A BAYESIAN APPROACH TO SHRINKAGE ESTIMATORS

Filipe das Neves RIZZO[1]
Cristiane Alvarenga GAJO[1]
Devanil Jaques de SOUZA[1]
Lucas Monteiro CHAVES[1]

- *ABSTRACT: Estimators obtained by shrinking the least squares estimator are becoming widely used since the work of Stein, in the early 60's, where it is presented an estimator for the mean of multivariate normal that dominates the sample mean, and the work of Hoerl and Kennard, in the early 70's, on ridge estimators. In this work we present an approach using Bayesian and empirical Bayesian procedures to obtain some important shrinkage estimators.*

- *KEYWORDS: shrinkage estimators, empirical Bayes, minimum mean square estimator.*

1 Introduction

Two landmark papers, one by James-Stein (1961), introducing the James-Stein estimators, and another by Hoerl and Kennard (1970), introducing ridge estimators, explore the idea that shrinking the usual least squares estimator, besides the fact of introducing some bias, has the advantage of decreasing the mean square error (MSE). Both these works explore the idea is to shrinking the estimates. This is a conservative procedure used in many areas of statistic. Nowadays, that is known as biased estimation theory and is used when the least squared estimator can´t be used in reason, for example, of multicollinearity. The theory of shrinkage estimators is a very active area of research and we can point, as examples, the LASSO theory (TIBSHIRANI, 1996) and Elastic net theory (ZOU-HASTIE, 2005).

In this short note, we present a systematic approach to shrinkage estimators using Bayesian and empirical Bayes procedures. Examples for ridge regression estimation and James-Stein estimator are present. These procedures are very well known and spread in the literature, but authors couldn't find any systematic similar approach. Since empirical Bayes procedure is not of common use, we review, in section 2, a classical example.

We end this work with a computational simulation showing the behavior of some shrinkage estimators.

Shrinkage estimators

[1] University Federal of Lavras – UFLA , Department of Exact Sciences, CP 37 , CEP: 37200 - 000, Lavras, MG, Brazil.: *rizzo.filipe@gmail.com; cristianegajo@yahoo.com.br;devaniljaques@dex.ufla.br; lucas@dex.ufla.br*

Unbiased estimators of a vector of parameter have a problem that, in general, is not mentioned. Consider $\hat{\beta} = \left(\hat{\beta}_1, ..., \hat{\beta}_k\right)$ an unbiased estimator for the vector $\beta = \left(\beta_1, ..., \beta_k\right)$,then,

$$E\left[\left\|\hat{\beta}\right\|^2\right] = E\left[\sum_{i=1}^{k} \beta_i^2\right] = \sum_{i=1}^{k} E\left[\beta_i^2\right] = \sum_{i=1}^{k}\left(\text{var}\left[\beta_i\right] + \left(E\left[\beta_i\right]\right)^2\right)$$

$$= \sum_{i=1}^{k} \text{var}\left[\beta_i\right] + \|\beta\|^2 .$$

We face here an unexpected fact: even though $\hat{\beta}$ is an *unbiased* estimator of β, $\left\|\hat{\beta}\right\|^2$ is a biased estimator for $\|\beta\|^2$, tending to overestimate it. This fact has motivated James-Stein to obtain his famous estimator. The idea was to shrink the least squares estimator by a factor that depends of the value observed. Hoerl-Kennard, working with multiple linear regressions in the presence of multicollinearity, proposed an estimation denoted ridge regression estimator, that is a kind of shrinkage estimator. Nowadays there is a huge theory about these estimators.

2 Empirical Bayes method

When the knowledge of the prior distribution parameters is incomplete or unknown, some information may be obtained from the data itself and, in this way, it is possible to make inference about these parameters. This is the so-called Bayesian empirical method**.**To illustrate the empirical Bayes methodology, consider the classical example: Let *X* a Poisson random variable with parameter μ and prior $\pi(\mu)$ completely unknown. Given a size one sample $X = x$, the corresponding posterior distribution, according to Bayes theorem, is

$$\pi(\mu \mid x) = \frac{\frac{e^{-\mu}\mu^x}{x!}\pi(\mu)}{g(x)} \text{ where } g(x) = \int_0^\infty \frac{e^{-\mu}\mu^x}{x!}\pi(\mu)\,d\mu .$$

In relation to the mean square error, the Bayes estimator $\hat{\mu}_B$ for μ is the mean of the posterior $\pi(\mu \mid x)$:

$$\hat{\mu}_B = E[\mu \mid x] = \int_0^\infty \mu \frac{e^{-\mu}\mu^x}{x!\,g(x)}\pi(\mu)\,d\mu$$

$$= \frac{(x+1)!\,g(x+1)}{x!\,g(x)} \underbrace{\int_0^\infty \frac{e^{-\mu}\mu^{x+1}}{(x+1)!\,g(x+1)}\pi(\mu)\,d\mu}_{1} = \frac{(x+1)!\,g(x+1)}{x!\,g(x)} .$$

Now, one can use the known past observations $x_1,...,x_n$ to estimate $g(x)$ and $g(x+1)$. Since $g(x)$ is a discrete distribution it can be estimated by the relative frequencies

$$\hat{g}(x) = \frac{card\{i;\ x_i = x\}}{n}$$

$$\hat{g}(x+1) = \frac{card\{i;\ x_i = x+1\}}{n}$$

where $card\{\mathrm{A}\}$ is the number of elements of set A.

