

**BAYESIAN REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO (RJMCMC)
UNTUK PEMODELAN MIXTURE SURVIVAL**

¹Najihatur Rejki, ²Nur Iriawan

^{1,2}Jurusan Statistika, FMIPA ITS, Surabaya

¹rezqi.najihatur@gmail.com, ²nuririawan@gmail.com

ABSTRAK. Analisis *survival* merupakan metode satistika yang tepat untuk menganalisis data waktu tempuh suatu objek sampai terjadinya suatu peristiwa atau kejadian tertentu terhadap objek tersebut yang telah ditetapkan. Banyaknya kasus perceraian di Pengadilan Agama merupakan masalah yang cukup mengkhawatirkan di masyarakat, pengamatan tentang lama suatu pernikahan dapat dipertahankan merupakan fenomena *survival* ini. Pengamatan dilakukan pada para pihak yang mendaftarkan gugatan perceraian di Pengadilan Agama Kabupaten Malang, sebagai unit penelitian. Makalah ini mendemonstrasikan kemampuan pemodelan *mixture survival* dalam suatu *cox proportional hazard* yang dipadukan dengan cara estimasi parameternya menggunakan metode *reversible jump markov chain monte carlo* (RJMCMC) pada data *survival* yang mempunyai pola multimodal. RJMCMC dapat membantu memodelkan permasalahan *mixture* secara bersamaan dengan penentuan banyaknya komponen penyusunan *mixture* yang optimal. Hasil pemodelan dan analisis menunjukkan bahwa model *survival* pernikahan di area Pengadilan Agama Kabupaten Malang terdiri atas 12 komponen *mixture*. Model *mixture survival* lama pernikahan disusun oleh 12 komponen, yaitu:

$$S_{mix}(t) = 0,95S_1(t) + 0,017S_2(t) + 0,008S_3(t) + 0,0055S_4(t) + 0,0035S_5(t) + 0,0035S_6(t) + 0,0025S_7(t) + 0,0025S_8(t) + 0,002S_9(t) + 0,002S_{10}(t) + 0,0015S_{11}(t) + 0,0015S_{12}(t)$$

Kata Kunci: Analisis Survival; Cox Proportional Hazard; Mixture Survival; Reversible Jump Markov Chain Monte Carlo

1. PENDAHULUAN

Analisis *survival* adalah analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu objek sampai terjadinya peristiwa tertentu yang telah ditetapkan. Metode regresi *survival* merupakan metode regresi yang digunakan untuk melihat faktor-faktor yang menyebabkan terjadinya suatu peristiwa dengan variabel responnya adalah waktu ketahanan hidup. Regresi *cox proportional hazard* memungkinkan untuk interpretasi pengaruh dari masing-masing variabel prediktor [1].

Analisis *survival* umumnya digunakan dalam bidang medis, namun analisis ini juga cocok digunakan dalam bidang lainnya seperti pada bidang teknik, kriminologi, sosiologi, dan pada bidang-bidang lainnya yang memiliki resiko, dalam penelitian ini penulis tertarik untuk mengamati kasus sosiologi yaitu lama pernikahan dengan ukuran para pihak yang telah melakukan pendaftaran perceraian.

Analisis *survival* kemudian dikembangkan dengan menggabungkan konsep *mixture* karena data yang diperoleh tidak selamanya dapat direpresentasikan terhadap satu distribusi saja, namun dapat pula diduga terdiri dari beberapa komponen distribusi penyusun. Penelitian mengenai *mixture survival* telah dilakukan oleh Ando, et al. [ii] tentang model *kernel mixture survival* pada beberapa jenis penyakit kanker, Muthen [iii] tentang *mixture survival* untuk variabel waktu diskrit, Hariyanto [iv] tentang kasus lama mencari kerja di pulau Jawa tahun 2007 dengan model *mixture survival*, dan Hasyim [v] tentang model *mixture survival* spasial dengan *frailty* berdistribusi *conditional autoregressive* (CAR) pada kasus kejadian demam berdarah (DBD) di Kabupaten Pamekasan.

