Statistics & Data Sciences: First Year Prelim Exam
May 2017

Instructions:

1. Do not turn this page until instructed to do so.
2. Start each new question on a new sheet of paper.
3. Attached is a page of pdf’s and cdf’s.
4. This is a closed book exam.
5. Show your working and explain your reasoning as fully as possible.
6. Switch off all electronic devices.
7. The time limit is 3 hours.
8. The exam has 5 questions – do as much as you can.
Question 1. Consider the mixture model

\[ g(x|\theta) = \sum_{j=1}^{M} w_j N(x|\mu_j, \sigma^2), \]

where \( M \) is finite and known. Prior distributions are assigned as follows: For \( w = (w_1, \ldots, w_M) \) there is a Dirichlet distribution with common parameter \( \alpha > 0 \). Hence

\[ f(w) \propto \prod_{j=1}^{M} w_j^{\alpha-1} \times 1(w_1 + \cdots + w_m = 1). \]

For each \( \mu_j \) the prior is normal with mean 0 and variance \( \tau^2 \), and for \( \lambda = 1/\sigma^2 \) there is a gamma prior with shape and scale parameters both \( \epsilon > 0 \); i.e.

\[ f(\lambda) \propto \lambda^{\epsilon-1} e^{-\lambda \epsilon}. \]

By introducing suitable latent variables, describe a Markov chain Monte Carlo algorithm for estimating the model. Provide all details so that the reader could implement the algorithm on a computer without needing to ask any further questions.

The aim is to estimate the density from which the samples \((x_1, \ldots, x_n)\) are coming. The Bayesian estimate is the predictive density function. Define this function and show how the output of the chain can be used to estimate it.
Question 2. Observations \( x_1, x_2, \ldots \) are i.i.d. from some density function \( g_0(x) = g(x|\theta_0) \). For short-hand, we write \( g_0(x) = g(x|\theta_0) \) and \( g_\theta(x) = g(x|\theta) \).

The maximum likelihood estimator \( \hat{\theta} \) exists and satisfies
\[
n^{-1} \log R_n(\hat{\theta}) \to 0
\]
where
\[
R_n(\theta) = \prod_{i=1}^{n} \frac{g(x_i|\theta)}{g(x_i|\theta_0)}, \quad \theta \in \Theta.
\]
The Hellinger distance \( H \) between densities \( g_0 \) and \( g_\theta \) is given by
\[
\frac{1}{2} H^2(g_0, g_\theta) = 1 - \int \sqrt{g(x|\theta_0) g(x|\theta)} \, dx.
\]

(i) Prove that
\[
E R_n(\theta)^{1/2} = \left[ 1 - \frac{1}{2} H^2(g_0, g_\theta) \right]^n
\]
where the expectation is with respect to the \( (x_i) \).

A Bayesian assigns prior \( f(\theta) \) to \( \theta \).

(ii) Using that \( R_n(\theta) \leq R_n(\hat{\theta}) \), prove that for large \( n \),
\[
\int_{A_\epsilon} R_n(\theta) \, f(\theta) \, d\theta \leq e^{\frac{1}{2}nc} e^{-\frac{1}{4}nc^2}
\]
for any \( c > 0 \), where
\[
A_\epsilon = \{ \theta : H(g_0, g_\theta) > \epsilon \}.
\]
You are given that for a sequence of non-negative random variables \( Z_n \); if
\[
E Z_n < e^{-n\delta}
\]
then for all large \( n \), \( Z_n < e^{-\frac{1}{2}n\delta} \). Also note that \( R_n(\theta) \leq R_n(\theta)^{1/2} R_n(\hat{\theta})^{1/2} \).

(iii) If it is known that for all large \( n \)
\[
\int_{\Theta} R_n(\theta) \, f(\theta) \, d\theta > e^{-nd}
\]
for any \( d > 0 \), prove that
\[
F(A_\epsilon|x_1, \ldots, x_n) \to 0 \quad \text{as} \quad n \to \infty
\]
where
\[
F(A_\epsilon|x_1, \ldots, x_n) = \frac{\int_{A_\epsilon} R_n(\theta) \, f(\theta) \, d\theta}{\int_{\Theta} R_n(\theta) \, f(\theta) \, d\theta}
\]
is the posterior mass assigned to \( A_\epsilon \).