Then, the empirical Bayes estimator is

$$\hat{\mu}_B = \frac{(x+1)\,\hat{g}(x+1)}{\hat{g}(x)}.$$

3 An approximation of the minimum mean squares error linear estimator for the mean

Consider independent observations Y_i, $i = 1,...,n$, with model

$$Y_i = \mu + \varepsilon_i,$$

where μ is the parameter to be estimated and the ε_i's are the associated independent errors The least square estimator for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y_n}.$$

Let´s consider all estimators of the form $\delta_c(Y_1,...,Y_n) = c\overline{Y}_n$. We want to determine the particular value of c that minimizes the mean squared error. To do that we take the function

$$\begin{aligned} MSE(c) &= E\left[\left(c\overline{Y_n} - \mu\right)^2\right] = c^2 E\left[\overline{Y_n}^2\right] - 2c\mu E\left[\overline{Y_n}\right] + \mu^2 \\ &= c^2\left(\frac{\sigma^2}{n} + \mu^2\right) - 2c\mu^2 + \mu^2. \end{aligned}$$

The minimum value of $MSE(c)$ defines the estimator

$$\hat{\mu} = \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2}\overline{Y}_n,$$

that depends on population parameters μ and σ^2, both unknown. The natural way to bypass this problem is to replace them by their natural unbiased estimators $\overline{Y_n}$ and S^2, resulting inand the estimator

$$\hat{\mu}_1 = \frac{\overline{Y}_n^{\,2}}{\frac{S^2}{n} + \overline{Y}_n^{\,2}} \overline{Y_n} = \frac{1}{\frac{S^2}{n\overline{Y}_n^{\,2}} + 1} \overline{Y_n} \tag{1}$$

that is an approximation of minimum mean square error linear estimator for the mean.

In section 8 we compare, via simulation, the behavior of these estimators.

4 Shrinkage estimators obtained from normality

The simplest example of a shrinkage estimator obtained from normality is: suppose $Y \sim N(\theta,1)$ and prior $\theta \sim N(0,1)$. If a sample $y = (y_1,...,y_n)$ is taken and $\overline{y}$, is it mean, the posterior distribution is

$$\begin{aligned} \pi(\theta \mid y) &\propto f(y \mid \theta)\pi(\theta) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}\exp\left\{-\frac{1}{2}\theta^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\frac{1}{n+1}}\left(\theta - \frac{\overline{y}}{1+\frac{1}{n}}\right)^2\right\}. \end{aligned}$$

Then the Bayes estimator is $\hat{\theta}_B = \frac{1}{1+\frac{1}{n}}\overline{Y}_n$, clearly a shrinkage estimator.

A more sophisticated argument is to obtain the estimator (1) using empirical Bayes approach.

If $Y \sim normal(\mu,\sigma^2)$ with priori $\mu \sim normal(0,\tau^2)$ and a sample $y = (y_1,...,y_n)$ is observed, the posteriori is normal (GRUBER, 2010 - p.50),

$$\pi(\mu \mid \overline{y}) \propto \exp\left\{-\frac{1}{2\frac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n}}\left(\mu - \frac{\tau^2\overline{y}}{\tau^2+\sigma^2/n}\right)^2\right\}.$$

Then, using the mean of the posteriori as the Bayes estimator for μ

$$\hat{\mu}_B = \frac{\tau^2 \overline{Y}_n}{\tau^2 + \sigma^2/n} = \frac{1}{\frac{\sigma^2}{n\tau^2} + 1} \overline{Y}_n \ .$$

We don't know τ^2 and the idea is to use empirical Bayes argument. The predictive distribution is a normal with mean zero and variance $\frac{\sigma^2}{n} + \tau^2$ (appendix). Observing that if $Z \sim normal\left(0, \sigma^2\right)$ as $\sigma^2 = E\left[Z^2\right] - \left(E[Z]\right)^2 = E\left[Z^2\right]$, by the methods of moments, σ^2 can be estimated by only one observation , $\hat{\sigma}^2 = Z^2$. Then we can estimate $\frac{\sigma^2}{n} + \tau^2$ by $\overline{Y}^2$. It is reasonable to suppose $\tau^2 >> \frac{\sigma^2}{n}$. Then we can use $\overline{Y}^2$ as estimator of τ^2. Replacing the parameters in $\hat{\mu}_B$ by its respective estimators we get again the estimator (1).

The estimator (1) can also be obtained using mixture. Let's suppose that, in the model $Y_i = \mu + \varepsilon_i$, $Y_i \sim normal(\mu, \sigma^2)$ and μ has a prior distribution $normal\left(0, \tau^2\right)$. We want to compute the mean square error $E\left[\left(c\overline{Y_n} - \mu\right)^2\right] = E\left[c^2 \overline{Y_n}^2 - 2c\overline{Y_n}\mu + \mu^2\right]$, considering both $\overline{Y_n}$ and μ as random variables.