Pada penelitian-penelitian sebelumnya, proses inferensi *bayesian* menggunakan algoritma *Markov Chain Monte Carlo* (MCMC), dimana metode ini terbatas pada kasus banyaknya komponen penyusun *mixture* diketahui. Hal ini akan menjadi tidak berlaku pada kondisi banyaknya komponen penyusun *mixture* tidak diketahui. Proses inferensi *bayesian* pada kondisi vektor parameter model tidak tetap dikembangkan oleh Green [vi] yang kemudian disebut *Reversible Jump Markov Chain Monte Carlo* (RJMCMC). Richardson dan Green [vii] memanfaatkan algoritma ini untuk pemodelan *mixture* dengan banyak komponen penyusun *mixture* yang tidak diketahui.

Penelitian mengenai analisis *survival* data lama pernikahan belum pernah ada yang menggabungkannya dengan konsep *mixture*, dimana penggunaan model *mixture distribution* dinilai lebih baik karena memperhatikan sifat data [viii], terlebih lagi penelitian yang telah dilakukan terbatas pada diketahui banyaknya komponen penyusun *mixture*. Kenyataannya banyak persoalan dimana banyaknya komponen penyusun *mixture* sangat sulit ditentukan atau dengan kata lain banyaknya komponen penyusun *mixture* tidak diketahui. Salah satu persoalan yang diterapkan dalam penelitian ini adalah bagaimana memodelkan data lama pernikahan dengan pemodelan *mixture survival* dalam suatu *cox proportional hazard* yang dipadukan dengan cara estimasi parameter menggunakan metode RJMCMC pada data *survival* yang mempunyai pola multimodal.

Di Indonesia, Kabupaten Malang menempati posisi kedua untuk tingginya kasus perceraian dan menempati posisi pertama di Provinsi Jawa Timur. Tingginya angka perceraian terlihat dalam kurun waktu Januari hingga November 2014 tercatat 6.945 perkara perceraian yang terdiri dari 4.592 cerai gugat dan 2.353 cerai talak. Berbagai alasan yang mendasari para pihak untuk kemudian mengajukan perceraian ke PA Kabupaten Malang, yaitu dengan alasan moral (krisis akhlak dan cemburu), meninggalkan kewajiban (ekonomi dan tidak adanya tanggung jawab), penganiayaan, cacat bilogis atau gila, dan terus menerus berselisih.

Pada akhirnya, penelitian ini diharapkan dapat menghasilkan model terbaik sehingga dapat dilakukan langkah pencegahan terjadinya kasus perceraian setelah adanya sebuah pernikahan dengan mengamati faktor-faktor yang berpengaruh pada lama pernikahan di Pengadilan Agama Kabupaten Malang.

2. METODE PENELITIAN

2.1 Analisis Survival

Dalam analisis *survival*, terdapat dua fungsi yang digunakan yaitu fungsi *survival* dan fungsi *hazard*. Fungsi *survival* $S(t)$ didefinisikan sebagai probabilitas seorang individu bertahan lebih lama dari waktu t [ix].

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

selanjutnya, Kelson [x] menyatakan fungsi *hazard* sebagai laju kegagalan (*failure*) sesaat dengan asumsi individu telah bertahan sampai waktu ke- t yang didefinisikan pada persamaan berikut:

$$h(t)dt = P(t \leq T < t + \Delta t | T \geq t) \quad (2)$$

Kemudian diperoleh hubungan antara fungsi *survival* dan fungsi *hazard* sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

