Available online at http://journal.uad.ac.id/index.php/Konvergensi

Jurnal Ilmiah Matematika 6 (2), 2019, 81-86

Estimasi Parameter Model *Survival* Distribusi *Mixture* Weibull

(Estimation of Weibull Mixture Distribution Survival Model Parameters)

Reskika Hasmayuni*[1], Joko Purwadi[2]

Program Studi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan Yogyakarta, Jalan Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul, Yogyakarta, 55191, Indonesia

E-mail: reskika1500015037@webmail.uad.ac.id

* Corresponding Author

ARTICLE INFO

Keyword (Bahasa Indonesia)
Survival
Distribusi Weibull
Distribusi *Mixture* Weibull
MCMC
Gibbs Sampling
Bayesian

Keywords (Bahasa Inggris)
Survival
Weibull Distribution
Weibull *Mixture* Distribution
MCMC
Gibbs Sampling
Bayesian

ABSTRACT

Penelitian ini membahas tentang estimasi model survival distribusi Mixture Weibull menggunakan inferensi Bayes dengan memperhatikan distribusi prior. Bertujuan untuk mengetahui model dari distribusi Mixture Weibull serta hasil estimasi dari model survival distribusi Mixture Weibull yang terdiri dari dimana masing-masing parameter memiliki distribusi prior menggunakan metode Bayes. Model untuk analisis disribusi Mixture Weibull menggunakan metode Bayesian. Dengan parameter Weibull dari tiap elemen distribusi serta adalah bobot dari tiap elemen. Hasil estimasi parameter model survival distribus Mixture Weibull dengan metode Bayes yang diterapkan pada studi kasus pasien penderita Tuberkulosis dengan dan maka diperoleh w1 sama dengan 0.333333, dan w2 sama dengan 0.333333 serta hasil nilai estimasi untuk komponen Weibull tunggal dengan lamda 1.232808 dan beta 1.751471

This study discusses the estimation of the Weibull Mixture distribution survival model using Bayes inference by considering the prior distribution. Aim to find out the model of the Weibull Mixture distribution and the estimation results of the Weibull Mixture distribution survival model which consists of each parameter having a prior distribution using the Bayes method. The model for the Weibull Mixture distribution analysis uses the Bayesian method. With the Weibull parameter of each distribution element and the weight of each element. The estimation results of the Weibull Mixture distribution survival model with the Bayes method were applied to case studies of tuberculosis patients with and it was obtained that w1 was equal to 0.333333, and w2 was equal to 0.333333 and the results of the estimated value for the single Weibull component with lamda 1.232808 and beta 1.751471

PENDAHULUAN

Analisis *survival* atau analisis kelangsungan hidup bertujuan menaksir probabilitas kelangsungan hidup, kekambuhan, kematian, dan peristiwa-peristiwa lainnya sampai pada periode waktu tertentu. Pemilihan model perlu memerhatikan hal-hal berikut : (1) Bentuk distribusi probabilitas kelangsungan hidup, apakah bersifat parametrik atau non-parametrik, sebab tiap penyakit dan keadaan-keadaan lainnya memiliki bentuk distribusi masing-masing; (2) Apakah faktor resiko yang mendapat perhatian hanya satu (univariat) ataukah majemuk (multivariat); (3) Ukuran sampel penelitian; dan (4) Apakah data mencakup pengamatan tersensor atau tak tersensor [1].

Distribusi Weibull merupakan salah satu metode untuk menganalisis waktu *survival* pada pasien penderita penyakit[2]. Namun, penggunaan Weibull sederhana tidak selalu tepat digunakan untuk populasi yang heterogen, oleh karena itu model *Mixture* Weibull perlu dipertimbangkan untuk digunakan dalam populasi heterogen. Pada kasus ini distribusi *Mixture* Weibull hanya terbatas pada jumlah komponen tetap.

Dalam estimasi parameter dengan menggunakan metode Bayes diperlukan distribusi posterior. Distribusi posterior ini dibentuk dari distribusi prior. Nilai estimasi parameter diperoleh dengan simulasi pengambilan sampel parameter dari distribusi posterior dengan menggunakan metode *Markov Chain Monte Carlo* (MCMC) [3]. Salah satu algoritma dari MCMC yang digunakan adalah algoritma Gibbs sampling. Algoritma Gibbs sampling dapat diterapkan apabila distribusi probabilitas bersama (condotional distribution) dari tiap-tiap variabel diketahui [4].

