STA 214 Assignment 3 Spring 1997

Readings: ...

Titterington et al handout, and papers on normal mixtures by C Robert for more background (also more of my "neuro" mixture paper) on normal mixtures. Next week is multivariate distributions and mixtures: paper by M Lavine & M West in the course-pack.

Please hand in solutions to the following at the start of class on Thursday 6th February.

Exercise 1

(A mixture of Poissons – a version of the mixture of normals from Assignments 1 and 2)

A series of counts of neutrinos is recorded at a detector used in astronomical observation. Over the area of sky monitored, background neutrino radiation is known to occur as a Poisson process with mean b particles per second, and b is known. So the number x_t of particles detected in t seconds is Poisson with mean bt, and so density $(bt)^{x_t}\exp(-bt)/x_t!$ over $x_t = 0, 1, \ldots$. Write $x_t \sim Po(x_t|bt)$ (cf. Gelman book: uses $Poisson()$).

A supernova event detectable on earth would increase neutrino flux. So observations would then follow a Poisson model with mean $(b+\mu)t$ for some $\mu > 0$. Astronomers are "looking" for increased fluxes as evidence for such supernova events. On the current experiment, our astronomer has a prior probability π that her recordings are influenced by such an event; if they are, she would have increased mean counts. So she models the data with the mixture

$$x_t \sim (1-\pi)Po(x_t|bt) + \pi Po(x_t|(b+\mu)t)$$

which says that either she measures just the background radiation, with chance $1-\pi$, or her data are affected by such an event, with chance π.

She also has an exponential prior for μ with specified mean m. (NOTE: of course, an exponential distribution is a special case of the gamma family: the exponential with mean m is the $Ga(\cdot|1, 1/m)$ distribution).

Given the observed number of counts x_t show that the posterior density is

$$p(\mu|x_t) = k\{(1-\pi)\exp(-\mu/m) + \pi(1+\mu/b)^{x_t}\exp(-\mu(t+1/m))\}$$

for some constant k. The astronomer fixes the timing interval so that b is small (and known) and hence interest lies in small counts x_t.

(a) If $x_t = 0$ so that no counts are recorded, show that the posterior is a mixture of two specific exponential distributions. Identify the two exponentials being mixed, and interpret the mixing probabilities.

(b) What is the posterior probability that data comes solely from background noise given that she observes $x_t = 0$?

(c) *This extra item is optional, NOT assessed: For your spare time ...*
Show that, whatever the value of x_t observed, the posterior is a mixture of gamma distributions.

Exercise 2

Program the fitting of a 2−component normal mixture with common unknown variance and unknown means and probabilities,

$$\pi N(\cdot|\mu_0, \sigma^2) + (1-\pi)N(\cdot|\mu_1, \sigma^2)$$

with all parameters $\theta = \{\pi, \mu_0, \mu_1, \sigma^2\}$ now unknown. Do this using Gibbs sampling with component classification indicators $z_i = 0$ or 1 according to whether y_i comes from $N(\cdot|\mu_0, \sigma^2)$ or $N(\cdot|\mu_1, \sigma^2)$ respectively. Rebuild the S-Plus I used for the special case $\mu_0 = 0$ and fixed σ. Use independent uniform priors of μ_0, μ_1 and σ. Reanalyse simulated data from that example and produce plots of posterior histograms etc for elements of θ. Produce a histogram representing the approximate predictive density for a single future observation from the mixture. Hand in your S-Plus code as well as (a small number of) summary graphs.

Exercise 3

Kernel density estimation. A standard technique for smoothing histograms of data is to estimate the underlying density by an equally-weighted mixture of normals (or other pdf) located at the observations. Given data $x_1, \ldots, x_n$ such an estimate has the form, as a function of the continuous y, given by

$$\hat{f}(y) = n^{-1} \sum_{i=1}^{n} p((y - x_i)/h)/h$$

where $p(\cdot)$ is the chosen kernel (density function) and h a specified scale factor (sometime called a *smoothing parameter*, or the *window-width* parameter.) For example, in normal kernel density estimation, $p(\cdot)$ is the standard normal density. It could be any other density symmetric about zero, such as a uniform or one of several others. Check the `density()` function in S-Plus. Use some of the 2-component normal mixture data generated in Assignment 2 to investigate the shapes of `density()` fits to the data for various choices of the smoothing/window-width parameter. Draw a few graphs, and summary conclusions about the effects of smoothing parameter values on your subjective impressions as to what the underlying density might really look like.

(*Later we'll see how kernel density estimates arise "naturally" in Bayesian models for uncertain density functions).*