Jika resiko gagal (*failure*) pada waktu tertentu bergantung pada nilai x_1, x_2, \dots, x_p dari p variabel prediktor X_1, X_2, \dots, X_p maka nilai variabel tersebut diasumsikan telah tercatat sebagai *time origin*. Kumpulan nilai variabel prediktor dalam model *hazard proportional* diwakili oleh vektor \mathbf{x} dengan $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$. Misalkan $h_0(t)$ sebagai fungsi *hazard* untuk setiap obyek dengan nilai dari semua variabel prediktor penyusun vektor \mathbf{x} adalah nol maka fungsi $h_0(t)$ dikatakan sebagai fungsi *baseline hazard* [xi]. Model *hazard proportional* sebagai berikut:

$$h(t) = h_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}) \quad (4)$$

2.2 Reversible Jump Markov Chain Monte Carlo (RJMCMC)

Penggunaan algoritma MCMC pada proses inferensi *bayesian* terbatas pada kasus dimana dimensi vektor parameter model bersifat tetap. Penggunaan metode RJMCMC dilakukan pada pemodelan *mixture* dengan banyak komponen yang tidak diketahui. Keuntungan dari penggunaan metode ini adalah dapat memodelkan suatu *mixture* secara bersama-sama dengan banyaknya komponen penyusun *mixture*. Algoritma RJMCMC dapat dilakukan melalui enam langkah berikut:

Algoritma 1.

1. Update w
2. Update θ dimana $\theta = (\mu, \sigma)$
3. Update z
4. Update β
5. Split/merge komponen-komponen *mixture*
6. Birth/death komponen-komponen *mixture*

Proses pada langkah 1 sampai 4 tidak akan mengubah dimensi vektor parameter model yang terdiri dari $(\beta, \theta, k, w, \text{ dan } z)$, sedangkan langkah 5 dan 6 akan mengubah banyaknya komponen penyusun *mixture* satu per satu.

2.3 Mixture Survival

Bernardo dan Giron [xii] menggambarkan suatu model *mixture* sebagai sebuah model peluang yang digambarkan dengan densitas sebagai berikut:

$$p(x | \lambda, \theta) = \sum_{j=1}^k \lambda_j p(x | \theta_j), \lambda_j > 0 \quad (5)$$

dimana $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_k\}$, $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ dan k adalah banyaknya komponen dalam *mixture*.

Model *mixture* regresi *survival* ini didasarkan pada persamaan (5) dengan fungsi densitasnya disusun dari distribusi data *survival*-nya. Persamaan dari model *mixture survival* adalah:

$$f(t | \lambda, \theta) = \lambda_1 f(t | \theta_1) + \lambda_2 f(t | \theta_2) + \dots + (1 - (\lambda_1 + \lambda_2 + \dots + \lambda_{k-1})) f(t | \theta_k) \quad (6)$$

dengan $f(t|\theta_k)$ adalah fungsi densitas untuk data *survival* komponen ke- k , λ_k adalah proporsi komponen distribusi *mixture* ke- k , dan k adalah komponen ke-1, 2, ..., k . Sehingga fungsi *survival* distribusi *mixture* dengan k -komponen adalah:

$$S(t) = \lambda_1 S_1(t) + \lambda_2 S_2(t) + \dots + \lambda_k S_k(t) \quad (7)$$

dengan $S_k(t)$ adalah fungsi *survival* dari komponen *mixture* ke- k . dan model *proportional hazard* untuk *mixture survival* adalah:

$$h_i(t) = \lambda_1 h_{i1}(t) + \lambda_2 h_{i2}(t) + \dots + \lambda_k h_{ik}(t) \quad (8)$$

dengan $h_{ik}(t)$ adalah fungsi *hazard* dari komponen *mixture* ke- k .

2.4 Langkah-Langkah Analisis

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari register perkara perceraian yang terjadi di Pengadilan Agama Kabupaten Malang pada bulan Januari – November 2014 sebanyak 6.945 kasus perceraian.

Failure event dalam analisis *survival* lama pernikahan pada penelitian ini adalah kejadian putusnya suatu hubungan pernikahan (perceraian). Sensor yang digunakan dalam penelitian ini adalah sensor kanan (*right censor*) yang berarti bahwa jika sepasang suami istri sampai dengan masa pendataan selesai belum mengalami *failure event* maka waktunya dibatasi hanya sampai dengan berakhirnya masa pendataan. Sensor juga diterapkan jika sepasang suami istri mencabut kasus perceraian (rujuk) atau kasus perceraian tersebut dinyatakan di tolak oleh PA Kabupaten Malang maka waktunya dibatasi hanya sampai dengan berakhirnya kasus tersebut.