METODE

Fungsi *Survival* dan Fungsi *Hazard*

Fungsi *survival* (*Survival Function*) adalah teknik statistik yang digunakan untuk menganalisis data yang bertujuan untuk mengetahui pengaruh dari variabel yang mempengaruhi suatu kejadian dari awal sampai akhir kejadian, misal waktu yang dicatat dalam hari, minggu, bulan, atau tahun [5]. Fungsi *survival S(t),* didefinisikan sebagai probabilitas seorang individu bertahan lebih besar dari waktu t sehingga:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

dengan $F(t)$ adalah *Cumulative Distribution Function* (CDF) dari disribusi data.

Fungsi *hazard* $h(t)$ merupakan probabilitas seseorang gagal setelah unit waktu yang ditentukan, seperti kebalikan dari fungsi *survival* $S(t)$. Formula *hazard* dapat diartikan probabilitas kondisional yaitu probabilitas terjadinya suatu kejadian pada interval waktu antara t dan $t + \Delta t$ dimana waktu *survival* (T) adalah lebih besar atau sama dengan t

$$h(t) = \frac{f(t)}{S(t)}$$

dengan kata lain, fungsi *hazard* $h(t)$ menaksir proporsi kematian individu atau individu mengalami suatu kejadian dalam waktu ke-t [5].

Distribusi *Survival* Weibull

Suatu variabel random kontinu X mempunyai distribusi Weibull dengan parameter bentuk $\beta > 0$ dan parameter skala $\theta > 0$, jika mempunyai fungsi densitas probabilitas:

$$w(x \mid \theta, \beta) = \begin{cases} \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}, & x > 0 \\ 0, & \end{cases}$$

Keuntungan yang memudahkan penerapan distribusi Weibull ini adalah fungsi *hazard* dengan berbagai nilai parameter bentuk θ mempunyai bentuk fungsi yang bermacam-macam [6].

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) adalah sebuah metode untuk membangkitkan peubah-peubah acak yang didasarkan pada rantai markov. Misalkan akan dilakukan simulasi dari suatu probabilitas π (atau disebut juga densitas target) permasalahan muncul karena simulasi secara langsung tidak mungkin dilakukan atau *infeasible* dikarenakan dimensi π yang tinggi. Maka MCMC membantu menyelesaikan masalah ini, Permasalahan semacam ini seringkali muncul dalam berbagai macam aplikasi sains contohnya pada bidang statistika, komputer sains, dan fisika statistik. MCMC memberikan solusi secara tidak langsung pada permasalahan yang didasarkan pada observasi dengan membangun sebuah rantai Markov ergodik dan π sebagai ukuran probabilitas stasioner. Ini lebih efektif dibandingkan melakukan simulasi secara langsung pada π. Disini akan digunakan salah satu dari algoritma MCMC yaitu algoritma Gibbs Sampling

Teorema Bayes

ISSN Print: 2087-8796, ISSN Online: 2774-3241

Misalkan peristiwa-peristiwa $A_1, A_2, \dots, A_k$ membentuk partisi dan saling lepas di dalam ruang sampel S sedemikian hingga $P(A_i) > 0$ untuk $i = 1, 2, \dots, k$ dan misalkan B sebarang peristiwa sedemikian hingga $P(B_i) > 0$, maka untuk $i = 1, 2, \dots, k$,

$$P(A_i \mid B) = \frac{P(A_i).P(B \mid A_i)}{\sum_{i=1}^{k} P(A_i).P(B \mid A_i)}$$

Prinsip – Prinsip Bayes

Definisi : Misalkan F adalah kelas dari fungsi densitas $f(y \mid \theta)$. Suatu kelas P dari distribusi Prior $\pi(\theta)$ disebut keluarga konjugat (jika fungsi peluang posterior memiliki distribusi yang sama dengan prior) untuk F bila distribusi Posterior $\pi(\theta \mid y)$ berada dalam kelas P untuk semua $\pi \in P$ [7]. Metode Bayes digunakan untuk memperbaiki distribusi Prior θ berdasarkan informasi sampel y, untuk memperoleh distribusi Posterior $\theta \mid y$ [7]. Distribusi Posterior tersebut digunakan untuk membuat inferensi tentang θ.

