# ST758, Homework 9 (Shrinkage)

## Due Dec 9, 2014 @ 11A (final exam date)

Shrinkage, an idea originating from Charles Stein, plays an important role in modern statistical estimation theory. At least two types of shrinkage estimators abound in current literature: regularized (or penalized) regressions and Bayesian shrinkage. In this homework, we compare the finite sample performance of some shrinkage estimators with their classical maximum likelihood estimation (MLE) counterparts using Monte carlo simulations.

1. Read carefully the following material about how to properly design Monte carlo studies for comparing statistical methods. These are available on course website.

    - M Davidian, Simulation studies in statistics, *slides for ST810A Spring 2005*.
    - JF Monahan, A guide for simulation studies in statistics, *unpublished note*.
    - WH Swallow and JF Monahan (1984), Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components, *Technometrics*, 26(1):47–57.

    Your report will be evaluated based on these guidelines.

2. (Estimate multiple multinomial parameters) In HW6/HW7, we implemented algorithms for the MLE of the Dirichlet-Multinomial distribution. The professor talks about possible use of this for empirical Bayes (EB) estimation of the multinomial parameters from multiple populations. That sounds crap! Why shall we "borrow information across populations" even when those populations are totally unrelated? Design and carry out a Monte carlo simulation study to convince yourself whether the empirical Bayes approach is a good idea or not.

    Following hints may be useful:

    - You may include following two factors in your simulation study: number of populations $I$ (say $I = 10, 20, 50$) and number of categories $d$ of multinomial (say $d = 2, 3, 5, 10$).
    - Choose a fixed batch size $N$ (say 20) for multinomial and a fixed number of simulation replicates $S$ (how large is good?) for your simulation study.
    - At each simulation replicate
        - Generate a "true" multinomial parameter $\boldsymbol{p}_i$ for each population $i = 1, \ldots, I$. You can make your own choice how to generate these $\boldsymbol{p}_i$. But make sure to generate them independently. That is these populations are indeed unrelated to each other.
        - Generate a count vector from each population $\boldsymbol{x}_i \sim \text{Multinomial}(N, \boldsymbol{p}_i)$ independently.
        - The MLE of multinomial parameters $\boldsymbol{p}_i$ in each population are

$$\hat{\boldsymbol{p}}_i^{\text{MLE}} = \frac{\boldsymbol{x}_i}{N}, \quad i = 1, \ldots, I.$$

- Estimate the Dirichlet-Multinomial parameter $\boldsymbol{\alpha}$ using the combined data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I)$ ($I$ observations). Then the empirical Bayes estimate of the multinomial parameters $\boldsymbol{p}_i$ for each population are

$$\hat{\boldsymbol{p}}_i^{\text{EB}} = \frac{\boldsymbol{x}_i + \hat{\boldsymbol{\alpha}}}{N + |\hat{\boldsymbol{\alpha}}|}, \quad i = 1, \ldots, I.$$

  You can use either your own functions or the functions in Professor's solution for fitting the Dirichlet-Multinomial distribution. Why the empirical Bayes estimate is called a shrinkage estimator?

- Record the estimation errors from MLE and empirical Bayes (EB) approach respectively

$$\begin{aligned} \text{Err}_s^{\text{MLE}} &= \frac{1}{I}\sum_{i=1}^{I} \|\hat{\boldsymbol{p}}_i^{\text{MLE}} - \boldsymbol{p}_i\|_{\text{TV}} = \frac{1}{2I}\sum_{i=1}^{I}\sum_{j=1}^{d} |\hat{p}_{ij}^{\text{MLE}} - p_{ij}| \\ \text{Err}_s^{\text{EB}} &= \frac{1}{I}\sum_{i=1}^{I} \|\hat{\boldsymbol{p}}_i^{\text{EB}} - \boldsymbol{p}_i\|_{\text{TV}} = \frac{1}{2I}\sum_{i=1}^{I}\sum_{j=1}^{d} |\hat{p}_{ij}^{\text{EB}} - p_{ij}|. \end{aligned}$$

- Summarize your estimated $\text{Err}^{\text{MLE}}$ and $\text{Err}^{\text{EB}}$ at each combination of $I$ and $d$ along with their standard errors by tables and/or figures.
- Draw proper conclusions based on your simulation results.

3. (Linear regression) Consider Gaussian model $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{y} \in \mathbb{R}^n$ is the response vector and $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is the covariate matrix. We compare three approaches – least squares (LS), ridge regression, and lasso regression – for estimating the regression coefficients $\boldsymbol{\beta}$ and prediction. The estimates being compared are

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{LS}} &= \arg\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \\ \hat{\boldsymbol{\beta}}_{\text{ridge}} &= \arg\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2 \\ \hat{\boldsymbol{\beta}}_{\text{lasso}} &= \arg\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1. \end{aligned}$$

Following hints may be useful:

- Why the ridge and lasso estimators $\hat{\boldsymbol{\beta}}_{\text{ridge}}$ and $\hat{\boldsymbol{\beta}}_{\text{lasso}}$ are called shrinkage estimators?
- Fix $p$ (say at 50). Possible factors in your simulation study include: sample size $n$ (say $n = 50, 100, 150, 200$), correlation $\rho$ (say $\rho = 0, 0.25, 0.5, 0.75$) between covariates, sparsity of $\boldsymbol{\beta}$, and noise level $\sigma^2$.
- You may use the `lm.ridge()` in the `MASS` package for fitting the ridge regression. Especially it outputs a generalized cross-validation (GCV) value for each input $\lambda$. GCV helps us decide which $\lambda$ value to use in the ridge regression. We choose the ridge solution at the $\lambda$ value with minimal GCV.

- You may use the `glmnet` package for fitting the lasso regression. You can use the Bayesian information criterion (BIC) for choosing the tuning parameter value $\lambda$. For example we may choose the lasso solution at the $\lambda$ value with minimal BIC.
- Compare the three estimates $\hat{\boldsymbol{\beta}}_{\text{LS}}$, $\hat{\boldsymbol{\beta}}_{\text{ridge}}$ and $\hat{\boldsymbol{\beta}}_{\text{lasso}}$ in terms of
  - mean squared error (MSE) of parameter estimate and
  - prediction errors.
- Summarize and present your results by tables and/or figures.

4. Summarize your results in a written report. Your report should follow

   - statement of problem
   - questions to be addressed
   - for each example (multinomial estimation and linear regression)
     - description of the design of the experiment
     - computational details
     - presentation of results (tables, figures)
     - analysis of results
   - conclusions
   - of course, include your code (with comments) and output, which will be graded too

   Be sure to temper the complexity of your study (e.g. the number of factors, levels, or responses (bias, MSE, etc.)) with your ability to address an issue clearly and completely.