Bayesian Analysis For Cure Rate Model Under log MGEV with COM-Poisson Regression

Shuang Yin
Department of Statistics
University of Connecticut

December 9, 2019

Abstract

This project introduces the log maxima generalized extreme value (GEV) distribution to analyze right-censored survival data with a cure fraction. The proposed GEV model can lead to very flexible hazard functions. We also propose a more general count data distribution–Conway Maxwell Poisson (Com-Poisson distribution) to describe the count data in the survival model. The advantage of COM-Poisson is that it can be able to handle both underdispersion and overdispersion through controlling one special parameter in the distribution. A problem in sampling COM-Poisson is the calculation of the normalized constant for which doesn't have a closed form so that there is no accurate estimate for the constant. So an alternative sampling method–Exchange algorithm instead of the common used MCMC sampling will be used in this project. The exchange sampling is specifically illustrated in this project and implemented through a simulation.

Keywords: COM-Poisson Distribution; Cure Rate Model;Exchange Algorithm; Maxima GEV Distribution; MCMC sampling.

1 Introduction

Following up the individuals always has time restriction in real experiments for which the cured fraction appears with censorship at the end of the study. Cure rate models concentrate on modeling cured proportion who survived long enough to be studied as the cured individuals. Meanwhile, the cure rate models also focus on the probability of those individuals who survived up to a certain time while were not considered as cured. The present cure rate models are basically divided into two main categories, mixture cure rate models and non-mixture rate models. The standard mixture cure rate model (Berkson and Gage (1952)) where a fraction of the population is considered "cured" and the rest proportion is non-cured, and incorporate a particular survival function for the non-cured group in this population. Chen et al. (1999) showed that the former standard cure rate model has several drawbacks, i.e, we might get the improper posterior distributions for many types of non-informative improper priors. The non-mixture cure rate models were first introduced by Yakovlev et al. (1993) and then discussed by Chen et al. (1999) and Chen et al. (2002).These models were under an assumption such that the number of cancer cells are growing after cancer treatment following a Poisson distribution. The hazard function, however, often is not monotone. Roy and Dey (2014) proved that the maxima GEV for the logarithm of the failure time/survival time of a subject, can be fit to flexible hazard functions through controlling the shape parameters. The number of cancel cells for the individual subjects, being described by the simple Poisson distribution in Chen et al. (1999), however, usually encounters the overdispersion or underdispersion. By incorporating a new parameter which controls the amount of dispersion, the number of cancer cells is following the COM-Poisson (Shmueli et al. (2005)) which will explain better in overdispersion and underdispersion. The number of cancer cells is not observed and considered to be a latent variable. The usual MCMC algorithm is not able to compute the accept ratio due to the normalized constant of COM-Poisson. Therefore, the exchange algorithm (Chanialidis et al. (2018)) is deployed, incorporating the density of auxiliary data sampled from the distribution estimated at the value of parameters from proposed distribution. With all assumptions satisfied, the posterior distribution for the parameters can be derived and based on which the Bayesian MCMC sampling will be performed.

In section 2, we will discuss the strengths and drawbacks of the previous work, and also we will provide the definition and derivation for the standard survival model. In section 3, we introduce the cure model with a standard log GEV for maxima as the survival function and derive the likelihood function with covariates under COM-Poisson regression. We also construct the exchange algorithm to perform the sampling in a Bayesian framework. In section 4 we present the results of recovery work of these methods applied to a simulated data set.

2 Previous Work

In the previously popular standard cure rate model (Berkson and Gage (1952)), the survival function can be written as $S(t) = \pi + (1 - \pi)S^*(t)$ where a fraction π of the population are considered "cured" and $1 - \pi$ is non-cured, and $S^*(t)$ denotes the survival function for non-cured group in this population. The common choices for $S^*(t)$ are Weibull and Gamma distributions. However, there are some disadvantages of the standard cure rate model, 1) it doesn't have a proportional hazard structure, which may be desirable for survival models, 2) The computation is time-consuming and we might get improper posterior distributions for many types of non-informative priors. An alternative cure rate model has been rised by Yakovlev et al. (1993), and a Bayesian formulation of this model is by *Chenet al.* (1999), which can be derived as follows:

- N: the number of clonogenic cells and is following Poisson (θ) distribution.
- $\mathbf{Z}$: the incubation time for the N cells, and $\mathbf{Z} = (Z_1, \ldots, Z_N)$.
- $T = \min(Z_1, \ldots, Z_N)$, and $P(Z_0 = \infty) = 1$.

- The survival function for the population is

$$\begin{aligned} S(t) &= P(\text{no cancer by time t}) = P(N=0) + P(Z_1 > t, \ldots, Z_N > t, N \geq 1) \\ &= \exp(-\theta) + \sum_{j=1}^{\infty} S(t)^j \frac{\theta^j}{j!} \exp(-\theta) \\ &= \exp(-\theta + \theta S(t)) \\ &= \exp(-\theta F(t)) \end{aligned} \tag{1}$$

$S(t) = \exp(-\theta F(t))$ is an improper survival function since $S(\infty) = \exp(-\theta) > 0$. Thus the cure rate is given by $P(N = 0) = \exp(-\theta) = \lim_{t\to\infty} S(t)$.

If we incorporate the covariates in the equation 1 through θ according to the relationship $\theta = \exp(x'\beta)$, where x is the covariates of interest and β is the corresponding regression coefficients. The regression coefficients now become interpretable for cured and non-cured proportions through the log link relating covariates and cured fraction. The likelihood function can be constructed as follows. Suppose that there are n subjects and N_i's are Poisson random variables with mean θ_i and N_i is not observed. Let $T_{i1}, \ldots, T_{iN_i}$ be i.i.d incubation time for the ith subject and distributed under the same c.d.f with a specific parametric form $F(\cdot|\xi)$. Set $t_i = \min(T_{i1}, \ldots, T_{iN_i})$ and we take $y_i = \min(t_i, c_i)$ where c_i is the censoring time. The observed data set $\boldsymbol{O} = (n, \boldsymbol{y}, \boldsymbol{\delta})$ where $\boldsymbol{\delta}$ is the censoring indicator. The complete data is $\boldsymbol{D} = (n, \boldsymbol{y}, \boldsymbol{N}, \boldsymbol{\delta})$ and the likelihood function of the parameter $(\boldsymbol{\beta}, \xi)$ can be written as

$$L(\boldsymbol{\beta}, \xi|\boldsymbol{D}) = \prod_{i=1}^{n} S(y_i|\xi)^{N_i - \delta_i} (N_i f(y_i|\xi))^{\delta_i} \times \exp\left\{ \sum_{i=1}^{n} [N_i \boldsymbol{x_i'\beta} - \log(N_i!\,) - \exp(\boldsymbol{x_i'\beta})] \right\} \tag{2}$$

By applying the collapsed Gibbs procedure, we have

- sample N_i from Poisson$(S(y_i|\xi)\exp(\boldsymbol{x_i'\beta})) + \delta_i$
- sample $\boldsymbol{\beta}$ from the conditional posterior distribution $\pi(\boldsymbol{\beta}|\xi, \boldsymbol{O}) \propto \exp(\sum_{i=1}^{n}(\delta_i \boldsymbol{x_i'\beta} - \exp(\boldsymbol{x_i'\beta}) + S(y_i|\xi)\exp(\boldsymbol{x_i'\beta}))) \times (\boldsymbol{\beta})$
- sample ξ from $\prod_{i=1}^{n} S(y_i|\xi)^{N_i - \delta_i} f(y_i|\xi)^{\delta_i} \times \pi(\xi)$.

The alternative model even through solves some problems in the previous model, it still has some drawbacks:

- The survival models do not include flexible hazard functions.
- The number of clonogenic cells N is assumed to be Poisson for simplicity and ease of implementation, however, the Poisson model usually fails to capture the phenomenon of dispersion, especially when the mean and variance differ significantly.

Therefore, a new cure rate model and a more flexible count distribution should be considered to overcome the drawbacks discussed above.

3 Model Description and Methods

3.1 Log Maxima Generalized Extreme Value Distribution and Likelihood Function

Suppose that $Y_1, \ldots, Y_n \overset{i.i.d}{\sim} F(y)$ and let $M_n = \max\{Y_1, \ldots, Y_n\}$. If there exits a non-degenerate distribution function $G(x)$ and a pair of sequence $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} P\{\frac{M_n - b_n}{a_n} \leq x\} = G(x)$$

on all points in the continuity set of $G(x)$, then $G(x)$ is a generalized extreme value distribution for maxima (denote as MGEV distribution).

$$G(x) = \begin{cases} \exp\left\{-(1+\xi\frac{x-\mu}{\sigma})_+^{-\frac{1}{\xi}}\right\} & \text{if } \xi > 0 \text{ or } \xi < 0 \\ \exp\left\{-\exp(-\frac{x-\mu}{\sigma})\right\} & \text{if } \xi = 0 \end{cases}$$

Furthermore, if we assume that $\log T$ is following a MGEV distribution, then $T \sim \log MGEV$. The cdf, suvival function and pdf of T are:

$$F_M(t|\xi) = \exp\left\{-(1+\xi \log t)^{-\frac{1}{\xi}}\right\} \tag{3}$$

$$S_M(t|\xi) = 1 - \exp\left\{-(1+\xi \log t)^{-\frac{1}{\xi}}\right\} \tag{4}$$

$$f_M(t|\xi) = \frac{\exp\left\{-(1+\xi \log t)^{-\frac{1}{\xi}}\right\}}{y_i(1+\xi \log t)^{\frac{1}{\xi}+1}} \tag{5}$$

for $t > \exp(-\dfrac{1}{\xi})$ if $\xi > 0$ or $t < \exp(-\dfrac{1}{\xi})$ if $\xi < 0$.

Suppose that N_i's are independently COM-Poisson distribution with parameters μ_i and ν_i, $i = 1, \ldots, n$, for which can be able to handle both of overdispersion and underdisperion. Then the complete data likelihood function of the parameters $(\boldsymbol{\theta}, \boldsymbol{\nu}, \xi)$ can be written as:

$$L(\boldsymbol{\theta},\boldsymbol{\nu}, \xi|\mathbf{D}) = \prod_{i=1}^{n} S(y_i|\xi)^{N_i-\delta_i}(N_i f(y_i|\xi))^{\delta_i} \times \exp\left\{\sum_{i=1}^{n} N_i \log(\theta_i) - \log(N_i!\,) - \log(Z(\theta_i,\nu_i))\right\}$$

with

$$P(N_i = n) = \frac{\theta_i^n}{(n!\,)^{\nu_i}} \frac{1}{Z(\theta_i,\nu_i)}, n = 0, 1, \ldots, \text{where } \; Z(\theta_i,\nu_i) = \sum_{j=0}^{\infty} \frac{\theta_i^j}{(j!\,)^{\nu_i}} \tag{6}$$

for $\theta_i > 0$ and $\nu_i \geq 0$. If we incorporate the covariates through θ_i, ν_i, for each subject i, let $\mathbf{x}_i' = [1, x_{i1}, \ldots, x_{ip}]$ be the $p \times 1$ vector of covariates, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ be the corresponding coefficients, $\log(\theta_i) = x_i'\boldsymbol{\beta}$ and $\log(\nu_i) = -x_i'\boldsymbol{\gamma}$.We can directly derive the posterior distribution of ξ:

$$\begin{aligned}\pi(\xi|\mathbf{D}, \boldsymbol{\beta},\boldsymbol{\gamma}) &\propto f(\mathbf{D}|\xi) \times \pi(\xi) \\ &\propto \prod_{i=1}^{n} S(y_i|\xi)^{N_i-\delta_i}(N_i f(y_i|\xi))^{\delta_i} \times \pi(\xi)\end{aligned}$$

Since the posterior distribution of parameter ξ has no closed form, so a standard MCMC sampling, Metropolis-Hastings algorithm will be deployed to sample ξ. However, the only problem is to sample $\boldsymbol{\beta},\boldsymbol{\gamma}$ since the normalized constant $Z(\theta_i, \nu_i)$ has no closed form so that it's difficult to calculate.

3.2 COM-Poisson Regression and Exchange Algorithm

Assume N_i's are independent Conway–Maxwell–Poisson (COM-Poisson) distribution with equation (6). ν governs the amount of dispersion:overdispersion ($\nu < 1$) and the underdispersion ($\nu > 1$)

$$\begin{cases} \nu = 1 & \text{Poisson} \\ \nu = 0, \theta < 1 & \text{Geometric} \\ \nu \to \infty \ \text{ with probability} \frac{\theta}{1+\theta} & \text{Bernoulli} \end{cases}$$

If we incorporate the covariates through θ_i, ν_i, for each subject i, let $\mathbf{x}_i' = [1, x_{i1}, \ldots, x_{ip}]$ be the $p \times 1$ vector of covariates, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)$ be the corresponding coefficients and through the log link functions, $\log(\theta_i) = x_i'\boldsymbol{\beta}$ and $\log(\nu_i) = -x_i'\boldsymbol{\gamma}$. Rewrite the parameters as $\mu_i = (\theta_i, \nu_i)'$ and $\theta_i = \eta(\beta, x_i) = \exp(x_i'\beta)$, $\nu_i = \eta(\gamma, x_i) = \exp(-x_i'\gamma)$ be the link functions, and rewrite the pmf $P(N_i = n_i) = \frac{h(n_i|\mu_i)}{Z_h(\mu_i)}$. The Bayesian parameter inference for β and γ in COM-Poisson regression models is a doubly-intractable problem, so the direct application of a standard MCMC methods is infeasible. For example, a Metropolis algorithm requires the calculation of the intractable ratios $\left\{\frac{Z_h(\mu_i)}{Z_h(\mu_i^*)}\right\}_{i=1}^{n}$ if it proposes a move from μ_i to μ_i^*. Now the acceptance ratio becomes

$$\rho(\mu, \mu^*) = \min\left\{1, \frac{\prod_{i=1}^{n} \frac{h(n_i|\mu_i^*)}{Z_h(\mu_i^*)}}{\prod_{i=1}^{n} \frac{h(n_i|\mu_i)}{Z_h(\mu_i)}} \frac{q(\mu^*, \mu)\pi(\mu^*)}{q(\mu, \mu^*)\pi(\mu)}\right\}.$$

Even through Shmueli et al. (2004) proposed an approximations by a truncated sum $Z_h(\mu_i) = \sum_{i=1}^{k} q_h(n|\mu_i)$ to estimate the normalized constant, there still exists some bias in the acceptance ratio.

The exchange algorithm proposed by Benson and Friel (2017) can solve this problem by augmenting the doubly-intractable posterior with auxiliary data. Same with standard Metropolis, the exchange algorithm update the parameter from the current state μ to proposed state μ^* using the proposal distribution $q(\mu, \mu^*)$, but in addition the posterior is augmented with n auxiliary draws $N^* = (n_1^*, \ldots, n_n^*)$ generated from the likelihood estimated at the values of the parameters μ^* which are just proposed from the proposal distribution. Therefore, the augmented posterior can be written as

$$\pi(\beta, \beta^*, \mathbf{n}^*|\mathbf{n}) \propto h(\mathbf{n}|\boldsymbol{\mu})\pi(\beta)q(\beta, \beta^*)h(\mathbf{n}^*|\boldsymbol{\mu}^*)$$

The acceptance ratio for the augmented posterior is now calculated as

$$\begin{aligned}\rho_{exchange}(\beta, \beta^*) &= \min\left\{1, \frac{\prod_{i=1}^{n} \frac{h(n_i|\mu_i^*)}{Z_h(\mu_i^*)}}{\prod_{i=1}^{n} \frac{h(n_i|\mu_i)}{Z_h(\mu_i)}} \frac{q(\beta^*, \beta)\pi(\beta^*)}{q(\beta, \beta^*)\pi(\beta)} \frac{\prod_{n=1}^{n} \frac{h(n_i^*|\mu_i)}{Z_h(\mu_i)}}{\prod_{n=1}^{n} \frac{h(n_i^*|\mu_i')}{Z_h(\mu_i')}}\right\} \\ &= \min\left\{1, \frac{\prod_{i=1}^{n} h(n_i|\mu_i^*)q(\beta^*, \beta)\pi(\beta^*) \prod_{i=1}^{n} h(n_i^*|\mu_i)}{\prod_{i=1}^{n} h(n_i|\mu_i)q(\beta, \beta^*)\pi(\beta) \prod_{i=1}^{n} h(n_i^*|\mu_i^*)} \xcancel{\frac{\prod_i \frac{1}{Z_h(\mu_i^*)} \prod_i \frac{1}{Z_h(\mu_i)}}{\prod_i \frac{1}{Z_h(\mu_i)} \prod_i \frac{1}{Z_h(\mu_i^*)}}}\right\}\end{aligned}$$

The cancellation of the normalizing constants in the acceptance ratio above is due to the *exchange* of parameters (μ_i, μ_I^*) associated with the data $\mathbf{N} = (n_1, ..., n_n)$ and the auxiliary data $\mathbf{N}^* = (n_1^*, ..., n_n^*)$, the auxiliary being discarded after each move. The acceptance ratio for the exchange algorithm becomes

$$a = \min\left\{1, \frac{\left\{\prod_i h_{\mu^*}(n_i)\right\} \pi(\beta^*)\pi(\gamma^*) \left\{\prod_i h_{\mu}(n_i^*)\right\}}{\left\{\prod_i h_{\mu}(n_i)\right\} \pi(\beta)\pi(\gamma) \left\{\prod_i h_{\mu^*}(n_i^*)\right\}}\right\}$$

The standard Metropolis-Hastings to sample ξ and N_i's and the Exchange algorithms for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ will be applied to a simulated data set.

4 Simulation and Recovery Work

In this project, we simulate a right-censored data set and the process is given below:

Step 1. Let the sample size $n = 1000$, and suppose that we have one covariate: the age the patients which is randomly sampled with replacement from 1 to 100 and then standardized this variable. The $n \times 2$ design matrix $\mathbf{X} = [\mathbf{1}, \text{standardized age}]$ and also set $\beta = (1, 0.2)'$ and $\gamma = (0.5, -0.3)' \Rightarrow \theta_i = \exp(X_i'\beta)$ and $\nu_i = \exp(-x_i'\gamma)$. For every $i = 1, ...n$ we draw a sample from CMP(θ_i, ν_i), denoted by N_i. And get N_i samples from the distribution log MGEV($\mu = 0$, $\sigma = 1$, $\xi = 0.3$), denote the samples as $Z_{i1}, ..., Z_{i,N_i}$. Set $t_i = \min(Z_{i1}, ..., Z_{i,N_i})$ and $t_i = \infty$ if $N_i = 0$.

Step 2. For every $i = 1, ...n$ we draw a sample from CMP(θ_i, ν), denoted by N_i. And get N_i samples from the distribution log MGEV($\mu = 0$, $\sigma = 1$, $\xi = 0.3$), denote the samples as $Z_{i1}, ..., Z_{i,N_i}$. Set $t_i = \min(Z_{i1}, ..., Z_{i,N_i})$ and $t_i = \infty$ if $N_i = 0$.

Step 3. Let $y_i = \min(t_i, c_i)$ and $\delta_i = I(t_i < c_i)$ where c_i is the censoring time and chosen to be a constant so that the censoring percentage to be 28%.

Step 4. We include the covariates in the model and perform a Bayesian analysis. The multivariate normal $\mathbf{N}_2(\mathbf{0}, 100 \times \mathbf{I})$ are assigned to both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the proper prior on ξ is $\pi(\xi) = \text{Uniform}(-1, 1)$. In this process, $20,000$ MCMC iterations are used for sampling the posterior estimation of the parameters and the burn-in is set to be $1,000$. We use trace plots, autocorrelations to check the convergence in the MCMC sampler.

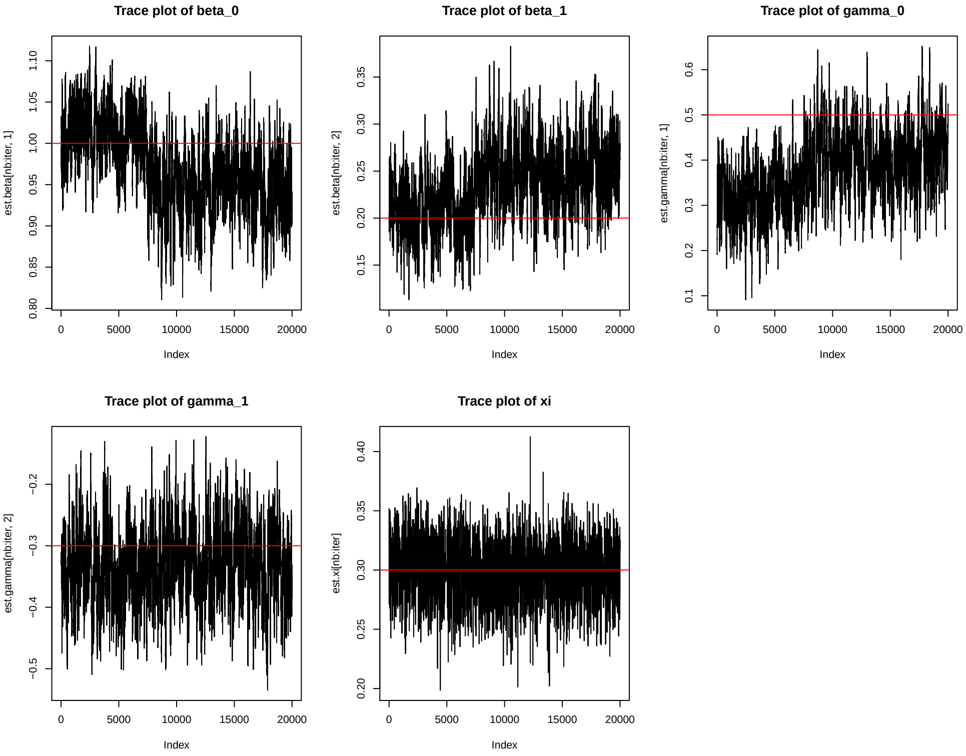

Figure 1: The trace plots for the parameters

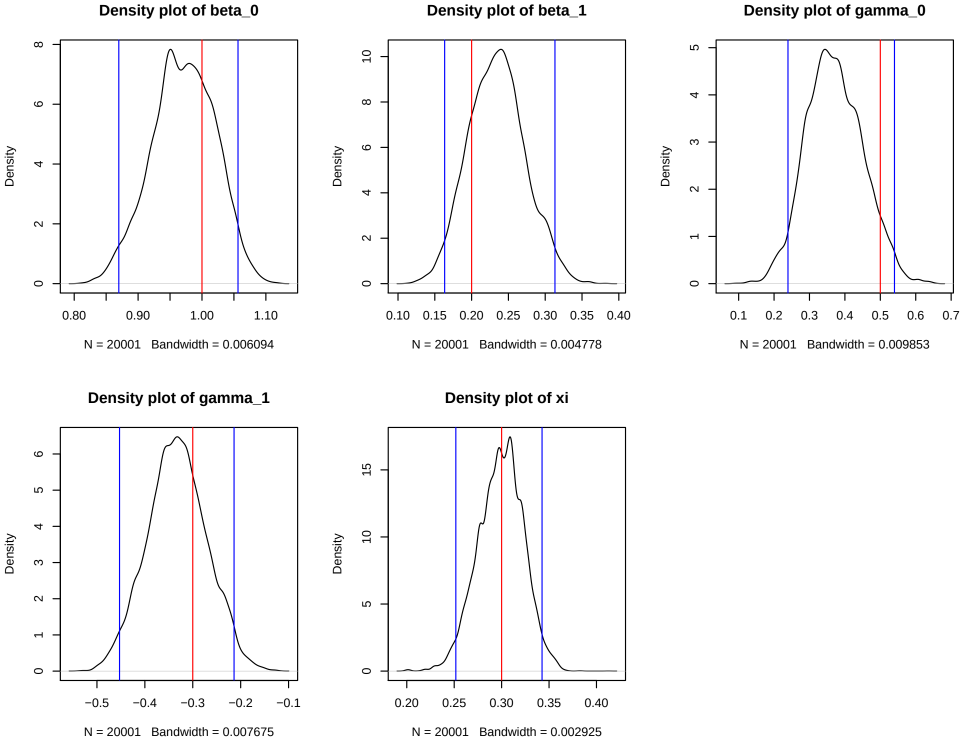

Figure 2: The density plots for the parameters

The trace plots for γ_1 and ξ seem to be mixing well, while there is a slight trend for other parameters, which might be due to the autocorrelation between states. Therefore, the thinning process would be taken to remove the strong autocorrelation. We can pick one sample for every 150 step.

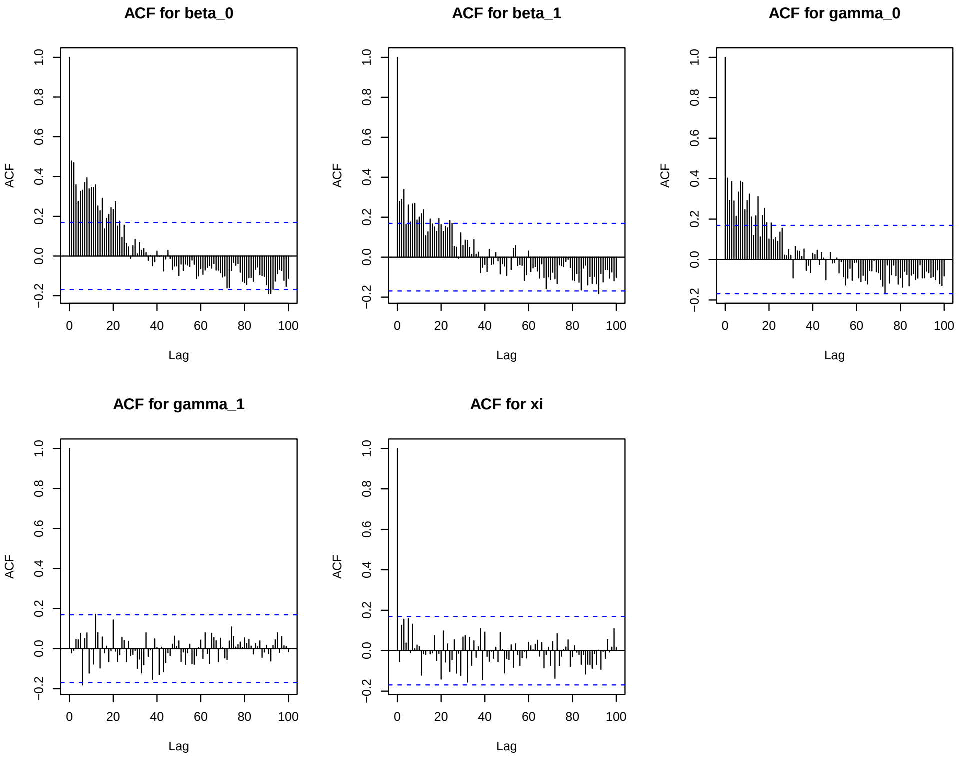

Figure 3: The density plots for the parameters

We also run the simulation for 10 times. Finally, 5 out 5 times the true values for the parameters are falling into the 95% HPD intervals.

Parameters					
	β_0	β_1	γ_0	γ_1	ξ
True	1	0.2	0.5	-0.3	0.3
Posterior Mean	0.971	0.234	0.372	-0.33	0.296
95% HPD lower	0.87	0.163	0.24	-0.453	0.252
95% HPD upper	1.06	0.313	0.54	-0.214	0.33

The averaged HPD 95% interval for the sampled parameters are calculated and can be see in table 4. And we can see that the true values are all within the HPD intervals and the posterior means are very close to the true values. Therefore, the sampling work is very stable and the recovery work performs well.

The model with Poisson cure rate is fit to the simulated survival data as well. The trace plots and density plots display the parameters sampled from the posterior distribution. Even though the true values of β_1 and ξ are falling into the 95% HPD interval, the true values deviate from the corresponding posterior means.

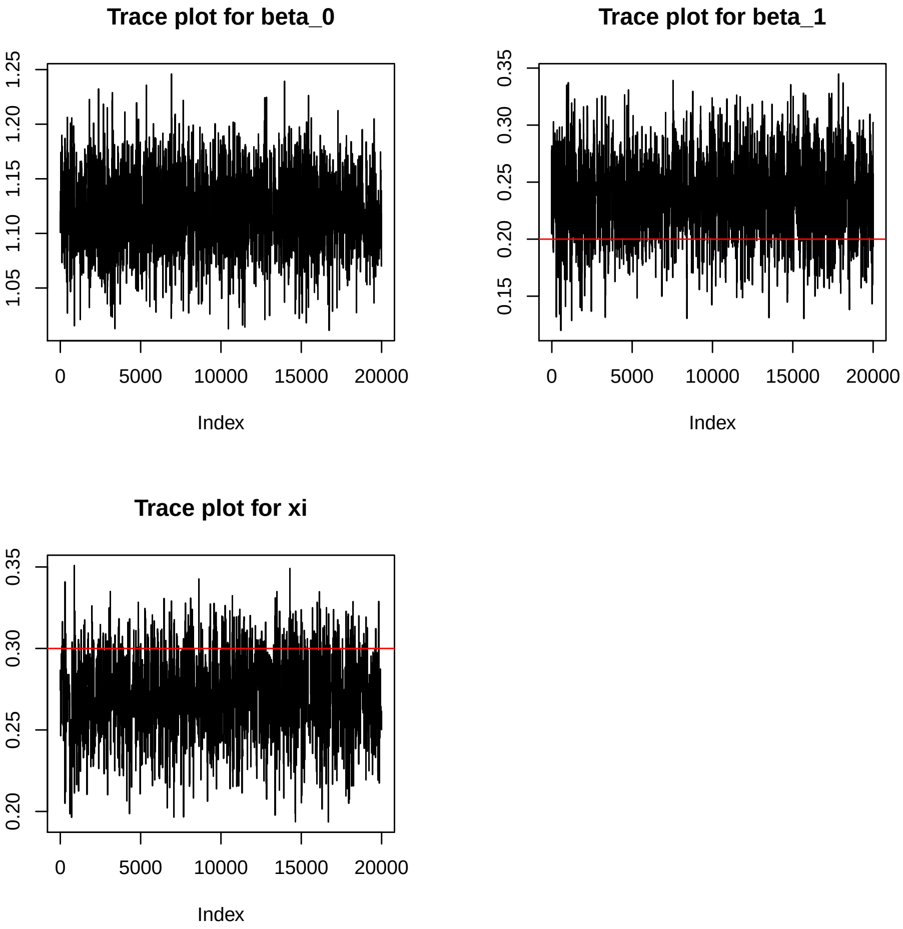

Figure 4: The trace plots for the parameters

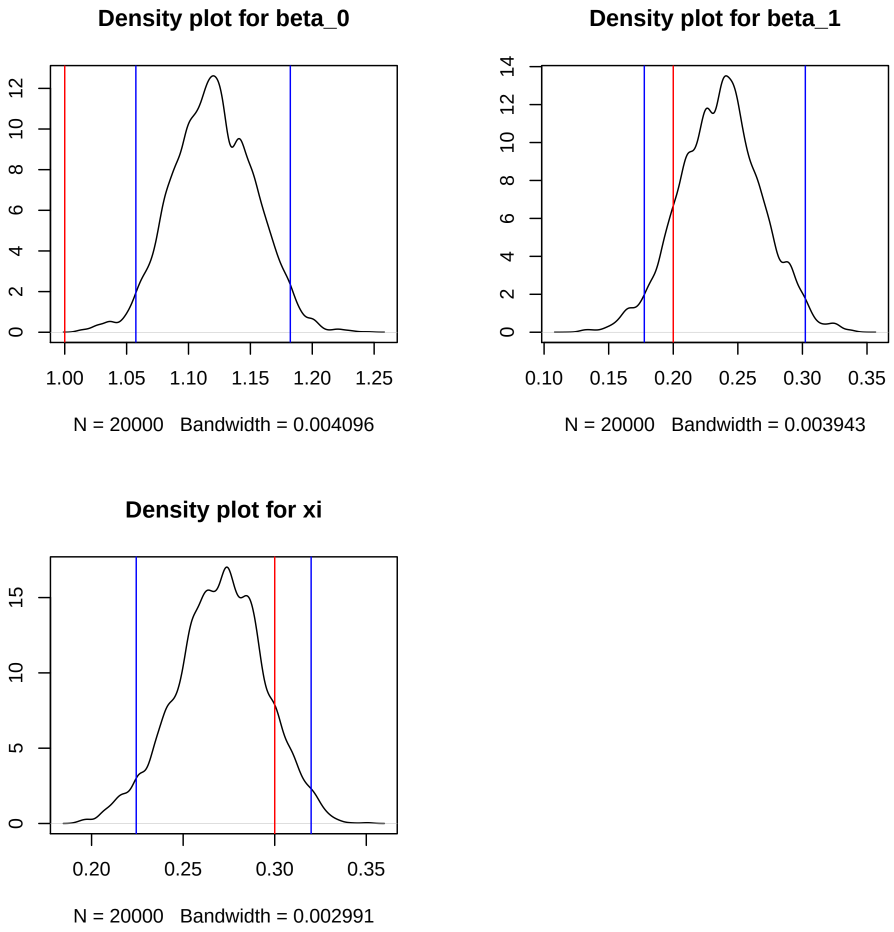

Figure 5: The density plots for the parameters

	β_0	β_1	ξ
True	1	0.2	0.3
Posterior Mean	1.12	0.237	0.27
95% HPD lower	1.057	0.178	0.224
95% HPD upper	1.82	0.302	0.319

5 Conclusions

- In this project, a computationally MCMC algorithm for COM-Poisson regression is presented to explain the count data in the cure rate model. Therefore, both of the survival function and the count data can be flexible to to general cases. It also displayed how the exchange algorithm worked, combined with the standard MCMC algorithm. Finally, the simulation results showed that those algorithm could recovery the true setups very well.

- The model selection and diagnostic are not carried out in this project. Since the normalized constant $Z(\theta, \nu)$ should be carefully and appropriately truncated to calculate the statistics of measure like Deviance Information Criteria or Log Pseudo Marginal Likelihood.

- The special MCMC sampling-Exchange algorithm nested with Metropolis algorithm works well in the simulated data set, further, those algorithms will be applied to describe the real data set. Meanwhile, the efficiency would also be a concentration in the future work.

References

Benson, A. and N. Friel (2017). Bayesian inference, model selection and likelihood estimation using fast rejection sampling: the conway-maxwell-poisson distribution. *arXiv preprint arXiv:1709.03471*.

Berkson, J. and R. P. Gage (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association 47*(259), 501–515.

Chanialidis, C., L. Evers, T. Neocleous, and A. Nobile (2018). Efficient bayesian inference for com-poisson regression models. *Statistics and Computing 28*(3), 595–608.

Chen, M.-H., J. G. Ibrahim, and S. R. Lipsitz (2002). Bayesian methods for missing covariates in cure rate models. *Lifetime Data Analysis 8*(2), 117–146.

Chen, M.-H., J. G. Ibrahim, and D. Sinha (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association 94*(447), 909–919.

Roy, V. and D. K. Dey (2014). Propriety of posterior distributions arising in categorical and survival models under generalized extreme value distribution. *Statistica Sinica*, 699–722.

Shmueli, G., T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright (2005). A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(1), 127–142.

Shmueli, G., R. P. Russo, and W. Jank (2004). Modeling bid arrivals in online auctions. *Robert H. Smith School Research Paper No. RHS-06-001*.

Yakovlev, A. Y., A. D. Tsodikov, and L. Bass (1993). A stochastic model of hormesis. *Mathematical Biosciences 116*(2), 197–219.