Cox Proportional Hazards Model with Biomarker Effect

Some baseline patient factors, such as biomarkers, are useful in predicting patients' responses to a new therapy. Identification of such factors is important in enhancing treatment outcomes. Many of the biomarkers, such as gene expression, are measured on a continuous scale. A threshold of the biomarker is often needed to define a sensitive subset for making easy clinical decisions. Chen B.E., Jiang W. and Tu D. have developed a hierarchical Bayesian model for estimating the biomarker threshold and treatment effects simultaneously.

Here we use the same model they developed but with a different methodology (they used a Gibbs sampler): the Approximate Bayesian Computation and Markov Chain Monte Carlo hybrid.

Reference:
Chen B.E. Jiang W. Tu D. 2013. A hierarchical Bayes model for biomarker subset effects in clinical trials.
Kypraios T. Neal P. Prangle D. 2016. A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation.

The hierarchical Bayes model:
Let T_i and C_i be the failure and censoring time of the i^{th} subject in a study. Let X_i =min(T_i, C_i) be the observed failure or censoring time, whichever occurs first. Let $\delta_i = I(T_i < C_i)$ be the survival status indicator. Let z_{1i}be the treatment indicator taking 1 if the i^{th} subject received the treatment of interest, and z_{2i}be a continuous biomarker variable which is normalized to be ranged between 0 and 1. Given z_{2i},we define a threshold parameter c (ranged between 0 and 1) for the biomarker variable to indicate the division of subjects into two subsets, the follow cox proportional hazards model with c as a latent variable is formulated:

$$h(t \mid z_{1i}, z_{2i}, c, \beta) = h_0(t)\exp\{\beta_1 z_{1i} + \beta_2 I(z_{2i} > c) + \beta_3 z_{1i} I(z_{2i} > c)\}$$

where $h(t)$ and $h_0(t)$ are the hazard function and baseline hazard function at time t.

The inference of this model cannot be handled as a typical cox regression problem because of the introduction of the latent variable c. We will use a Bayesian paradigm to estimate β's and c.

The Bayesian paradigm

To make the generation of c from its prior distribution simpler, we introduce a variable q, $q > 1$, such that the conditional prior of c be

$$p_2(c \mid q) \propto q(q+1)c(1-c)^{q-1}$$

which is a Beta(2,q) distribution, where the prior of q is

$$p_1(q) \propto \frac{(q-1)}{q(q+1)}, q > 1$$

The conditional prior of $\beta \mid c, q$ is straightforward: ordinary cox proportional hazards model:

$$p_3(\beta \mid c, q) = \prod_i \left[\frac{\exp\{Z_i'(c)\beta\}}{\sum_{j \in R(x_i)} \exp\{Z_i'(c)\beta} \right]^{\delta_i}$$

where $R(t)$ is the index set of subjects that are at risk at time t.

Based on these (conditional) priors, the joint posterior distribution of c, β, q is

$$\begin{aligned} p(\beta, c, q \mid data) &\propto p_1(q)p_2(c \mid q)p_3(\beta \mid c, q) \\ &= \prod_i \left[\frac{\exp\{Z_i'(c)\beta\}}{\sum_{j \in R(x_i)} \exp\{Z_i'(c)\beta} \right]^{\delta_i} c(1-c)^{q-1}(q-1) \end{aligned}$$

Conditional distribution of $q \mid c, \beta$:
Let $v = q - 1$ and λ =-log$(1 - c)$,