Computing these expected values:

$$f_{\overline{Y}_n}(\overline{y}) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\overline{y} - \mu)^2\right\}$$

$$f_{\mu}(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\mu^2\right\}$$

$$E\left[c^2\overline{Y}^2 - 2c\overline{Y}\mu + \mu^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(c^2\overline{y}^2 - 2c\overline{y}\mu + \mu^2\right) f_{\overline{Y}}(\overline{y})\, f_{\mu}(\mu)\, d\mu\, d\overline{y} \ .$$

Computing each integral:

$$E\left[c^2\overline{Y}^2\right] = c^2 \int_{-\infty}^{\infty} \overline{y}^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\overline{y} - \mu)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\mu^2\right\} d\mu d\overline{y}$$

$$= c^2 \int_{-\infty}^{\infty} \overline{y}^2 \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n}+\tau^2\right)}} \exp\left\{-\frac{1}{2\left(\frac{\sigma^2}{n}+\tau^2\right)}\overline{y}^2\right\} d\overline{y}$$

$$= c^2\left(\frac{\sigma^2}{n}+\tau^2\right).$$

$$E\left[\mu\overline{Y_n}\right] = \int_{-\infty}^{\infty} \mu \left[\int \overline{y} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\overline{y}-\mu)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\mu^2\right\} d\overline{y}\right] d\mu$$

$$= \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\mu^2\right\} \underbrace{E[\overline{y}]}_{\mu} d\mu.$$

$$E\left[\mu^2\right] = Var[\mu] + \underbrace{\left(E[\mu]\right)^2}_{0} = \tau^2.$$

All together:

$$E\left[c^2\overline{Y}^2 - 2c\overline{Y}\mu + \mu^2\right] = c^2\left(\frac{\sigma^2}{n}+\tau^2\right) - 2c\,\tau^2 + \tau^2.$$

So, the minimum value of the mean squared error $E\left[\left(c\overline{Y}-\mu\right)^2\right]$ is obtained when $c_0^* = \frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}$. Then, the MSE estimator for μ is:

$$\hat{\mu}_2 = \frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\overline{Y}_n .$$

Replacing the parameters by its estimators we, again, get the estimator (1).

5 A Bayesian interpretation of ridge regression estimators

Consider the multiple linear regression

$$\mathbf{Y}_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1} \quad , \quad \varepsilon_{n\times 1} \sim N_n\left(0, \sigma^2 I_{n\times n}\right).$$

We will suppose σ^2 know and a priori distribution for the parameter vector $\beta_{p\times 1}$ given by a *p*-variate normal

$$\beta_{p\times 1} \mid \sigma_\lambda^2 \sim N_p\left(\mathbf{0}, \frac{\sigma^2}{\lambda} I\right).$$

The posterior distribution given data ***Y*** and design matrix X is

$$f_\beta\left(\beta, \sigma^2 \mid \mathbf{Y}, X, \lambda\right) \propto f_{\mathbf{Y}}\left(\mathbf{Y} \mid X, \beta, \sigma^2\right) f_\beta\left(\beta \mid \sigma^2, \lambda\right)$$

$$\propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta)\right]$$

$$\propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta)\right] \sigma^{-p} \exp\left[-\frac{\lambda}{2\sigma^2}\beta'\beta\right]$$

$$\propto \sigma^{-(n+p)} \exp\left[-\frac{1}{2\sigma^2}\left[(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta) - \lambda\beta'\beta\right]\right]$$

$$\propto \sigma^{-(n+p)} \exp\left[-\frac{1}{2\sigma^2}\left[\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'X\beta - \beta'X'\mathbf{Y} + \beta'X'X\beta + \lambda\beta'\beta\right]\right]$$

$$\propto \sigma^{-(n+p)} \exp\left[-\frac{1}{2\sigma^2}\left[\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'X\beta - \beta'X'\mathbf{Y} + \beta'\left[X'X\beta + \lambda I\right]\beta\right]\right].$$

To explicit the distribution it is necessary to complete the square in

$$(Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta + \lambda\beta'\beta$$
$$= Y'Y - 2Y'X\beta + \beta'(X'X + \lambda I)\beta.$$

So, we need to determine W such that

$$(\beta - W)'(X'X + \lambda I)(\beta - W) = Y'Y - 2Y'X\beta + \beta'(X'X + \lambda I)\beta$$
$$= \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta + W'(X'X + \lambda I)W.$$

Then $W = (X'X + \lambda I)^{-1} X'Y$ and $Y'X\beta = W'(X'X + \lambda I)\beta$.

$$(Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta = Y'Y + \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta$$
$$= Y'Y + \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta$$
$$+ W'(X'X + \lambda I)W - W'(X'X + \lambda I)W$$
$$= (\beta - W)'(X'X + \lambda I)(\beta - W) + Y'Y - W'(X'X + \lambda I)W.$$

Therefore

$$f_\beta\left(\beta,\sigma^2 \mid \mathbf{Y}, X\right) \propto exp\left[(\beta - W)'(X'X + \lambda I)(\beta - W) + Y'Y - W'(X'X + \lambda I)W\right]$$

$$\propto (\beta - W)'(X'X + \lambda I)(\beta - W).$$

The posterior is a multivariate normal with mean $W = \left(X'X + \lambda I\right)^{-1} X'Y$, therefore the Bayes estimator is

$$\hat{\beta}(\lambda) = \left(X'X + \lambda I\right)^{-1} X'Y .$$

This is the ridge regression estimator with ridge parameter λ (Hoerl-Kennard, 1970) . For $\lambda = 0$ we have the ordinary least square estimator and $\left\|\hat{\beta}(\lambda)\right\| < \left\|\hat{\beta}(0)\right\|$ and, for $\lambda > 0$ $\hat{\beta}(\lambda)$ is a shrinkage estimator.