Estimator Bayes

Selanjutnya dapat dikatakan bahwa estimator terbaik adalah θ yang meminimumkan $R(T(y), \theta)$ dan $r(T(y) = \int R(T(y), \theta).\pi(\theta)d\theta$ disebut sebagai resiko Bayes $T(y)$ relatif terhadap Prior $\pi(\theta)$. $T(y)$ disebut sebagai estimator Bayes bila $r(T(y) \leq r(T'(y))$ untuk setiap $T'(y)$. Estimator Bayes untuk θ dibawah fungsi kerugian kuadratik yang merupakan mean posterior adalah

$$\hat{\theta} = E(\theta \mid y) = \begin{cases} \int \theta.\pi(\theta \mid y)d\theta, & \text{jika } \theta \text{ kontinu} \\ \sum \theta.\pi(\theta \mid y)d\theta, & \text{jika } \theta \text{ diskrit} \end{cases}$$

Tipe – Tipe Penyensoran

Penyensoran digunakan untuk menunjukkan bahwa periode pengamatan terputus sebelum peristiwa terjadi. Data dikatakan tersensor jika observasi waktu *survival* hanya sebagian, tidak sampai *failure event*. Penyebab terjadinya data tersensor antara lain :

1. Menghilang, terjadi bila objek pindah, meninggal atau menolak untuk berpartisipasi.
2. Keluar, terjadi bila perlakuan dihentikan karena alasan tertentu.
3. Pengakhiran penelitian, terjadi bila masa penelitian berakhir semantara objek yang diobservasi belum mencapai failure event (kegagalan).
4. Kematian, jika penyebab kematian bukan dibawah penyelidikan (misalnya bunuh diri).

Ada 2 jenis penyensoran, yaitu:

1. Sensor kanan, terjadi apabila orang yang kita amati tidak mengalami *event*, orang yang kita amati hilang dari pengamatan (*lost to follow up*), orang yang kita amati meninggal yang terjadi bukan karena *event*.
2. Sensor kiri, terjadi apabila kita tidak mengetahui dengan pasti waktu dari keadaan sebelum pengamatan.

Distribusi *Survival* Weibull Tunggal

Suatu variabel random kontinu X mempunyai distribusi Weibull dengan parameter bentuk $\beta > 0$ dan parameter skala $\theta > 0$, jika mempunyai fungsi densitas probabilitas [8]:

$$f(x \mid \theta, \beta) = \begin{cases} \frac{\beta}{\theta} x^{\beta-1} e^{-(\frac{x}{\theta})^{\beta}}, \ x > 0 \\ 0, \ untuk\ yang\ lain \end{cases}$$

Maka diperoleh Variansi dari Distribusi Weibull tunggal sebagai berikut,

$$Var(x) = \theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2(1 + \frac{1}{\beta})\right]$$

Distribusi *Mixture* Weibull

Fungsi densitas distribusi *Mixture* Weibull untuk k komponen dimana dalam kasus ini k komponen diketahui, sebagai berikut [8][9]:

ISSN Print: 2087-8796, ISSN Online: 2774-3241

$$f(x \mid k, w, \theta, \beta) = \sum_{j=1}^{k} w_j f_j(x); \quad 0 < w_j < 1; j = 1, 2, \ldots, k$$

fungsi *survival* dari distribusi *Mixture* Weibull didefinisikan sebagai berikut [9]:

$$S_m(x \mid k, w, \lambda, \beta) = \sum_{j=1}^{k} w_j . e^{-\lambda_j x^{\beta j}}$$

fungsi *Hazard* dari model distribusi *Mixture* Weibull sebagai berikut[10]:

$$h_m(x \mid k, w, \lambda, \beta) = \frac{\sum_j^k w_j f_j(x \mid k, w, \lambda, \beta)}{\sum_j^k w_j S_j(x \mid k, w, \lambda, \beta)}$$

Fungsi *Likelihood* dari model campuran dengan data tersensor kanan disajikan sebagai berikut :

$$L(k, w, \lambda, \beta \mid data) \propto \prod_{i=1}^{n} \sum_{j=1}^{k} w_j . (\lambda_j \beta_j)^{\delta_i} . x_i^{(\beta_j - 1)^{\delta_i}} . e^{(-\lambda_j x^{\beta_j})}$$