(iv) Explain the significance of the outcome in (iii).
**Question 3.** Suppose observations \((y_i)\) for \(i = 1, \ldots, n\) follow a normal distribution with unknown mean parameter \(\theta\) and known variance \(\sigma^2\). Consider an estimator of \(\theta\) of the form

\[
\hat{\theta} = c \bar{y}
\]

where \(\bar{y}\) is the sample mean.

(i) Find the mean square error for \(\hat{\theta}\) in terms of the true value \(\theta_0\). What is the optimal choice for \(c\) in terms of \(\theta_0\) under this mean square error criterion.

(ii) Regardless of the value of \(\theta_0\), save \(\theta_0 \neq 0\), explain why whatever sequence of \(c_n\) is chosen, it must be that \(c_n \to 1\) as \(n \to \infty\) in order for \(\hat{\theta}\) to converge to \(\theta_0\).

(iii) If \(\tilde{\theta}\) is an unbiased estimator of \(\theta\); e.g. \(\tilde{\theta} = \bar{y}\), then Cramér–Rao says that

\[
\text{Var} \tilde{\theta} \geq \frac{\sigma^2}{n}.
\]

On the other hand, if \(\hat{\theta}\) is a biased estimator of \(\theta\) with bias \(b(\theta_0)\), then

\[
\text{Var} \hat{\theta} \geq (1 + b'(\theta_0))^2 \frac{\sigma^2}{n}.
\]

Confirm that the variance of \(\hat{\theta}\), where \(\hat{\theta} = c\bar{y}\), does have a lower bound given in (1).

(iv) Try and prove the lower bound in (iii). That is, if \(\hat{\theta}\) has bias \(b(\theta_0)\) then (1) holds. In order to attempt this you would need to know how the proof of the Cramér–Rao lower bound works.

Hint: Take \(T\) as the biased estimator, so \(E T = \theta + b(\theta)\) where the expectation is with respect to the sample \(y\). Then take

\[
Z = \frac{\partial}{\partial \theta} \log p(y|\theta)
\]

where \(p(y|\theta)\) is normal with mean \(\theta\) and variance \(\sigma^2/n\). First show that \(E Z = 0\). Now use the fact that a correlation is bounded between \(-1\) and \(+1\), so

\[
[E \{(T - \theta) Z\}]^2 \leq \text{Var}(T - \theta) \times E Z^2.
\]
Question 4. Suppose observations \((y_i)\) for \(i = 1, \ldots, n\) have a normal distribution with unknown mean \(\theta\) and known variance \(\sigma^2\).

The aim is to test

\[
H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \neq 0. 
\]

The test statistic is \(T = \bar{y}\).

(i) Describe the test in full, such as the critical value for a Type I error of \(\alpha\), and derive the power function for the test.

(ii) A Bayesian assigns a prior \(f(\theta)\) which is normal with mean 0 and variance \(\tau^2\). Find the Bayes factor for the test, given by

\[
B = \frac{m(\text{data}|H_1)}{m(\text{data}|H_0)}
\]

where \(m(\text{data}|H)\) is the marginal likelihood of the data given hypothesis \(H\).

(iii) The Bayesian rejects \(H_0\) if \(B > \lambda\) for some \(\lambda\). How would the Bayesian set \(\lambda\) so that the Type I error of the test is \(\alpha\).

(iv) With this choice of \(\lambda\), find the corresponding power function for this Bayesian test. Comment on your findings.
**Question 5.** Let $X_1, X_2, \ldots$ be non-negative independent and identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Also, for the positive integer valued random variable $N \in \{1, 2, \ldots\}$ it is that $E(N) = m$ and $\text{Var}(N) = \tau^2$. Define

$$S_N = \sum_{i=1}^{N} X_i.$$  

(i) Find $E(S_N)$ and $\text{Var}(S_N)$.

(ii) Suppose $E \phi^N = H(\phi)$ for $0 < \phi < 1$. Show that

$$E e^{-\theta S_N} = H(G(\theta))$$

where

$$G(\theta) = E e^{-\theta X}.$$ 

(iii) Now suppose $Z_0 = 1$ and recursively it is that

$$Z_n = X_1 + \cdots + X_{Z_{n-1}}.$$ 

If each $X_i$ is an independent Bernoulli random variable with $P(X_i = 1) = p$ for all $i$, prove that for $0 < s < 1$ it is that

$$G_n(s) = G_{n-1}(1 - p + ps),$$

where $G_n(s) = E s^{Z_n}$. Hence, prove that $G_n(s) = 1 - p^n + p^n s$.

(iv) Use (iii) to show that $P(Z_n = 0) \to 1$. 
