## Foundations of Statistical Inference, BS2a, Exercises 4

Some questions may require you to look up material fromlectures 14 and 15.

1. Suppose $x_i$, $i = 1, 2, \ldots, k$ has a Binomial distribution $B(n, \theta_i)$ and we wish to construct an EB estimator for $\theta_i$ using the prior $\theta_i \sim \text{Beta}(\alpha, \beta)$. Using the identities

$$\mathbb{E}(X_i) = \mathbb{E}(\mathbb{E}(X_i|\theta_i))$$

and

$$\text{Var}(X_i) = \mathbb{E}\left[\text{Var}(X_i \mid \theta_i)\right] + \text{Var}\left[\mathbb{E}[X_i \mid \theta_i]\right]$$

give method-of moments estimators for $\alpha$ and $\beta$ (use the first and second moment). Show estimates of $\alpha$ and $\beta$ are

$$\alpha = \frac{n\bar{x}^2 - \bar{x}^3 - \bar{x}s^2}{ns^2 - n\bar{x} + \bar{x}^2}, \; \beta = \frac{(n - \bar{x})(n\bar{x} - \bar{x}^2 - s^2)}{ns^2 - n\bar{x} + \bar{x}^2}$$

where $\bar{x}$ and $s^2$ are the sample mean and variance of $x_1, \ldots, x_k$. Given an Empirical Bayes estimator for $\theta_i$, how do the EB estimators for $\theta_i, i = 1, 2, ..., k$ differ from the corresponding MLE's? (If you need to use a loss function you can assume that it is quadratic).

2. (Adapted from 2016) Suppose that we have a parametric family of distributions $f(x|\theta)$ with $\theta \in \mathbb{R}$ for an observation $X$ and that we want to estimate $\theta$ using a loss function $L(d, \theta)$ for the estimator $d$.

   (a) i. Define the notions of *risk function, Bayes risk* and *Bayes estimator.*

   ii. State the Bayes estimator when $L(d, \theta) = (d - \theta)^2$? When $L(d, \theta) = |d - \theta|$? What is the limit of the Bayes estimator when $L(d, \theta) = c\mathbb{I}\{|d - \theta| > a\}$ and $a \to 0$?

   (b) Let $\underline{X} = (X_1, \ldots, X_n)$ be a sample such that $X_i \sim \text{Poisson}(\lambda_i), i = 1, \ldots, n$ and suppose that the $\lambda_i$ themselves are i.i.d. with common distribution $\text{Gamma}(a, b)$ where $a, b > 0$. Recall that the density of the $\text{Gamma}(a, b)$ distribution is

$$\frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, x > 0$$

   and that the mean of a $\text{Gamma}(a, b)$ variable is $a/b$.

i. Assume first that $(a, b)$ are known. For a fixed value of $i \in \{1, \ldots, n\}$, write down the posterior distribution of $\lambda_i$ and deduce the Bayes estimator of $\lambda_i$ assuming that we are using a quadratic loss function.

ii. Continuing to assume that $(a, b)$ are known, show that the marginal distribution of $X_i$ is negative binomial with some parameters that you must identify. Hence deduce the marginal distribution $m(\underline{x}|a, b)$ of $\underline{X} = (X_1, \ldots, X_n)$.

iii. Suppose that only $a$ is known. Using b.(ii), find the marginal MLE $\hat{b}$ of $b$.

iv. Continuing to assume that $a$ is known but not $b$, what is the empirical Bayes estimator of $\lambda_i$ using quadratic loss? Show that this empirical Bayes estimator is a *shrinkage* estimator.

(c) We now wish to analyze the same situation (i.e. the $X_i$ are independent Poisson($\lambda_i$) variables, the $\lambda_i$ are i.i.d. Gamma$(a, b)$ variables with $a$ known) but this time we want to put a prior $\psi$ on $b$ to build a hierarchical model.

i. Give an expression of the hyper-posterior $\pi(b|\underline{x})$ of $b$ of the form

$$\pi(b|\underline{x}) \propto \phi(b, \underline{x})\psi(b)$$

where the constant of propotionality does not depend on $b$. (Hint: use part b.ii above.)

ii. Suppose that $\psi$ has an F-distribution

$$\psi(b) \propto \frac{b^{\alpha-1}}{(1+b)^{\alpha+\beta}}, b > 0,$$

where $\alpha, \beta$ are positive constants. Compute the normalizing constant $\int_0^\infty \phi(b, \underline{x})\psi(b)db$ (Hint: the change of variable $t = b/(1+b)$ can be used).

iii. Hence deduce the hierarchical Bayes estimator of $\lambda_i$ (still assuming a quadratic loss function).

iv. Are there parameters $\alpha, \beta$ for which the hierarchical Bayes estimator coincides with the empirical Bayes estimator? What is $\psi$ for these parameters? Is it a proper prior?

3. Show that the Kullback-Liebler divergence between two continuous probability densities $q(x)$ and $p(x)$ has the properties

(a) $KL(q|p) \geq 0$

(b) Show that $KL(q|p) = 0$ if $q = p$.

4. Consider a Normal pdf $q(x) = N(x; \mu, \sigma^2)$ and Gamma pdf $p(x) = \text{Gamma}(x; a, b)$

   (a) Derive an approximate expression for $KL(q|p)$ (see Hint below).

   (b) Show that this approximate $KL(q|p)$ is minimized by

$$\mu = \frac{a}{b}, \quad \sigma^2 = \frac{\mu^2}{a-1}$$

   Hint : use the following approximation for the expectation of the log of a Normal random variable

$$X \sim N(\mu, \sigma^2) \Rightarrow \mathbb{E}(\log X) \approx \log \mu - \frac{\sigma^2}{2\mu^2}$$

5. Consider applying Variational Bayes to the hierarchical model

$$D_i \sim N(\mu, \tau^{-1}) \quad i = 1, \ldots, P$$

   with priors

$$\mu \sim N(m, \beta^{-1}) \quad \tau \sim \Gamma(a, b)$$

   Assuming a mean-field approximation to the posterior distribution $Q(\mu)Q(\tau) \approx \pi(\mu, \tau|D)$ derive the form of the components $Q(\mu)$ and $Q(\tau)$.

6. Let $X_1, X_2, X_3, X_4$ be multinomial with total count $N = \sum_i X_i$ and probabilities $(\frac{1}{2} - \frac{1}{2}\theta, \frac{1}{4}\theta, \frac{1}{4}\theta, \frac{1}{2})$. Suppose we observe counts $Y_1, Y_2, Y_3$ where $Y_1 = X_1$, $Y_2 = X_2$ and $Y_3 = X_3 + X_4$. Therefore the missing data is $X_3$ and $X_4$.

   (a) Write down the full data log-likelihood function.

   (b) Calculate the conditional distribution of the missing data given the observed data and a current estimate $\theta_t$.

   (c) Derive the expected value of the full data log-likelihood function using the conditional distribution in (b).

   (d) Derive the M-step of an EM algorithm to estimate $\theta$ (this will be in next week lectures, see slides)o.