6 The James-Stein estimator as an empirical Bayes estimator

Consider the random normal vector $X = \left(X_1, \ldots, X_p\right)$, $X \sim N_p(\theta, I)$ and suppose independent normal priors distributions for the parameters: $\theta_i \sim N(0, \sigma^2)$. If one observation $X = \left(X_1, ..., X_p\right)$ is taken, the Bayes estimator is

$$\hat{\theta}_i = \frac{\sigma^2}{\sigma^2 + 1} X_i = \left(1 - \frac{1}{\sigma^2 + 1}\right) X_i .$$

As σ^2 is unknown, using an empirical Bayes procedure it can be estimate from the data $X = \left(X_1, ..., X_p\right)$. The predictive distribution is a multivariate independent normal with variance $\sigma^2 + 1$, $N_p\left(0, \left(\sigma^2 + 1\right) I\right)$. Consider the random variables

$$Z_i = \frac{X_i}{\sqrt{\sigma^2 + 1}} \Rightarrow Z_i \sim N(0,1)$$

$$\sum_{i=1}^{p} Z_i^2 = \frac{\sum_{i=1}^{p} X_i^2}{\left(\sqrt{\sigma^2 + 1}\right)^2} = \frac{X'X}{\sigma^2 + 1} .$$

As sum of square *p* independent standard normal, it has chi-square distribution with parameter *p* , that is $Y = \frac{X'X}{\sigma^2 + 1} \sim \chi^2(p)$.

As $Y = \frac{X'X}{\sigma^2 + 1} \Leftrightarrow \frac{1}{X'X} = \frac{1}{Y}\frac{1}{\sigma^2 + 1}$, taking expectation,

$$E\left[\frac{1}{X'X}\right] = E\left[\frac{1}{(\sigma^2 + 1)Y}\right] = \frac{1}{\sigma^2 + 1} E\left[\frac{1}{Y}\right].$$

As

$$Y \sim Gama\left(\alpha=\frac{p}{2},\beta=\frac{1}{2}\right)$$

$$E\left[\frac{1}{Y}\right]=\int_0^\infty \frac{1}{y}\underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)}y^{\left(\frac{p}{2}-1\right)}e^{-\frac{1}{2}y}dy}_{f_Y(y)}$$

$$=\frac{\Gamma\left(\frac{p}{2}-1\right)}{\Gamma\left(\frac{p}{2}\right)\left(\frac{1}{2}\right)^{-1}}\int_0^\infty \underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{p}{2}-1}}{\Gamma\left(\frac{p}{2}-1\right)}y^{\left(\frac{p}{2}-1\right)-1}e^{-\frac{1}{2}y}dy}_{Gama\left(\alpha=\frac{p}{2}-1,\beta=\frac{1}{2}\right)}$$

$$=\frac{\Gamma\left(\frac{p}{2}-1\right)}{\Gamma\left(\frac{p}{2}\right)\left(\frac{1}{2}\right)^{-1}}=\frac{\Gamma\left(\frac{p}{2}-1\right)}{2\left(\frac{p}{2}-1\right)\Gamma\left(\frac{p}{2}-1\right)}=\frac{1}{2\left(\frac{p}{2}-1\right)}=\frac{1}{p-2}.$$

It follows that,

$$E\left[\frac{1}{X'X}\right]=\frac{1}{\sigma^2+1}\frac{1}{p-2}$$

and, therefore,

$$E\left[\frac{p-2}{X'X}\right]=\frac{1}{\sigma^2+1}.$$

By the methods of moments we can estimate $\frac{1}{\sigma^2+1}$ by $\frac{p-2}{X'X}$. As

$$\hat{\theta}_i=\frac{\sigma^2}{\sigma^2+1}x_i=\left(1-\frac{1}{\sigma^2+1}\right)x_i,$$

replacing $\frac{1}{\sigma^2+1}$ by $\frac{p-2}{X'X}$, the result is the famous James - Stein estimator (James-Stein, 1961)

$$\left(\hat{\theta}_{JS}\right)_i=\left(1-\frac{p-2}{X'X}\right)X_i$$

in vectorial form

$$\hat{\theta}_{JS}=\left(1-\frac{p-2}{X'X}\right)X.$$

Observe that the James – Stein estimator is a shrinkage estimator for $p>2$.

