Computational Statistics
8. exercise session, 16.4.2010

1. Consider *simple linear regression*, where

$$[Y_i \mid \alpha, \beta] \overset{\text{ind}}{\sim} N(\alpha + \beta x_i, \sigma^2), \qquad i = 1, \ldots, n.$$

The covariates $x_1, \ldots, x_n$ as well as the error variance $\sigma^2 > 0$ are assumed to be known quantities. The prior for $\theta = (\alpha, \beta)$ is the bivariate normal distribution $N(\mu_0, Q_0^{-1})$. (The posterior distribution can be obtained from problem 5 of session 5. The design matrix X consists of a column of ones and a column consisting of the covariate values.)

a) Show that the components of $\theta = (\alpha, \beta)$ are usually dependent in their joint posterior, even if they are independent in the prior.

b) If we reparametrize by centering the covariates so that

$$[Y_i \mid \phi, \beta] \overset{\text{ind}}{\sim} N(\phi + \beta t_i, \sigma^2), \qquad i = 1, \ldots, n,$$

where the centered covariates are

$$t_i = x_i - \bar{x} = x_i - \frac{1}{n} \sum_{j=1}^{n} x_j$$

and $\phi = \alpha + \beta \bar{x}$, and take ϕ and β to be independent in their joint prior,

$$p(\phi, \beta) = N(\phi \mid \mu_\phi, \sigma_\phi^2)\, N(\beta \mid \mu_\beta, \sigma_\beta^2),$$

then ϕ and β are independent in their joint posterior.

(Hint: two random variables with a joint bivariate normal distribution are independent if and only if the covariance matrix of the distribution is diagonal, and this is the case if and only if the precision matrix of the distribution is diagonal.)

2. Consider simple linear regression with the following data.

x_i	11	12	13	14	15
y_i	1.7	1.3	2.3	3.7	4.0

The error variance $\sigma^2 = 1$ is assumed known. Derive the Gibbs sampler for the non-centered (α, β) parametriztion defined in the previous problem. Take α and β to have independent normal distributions $N(0, 100)$ in the prior.

Run the Gibbs sampler for 2000 steps, discard the burn-in of 1000 first iterations, and produce trace plots and autocorrelation plots for the two parameters. (In simple linear regression, it would be better to use the centered parametrization and independent normal priors, since then the Gibbs sampler would produce i.i.d. samples from the posterior of (ϕ, β).)

3. If the prior is the logistic distribution and the likelihood corresponds to n observations from $N(\theta, 1)$, then the posterior is proportional to

$$
\begin{aligned}
p(\theta \mid y) &\propto q_1(\theta)\, q_2(\theta), \quad \text{where} \\
q_1(\theta) &= \frac{\exp(\theta)}{(1+\exp(\theta))^2}, \qquad q_2(\theta) = \exp\left(-\frac{1}{2} n(\theta - \bar{y})^2\right)
\end{aligned}
$$

This is equal to the marginal distribution of θ in the augmented model where

$$
p(\theta, z_1, z_2) \propto 1(0 < z_1 < q_1(\theta))\, 1(0 < z_2 < q_2(\theta))
$$

Derive formulas for the slice sampler (i.e., the Gibbs sampler in the augmented model). Try it when $n = 5$ and $\bar{y} = 1$. (Hints: the sets $\{\theta : q_i(\theta) > z_i\}$, $i = 1, 2$ turn out to be intervals; the intersection of two overlapping intervals, $(a_1, b_1) \cap (a_2, b_2) = (a, b)$, where $a = \max(a_1, a_2)$ and $b = \min(b_1, b_2)$).

4. (A changepoint model) Consider the following statistical model,

$$
\begin{aligned}
[Y_i \mid \mu_1, \mu_2, k] &\overset{\text{i.i.d.}}{\sim} \mathrm{Poi}(\mu_1), \qquad i = 1, \ldots, k \\
[Y_i \mid \mu_1, \mu_2, k] &\overset{\text{i.i.d.}}{\sim} \mathrm{Poi}(\mu_2), \qquad i = k+1, \ldots, n.
\end{aligned}
$$

The first k observations come from a Poisson distribution with mean μ_1 and the rest from a Poisson distribution with mean μ_2. The parameters μ_1, μ_2 and the changepoint k are unknown.

In the prior, we take the parameters to be independent with the following marginal distributions,

$$
\begin{aligned}
p(\mu_1) &= \mathrm{Gam}(\mu_1 \mid a_1, b_1), \qquad p(\mu_2) = \mathrm{Gam}(\mu_2 \mid a_2, b_2) \\
p(k) &= \frac{1}{n-1}, \qquad k = 1, 2, \ldots, n-1.
\end{aligned}
$$

The prior of the changepoint is the discrete uniform on the indicated set. Here a_1, b_1, a_2 and b_2 are known hyperparameters.

In order to derive the Gibbs sampler, calculate the posterior full conditional distributions of all the parameters μ_1, μ_2 and k.

Hint: $p(\mu_1 \mid \cdot)$ and $p(\mu_2 \mid \cdot)$ are standard distributions, but $p(k \mid \cdot)$ is not. However, the full conditional of k is a discrete distribution on a finite set, and therefore you only need to know how to calculate its pmf in order to be able to simulate it.

5. Consider the EM algorithm for the genetic linkage example, where we use the auxiliary variable defined in Example 8.2. (Because the prior is uniform, the MAP and the ML estimates are the same.) Calculate a formula for $Q(\theta \mid \theta_0)$, and obtain a formula for the point θ_1 which maximizes it with respect to θ.