ST 790: Advanced Bayesian Inference
Midterm Project Description

Spring, 2013

Project report Due: March 1, 2013, 5:00PM

40% of your course grade will depend upon successful, on time completion
of the midterm project. Project reports are to result in thorough but concise,
professional quality technical reports of double spaced pages. Projects are to
be turned in as a single PDF attachment with following filename Formatting:
Lastname-ST790-MidtermReport.pdf by the deadline stated above. Codes (preferably written in R)
should also be submitted as a single plain
text attachment with filename Lastname-ST790-MidtermCodes.txt

Sign the Honor Pledge: I have neither given nor received unauthorized
aid on this assignment on your cover page, before you submit the project.

Consider a Markov Chain \( \{ \theta^{(l)}; \ l = 0, 1, 2, \ldots \} \) with transition kernel
density (TKD) \( T(\cdot, \cdot) : \Theta \times \Theta \to [0, \infty) \) and initial density \( p_0(\cdot) \) (wrt some
\( \sigma \)-finite measure \( \mu \)) where \( \Theta \subseteq \mathbb{R}^d \) for some \( d \geq 1 \). In other words, for
each \( \theta' \in \Theta, T(\theta', \cdot) \) is a density function (wrt the measure \( \mu \)) and for each
\( \theta \in \Theta, T(\cdot, \theta) \) is a measurable function (wrt the measure \( \mu \)) and we have the
following generating scheme:

\[
\theta^{(0)} \sim p_0(\cdot) \quad \text{and} \quad \theta^{(l)} \sim T(\theta^{(l-1)}, \cdot) \quad \text{for} \ l = 1, 2, \ldots
\]

The \( k \)-step TKD is defined recursively,

\[
T_k(\theta', \theta) = \int T_{k-1}(\theta', \tilde{\theta})T(\tilde{\theta}, \theta)d\mu(\tilde{\theta}) \quad \text{for} \ k = 1, 2, \ldots
\]

where \( T_0(\theta', \tilde{\theta}) \) = 0 if \( \theta' \neq \tilde{\theta} \) and is 1 otherwise. Assume that \( p(\cdot) \)
is the invariant density (wrt the measure \( \mu \)) and satisfies the following identity

\[
p(\theta) = \int p(\theta')T(\theta', \theta)d\mu(\theta') \quad \forall \ \theta \in \Theta
\]

In Bayesian analysis, we would know \( p(\theta) \) only upto its kernel \( K(\theta) = cp(\theta) \)
where \( c = \int K(\theta)d\mu(\theta) \) may depend on the observed data.
1. Let $T_0(\cdot, \cdot)$ be a TKD with invariant density $q(\cdot)$ satisfying the detailed balance condition (DBC) $q(\theta')T_0(\theta', \theta) = q(\theta)T_0(\theta, \theta')$. Consider the following variant of the Metropolis-Hastings algorithm: $\theta^{(0)} \sim p_0(\cdot)$ and for $l = 1, 2, \ldots$

$$\tilde{\theta} \sim T_0(\theta^{(l-1)}, \cdot)$$

$$\theta^{(l)} = \begin{cases} 
\tilde{\theta} & \text{with probability } \rho(\theta^{(l-1)}, \tilde{\theta}) \\
\theta^{(l-1)} & \text{otherwise}
\end{cases}$$

where $\rho(\theta', \theta) = \min\left\{ \frac{p(\theta')q(\theta')}{p(\theta)q(\theta)}, 1 \right\}$

(a) Obtain the TKD of the above Markov chain.
(b) Show that the TKD (in part (a)) satisfies the DBC with the invariant density $p(\cdot)$. 

[5+5=10 points]
2. Consider a resolvant \( T_\lambda(\theta', \theta) = e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k T_k(\theta', \theta)/k! \) where \( T_k(\theta', \theta) \) is as defined on p.1 for \( \theta, \theta' \in \Theta \) and assume \( \lambda > 0 \).

(a) Consider the following generating scheme:

\[
\begin{align*}
\theta^{(0)} & \sim p_0(\cdot) \\
\theta^{(l)} & \sim T^{(K_l)}(\theta^{(l-1)}, \cdot) \text{ where } K_l \sim \text{Pois}(\lambda) \text{ for } l = 1, 2, \ldots
\end{align*}
\]

where \( \text{Pois}(\lambda) \) denotes the Poisson distribution with mean \( \lambda \). Show that the above scheme generates a Markov Chain with TKD \( T_\lambda(\cdot, \cdot) \) for any \( \lambda > 0 \).

