Nonlinear Gaussian regression

Let us consider the context of Example 6.1 (pages 192-194) of Gamerman and Lopes (2006), where the response variable $y$, is the velocity of an enzymatic reaction (in counts/min/min) and the regressor, $x$, is substrate concentration (in ppm). Check the book webpage at [http://www.dme.ufrj.br/mcmc/chapter6.html](http://www.dme.ufrj.br/mcmc/chapter6.html). Here is the data and a scatter plot showing the nonlinear relationship between $y$ and $x$:

$x = c(0.02,0.02,0.06,0.06,0.11,0.11,0.22,0.22,0.56,0.56,1.10,1.10)$

$y = c(76,47,97,107,123,139,159,152,191,201,207,200)$

plot(x,y)

**Model.** We would like to entertain a Gaussian nonlinear model, for $i = 1, \ldots, n$ ($(n = 12)$),

$$y_i | x_i, \beta, \gamma, \sigma^2 \sim N(\beta_0 + \beta_1 g(x_i, \gamma), \sigma^2),$$

where $\beta = (\beta_0, \beta_1)'$ and $g(x_i, \gamma) = x_i/(\gamma + x_i)$, for $\beta \in \mathbb{R}^2$, $\gamma \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$. 

**Prior.** Let us consider the prior for $\beta$ and $\sigma^2$ as follows:

$$p(\beta, \gamma, \sigma^2) = p(\beta|\sigma^2)p(\gamma)p(\sigma^2)$$

$$\beta|\sigma^2 \sim N(b_0, \sigma^2 B_0), \quad b_0 = (50,170)', \quad B_0 = 3I_2$$

$$\gamma \sim N(g_0, \gamma_0^2), \quad g_0 = 0, \gamma_0^2 = 1$$

$$\sigma^2 \sim IG(\nu_0/2, \nu_0 \sigma_0^2/2), \quad \nu_0 = 5, \sigma_0^2 = 10.$$ 

Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Answer the following questions.

(a) Show that $p(\beta|x,y,\sigma^2,\gamma)$ is a Gaussian distribution.

(b) Show that $p(\sigma^2|x,y,\beta,\gamma)$ is an Inverse-Gamma distribution.

(c) Show that $p(\sigma^2|x,y,\gamma)$ is also an Inverse-Gamma distribution.

*Note 1:* (c) is possible because we made the prior of $\beta$ conditional on $\sigma^2$, $p(\beta|\sigma^2)$, which results in $p(\sigma^2|x,y,\gamma)$ also being an Inverse-Gamma distribution. What is the catch? Well, the catch is that multiplying (a) and (c) is exactly $p(\beta,\sigma^2|x,y,\gamma)$, while iterating between (a) and (b) is approximately $p(\beta,\sigma^2|x,y,\gamma)$, which is the standard Gibbs (or, more generally, MCMC) theorem.
Note 2: It is not hard to see that \((\gamma|x, y, \beta, \sigma^2)\) comes from a distribution of no known form, since \(\gamma\) from the prior does not conjugate with \(\gamma\) from the likelihood function. Nonetheless, it can be point-wise evaluated up to a normalizing constant as

\[
p(\gamma|x, y, \beta, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left( \frac{\gamma^2}{2} - 2\gamma m_0 \right) \right\} \times \exp \left\{ -\frac{1}{2\sigma^2} \left( \beta_1^2 \sum_{i=1}^n g(x_i, \gamma) - 2 \sum_{i=1}^n (y_i - \beta_0)g(x_i, \gamma) \right) \right\}.
\]

(d) Algorithm 1: Implement an MCMC algorithm that cycles through the full conditionals:

(d1) Gibbs step: Sample \(\sigma^2\) from \(p(\sigma^2|x, y, \beta, \gamma)\);
(d2) Gibbs step: Sample \(\beta\) from \(p(\beta|x, \sigma^2, \gamma)\);
(d3) Metropolis-Hastings step: Sample \(\gamma\) from \(p(\gamma|x, y, \beta, \sigma^2)\).

(e) Algorithm 2: Replace step (d1) from Algorithm 1 by

(e1) Gibbs step: Sample \(\sigma^2\) from \(p(\sigma^2|x, y, \gamma)\).

Conceptually, what is the difference between the above MCMC algorithms?

(f) Compare the algorithms in terms of MCMC mixing, sample autocorrelation functions and effective sample sizes, as well as by comparing the approximate marginal posterior distributions for \(\beta_0, \beta_1, \gamma\) and \(\sigma^2\).

(g) Finally, on the top of the scatterplot of \(x\) against \(y\), add the posterior predictive curve. More precisely, for a grid of new values of \(x\), say \(x_{n+1}\), in \(\{0.02, 0.03, \ldots, 1.10\}\), a 109-point grid, draw the 2.5th, 50th and 97.5th percentiles the posterior predictive densities

\[
p(y_{n+1}|x_{n+1}, x, y) = \int p(y_{n+1}|x_{n+1}, \beta, \gamma, \sigma^2)p(\beta, \gamma, \sigma^2|x, y)d\beta d\gamma d\sigma^2.
\]

Recall that, by Monte Carlo integration,

\[
\hat{p}_{mc}(y_{n+1}|x_{n+1}, x, y) = \frac{1}{M} \sum_{i=1}^M p(y_{n+1}|x_{n+1}, \beta^{(i)}, \gamma^{(i)}, \sigma^{2(i)}),
\]

where \(\{\beta^{(i)}, \gamma^{2(i)}\}_{i=1}^M\) are draws from the posterior \(p(\beta, \gamma, \sigma^2|x, y)\), which could be obtained from algorithm 1 or 2 above. In fact, draws \(\{(\beta^{(i)}, \gamma^{2(i)})\}_{i=1}^M\) can also be used to generate draws \(\{y_{n+1}^{(i)}\}_{i=1}^M\) from \(p(y_{n+1}|x_{n+1}, x, y)\) by sampling \(y_{n+1}^{(i)}\) from \(p(y|x_{n+1}, (\beta^{(i)}, \gamma^{(i)}, \sigma^{2(i)})\), for \(i = 1, \ldots, M\).