# 18.650. Statistics for Applications
# Fall 2016. Problem Set 10

Due Friday, Dec. 2 at 12 noon

## Problem 1 Bayesian Estimation

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \theta)$, for some unknown positive number $\theta$.

1. Compute the maximum likelihood estimator of $\theta$.
2. Prove that the MLE is asymptotically normal and find its asymptotic variance.
3. In a Bayesian approach:
   a) Compute Jeffreys prior ? Is it proper ?
   b) Use Bayes' formula in order to compute the posterior distribution. Is it well defined ?
   c) Compute the Bayesian estimator of $\theta$ associated with Jeffreys prior.
   *Recall that the inverse Gamma distribution with parameters $\alpha > 1, \beta > 0$ has density*
   $$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{e^{-\beta/x}}{x^{\alpha+1}}, x > 0.$$
   *Its expectation is given by* $\dfrac{\beta}{\alpha - 1}$.

## Problem 2 Bayesian Estimation and Linear Regression

Let $X_1, \ldots, X_n$ be $n$ deterministic vectors in $\mathbb{R}^p$ and let $\boldsymbol{X}$ be the $n \times p$ matrix whose rows are $X_1', \ldots, X_n'$. Consider a sample $Y_1, \ldots, Y_n$, such that conditionally on some $p$ dimensional random vector $\boldsymbol{\beta}$, $Y_1, \ldots, Y_n$ are independent and for each $i = 1, \ldots, n$,

$$Y_i - X_i'\boldsymbol{\beta} \sim \mathcal{N}(0, \sigma^2),$$

where $\sigma^2$ is a given positive number. Denote by $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$.

1. Conditionally on $\boldsymbol{\beta}$, what is the distribution of $\boldsymbol{Y}$ ?
2. Assume that the prior distribution on $\boldsymbol{\beta}$ is $\mathcal{N}_p(0, \tau^2 I_p)$, where $\tau^2$ is a fixed positive number.
   a) Prove that the posterior distribution of $\boldsymbol{\beta}$ (i.e., the distribution of $\boldsymbol{\beta}$ conditional on $\boldsymbol{Y}$) has a density proportional to $\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\beta\|^2\right)\right)$, for some $\lambda$ to be determined.

b) Conclude that the posterior distribution of $\boldsymbol{\beta}$ is Gaussian and determine the corresponding parameters.
*Hint: The density of the Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ is given by*

$$g(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right).$$

c) What is the posterior mean of $\boldsymbol{\beta}$ ?

3. Consider a frequentist approach: Assume that the linear regression of $Y_i$ on $X_i$ is $Y_i = X_i'\boldsymbol{\beta} + \varepsilon_i$, for some unknown vector $\boldsymbol{\beta} \in \mathbb{R}^p$, where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma^2)$, for some $\sigma^2 > 0$. The *Ridge* estimator of $\boldsymbol{\beta}$ is defined as:

$$\hat{\boldsymbol{\beta}}^{(R)} \in \underset{\boldsymbol{t}\in\mathbb{R}^S}{\text{argmin}} \ \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{t}\|^2 + \lambda\|\boldsymbol{t}\|^2,$$

where $\lambda$ is a *tuning parameter*, i.e., a given number chosen by the statistician. Note that similarly to the BIC or the Lasso estimators, the Ridge estimator minimizes a penalized version of the sum of squared errors.

a) Compute the Ridge estimator.
b) Using the previous questions, prove that there exists a value of $\tau^2$ such that $\hat{\boldsymbol{\beta}}^{(R)}$ is equal to the Bayesian estimator with prior distribution $\mathcal{N}_p(0, \tau^2 I_p)$.
c) What is the distribution of $\hat{\boldsymbol{\beta}}^{(R)}$ ?
d) Compute the quadratic risk of $\hat{\boldsymbol{\beta}}^{(R)}$ in terms of the matrix $\boldsymbol{X}$, $\lambda$ and $\boldsymbol{\beta}$.

## Problem 3 Covariance Matrices

1. Recall the definition of the covariance matrix of a $p$-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$. What are its dimensions ?
2. Given a sample of $n$ random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of size $p$, recall the definition of the sample covariance matrix.
3. Prove that the covariance matrix of a random vector is positive semi-definite.
4. Prove that the sample covariance matrix of a sample of random vectors is positive semi-definite.
5. Let $\boldsymbol{X}$ be a $p$-dimensional random vector and let $\Sigma$ be its covariance matrix. Let $A$ be a $q \times p$ matrix.
   a) What is the covariance matrix of the random vector $A\boldsymbol{X}$ ? Prove your answer.
   b) If $q > p$, can the covariance matrix of $A\boldsymbol{X}$ be invertible ? Why ?
   c) If $\boldsymbol{u} \in \mathbb{R}^p$, what is the variance of $\boldsymbol{u}'\boldsymbol{X}$ ?

6. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be $p$-dimensional random vectors and let $\hat{\Sigma}$ be the corresponding sample covariance matrix.
   a) Let $B$ be a $q \times p$ matrix. What is the sample covariance matrix of $B\boldsymbol{X}_1, \ldots, B\boldsymbol{X}_n$ ? Prove your answer.
   b) If $\boldsymbol{u} \in \mathbb{R}^p$, what is the sample covariance of $\boldsymbol{u}'\boldsymbol{X}_1, \ldots, \boldsymbol{u}'\boldsymbol{X}_n$ ?
7. Let $\boldsymbol{X}$ be a $d$-dimensional random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. Let $A \in \mathbb{R}^{k \times d}$. Prove that

$$\mathbb{E}\left[\boldsymbol{X}'A'A\boldsymbol{X}\right] = \|A\boldsymbol{\mu}\|_2^2 + Tr(A\Sigma A').$$

MIT OpenCourseWare
https://ocw.mit.edu

18.650 / 18.6501 Statistics for Applications
Fall 2016

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.