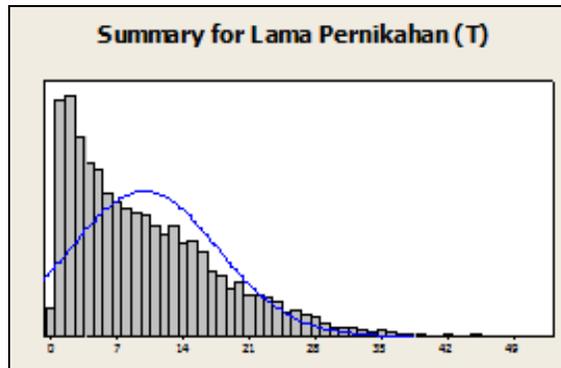
Variabel yang digunakan dalam penelitian ini yaitu variabel respon yang merupakan lama pernikahan (dalam tahun), sedangkan variabel prediktor meliputi jenis cerai, umur penggugat, pendidikan penggugat, pekerjaan penggugat, umur tergugat, pendidikan tergugat, pekerjaan tergugat, jumlah anak, dan alasan perceraian.

Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut:

1. Identifikasi awal data lama pernikahan berdistribusi *mixture*.
 - 1) Membuat histogram.
 - 2) Melakukan uji *goodness of fit*.
2. Menentukan banyaknya komponen penyusun *mixture*.
Tidak diketahuinya banyak komponen penyusun *mixture*, maka akan dibuat algoritma RJMCMC. Algoritma ini digunakan untuk mengestimasi banyak komponen model *mixture*.
3. Pembentukan model *mixture* dengan banyak komponen tidak diketahui.
 - 1) Berdasarkan hasil estimasi parameter regresi *cox proportional hazard* kemudian digunakan sebagai nilai initial untuk menyusun *mixture* regresi *survival*.
 - 2) Membentuk model *mixture survival*.

3. HASIL PENELITIAN DAN PEMBAHASAN

Dalam penelitian ini untuk mengidentifikasi distribusi *mixture*, tahapan awal yang dilakukan adalah dengan melihat distribusi data lama pernikahan untuk setiap pasangan yang mendaftarkan perceraian di PA Kabupaten Malang dengan menggunakan metode histogram dan melakukan pengujian *Anderson-Darling*.



Gambar 1. Histogram Data Lama Pernikahan

Histogram pada Gambar 1 menunjukkan bahwa data lama pernikahan tidaklah simetris dan patut diduga bahwa data waktu *survival* tidaklah berasal dari satu distribusi (*uni-modal*) melainkan disusun oleh beberapa distribusi (*multimodal*). Hal ini dipertegas dengan uji *Anderson-Darling*.

Tabel 1. Uji Distribusi Waktu *Survival*

Distribusi	Statistik uji (A_n^2)	Nilai Kritis ($a_{n,1-\alpha}$)	Keputusan
Lognormal	88,308	2,5018	Tolak H_0
Weibull	20,898	2,5018	Tolak H_0
Weibul 3P	17,288	2,5018	Tolak H_0
Normal	155,630	2,5018	Tolak H_0
Ekspensial	87,980	2,5018	Tolak H_0
Loglogistik	103,120	2,5018	Tolak H_0

Tabel 1 menunjukkan uji dari beberapa distribusi yang umumnya digunakan dalam analisis *survival*, namun berdasarkan uji *Anderson-Darling* tidak ada yang sesuai dengan distribusi-distribusi dugaan tersebut karena nilai statistik uji *Anderson-Darling* > nilai kritis pada $\alpha = 0,05$. Hasil uji ini menegaskan kesimpulan yang didapatkan dari *visual* histogram pada Gambar 1.

