Third homework assignment - SOLUTION

PhD in Business Economics — Advanced Bayesian Econometrics

Hedibert Freitas Lopes — Due at 9am, February 18th, 2021.

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`. 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,ylim=range(y,q1,q2),pch=16,xlab="Substrate concentration (ppm)",
     ylab="Velocity of enzymatic reaction (counts/min/min)",col=2)
```

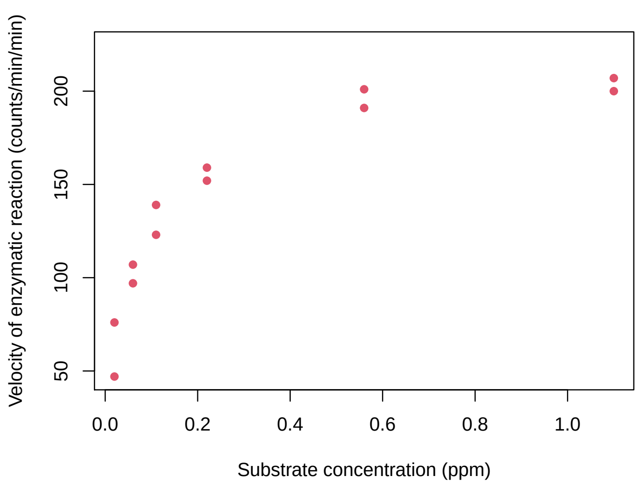

Figura 1: Observe the nonlinear nature of the relationship between y and x.

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 $\sigma^2 \in \mathbb{R}^+$. The first thing to realize here is that, conditional γ the above model is a Gaussian linear regression of y on $g(x, \gamma)$. That being said, conditional on γ, posterior inference for β and σ^2 might be straightforward should one use a conjugate Normal-Inverse Gamma prior on (β, σ^2). In other words, the bottleneck is the parameter γ, since it appears in a nonlinear fashion when linking y to x. This is the source of the nonlinearity, not the fact that x appears as $x/(\gamma + x)$.

Prior. Let us consider the prior for (β, σ^2) and γ as follows:

$$\begin{aligned} p(\beta,\gamma,\sigma^2) &= p(\beta|\sigma^2)p(\gamma)p(\sigma^2) && \\ \beta|\sigma^2 &\sim N(b_0, \sigma^2 B_0), && b_0 = (50, 170)', B_0 = 3I_2 \\ \gamma &\sim N(g_0, \tau_0^2), && g_0 = 0, \tau_0^2 = 1 \\ \sigma^2 &\sim IG(\nu_0/2, \nu_0\sigma_0^2/2), && \nu_0 = 5, \sigma_0^2 = 10. \end{aligned}$$

The prior of (β, σ^2) is Normal-Inverse Gamma, as pointed out above. This might be useful when designing an MCMC scheme that either cycles through $p(\beta, \sigma^2|x, y, \gamma)$ and $p(\gamma|x, y, \beta, \sigma^2)$ or cycles through $p(\beta|x, y, \sigma^2, \gamma)$, $p(\sigma^2|x, y, \beta, \gamma)$ and $p(\gamma|x, y, \beta, \sigma^2)$. It is worth noticing cycling through $p(\beta|x, y, \sigma^2, \gamma)$ and $p(\sigma^2|x, y, \beta, \gamma)$ will produce draws from $p(\beta, \sigma^2|x, y, \gamma)$, BUT only in the limit and as a result of a Markov chain argument. The lesson here is to always derive your posterior distributions analytically and resort to MC and MCMC schemes only when necessary.