HASIL DAN PEMBAHASAN

Estimasi Parameter Distribusi Weibull Tunggal

Fungsi densitas prior yang dihasilkan mengikuti distribusi Gamma dengan parameter $\left(\frac{b}{bT+1}, r + d\right)$ dengan demikian kita dapatkan *Mean* Posterior dan Variansi Posterior sebagai berikut :

$$E(\lambda \mid T, r, b, d) = \frac{b(r + d)}{bT + 1}$$

dan

$$Var(\lambda \mid T, r, b, d) = \frac{b^2(r + d)}{(bT + 1)^2}$$

Dengan $T = \sum_{i=1}^{r} x_{i:n}^{\beta} + (n - r)x_{r:n}^{\beta}$ diperoleh estimatornya sebagai berikut :

$$\hat{\lambda} = E(\lambda \mid T, r, b, d) = \frac{b(r + d)}{b\left[\sum_{i=1}^{r} x_{i:n}^{\beta} + (n - r)x_{r:n}^{\beta}\right] + 1}$$

Estimasi Parameter Distribusi *Mixture* Weibull

Data dideskripsikan memiliki campuran Distribusi Weibull dengan 3 komponen $(k = 3)$ yang didasarkan pada data penelitian. Jumlah total data sebanyak 51 pasien $(n = 51)$ dengan data tersensor berjumlah 13 data $r = 13.$ Maka model dari Distribusi *Mixture* Weibull diberikan oleh persamaan berikut :

$$f_m(x \mid .) = w_1 f(x \mid \lambda_1, \beta_1) + w_2 f(x \mid \lambda_2, \beta_2) + w_3 f(x \mid \lambda_3, \beta_3)$$

Dengan $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ dan $\beta = (\beta_1, \beta_2, \beta_3)$ adalah parameter Weibull dari tiap elemen distribusi serta $w = (w_1, w_2, w_3)$ adalah bobot dari tiap elemen. Kemudian akan dicari nilai estimasi dari parameter campuran yaitu w, λ, β.

Uji Kesesuaian Data

Dari data pasien rawat inap penyakit Tuberkolosis akan diuji apakah data berdistribusi Weibull atau tidak. Digunakan pengujian dengan uji statistik Probabilitas Weibull sebagai berikut :

1. H_0: Data Berdistribusi Weibull
   H_1: Data Tidak Berdistribusi Weibull
2. Dengan tingkat Signifikansi
   $\alpha = 5\%$
3. Statistik Uji
   Dengan menggunakan uji statistik probabilitas Weibull melalui program R maka diperoleh :
   $p - value = 0.8806$
4. Daerah Kritis
   H_0 ditolak jika $p - value < \alpha$

ISSN Print: 2087-8796, ISSN Online: 2774-3241

5. Kesimpulan
   Karena $p - value = 0.8806 > \alpha = 0.05$, maka dapat ditarik kesimpulan bahwa H_0 diterima yang berarti data pasien Tuberkolosis yang diperoleh berdistribusi Weibull untuk $\alpha = 5\%$.

Dengan menggunakan software R, maka diperoleh hasil estimasi dari parameter sebagai berikut:

Tabel 1. Hasil Estimasi Parameter *Mixture* Weibull

w_1	w_2	w_3
0.333333	0.333333	0.333333
λ_1	λ_2	λ_3
0.632666	1.378291	0.700737
β_1	β_2	β_3
1.357401	2.663059	0.767066

Dapat diketahui nilai estimasi untuk komponen Weibull pertama memiliki bobot 0.333333 dan nilai parameter skala 0.632666 dengan nilai parameter bentuk 1.357401, begitu pula dengan komponen dua dan tiga.

Sedangkan dengan menggunakan metode Weibull tunggal maka hasil yang diperoleh sebagai berikut :

Tabel 2. Hasil Estimasi Parameter Weibull Tunggal ($k = 1$)

λ	β
1.232898	1.751471

Tabel 3. Kesalahan Standar

S_x ***Mixture* Weibull**	S_x **Weibull Tunggal**
0.066329	0.518573

Dari hasil estimasi parameter Weibull tunggal diperoleh nilai parameter skala $\lambda = 1.232898$ dan niai parameter bentuk $\beta = 1.751471$. Kemudian kesalahan standar untuk model *Mixture* Weibull sebesar 0.066329 sedangkan untuk Weibull tunggal sebesar 0.518573. Karena kesalahan standar model *Mixture* Weibull lebih kecil dari kesalahan standar model Weibull tunggal maka pemilihan model *Mixture* Weibull dalam menyelesaikan estimasi pada data pasien Tuberkolosis dapat dikatakan lebih baik dibandingkan dengan penggunaan model Weibull tunggal.