Dengan menggunakan algoritma RJMCMC, kemudian akan ditentukan banyaknya komponen penyusun *mixture* untuk data lama pernikahan dengan *listing* program sebagai berikut:

Program 1:

```

model {
  for (i in 1:n) {
    Z[i] ~ dlnorm(psi[i], tau)I(t.cen[i],)
  }
  psi[1:n] <- jump.lin.pred(X[1:n, 1:Q], k, beta.prec)
  id<- jump.model.id(psi[1:n])
  beta.prec <- tau / lambda
  tau ~ dgamma(a, b)
  k ~ dbin(0.5, Q)
}
list(tau = 0.01)

```


dengan

$$l(x | \mu, \sigma, \lambda) = \prod_{i=1}^n \frac{1}{x_i \sigma \sqrt{2\pi}} \exp \left[\frac{-(\log x_i - \mu)^2}{2\sigma^2} \right]$$

$$p(\sigma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{\alpha-1} e^{-\beta\sigma} \text{ adalah distribusi prior bagi } \sigma \text{ yaitu Gamma } (\alpha, \beta)$$

$$p(\lambda) = \frac{\left(\prod_{i=1}^k x_i^{\alpha_i-1} \right) \left(\Gamma \left(\sum_{i=1}^k \alpha_i \right) \right)}{\prod_{i=1}^k \alpha_i} \text{ adalah distribusi prior bagi } \lambda \text{ yaitu Dirichlet } (\alpha)$$

Model *mixture survival* dengan metode regresi *cox proportional hazard* dapat digambarkan sebagai persamaan berikut:

$$p(x | \lambda, \theta) = \lambda_1 \left(\exp(\beta_1 x_{i1}) \right) + \lambda_2 \left(\exp(\beta_2 x_{i2}) \right) + \dots + \lambda_{12} \left(\exp(\beta_{12} x_{i12}) \right)$$

Dalam regresi *cox proportional hazard*, yang disusun dalam model adalah $\exp(\beta x)$ karena pada kenyataannya *baseline hazard* $h_0(t)$ tidak perlu diketahui dan dimasukkan dalam model jika data berasal dari populasi yang sama (Collet, 2003). Selanjutnya, distribusi *prior* untuk nilai *beta* pada masing-masing variabel *prediktor* menggunakan *prior informatif* yang mengikuti distribusi normal. *Prior beta* didasarkan pada hasil pengolahan regresi *cox proportional hazard*.

Model *mixture survival* lama pernikahan disusun oleh 12 komponen, yaitu:

$$S_{mix}(t) = 0,95S_1(t) + 0,017S_2(t) + 0,008S_3(t) + 0,0055S_4(t) + 0,0035S_5(t) + 0,0035S_6(t) + \text{ da} \\ 0,0025S_7(t) + 0,0025S_8(t) + 0,002S_9(t) + 0,002S_{10}(t) + 0,0015S_{11}(t) + 0,0015S_{12}(t)$$

n model *proportional hazard* untuk *mixture survival* adalah:

$$h_i(t) = 0,95(h_{i1}(t)) + 0,017(h_{i2}(t)) + 0,008(h_{i3}(t)) + 0,0055(h_{i4}(t)) + 0,0035(h_{i5}(t)) + \\ 0,0035(h_{i6}(t)) + 0,0025(h_{i7}(t)) + 0,0025(h_{i8}(t)) + 0,002(h_{i9}(t)) + 0,002(h_{i10}(t)) + \\ 0,0015(h_{i11}(t)) + 0,0015(h_{i12}(t))$$