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 β conditional on σ^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 γ from the prior does not conjugate with γ from the likelihood function. Nonetheless, it can be point-wise evaluated up to a normalizing constant as

$$\begin{aligned} p(\gamma|x, y, \beta, \sigma^2) &\propto \exp\left\{-\frac{1}{2\tau_0^2}(\gamma^2 - 2\gamma m_0)\right\} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}\left[\beta_1^2\sum_{i=1}^n g^2(x_i,\gamma) - 2\sum_{i=1}^n (y_i - \beta_0)g(x_i,\gamma)\right]\right\}. \end{aligned}$$

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

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

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

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

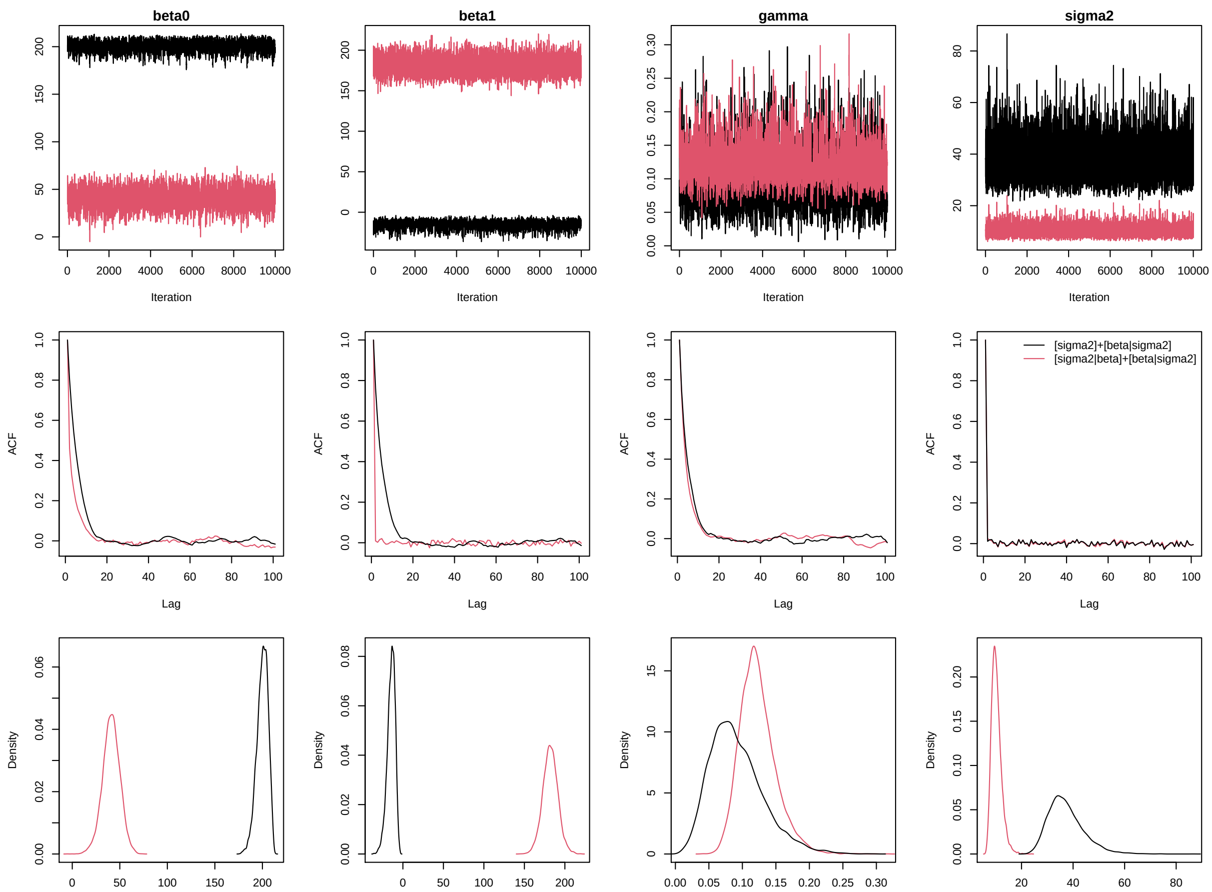

Figura 2: MCMC chains (top row), autocorrelations of the chains (middle row) and marginal posterior densities (bottom row). Algorithm 1 in red and algorithm 2 in black.

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 β_0, β_1, γ and σ^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,

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

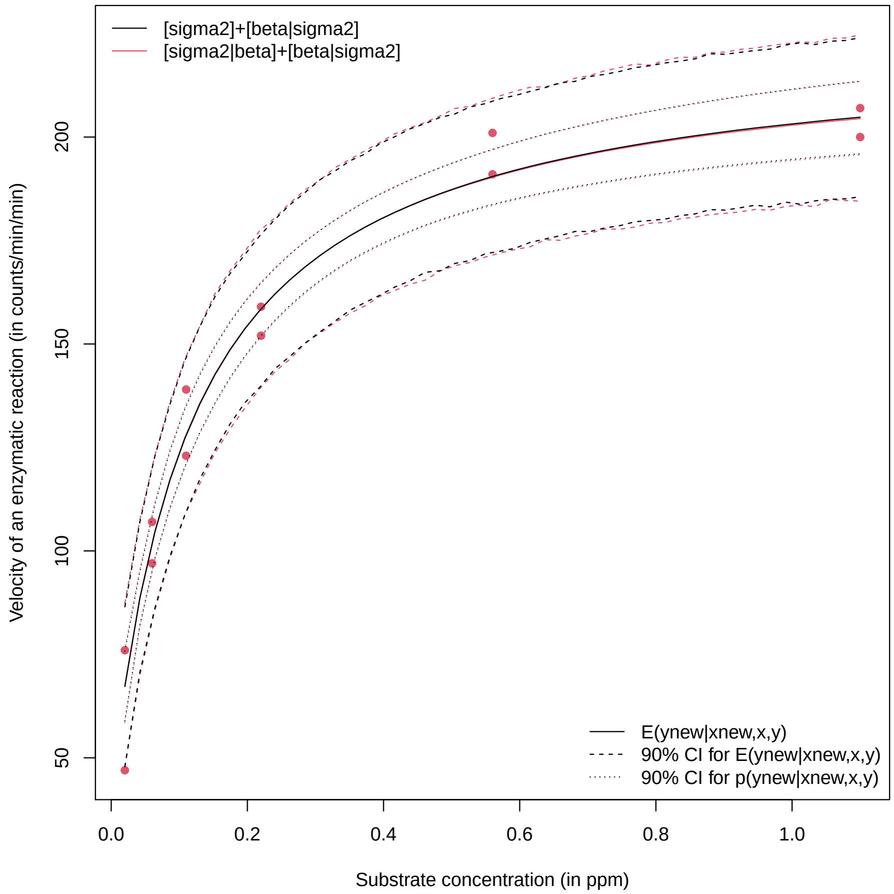

Figura 3: Posterior of $E(y_{new}|x_{new}, x, y) = \beta_0 + \beta_1 g(x_{new}, \gamma)$ and $p(y_{new}|x_{new}, x, y)$.

where $\{(\beta, \gamma, \sigma^2)^{(i)}\}_{i=1}^M$ 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 $\{(\beta, \gamma, \sigma^2)^{(i)}\}_{i=1}^M$ 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 $p(y|x_{n+1}, (\beta, \gamma, \sigma^2)^{(i)})$, for $i = 1, \ldots, M$.