7 A non-normal example

In the example (CASELLA-BERGER, 2010, pp. 295), suppose $X_1,...,X_n$ *i.i.d.* $Bernoulli(p)$. It is well known that the maximum likelihood and unbiased estimator for the parameter p is

$$\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

The MSE for $\hat{p}$ is

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

$$= \text{var}\left[\overline{X}\right] = \frac{p(1-p)}{n}.$$

To compute the Bayes estimator we consider that $Y = \sum_{i=1}^{n} X_i$ is $Binomial(n, p)$, and use a $Beta(\alpha, \beta)$ as the priori for *p*.Then, the posteriori is given by:

$$\pi(p \mid y) = \frac{1}{B(y+\alpha, n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}.$$

So, the posterior is a $Beta(y+\alpha, n-y+\beta)$ and the Bayes estimator for the parameter *p*is its mean:

$$\hat{p}_B = \frac{y+\alpha}{\alpha+\beta+n}.$$

The mean squared error for this estimator is:

$$E\left[\left(\hat{p}_B - p\right)^2\right] = Var\left[\hat{p}_B\right] + \left(E\left[\hat{p}_B - p\right]\right)^2 = \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2.$$

Observe that, taking $\alpha = \beta = \sqrt{\frac{n}{4}}$, we get that the MSE is constant, that is, it doesn't depend of the parameter *p*. In this case the estimator $\hat{p}_B = \frac{y+\sqrt{\frac{n}{4}}}{\sqrt{n}+n}$ is minimax (MOOD et al, 1963, theorem 14, page 350) and

$$E\left[\left(\hat{p}_B - p\right)^2\right] = \frac{np(1-p)}{\left(\sqrt{n}+n\right)^2} + \left(\frac{np+\sqrt{\frac{n}{4}}}{\sqrt{n}+n} - p\right)^2$$

$$= \frac{np(1-p)}{\left(\sqrt{n}+n\right)^2} + \frac{\left[\left(np+\sqrt{\frac{n}{4}}\right) - 2p\sqrt{\frac{n}{4}} - np\right]^2}{\left(\sqrt{n}+n\right)^2} = \frac{n}{4\left(n+\sqrt{n}\right)^2}.$$

Now, as $E\left[\left(\hat{p}_B - p\right)^2\right] = \frac{n}{4\left(n+\sqrt{n}\right)^2}$ and $E\left[\left(\hat{p} - p\right)^2\right] = \frac{p(1-p)}{n}$.

Figure 1 helps to choose between these two estimators.

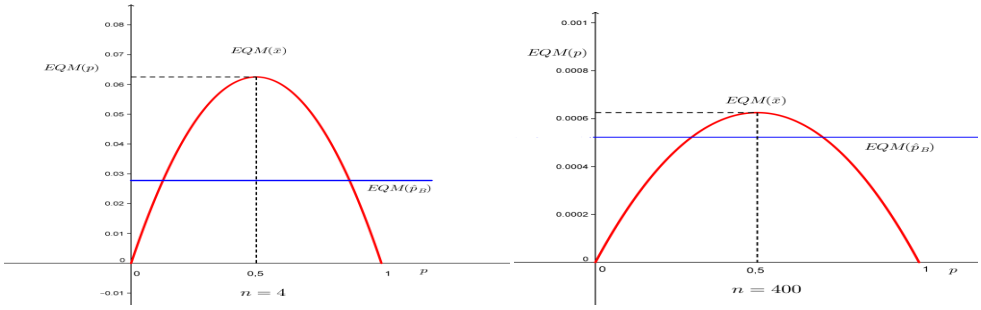

Figure 1 - Comparison of MSE of $\hat{p}$ (blue) and $\hat{p}_B$ (red) for different sample sizes

For small values of *n*, clearly $\hat{p}_B$ is a better choice, except if there is a strong reason to believe that *p* is near 0 or 1. For high values of *n* , the choice is the estimator $\hat{p}$, except for an strong believe that *p* is near $1/2$.

8 Comparing the estimator $\hat{\mu}_1$ and the pseudo-estimator $\hat{\mu}$

To emphasize the level of complexity implied by the shrinking procedure we compare the behavior of the pseudo estimator $\hat{\mu} = \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2}\bar{Y}_n$ and the shrinkage estimator

$$\hat{\mu}_1 = \frac{1}{\frac{S^2}{n\overline{Y}_n^2} + 1}\overline{Y}_n \ .$$

Considering the distributions Normal, Uniform and Double Exponential, we compute the mean squared error using 10.000 samples of sizes 20, 40 and 100 with variances 4, 16 and 36. The results are shown in Figures 2, 3 and 4.