SIMPULAN

Kesimpulan yang dapat diambil dari hasil analisis dan pembahasan pada penelitian ini adalah sebagai berikut:

1. Model untuk analisis disribusi *Mixture* Weibull menggunakan metode Bayesian adalah
   $$f_m(x \mid .) = w_1 f(x \mid \lambda_1, \beta_1) + w_2 f(x \mid \lambda_2, \beta_2) + w_3 f(x \mid \lambda_3, \beta_3)$$
   Dengan $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ dan $\beta = (\beta_1, \beta_2, \beta_3)$ adalah parameter Weibull dari tiap elemen distribusi serta $w = (w_1, w_2, w_3)$ adalah bobot dari tiap elemen.
2. Hasil estimasi parameter model *survival* distribus *Mixture* Weibull dengan metode Bayes yang diterapkan pada studi kasus pasien penderita Tuberkulosis dengan $n = 51$ dan $r = 13$ maka diperoleh
   $$w_1 = 0.333333,\ \lambda_1 = 0.632666, \beta_1 = 1.357401, w_2 = 0.333333, \lambda_2 = 1.378291,\ \beta_2 = 2.663059, w_1 = 0.333333,\ \lambda_3 = 0.700737, \beta_3 = 0.767066$$
   serta hasil nilai estimasi untuk komponen Weibull tunggal dengan $\lambda = 1.232808$ dan $\beta = 1.751471$.
3. Hasil yang diperoleh dengan menggunakan model *Survival* Distribusi *Mixture* Weibull dapat dikatakan lebih baik dibandingkan dengan menggunakan model *Survival* Distribusi Weibull Tunggal.
4. Pendekatan model *Survival* Distribusi *Mixture* Weibull dengan Metode Bayes khususnya dengan Algoritma Gibbs Sampling dapat digunakan sebagai pendekatan model parametrik dan non

ISSN Print: 2087-8796, ISSN Online: 2774-3241

parametrik, dilihat dari hasil posterior β yang diperoleh, tidak mengikuti salah satu dari distribusi yang dikenal. Tetapi pada hasil Posterior yang lain mengikuti beberapa distriubsi lain.

DAFTAR PUSTAKA

[1] B. Murti, "Prinsip dan metode riset epidemiologi," *Yogyakarta UGM*, 1997.
[2] Y. Miyata, "Maximum likelihood estimators in finite mixture models with censored data," *J. Stat. Plan. Inference*, vol. 141, no. 1, pp. 56–64, 2011.
[3] W. R. Gilks and P. Wild, "Adaptive rejection sampling for Gibbs sampling," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 41, no. 2, pp. 337–348, 1992.
[4] C. M. Carlo, "Markov chain monte carlo and gibbs sampling," *Lect. notes EEB*, vol. 581, 2004.
[5] D. G. Kleinbaum and M. Klein, "Survival Analysis: A Self-Learning Text (Statistics for Biology and Health). 2005." New York: Springer Science+ Business Media.
[6] J. P. Klein and M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.
[7] J. O. Berger, *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.
[8] E. G. Tsionas, "Bayesian analysis of finite mixtures of Weibull distributions," *Commun. Stat. Methods*, vol. 31, no. 1, pp. 37–48, 2002.
[9] J. M. Marín, M. T. Rodriguez-Bernal, and M. P. Wiper, "Using weibull mixture distributions to model heterogeneous survival data," *Commun. Stat. Comput.*, vol. 34, no. 3, pp. 673–684, 2005.
[10] J. M. Marín, M. R. Rodríguez Bernal, and M. P. Wiper, "Using Weibull mixture distributions to model heterogeneous survival data. Universidad Carlos III de Madrid," Working Paper 03-32, Statistics and econometrics series 08, 2003.

ISSN Print: 2087-8796, ISSN Online: 2774-3241