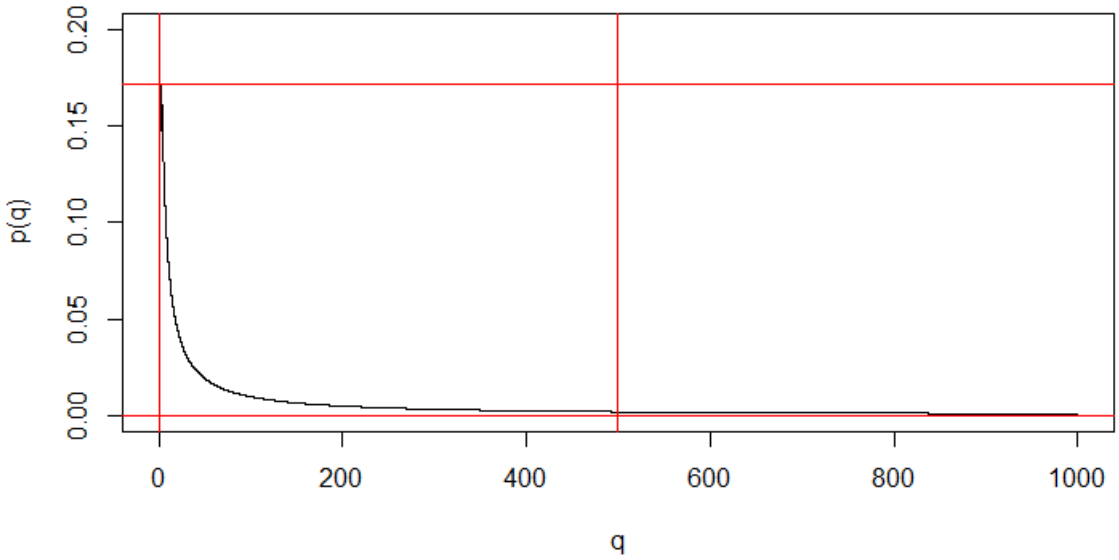
$$\begin{aligned} p(v \mid c, \beta) &\propto v(1-c)^v \log(\frac{1}{1-c}) \\ &= \lambda v \exp\{-\lambda v\} \end{aligned}$$

From this, we see that $v \sim \Gamma(2,\lambda)$, setting $q = v + 1$ we could simulate q based on c and β's.

Generation of q from its prior:

Since the prior of q $\propto \frac{(q-1)}{q(q+1)}$ is not easy to sample from, we propose a rectangular rejection sampling scheme to do so. The idea is to enclose our target density function within a rectangular box and generate points uniformly over

this region. We then reject any points lying outside the density function and retain otherwise in our sample.