(b) Show that \( T_\lambda(\cdot, \cdot) \) and \( T(\cdot, \cdot) \) have the same invariant density \( p(\cdot) \).

(c) Suppose that \( T(\cdot, \cdot) \) satisfies the DBC with \( p(\cdot) \) as the invariant density. Show that \( T_\lambda(\cdot, \cdot) \) also satisfies the DBC with \( p(\cdot) \) as the invariant density.

(d) Let \( T^{(l)}_\lambda(\theta', \theta) = \int T^{(l-1)}_\lambda(\theta', \tilde{\theta}) T_\lambda(\tilde{\theta}, \theta) d\mu(\tilde{\theta}) \) for \( l = 1, 2, \ldots \) where \( T^{(0)}_\lambda(\cdot) \equiv T_0(\cdot) \). Show that

\[
T^{(l)}_\lambda(\theta', \theta) = e^{-\lambda l} \sum_{k=0}^{\infty} (\lambda l)^k T_k(\theta', \theta)/k!
\]

(e) Let \( p_l(\theta) = \int T_l(\theta', \theta) p_0(\theta') d\mu(\theta') \) for \( l = 1, 2, \ldots \). Suppose that \( ||p_l - p||_1 \equiv \int |p_l(\theta) - p(\theta)| d\mu(\theta) \leq M \rho^l \) for some \( M > 0 \) and \( \rho \in (0, 1) \). Let \( p^\lambda_l(\theta) = \int T^{(l)}_\lambda(\theta', \theta) p_0(\theta') d\mu(\theta') \). Show that

\[
||p^\lambda_l - p||_1 \leq M e^{-\lambda(1-\rho)} \quad \text{for } l = 1, 2, \ldots
\]

Hence, conclude that the Markov Chain with TKD \( T_\lambda(\cdot, \cdot) \) converges at a faster rate than that of \( T(\cdot, \cdot) \) if \( \lambda > -\frac{1}{1-\rho} \ln \rho \). In particular, justify that one should always choose \( \lambda \geq 1 \).

[5+5+5+10+10=35 points]
3. Let $Z_i \overset{iid}{\sim} f(z|\theta) \equiv \sum_{k=1}^{m} \theta_k f_k(z)$ for $i = 1, 2, \ldots, n$ where $f_k(\cdot)$’s are known density functions (wrt a $\sigma$-finite measure $\mu$) with common support $Z \subseteq \mathbb{R}^d$ and $\theta = (\theta_1, \ldots, \theta_m) \in S_m = \{(t_1, \ldots, t_m) \in [0, 1]^m : \sum_{k=1}^{m} t_k = 1\}$.

(a) Show any improper prior on $\theta$ leads to an improper posterior if $f_k(z) > 0$ for all $k \in \{1, \ldots, m\}$ and $z \in Z$.

(b) Suppose $\theta \sim \text{Dir}(a_1, \ldots, a_m)$ where $a_k > 0 \ \forall \ k$. Consider (unobserved) independent random variables $K_i$’s such $\Pr[K_i = k|\theta] = \theta_k$ for $k = 1, \ldots, m$ and $i = 1, \ldots, n$. Starting with $\theta^{(0)} = \frac{1}{m}(1, 1, \ldots, 1)'$, consider the Gibbs sampler, for $l = 1, 2, \ldots$

$$K^{(l)}_i \sim \frac{\theta_k^{(l-1)} f_k(Z_i)}{\sum_{k=1}^{m} \theta_k^{(l-1)} f_k(Z_i)} I(K_i = k) \text{ for } i = 1, \ldots, n$$

$$K^{(l)} = (K^{(l)}_1, \ldots, K^{(l)}_n)$$

$$\theta^{(l)} \sim \text{Dir}(n_1(K^{(l)}) + a_1, \ldots, n_m(K^{(l)}) + a_m)$$

where $n_k(K^{(l)}) = \sum_{i=1}^{n} I(K^{(l)}_i = k)$ for $k = 1, \ldots, m$. Show that the Markov chain $\{\theta^{(l)}, l = 0, 1, 2, \ldots\}$ satisfies the DBC with the posterior density, $p(\theta|Z)$ of $\theta$ (given the observed data $Z = (Z_1, \ldots, Z_n)$) as the invariant density.