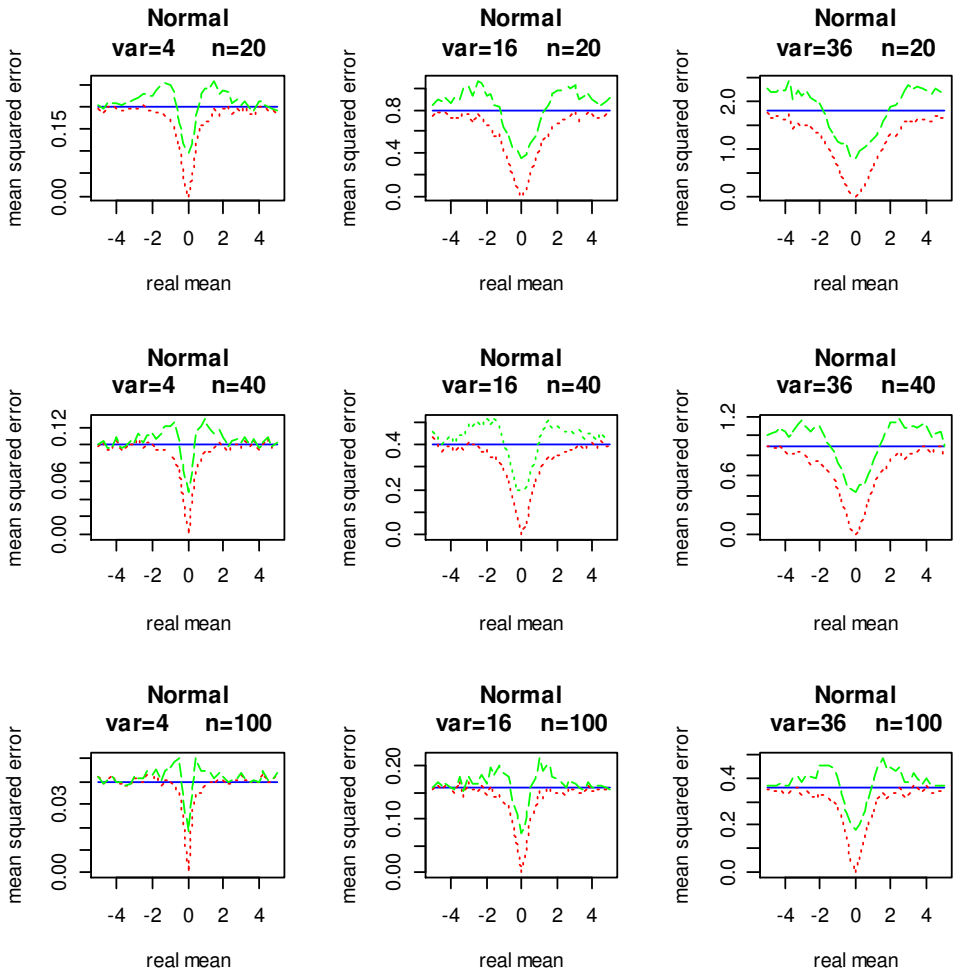

Figure 2 - MSE for $\overline{Y}_n$ (blue – solid), $\hat{\mu}$ (red – dotted) and $\hat{\mu}_1$ (green – long dash) for Normal distribution with different variances and sample sizes.

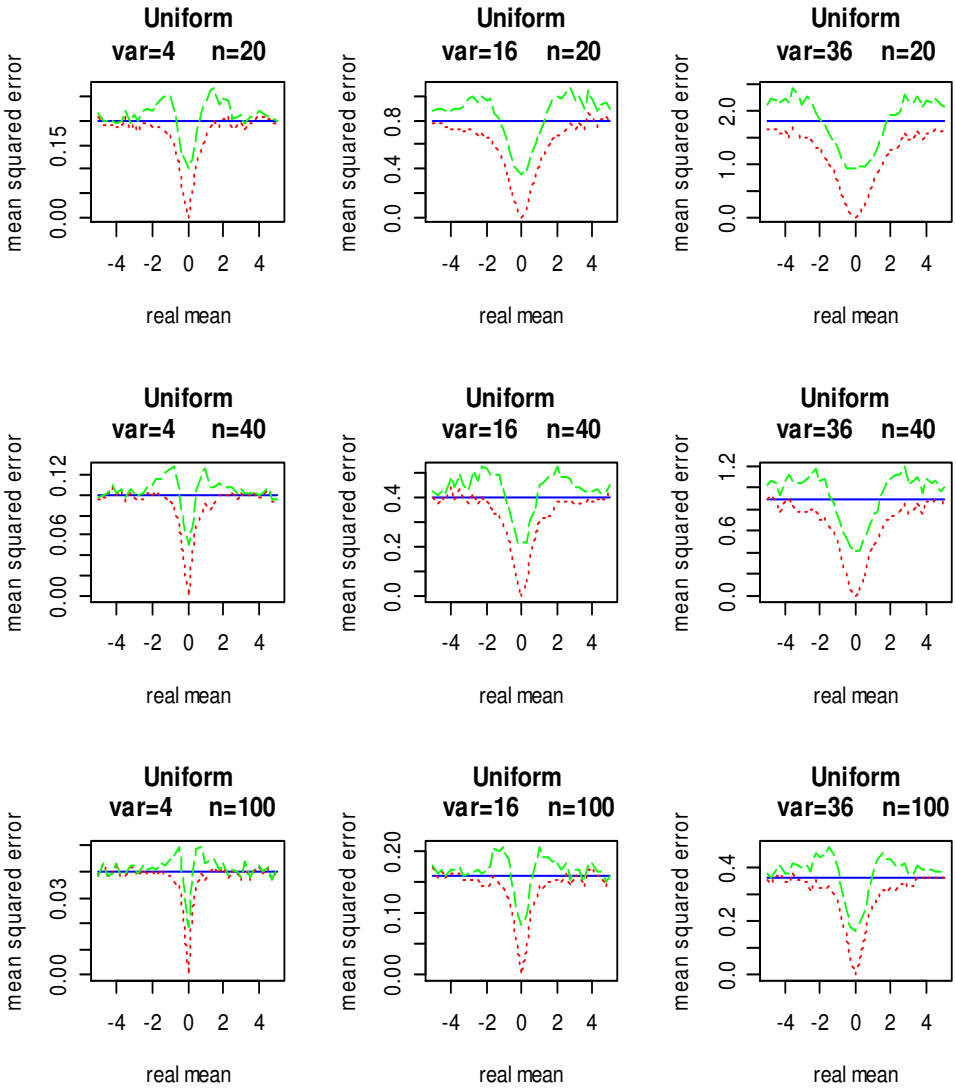

Figure 3 - MSE for $\overline{Y}_n$ (blue – solid), $\hat{\mu}$ (red – dotted) and $\hat{\mu}_1$ (green – long dash) for Uniform distribution with different variances and sample sizes.