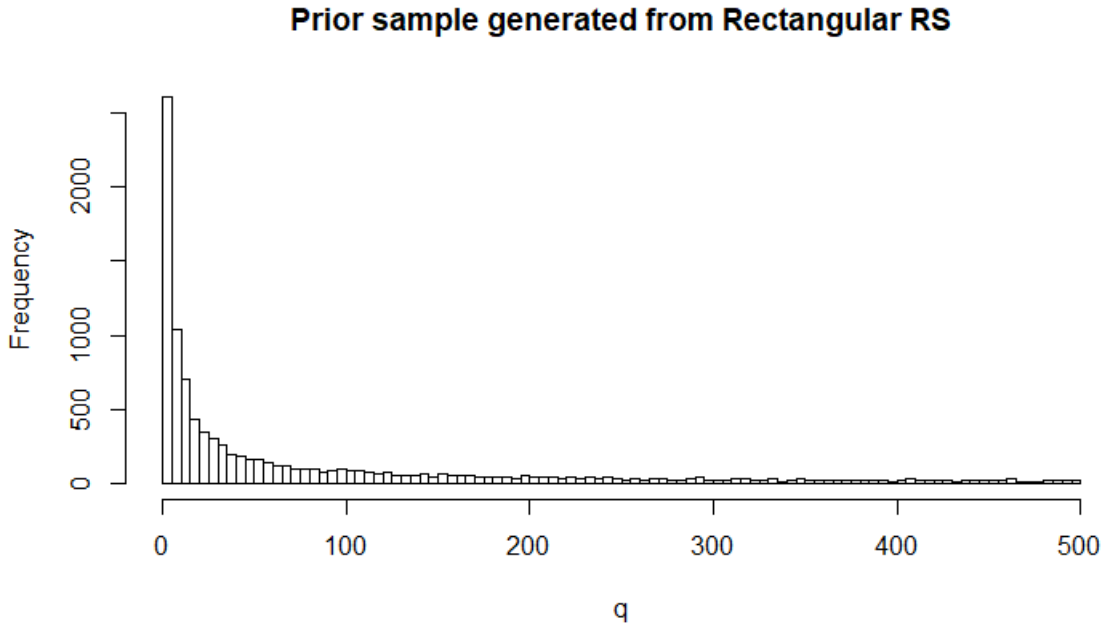


From the plot, we can safely take the x-range of the rectangular box to be [1,500] (Note: the vertical line of x=1 overlaps at x=0 because of graphing) and the y-range to be [0,0.1715725].

The sampling scheme is outlined as below:
1. Generate $U \sim$ U(1,500) and let $X = U$
2. Generate $V \sim$ U(0,1) and set $Y = 0.1715725 * V$
3. If $Y > p_1(X)$ then go back to step 1 and repeat, otherwise accept X.

An R file "prior_of_q" on the codes to implement this scheme is included in the repository.

The histogram of the obtained sample:



With the prior sample of q, we could implement Approximate Bayesian Computation (ABC) as follows. This is a Markov Chain Monte Carlo ABC hybrid algorithm (ABC-MCMC).

Input: observed survival data D, tolerance level ϵ, distance function $d(\cdot,\cdot)$, summary statistic $s(\cdot)$, proposal distributions $q_2(c^{s+1} \mid c^s)$and $q_3(\beta^{s+1} \mid \beta^s)$, a number of burn-in samples (B) to be discarded before our MCMC chain reached stationarity, total number of iterations in our chain (S).

Our choices: $\epsilon = 0.01, d(\cdot) = |\cdot - \cdot|$ (absolute difference), $s(\cdot)$= sample mean, $q_2(c^{s+1} \mid c^s) = U(c^s - 0.3, c^s + 0.3), q_3(\beta^{s+1} \mid \beta^s) = N(\beta^s, \hat{\Sigma}_{mle}), B = 1000, S = 3000$

Sampling algorithm:

Initial estimate:
1. Sample q from its prior sample
2. Sample c from $p_2(c \mid q)$ (if $c < 0.05$ or $c > 0.95$, repeat step 1 and 2)
3. Compute $I(z_{2i} > c)$ and $z_{2i}I(z_{2i} > c)$

4. Fit a cox proportional hazards model to obtain maximum likelihood estimates (mle) $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$
5. Perturb $\hat{\beta}$'s based on a normal distribution with mean $\hat{\beta}$'s and covariance the mle covariance estimated.
6. Simulate a survival time based on c and the finalized β's
7. Compute the absolute difference between the means of the observed and simulated survival times
8. If the distance is smaller than 0.01, accept the set of q, c, β's, otherwise repeat from step 1 until we obtain an initial set of q, c, β's.

Iteration 2 onwards:
for (s in 1:S){
1. Compute $c^{s-1}-0.3$ and $c^{s-1}+0.3$, set lower bound lb=max$(0.05, c^{s-1}-0.3)$ and upper bound ub=min$(0.95, c^{s-1}+0.3)$
2. Propose $c.p$ from $U(lb, ub)$
3. Compute $I(z_{2i} > c.p)$ and $z_{2i}I(z_{2i} > c.p)$
4. Fit a cox proportional hazards model to obtain mle covariance estimates $\hat{\Sigma}_{mle}$
5. Propose $\beta.p$'s based on a normal distribution with mean β^{s-1}'s and covariance the mle covariance estimated.
6. Compute λ = -log$(1 - c.p)$
7. Generate $v.p \mid c.p, \beta.p$'s by $v.p \mid c.p, \beta.p \sim \Gamma(2, \lambda)$, set $q.p = 1 + v.p$
8. Simulate a survival time based on $c.p$ and $\beta.p$'s
9. Compute the absolute difference between the means of the observed and simulated survival times
10. If the distance is smaller than 0.01, accept the set of $q.p, c.p, \beta.p$'s, otherwise repeat from step 1.
11. Compute the ratio $r = \dfrac{p(\beta.p, c.p, q.p \mid D)}{p(\beta^{s-1}, c^{s-1}, q^{s-1} \mid D)}$
12. Generate $u \sim U(0, 1)$, if $u < r$, then accept $[q^s, c^s, \beta^s] = [q.p, c.p, \beta.p]$, otherwise repeat from step 1.
13. After S iterations, discard B samples to obtain our finalized sample.
}

