Stats 210A, Fall 2017
Homework 5

Due date: Thursday, Oct. 5

You may disregard measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, “all functions” vs. “all measurable functions,” etc. (unless the problem is explicitly asking about such issues).

1. Admissibility and Bayes estimators

One of the frequentist motivations for Bayes estimators is their connection to admissibility.

(a) Suppose that the Bayes estimator $δ_Λ$ for the prior $Λ$ is unique up to $P$-almost-sure equality. That is, for any other Bayes estimator $\tilde{δ}_Λ$, we have $δ_Λ(X) = \tilde{δ}_Λ(X)$ almost surely, for every $θ \in Ω$. Show that $δ_Λ$ is admissible.

(b) A randomized estimator is an estimator that is a random function of the data. We can formalize it generically as $δ(X,W)$ where $X \sim P_θ$ as usual and $W$ is some auxiliary random variable generated by the analyst. For this part, “admissible” and “Bayes” are defined with respect to all estimators including randomized ones.

Now consider a model with a finite parameter space, $|Ω| = n < ∞$ and assume the loss function is non-negative ($L : Ω \times R → [0, ∞)$). Show that every admissible estimator is a (possibly randomized) Bayes estimator for some prior.

Hint: consider the set $A$ of all achievable risk functions, and the set $D_δ$ of all (possibly unachievable) risk functions that would dominate a given estimator $δ$. Recall the hyperplane separation theorem: for any two disjoint non-empty convex subsets $A,B \subseteq R^n$ there exist $v \in R^n$ and $c \in R$ such that $v^′a ≤ c ≤ v^′b$ for all $a \in A,b \in B$.

2. Stationary distribution for MCMC algorithms

This problem considers MCMC sampling from a generic posterior density $λ(θ | x)$ where $θ ∈ R^d$.

The Metropolis–Hastings algorithm is a Markov chain using the following update rule: First, sample $Z \sim f(· | Θ^{(t)})$ according to some “proposal distribution” $f(ζ | θ) : Ω × Ω → (0, ∞)$, where $f(· | θ)$ is a probability density for each $θ$ (assume $λ$ and $f(· | θ)$ are densities w.r.t. the same dominating measure). Next, compute the “accept probability” as

$A = \min \{ 1, \frac{λ(Z | X)f(Θ^{(t)} | Z)}{λ(Θ^{(t)} | X)f(Z | Θ^{(t)})} \}.$

Finally, let $Θ^{(t+1)} = Z$ with probability $A$ and $Θ^{(t+1)} = Θ^{(t)}$ with probability $1 − A$. Show that $λ(θ | X)$ is stationary for the Metropolis–Hastings algorithm.

Note: this algorithm is computationally attractive because we can can always implement it using only the unnormalized posterior $p_θ(X)λ(θ)$ (or any function $g(θ)$ that is proportional to it), which is often much easier to compute than the normalized posterior.

3. Empirical Bayes for exponential families

Consider an $s$-parameter exponential family model in canonical form:

$p_θ(x) = e^{θ^′T(x) − A(θ)h(x)}$

where $x = (x_1, \ldots, x_n)$ and the random vector $Θ$ has prior density $λ_γ(θ)$. Here $γ ∈ Γ$, an open subset of $R$. Let $λ_γ(θ | x)$ and $q_γ(x)$ denote the posterior and marginal, respectively, for a given choice of $γ$.

Assume throughout this problem that all relevant quantities are suitably differentiable and/or integrable, and derivatives can always be taken inside the integral sign.
(a) Show that for $i = 1, \ldots, n$, we have

$$E_{\gamma} \left[ \sum_{j=1}^{n} \Theta_j \frac{\partial T_j(x)}{\partial x_i} \mid X = x \right] = \frac{\partial}{\partial x_i} \log q_{\gamma}(x) - \frac{\partial}{\partial x_i} \log h(x),$$

(b) Now assume $n = s$ with $T(x) = x$:

$$p_\theta(x) = e^{\theta' x - A(\theta)} h(x).$$

Let $\hat{\gamma}(X)$ denote the maximum likelihood estimator (MLE) of $\gamma$ based on the observed data:

$$\hat{\gamma}(X) = \arg\max_{\gamma \in \Gamma} q_{\gamma}(X),$$

which we assume always exists.

Show that the empirical posterior mean of $\Theta$, using $\hat{\gamma}$ to estimate $\gamma$, is

$$E_{\hat{\gamma}} [\Theta \mid X = x] = \nabla (\log q_{\hat{\gamma}(x)}(x) - \log h(x)).$$

4. **Gamma-Poisson empirical Bayes model**

Consider the empirical Bayes model with

$$\theta_i \sim \text{Gamma}(k, \beta)$$

$$X_i \mid \theta_i \sim \text{Pois}(\theta_i),$$

independently for $i = 1, \ldots, n$, and assume $k$ (shape parameter) is known and $\beta$ (rate parameter, inverse of the scale parameter) is unknown and estimated via the MLE. Show that the empirical Bayes posterior mean for $\theta_i$ is

$$\frac{\bar{X}}{\bar{X} + k}(k + X_i), \text{ where } \bar{X} = n^{-1} \sum_{i} X_i.$$

You may use without proof the fact that the marginal distribution of $X_i$ is negative binomial.

5. **Gibbs Sampler for Gamma-Poisson model**

Consider the hierarchical Bayes model we have discussed in class, where

$$\beta \sim \text{Gamma}(m, \zeta)$$

$$\theta_i \mid \beta \overset{\text{i.i.d.}}{\sim} \text{Gamma}(k, \beta), \ i = 1, \ldots, n$$

$$X_i \mid \beta, \theta \overset{\text{ind.}}{\sim} \text{Pois}(\theta_i), \ i = 1, \ldots, n,$$

where $\beta$ and $\zeta$ are rate parameters, and $m, \zeta$, and $k$ are fixed and known.

Give an explicit algorithm for one full iteration of the Gibbs sampler.