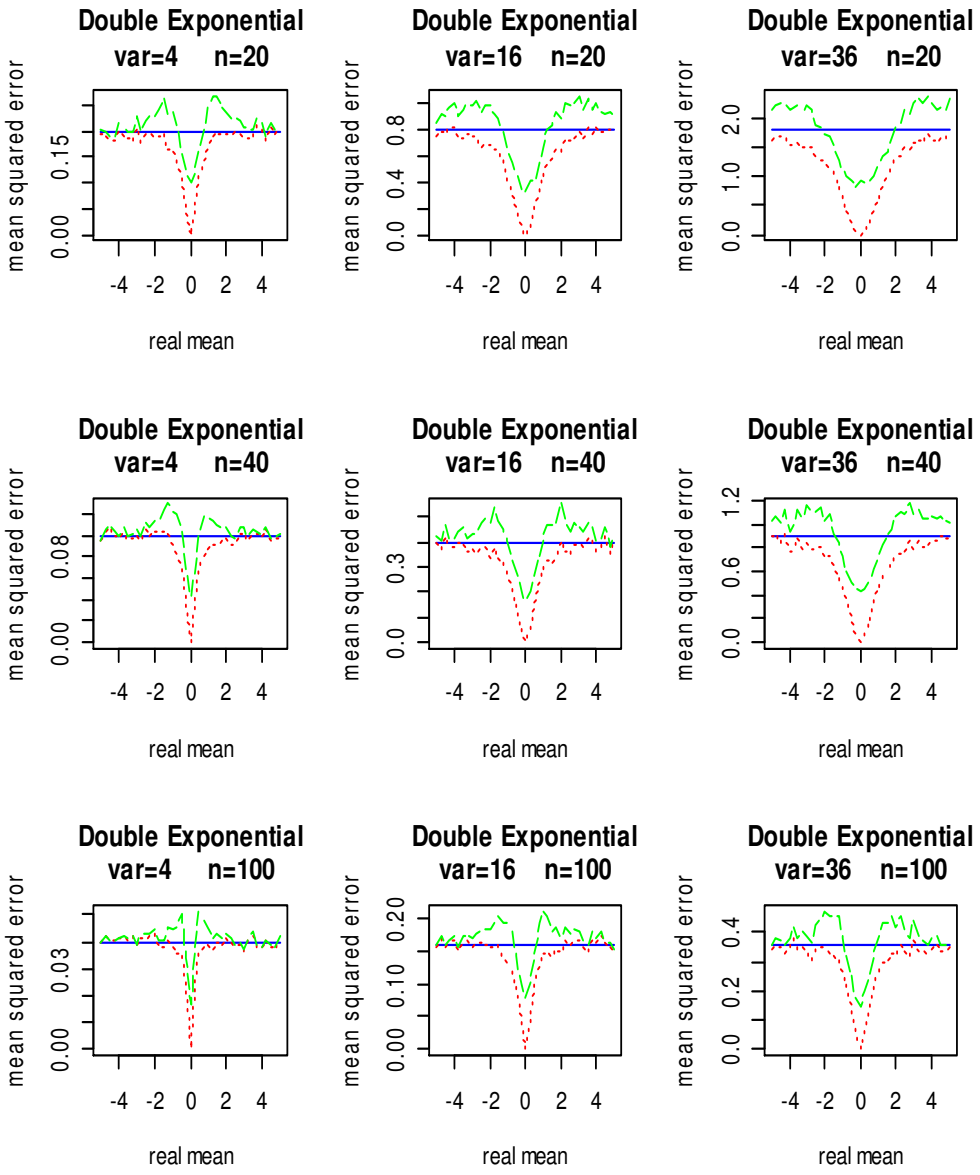

Figure 4 - MSE for $\overline{Y}_n$ (blue – solid), $\hat{\mu}$ (red – dotted) and $\hat{\mu}_1$ (green – long dash) for Double Exponential distribution with different variances and sample sizes.

About figures 2, 3 and 4, observe that the pseudo-estimator $\hat{\mu}$ behaves as it is supposed to, that is, except for some fluctuation when the real mean is far from the origin, it has MSE lower than the estimator $\overline{Y}_n$. On the other hand, the estimator $\hat{\mu}_1$ has a complex behavior and only for the real mean near the origin it has MSE lower than the estimator $\overline{Y}_n$. We couldn't find or guess a good reason for that strange behavior.

