other observations.

The "true" vector of mean radon levels by county, $\boldsymbol{\mu}$, are calculated using the test data variable *activity*. Our goal is to estimate this population-level parameter $\boldsymbol{\mu}$. We adopt the standard independent-normals assumption here, with a slightly different notation due to the multiple observations for each county.

For county i, $(i = 1, \ldots, N, N = 17)$, observation j, $(j = 1, \ldots, n, n = 5)$,

$$x_{ij} \sim^{i.i.d} \mathcal{N}(\mu_i, \tau^2)$$

So the MLE estimates are

$$\hat{\mu}_i^{(MLE)} = z_i = x_{i.} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

and therefore,

$$\hat{\boldsymbol{\mu}}^{(MLE)} = \boldsymbol{z} \sim \mathcal{N}_N(\boldsymbol{\mu}, \frac{\tau^2}{n}\boldsymbol{I})$$

As for the James-Stein estimates, we use the pooled-variance technique for estimating τ^2, because the same number of observations in each county aids the assumption of common standard error (SE) among MLE estimators.

$$\hat{\mu}_i^{(JS)} = \bar{z} + (1 - \frac{(N-3)\cdot\hat{\tau}^2/n}{S})\cdot(z_i - \bar{z})$$

where

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$$

$$\hat{\tau}^2 = \frac{1}{N(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{n}(x_{ij} - z_i)^2$$

$$S = \sum_{i=1}^{N}(z_i - \bar{z})^2$$

So, the estimates we get are shown in the table below.

County	μ_i	$\hat{\mu}_i^{(MLE)}$	$\hat{\mu}_i^{(JS)}$
ANOKA	2.85	5.08	5.31
BLUE EARTH	8.10	6.06	5.87
CLAY	9.30	12.82	9.72
DAKOTA	4.86	5.56	5.59
GOODHUE	12.19	4.10	4.75
HENNEPIN	4.89	3.94	4.66
ITASCA	2.60	3.52	4.42
MOWER	9.00	7.16	6.50
OLMSTED	5.12	3.86	4.62
RAMSEY	3.48	3.66	4.50
RICE	6.54	8.24	7.11
ST LOUIS	3.35	1.78	3.43
STEARNS	5.89	4.32	4.88
STEELE	4.68	6.54	6.14
WASHINGTON	5.17	4.88	5.20
WINONA	8.92	6.40	6.06
WRIGHT	5.01	7.60	6.75
TSE		118.56	92.57

So, the ratio of the total square error is $TSE^{(MLE)}/TSE^{(JS)} = 1.28$. And we can conclude that the Stein shrinkage in this application help moderately reduce the total square error.