(c) [Optional bonus problem] Show that Markov chain $\{\theta^{(l)}, l = 0, 1, 2, \ldots\}$ is geometrically ergodic if $f_k(z) > 0$ for all $k \in \{1, \ldots, m\}$ and $z \in Z$. 

$$[5+5+(10)=10+(10)]$$
4. Consider again the finite mixture model described in the previous problem with \( Z_i = (X_i, Y_i) \) for \( i = 1, \ldots, n \) and let the component densities for \( z = (x, y) \) be given by

\[
f_k(z) = \frac{1}{\lambda_1} \psi_1 \left( \frac{x - x_k}{\lambda_1} \right) \frac{1}{\lambda_2} \psi_2 \left( \frac{y - y_k}{\lambda_2} \right)
\]

for \( k = 1, \ldots, m \)

where \( \lambda_j > 0 \) are the bandwidth and \( \psi_j(\cdot) \)'s are kernel functions satisfying \( \int \psi_j(u)du = 1 \) and \( \int u \psi_j(u)du = 0 \) for \( j = 1, 2 \). The knot points \( z_k = (x_k, y_k) \) for \( k = 1, \ldots, m \) are to be chosen suitably from the support \( Z \subseteq \mathbb{R}^2 \).

(a) Show that the conditional mean function of \( Y \) given \( X = x \) (i.e., the regression function) is given by

\[
\mu(x, \theta) = E[Y|X = x, \theta] = \frac{\sum_{k=1}^m \theta_k y_k \psi_1 \left( \frac{x - x_k}{\lambda_1} \right)}{\sum_{k=1}^m \theta_k \psi_1 \left( \frac{x - x_k}{\lambda_1} \right)}
\]

(b) Derive an expression for \( E[Y^2|X = x, \theta] \) and hence the conditional variance function, \( \sigma^2(x, \theta) = Var[Y|X = x, \theta] \).

(c) Show that the (pointwise) Bayes estimator of \( \mu(x, \theta) \) with respect to the squared error loss: \( (\mu(x, \theta) - \mu(x))^2 \) is given by

\[
\hat{\mu}(x) = E[\mu(x, \theta)|Z_1, \ldots, Z_n]
\]

(d) [Optional bonus problem] Show that the (pointwise) Bayes estimator of \( \sigma(x, \theta) \) with respect to the relative error loss: \( \left( \frac{\sigma(x)}{\sigma(x, \theta)} - 1 \right)^2 \) is given by

\[
\hat{\sigma}(x) = \frac{E[\sigma^{-1}(x, \theta)|Z_1, \ldots, Z_n]}{E[\sigma^{-2}(x, \theta)|Z_1, \ldots, Z_n]}
\]

\([5+5+5+(5)=15+(5)]\)
5. Consider again the mixture model described in the previous two problems along the data set mixreg.txt posted online. Notice that the sample size is \( n = 100 \) and \( d = 2 \). Fix \( m = \lceil \sqrt{n} \rceil \), \( \lambda_1 = sd(X)/n^{1/5} \) and \( \lambda_2 = sd(Y)/n^{1/5} \) where \( sd(X) \) and \( sd(Y) \) denote the sample standard deviations of \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_n)\), respectively.

Choose the knot points \( x_k \)'s as the \((k/(m+1))-th \) quantiles of \( X_1, \ldots, X_n \) and the knot points \( y_k \)'s as the values of \( Y_1, \ldots, Y_n \) for \( k = 1, \ldots, m \) that correspond to the ordered values of quantiles of the \( X \)'s. Let \( \psi_j(u) = \phi(u) = e^{-u^2/2}/\sqrt{2\pi} \) be the standard normal density for \( j = 1, 2 \).

(a) Using the data set mixreg.txt implement the Gibbs sampler described in part 3(b) with \( a_k = 1 \ \forall \ k \) and display the \((m)\) trace and autocorrelation function (ACF) plots of the Markov Chain samples of \( \theta \).

(b) Consider the resolvant approach of problem 2. Choose an appropriate \( \lambda > 0 \) and implement the Markov chain sampler using the resolvant (as described in problem 2) with the Gibbs sampler of previous part as the TKD \( T(\cdot, \cdot) \). Compare the ACF plots of resolvant based sampler with that of the Gibbs sampler in the previous part.

(c) Using the samples generated by the two Markov chains (in parts (a) and (b) above) compute the Bayes estimator (as given in part 4(c)) and plot the pointwise Bayes estimator \( \hat{\mu}(x) \) vs. \( x \) on a grid of points within the observed range of the x-values.

\[ 10+10+10=30 \]