Statistics 221 – Assignment 4

Due: Wednesday, May 19, 2004

1. Consider the joint distribution of X and Y,

$$f(x,y) \propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1};\ x = 0,1,\ldots,n, 0 \le y \le 1. \tag{1}$$

   Suppose we are interested in characteristics of the marginal distribution $f(x)$ of X.

   (a) Derive a Gibbs sampling algorithm. The generate a sample of size $m = 1000$ with 100 burn-in scans. The parameter values are taken to be $n = 16, \alpha = 2$, and $\beta = 4$.

   (b) In fact, Gibbs sampling is not needed here, since $f(x)$ can be obtained analytically. Find $f(x)$ and generate a sample (independent) of size $m = 1000$. The parameter values are to be specified as in part (a).

   (c) Compare the histograms (in a single plot) of the two samples obtained in (a) and (b) and comment on features of the plot.

   (d) The probability function of X can be estimated by

$$\hat{P}[X = x] = \frac{1}{m}\sum_{i=1}^{m} P[X = x|Y_i = y_i]$$

   Compute the estimated probabilities from the sample generated in (a). Plot these estimated probabilities and compare them (in the same plot) to the exact probabilities.

   (e) In the distribution specified in (1), now let n be the realization of a Poisson random variable with mean λ. Repeat (a) for $\lambda = 16, \alpha = 2$, and $\beta = 4$. Then estimate $P[X = x]$ as in (d).

2. Consider the following hidden Markov model

$$\begin{aligned} Z_1 &\sim Bern(0.5) \\ Z_k|Z_{k-1} &\sim Z_{k-1} + Bern(p) \bmod 2;\ k = 2,\ldots,K \\ X_k|Z_k &\sim N(\mu_1 Z_k + \mu_0(1-Z_k), \sigma^2);\ k = 2,\ldots,K \end{aligned} \tag{2}$$

   where $Z = \{Z_1,\ldots,Z_K\}$ is the vector of unobserved states and $X = \{X_1,\ldots,X_K\}$ is the vector of observed data. Note that the second line of this model corresponds to $P[Z_k = Z_{k-1}] = p$.

   Suppose we are interested on inference on p and Z conditional on a given p, assuming that μ_1, μ_0, and σ^2 are known. This can be done using a Sequential Importance Sampler (SIS) for generating Z given the observed data X.

   (a) Write as SIS sampler to generate Z given X for this model allowing for an arbitrary m, the number of imputations, p_{sim}, the value of p used for imputation, K, μ_1, μ_0, and σ^2.

(b) Use the code written in part (a) to analyze the data in the file hmm.txt on the Assignments page of the course web site to find the maximum likelihood estimate $\hat{p}$ of p, the regime switching parameter, with $m = 2000$ imputations and $p_{sim} = 0.5$. This data set was generated with $K = 20, \mu_1 = 1, \mu_0 = -1$, and $\sigma^2 = 1$.

(c) Find $E_p[Z_k|X]$ for $k = 1, 2, 3, 4, 5, 10, 15, 20$ when $p = 0, 0.25, 0.5$, and $\hat{p}$. Also give the standard error for each estimator.

(d) The estimates for one choice of p in the previous part probably act weird. Give a probable explanation for this.