STA 214 Assignment 5 Spring 1997

Readings: ...

We're getting into mixtures of Dirichlet processes – fancy name for a "simple" class of mixtures models (mixtures of normals and whatever else you want to mix.) Check the appendix of the paper by Turner and West in the coursepack for some easy reading on this. The paper by West and Turner there delivers a full-blown analysis of some neurological mixture data. We won't dwell on (or even get exposed to) the real theory of Dirichlet processes, but we will learn a lot about some aspects of these models and their uses in non-parametric density estimation (and elsewhere).

Please hand in solutions to the following at the start of class on Thursday 27th February **(as usual, get started now!)**

Exercise 1
Mixtures of exponentials.

Times to remission of individuals under treatment for a clinical condition are exponentially distributed (the "canonical" model for survival data!). A study group has $n = 65$ cases with observed remission times $x_1, \ldots, x_n$. Of this group, cases $i = 1, \ldots, 5$ are known to come from an "at risk" group with exponential mean μ_1, and cases $i = 6, \ldots, 15$ come from a "normal" group with exponential mean μ_0, and where $\mu_0 < \mu_1$. The other 50 are unclassified, so come from the mixture

$$(1-\pi)Exp(x_i|1/\mu_0) + \pi Exp(x_i|1/\mu_1)$$

where π is the population proportion of "at risk" cases. *(n.b., The Gelman book notation for exponentials is Expon, not Exp as here.)*

Starting with the (improper!) prior $p(\mu_0, \mu_1) = constant$ on $0 < \mu_0 < \mu_1$, and by introducing latent indicators z_i that classify the cases as at risk or normal, develop the collection of conditional posterior distributions for each of μ_0, μ_1, π and each of the z_i in turn, in each case conditioning on the data and all other parameters and latent variables.

Write S-Plus MCMC code to iteratively simulate the posteriors, and analyse the data in the file linked to the web page, titled "Remission times data". Produce histograms of posteriors for π, μ_0 and μ_1, and plots supporting the convergence of the MCMC iterations.

Comment on any difficulties encountered, and similarities or differences between this exponential mixture problem and your earlier experiments with normal mixtures. In particular, comment on the difficulties of discriminating between exponential distributions for the unclassified data, supporting your comments with any graphs you feel relevant.

Exercise 2

A change-point problem.

Data arising from a nuclear physics experiment follow a Poisson process of constant mean per unit time, so that observations x_t recorded in each time interval $t = 1, 2, \ldots,$ are independent Poisson. The experiment is designed to detect and infer a change-point in the Poisson rate, which occurs at an unknown time s. T independent observations are made over times $t = 1, \ldots, T$. Observations before the change-point are $Po(x_t|\mu)$, and those after the change-point are $Po(x_t|\theta)$, for some means μ and θ. Assume that a change-point does indeed occur in the interval $1, \ldots, T$, and that the experimenter assumes the discrete uniform prior on the change-point; i.e., s has prior density $p(s) = 1/T$ over $s = 1, \ldots, T$.

(a) Show that, in the absence of knowledge of the change-point s, each x_t follows a mixture of two Poisson distributions, and write down the form of this mixture.

(b) What is the joint density of the data given s, μ and θ?

(c) Assume prior independence among μ, θ and s, and suppose priors $\mu \sim Exp(\mu|10)$ and $\theta \sim Exp(\theta|50)$. By integrating $p(x|s, \mu, \theta)$ with respect to $p(\mu, \theta)$, find $p(s|x)$.

(c) How would you suggest that we simulate the posterior $p(\mu, \theta, s|x)$? *(just suggest, don't do it!)*.
