A Galtonian Perspective on Shrinkage Estimators

He Li, ***hl3420***

Consider n independent $X_1, \cdots, X_n$, each is normally distributed $X_i \sim \mathcal{N}(\theta_i, 1), \forall i = 1, \cdots, n$. We would like to estimate θ_i's measured by the risk function:

$$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \mathbb{E}L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \mathbb{E}\left[\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2\right]. \tag{1}$$

James and Stein (1961) show that for $n \geq 3$, the James-Stein estimators of the form,

$$\hat{\theta}_i^{JS} = \left(1 - \frac{c}{\sum_{j=1}^{n} X_j^2}\right) X_i,\ \forall\ 0 < c < 2(n-2), \tag{2}$$

(with $c = n - 2$ the best choice) dominate the ordinary estimator $\hat{\theta}_i = X_i$ under risk function 1.

Efron and Morris (1973) provide the Efron-Morris estimators of the form,

$$\hat{\theta}_i^{\mathrm{EM}} = \bar{X} + \left(1 - \frac{c}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right)\left(X_i - \bar{X}\right),\ \forall\ 0 < c < 2(n-3), \tag{3}$$

(with $c = n - 3$ the best choice) beats the ordinary estimators when $n \geq 4$.

This paper gives an novel explanation of why those shrinkage estimators work better using a regression setup. To be specific, consider the regression task of $\mathbb{E}[\theta|X]$, the least square estimator with intercept is,

$$\hat{\theta}_i = \bar{\theta} + \hat{\beta}\left(X_i - \bar{X}\right),\ \text{where } \hat{\beta} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(\theta_i - \bar{\theta}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}. \tag{4}$$

The Efron-Morris estimator is conducted by replacing $\bar{\theta}$ and $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(\theta_i - \bar{\theta}\right)$ by $\bar{X}$ and $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 - (n-1)$, their unbiased estimators, respectively.

The James-Stein estimator can be derived similarly with the least square estimator without intercept,

$$\hat{\theta}_i = \hat{\beta}\left(X_i - \bar{X}\right),\ \text{where } \hat{\beta} = \frac{\sum_{i=1}^{n} X_i\theta_i}{\sum_{i=1}^{n} X_i^2}. \tag{5}$$

The James-Stein estimator is conducted by replacing $\sum_{i=1}^{n} X_i\theta_i$ by its unbiased estimators $\sum_{i=1}^{n} X_i^2 - n$.

A simple proof of both Efron-Morris and James-Stein estimators can be derived from this regression perspective, using the fact that the orthogonality of the least square fitted values and the residuals can be exploited to seperate the least square loss RSS_{LS} from the loss function $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ for any estimator $\hat{\boldsymbol{\theta}}$.

This paper gives a clear and simple argument on why ordinary estimator fails in this estimation task as it is based on th wrong regression line $\mathbb{E}[X|\theta]$ instead of $\mathbb{E}[\theta|X]$ and therefore the least square estimator, (along with its unbiased estimators) dominates the ordinary estimator.

References

Efron, B. and C. Morris (1973). Stein's estimation rule and its competitors—an empirical bayes approach. *Journal of the American Statistical Association 68*(341), 117–130.

James, W. and C. Stein (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. 1*, 361–379.

Stigler, S. M. et al. (1990). The 1988 neyman memorial lecture: a galtonian perspective on shrinkage estimators. *Statistical Science 5*(1), 147–155.