Conclusion

The shrinkage procedure shows to be fully compatible with Bayesian and empirical Bayesian method. The major problem, when shrinking an estimator, is to decide the amount of reduction to be applied. The Bayesian method can help, given an interpretation of the necessary reduction to be used. The complexity introduced by shrinkage is far from been completely understood.

RIZZO, F. N.; GAJO, C. A.; SOUZA, D. J.; CHAVES, L. M. Uma abordagem Bayesiana do estimador de encolhimento. *Rev. Bras. Biom*. São Paulo, v.33, n.4, p.485-602, 2015.

- RESUMO: Estimadores obtidos por encolhimento do estimador de mínimos quadrados usual têm sido amplamente utilizados, desde os trabalhos de James-Stein (1961) e Hoerl e Kennard (1970), sobre estimadores de cumeeira. Neste trabalho, apresentamos uma abordagem para alguns importantes estimadores de encolhimento, utilizando procedimentos Bayesianos e Bayesianos empíricos.

- PALAVRAS-*CHAVE:* Estimador de encolhimento, Bayes empírico, estimador de quadrados mínimos.

References

CARLIN, B. P.; LOUIS, T. A.. *Bayes and empirical Bayes methods for data analysis*. New York:Chapman & Hall/CRC, 2000.

CASELLA, G.; BERGER, R. L. *Inferência estatística*. 2. ed. São Paulo: Cengage Learning, 2010. 573p.

GRUBER, M. H. J. *Improvingefficiency by shrinkage: the James-Stein and ridge regression estimators*. New York: CRC, 1998. 648 p.

GRUBER, M. H. J. *Regression estimators: a comparative study*. Baltimore: The Johns Hopkins Press/CRC, 2010. 412p.

HOERL, A. E.; KENNARD, R. W. , Biased Estimation for Nonorthogonal Problems, *Technometrics*, v. 12, n. 1, p. 55-67, 1970.

MOOD, A. M.; BOES, F. A.; GRAYBILL, F. A. *Introduction to the theory of statistics*. 3rd ed. Columbus: McGraw-Hill, 1974. 564p.

JAMES, W.; STEIN, C. Estimation with Quadratic Loss. Proc. 4th Berkeley Sympos. Math. Statist. And Prob., v. 1, 1961, Univ. California Press, Berkeley, Calif., p.361-379, 1961.

TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. *J.R. Statis. Soc*., v.58, n.1, p.267-288, 1996.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *J.R. Statis. Soc*., v. 68, n.2, p.301-320, 2005.

Received in 11.05.2015

Approved after revised in 12.08.2015

Appendix

The predictive for normal and mean prior normal

$$f(\bar{y}) = \int_{-\infty}^{\infty} (\bar{y} \mid \mu)\pi(\mu)d\mu$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2/n}} exp\left\{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} exp\left\{-\frac{1}{2\tau^2}\mu^2\right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi/n}\sigma\tau} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{n}{2\sigma^2}\left(\bar{y}^2 - 2\bar{y}\mu + \mu^2\right) - \frac{1}{2\tau^2}\mu^2\right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi/n}\sigma\tau} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\left\{\mu^2\left[\left(-\frac{1}{2}\right)\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\right] + \frac{2n\bar{y}\mu}{2\sigma^2} - \frac{n\bar{y}^2}{2\sigma^2}\right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi/n}\sigma\tau} \exp\left\{-\frac{n\bar{y}^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{1}{\left(\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)}\left[\mu^2 - \frac{2n\bar{y}\mu\tau^2}{\left(\sigma^2+n\tau^2\right)}\right]\right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi/n}\sigma\tau} \exp\left\{-\frac{n\bar{y}^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{1}{\left(\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)}\left\{\left[\mu - \frac{n\bar{y}\tau^2}{\left(\sigma^2+n\tau^2\right)}\right]^2 - \left(\frac{n\bar{y}\tau^2}{\left(\sigma^2+n\tau^2\right)}\right)^2\right\}\right\} d\mu$$

$$= \frac{1}{\sqrt{\frac{2\pi\sigma^2\tau^2}{n}}} \exp\left\{-\frac{n\bar{y}^2}{2\sigma^2} + \frac{n^2\bar{y}^2\tau^4}{\left(\sigma^2+n\tau^2\right)2\sigma^2\tau^2}\right\} \sqrt{\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}}$$

$$\int \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)}} \exp\left\{-\frac{1}{2}\frac{1}{\left(\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)}\left[\mu - \frac{n\bar{y}\tau^2}{\left(\sigma^2+n\tau^2\right)}\right]\right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2 + n\tau^2}{n}\right)}} \exp\left\{-\frac{1}{2}\left[\frac{n\sigma^2\tau^2 + n^2\tau^4 - n^2\tau^4}{\left(\sigma^2 + n\tau^2\right)\sigma^2\tau^2}\right]\bar{y}^2\right\}$$

$$= \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2 + n\tau^2}{n}\right)}} \exp\left\{-\frac{1}{2}\left[\frac{1}{\left(\frac{\sigma^2 + n\tau^2}{n}\right)}\right]\bar{y}^2\right\}$$

$$= \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n} + \tau^2\right)}} \exp\left\{-\frac{1}{2\left(\frac{\sigma^2}{n} + \tau^2\right)}\bar{y}^2\right\}.$$