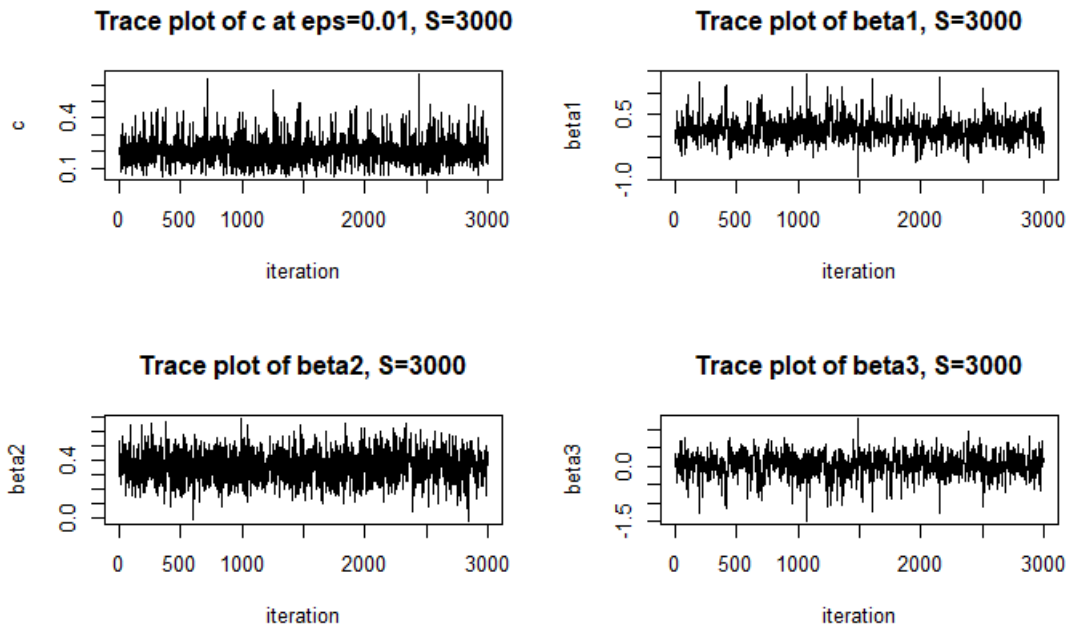
Simulation study

We simulate a survival date based on the following conditions to study the performance of our algorithm:
Number of subjects = 300. Half of them received the treatment ($z_{1i} = 1$), half did not ($z_{1i} = 0$). The values of their biomarker (z_{2i}) follows a $U(0, 1)$.

True $c = 0.2$. True $\beta = [0, \log(1.5), 0]$. Baseline hazards rate $h_0 = 1$.

Detailed codes on my implementation of the simulation study "ABC-MCMC_algorithm" is available in the repository.

Results of our model:



As seen from the trace plots of our posterior sample, stationarity of the MCMC chain is acheived.

Inference

Based on our samples, the point estimates of (c, β_1, β_2, β_3) is (0.206, 0.152, 0.358, 0.013), based on the posterior sample means.

The distribution of our posterior samples are shown below: