Survival function model estimation for parkinson disease using independent metropolis-hastings algorithm with uniform proposal distribution in bayesian inference

R Setiawan1,a, S Abdullah1 and A Bustamam1
1Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Kampus Baru UI, Depok 16424, Depok
a)E-mail: rohmat.setiawan@sci.ui.ac.id

Abstract. In medicine study, one of the things that is interesting enough to be studied is “time-to-event”. In general, time-to-event is used in doing survival analysis, such as analysis of Parkinson disease. Parkinson disease is one of the diseases which affects dopamine producer in brain area that is called by substantia nigra. The symptom of Parkinson disease is measured specifically by stages that are called by Hoehn and Yahr stages. This stages are distributed on integers between 0 to 5 with stage 0 is stage that does not have big impact and stage 5 is the most severe level. In this study, the survival function will be constructed from the time that the patient has the Hoehn and Yahr stages at A until increase to stage B with A < B. With A = 1, 2 and B = 3, 4, 5, overall it will be estimated six graphs of survival function. The process of construction survival function is using the Independent Metropolis-Hastings algorithm in Markov Chain Monte Carlo Methods on Bayesian Inference with uniform proposal distribution and the results are compared with Kaplan-Meier estimator for survival function. The result that is obtained through this algorithm is more represents the actual survival function if it is compared with Kaplan-Meier estimator, although there are so many censored data in the dataset.

1. Introduction
“Time-to-event” is one of the things that is interested enough to be studied in medicine study. “Time-to-event” means time until event happens that usually uses in doing survival analysis, such as analysis of Parkinson disease.

Parkinson disease is one of disease which affects dopamine producer in brain area that is called by substantia nigra (Parkinson’s Foundation, nd.). In general, this disease has unknown cause and does not have the specific treatment, so the patient who infected this disease is recommended to do some therapies to increase the production of dopamine.

Regarding the timing of Parkinson disease and the time until the increment of the stage of Parkinson disease happens, it can be done with a survival analysis approach. In this case, the event that want to be observed is the increment of the stage of Parkinson disease.

The increment of the stage of Parkinson disease can be different for each patient (Flensborg Damholt et al., 2012; Kehagia et al., 2010; Lewis et al. 2005; Marras and Lang, 2013). The symptoms and the increment of the disease can be different for each patient although the patients are in the same stage of disease. The interested thing to be observed is how the time distribution until the Parkinson patient reaches a certain stages of the disease.
The parametric approach is used in this research and then it will be generalized with Bayesian inference. The parametric approach is selected because in data analysis model, this approach does not need specification to the distribution of data. Next, because the data that is used is empiric, it is reasonable if the doubt arises to the result. In this condition, inference with Bayesian approach can answer the problem, which the Bayesian approach can allow the uncertainty. So, the Bayesian approach is known as the approach that is flexible and can represent the actual condition.

In this method, from the empirical data will be constructed likelihood function, then the prior distribution. Likelihood function is the probability function that is obtained from the before experience, while prior distribution is the distribution of the expert belief about the trial.

From likelihood function and prior distribution, then the posterior distribution will be obtained. Posterior distribution is the renewal distribution from prior distribution that counts the before experience. So, the results are more flexible and more representative in estimating the distribution.

Then the sampling will be sampled from posterior distribution using Markov Chain Monte Carlo. Markov Chain Monte Carlo is one of method that popular to sample from multidimensional distribution with various application (Wu and Chen, 2018). The criteria will be based on Independent Metropolis-Hastings algorithm with uniform proposal distribution. This algorithm is used because in general the direct sampling from posterior distribution is too hard to do.

The purposes of this research are construct the Metropolis-Hastings algorithm to estimate survival function and explain the increment of Parkinson disease stages using the survival function estimation. And the benefits of this research are as characteristic image of Parkinson disease stages based on time, give the contribution in the science development in health, especially about Parkinson disease.

2. Methodology
2.1. The designing process of posterior distribution
Previously, it has been explained that the probability density function of posterior distribution is proportional to multiplication of likelihood function and probability density function of prior distribution. Mathematically, it can be written as

\[ \text{posterior} \propto \text{likelihood} \times \text{prior}. \]

2.1.1. Likelihood function
For the i-th individual, \( X_i \) is defined as a random variable that states time until an event happened and \( C_i \) is defined as a random variable that states censored time. The data about people with Parkinson’s disease is the data in the form of pairs \((Y_i, \delta_i)\) with \( Y_i \) is the time for the i-th individual until the first experienced an event or until the censored happen and \( \delta_i \) is an event indicator for the i-th individual with \( \delta_i = 1 \) means the event happened and \( \delta_i = 0 \) means the event did not happen or censored happen. In this case, \( Y_i = \min(X_i, C_i) \) and \( \delta_i = I_{(X_i \leq C_i)} \).

Then the continuous functions \( f(\cdot), F(\cdot), g(\cdot), G(\cdot) \) are defined as the probability density function of \( X \), cumulative density function of \( X \), probability density function of \( C \), and cumulative density function of \( C \), respectively, with \( X \) and \( C \) are assumed independent.

With this assumption, the likelihood function construction can be written as

\[ L = \prod_{i=1}^{n} \left( f(y_i) \right)^{\delta_i} \left(S(y_i)\right)^{1-\delta_i}. \]

On the data about time-to-event, there are \( D \) different time points \( t_1, t_2, ..., t_D \) with \( 0 < t_1 < t_2 < \cdots < t_D \). So, the likelihood function construction from equation (1) can be written as

\[ L = \prod_{l=1}^{D} \left( -\Delta S(t_l) \right)^{d_l} \left( \sum_{j=l+1}^{p+1} -\Delta S(t_j) \right)^{c_l}, \]

with \( c_i \) is the number of censored data in interval \([t_i, t_{i+1})\) and \( d_i \) is the number of event that happened at \( t_i \).

2.1.2. Prior distribution
Survival function, \( S(y) \) is a function with \( S(y) = \Pr(Y > y) = \Pr(Y \in X_y) \) with \( X_y = (y, \infty) \). So, if there are \( D \) different time points \( t_1, t_2, ..., t_D \) with \( 0 < t_1 < t_2 < \cdots < t_D \), so it will be obtained \( D + 1 \) partitions \( A_1, A_2, ..., A_{D+1} \) in positive real line with \( A_i = [t_{i-1}, t_i) \), \( t_0 = 0 \), and \( t_{D+1} = \infty \).

So, the probability of an event happen in interval \( A_i \) can be defined as

\[
\Pr(Y \in A_i) = -\Delta S(t_i).
\]  

(3)

Then \(-\Delta S(t)\) will be Dirichlet distributed with the parameter \( \alpha = k(-\Delta S(t)) \) with \( k \) is the weight in prior estimation. So, the probability density function of prior distribution can be written as

\[
p(-\Delta S(t)) \propto \prod_{i=1}^{D+1} (-\Delta S(t_i))^{\alpha_i-1}
\]  

(4)

2.1.3. Posterior distribution

Since the probability density function of posterior distribution is proportional to multiplication of likelihood function and probability density function of prior distribution. So, from equation (2) and equation (4), there are can be derived the probability density function distribution that can be written as

\[
p(-\Delta S(t) | t) \propto (-\Delta S(t_{D+1}))^{D+1-1} \prod_{i=1}^{D} (-\Delta S(t_i))^{d_i+\alpha_i-1} \left( \sum_{j=i+1}^{D+1} -\Delta S(t_j) \right)^{\xi_i} \]  

(5)

The form \(-\Delta S(t)\) is the parameter that can be written as \( \theta \). So, the equation (5) can be updated as

\[
p(\theta | t) \propto (\theta_{D+1})^{D+1-1} \prod_{i=1}^{D} (\theta_i)^{d_i+\alpha_i-1} \left( \sum_{j=i+1}^{D+1} \theta_j \right)^{\xi_i} \]  

(6)

3. Simulation

3.1. Metropolis-Hastings algorithm

Simulation with Metropolis-Hastings algorithm is based on posterior distribution with probability density function \( p(\theta | y) \). It named after Metropolis et al. (1953) and Hastings (1970) proposes the way to construct a Markov chain that is stationary. Stages of the Metropolis-Hastings algorithm are described as follows:

1. Initialization the initial point \( \theta^{(0)} \) from arbitrary proposal distribution with probability density function \( q(\theta) \).
2. Then for \( i = 1 \), generate the point \( \theta^* \) from \( q(\theta^{(i-1)} | \theta^{(i-1)}) \).
3. Calculate \( P \) as the probability of acceptance with the formula is written as

\[
P = \min \left\{ 1, \frac{q(\theta^{(i-1)} | \theta^*) p(\theta^*)}{q(\theta^* | \theta^{(i-1)}) p(\theta^{(i-1)})} \right\}.
\]  

(7)

4. Generate the value \( u \) from uniform \( [0,1] \).
5. If \( u < P \), then \( \theta^{(i)} = \theta^* \). But if \( u \geq P \), then \( \theta^{(i)} = \theta^{(i-1)} \).
6. Do the step 2 – 6 for \( i = 2,3, ... \).

3.1.1. Independent Metropolis-Hastings algorithm

Independent Metropolis-Hastings algorithm is a Metropolis-Hastings algorithm with independent proposal distribution or \( q(x | y) = q(x) \). So, the probability of acceptance from equation (7) can be updated as

\[
P = \min \left\{ 1, \frac{q(\theta^{(i-1)} | \theta^*) p(\theta^*)}{q(\theta^* | \theta^{(i-1)}) p(\theta^{(i-1)})} \right\}.
\]  

(8)

In this case, the proposal distribution that will be used to generate the point \( \theta \) is uniform \([0,1]\). Since the \( \theta \) chosen with Dirichlet distribution, so it has properties that \( \theta_i \geq 0 \) for \( i = 1,2,\ldots, D+1 \) and \( \sum_{i=1}^{D+1} \theta_i = 1 \). Because the uniform \([0,1]\) generate number of \( \theta \) that is not necessarily \( \sum_{i=1}^{D+1} \theta_i = 1 \), so before the point \( \theta \) is used, it must be transformed as follow:
3.2. Data specification

Data that is used in this research is the data of people with Parkinson’s disease that is released by The Parkinson’s Progression Markers Initiative (2018) that is updated on 28 March 2018. The data is based on dataset from Movement Disorder Society-Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) part III, that is motor examination, especially Hoehn and Yahr (HY) stages. HY stages is stages to measure the Parkinson’s disease progression that is consist of 6 stages as follows:

**Table 1. Hoehn and Yahr stages description**

| Stages | Description |
|--------|-------------|
| 0      | Asymptomatic. |
| 1      | Unilateral involvement only. |
| 2      | Bilateral involvement without impairment of balance. |
| 3      | Mild to moderate involvement; some postural instability but physically independent; needs assistance to recover from pull test. |
| 4      | Severe disability; still able to walk or stand unassisted. |
| 5      | Wheelchair bound or bedridden unless aided. |

This dataset has 11985 records and 51 features. Even though, in this research only use 3 features, there are patient number (PATNO), time of examination (INFODT), and HY stages (NHY).

3.3. Data processing

In this research, it will be used the people with Parkinson’s disease have experienced the event with HY stages in stage $A$ for further is constructed survival function until time-to-event with HY stages in stage $B$ happen. With $A = 1,2$ and $B = 3,4,5$ it will be generated survival functions for six conditions.

Data processing is performed for 10000 Independent Metropolis-Hastings simulations with uniform proposal distribution which has been transformed using equation (9). Then, the last 1000 simulations will be taken to determine the parameter $\theta$.

The results of the simulation can be seen through figure below:
Figure 1. Survival function estimation for progression of HY stages from stage 1 to stage 3.

Figure 2. Survival function estimation for progression of HY stages from stage 1 to stage 4.

Figure 3. Survival function estimation for progression of HY stages from stage 1 to stage 5.

Figure 4. Survival function estimation for progression of HY stages from stage 2 to stage 3.

Figure 5. Survival function estimation for progression of HY stages from stage 2 to stage 4.

Figure 6. Survival function estimation for progression of HY stages from stage 2 to stage 5.

From the results, this simulation can generate the survival function in the good way corresponding to the properties of the survival function that are decrease function, and the value of function towards to 0 when it toward to infinity.

3.4. Comparison with Kaplan-Meier estimator

Kaplan-Meier estimator is one of the estimators that is powerful to estimate the survival function. For this case, from the Kaplan-Meier estimator that is obtained the results as follows:

Figure 7. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 1 to stage 3.

Figure 8. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 1 to stage 4.
Figure 9. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 1 to stage 5.

Figure 10. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 2 to stage 3.

Figure 11. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 2 to stage 4.

Figure 12. Survival function estimation with Kaplan-Meier estimator for progression of HY stages from stage 2 to stage 5.

It can be seen that the results from Kaplan-Meier estimator does not represent the actual survival function because one of the properties is not fulfilled, that is the value of function didn’t toward to 0 when it towarded to infinity. It happened because the dataset has a lot censored data.

4. Conclusion
Independent Metropolis-Hastings algorithm with uniform proposal distribution can be used to generate the survival function estimator. It is more represent than Kaplan-Meier estimators. Kaplan-Meier estimator is powerful in estimate the survival function in many cases. But, in the case with a lot censored data, Kaplan-Meier estimator is not powerful as known. So, for a lot censored data, the Kaplan-Meier estimator for survival function is not too good to build the model for survival function.

In this case, the more powerful estimator that can be used is Independent Metropolis-Hastings algorithm. This algorithm works based on Markov Chain Monte Carlo method principles. Monte Carlo method explains about how the point moves randomly, and the Markov Chain explains about the dependency of next point with the current point.

As the result, Independent Metropolis-Hastings algorithm can be used to estimate the survival function with a lot censored data in dataset and it can be used to explain the probability of people with Parkinson’s Disease with HY stages at stage A = 1,2 will (or will not) experience the progression of HY stages to stage B = 3,4,5 in during further month.

References
[1] Flensborg Damholdt M, Shevlin M, Borghammer P, Larsen L and Ostergaard K 2012 Acta Neurologica Scandinavica 125 311
[2] Hastings W K 1970 Biometrika 57 97
[3] Kehagia A A, Barker R A and Robbins T W 2010 The Lancet Neurology 9 1200
[4] Lewis S J, Foltynie T, Blackwell A D, Robbins T W, Owen A M and Barker R A 2005 Journal of Neurology, Neurosurgery & Psychiatry 76 343
[5] Marras C and Lang A 2013 Journal of Neurology, Neurosurgery & Psychiatry 84 409
[6] Metropolis N, Rosenbluth A W, Rosenbluth M N and Teller A H 1953 The Journal of Chemical Physics 21 1087
[7] Parkinson’s Foundation nd. What is Parkinson’s?
[8] RStudio Team 2016 Rstudio: Integrated development for R
[9] The Parkinson’s Progression Markers Initiative 2018 Motor Assessments Data: MDS UPDRS Part III
[10] Wu C H and Chen T L 2018 On the asymptotic variance of reversible markov chain without cycles Statistics and Probability Letters 137 224