4. SIMPULAN

Identifikasi data memiliki indikasi multimodal dan berdistribusi *mixture log normal* yang ditunjukkan oleh *plot histogram* dan uji *Anderson-Darling* yang menunjukkan bahwa data tidak dapat didekati dengan satu distribusi univariat. Tidak diketahuinya banyak komponen *mixture* yang membentuk data, selanjutnya dengan menggunakan algoritma RJMCMC didapatkan sebanyak 12 komponen penyusun *mixture*. Model *mixture survival* lama pernikahan disusun oleh 12 komponen, yaitu:

$$S_{mix}(t) = 0,95S_1(t) + 0,017S_2(t) + 0,008S_3(t) + 0,0055S_4(t) + 0,0035S_5(t) + 0,0035S_6(t) + \text{ da} \\ 0,0025S_7(t) + 0,0025S_8(t) + 0,002S_9(t) + 0,002S_{10}(t) + 0,0015S_{11}(t) + 0,0015S_{12}(t)$$

n model *proportional hazard* untuk *mixture survival* adalah:

$$h_i(t) = 0,95(h_{i1}(t)) + 0,017(h_{i2}(t)) + 0,008(h_{i3}(t)) + 0,0055(h_{i4}(t)) + 0,0035(h_{i5}(t)) + \\ 0,0035(h_{i6}(t)) + 0,0025(h_{i7}(t)) + 0,0025(h_{i8}(t)) + 0,002(h_{i9}(t)) + 0,002(h_{i10}(t)) + \\ 0,0015(h_{i11}(t)) + 0,0015(h_{i12}(t))$$

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Prof. Drs. Nur Iriawan, M.Ikomp., Ph.D. yang telah bersedia meluangkan waktunya membimbing penulis dalam penulisan makalah ini. Terimakasih pula kepada Pengadilan Agama Kabupaten Malang, Jawa Timur yang telah bersedia memberikan data guna kelancaran penyelesaian makalah ini dan juga tidak lupa kepada para pihak yang telah membantu penulis dalam menyelesaikan makalah ini.

DAFTAR PUSTAKA

- [1] Lee, E. T., & Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis*. 3rd ed. New York: John Wiley and Sons, Inc.
- [2] Ando, T., Imoto, S. & Miyano, S. 2004. Kernel Survival Models for Identifying Cancer Subtypes, Predicting Patient's Cancer Types and Survival Probabilities. *Genome Informatics*, 15(2), pp. 201-210.
- [3] Muthen, B. & Masyn, K. 2005. Discrete-Time Survival Mixture Analysis. *Journal of Educational and Behavioral Statistics*, 5(2), pp. 63-74.
- [4] Hariyanto, S. 2009. *Model Mixture Survival pada Kasus Lama Kerja di Pulau Jawa Tahun 2007*. Tesis (Tidak Dipublikasikan). Surabaya: Institut Teknologi Sepuluh Nopember.
- [5] Hasyim, M. 2012. *Model Mixture Survival Spasial dengan Frailty Berdistribusi Conditionally Autoregressive (CAR) pada Kasus Kejadian Demam Berdarah Dengue (DBD) di Kabupaten Pamekasan*. Tesis (Tidak Dipublikasikan). Surabaya: Institut Teknologi Sepuluh Nopember.
- [6] Green, P. J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4), pp. 711-732.
- [7] Richardson, S. & Green, P. J. 1997. On Bayesian Analysis of Mixture with an Unknown Number of Components. *Journal of the Royal Statistical Society*, 59(4), pp. 731-792.
- [8] Iriawan, N. 2001. *Studi tentang Bayesian Mixture Normal dengan Menggunakan Metode MCMC*. Surabaya: Lemlit ITS.
- [9] Le, C. T. 1997. *Applied Survival Analysis*. New York: John Wiley and Sons. Inc.
- [10] Kleinbaum, D. G. & Klein, M. 2005. *Survival Analysis: A Self Learning*. 3rd ed. New York: Springer.

- [11] Collet, D. 2003. *Modelling Survival Data in Medical Research*. London: Chapman and Hall.
- [12] Bernardo, J. M. & Giron, F. J. 1988. A Bayesian Analysis of Simple Mixture Problem. *Bayesian Statistics*, Volume 3, pp. 67-68.

^{xii} Bernardo