

PENDEKATAN METODE BAYESIAN SELF DAN BAYESIAN GELF UNTUK ESTIMASI PARAMETER MODEL SURVIVAL EKSPONENSIAL DENGAN PRIOR JEFFREYS

Dewi Puspitawati, Setyo Wira Rizki, Nurfitri Imro'ah

INTISARI

Data survival adalah data yang menunjukkan waktu suatu individu atau objek dapat bertahan hidup hingga terjadinya suatu kegagalan atau kejadian tertentu. Tujuan pada penelitian ini adalah menentukan estimasi parameter model survival berdistribusi eksponensial dengan metode Bayesian SELF dan metode Bayesian GELF menggunakan prior Jeffreys, serta menerapkan pada kasus data penderita kanker paru-paru. Setelah diperoleh estimator dengan menggunakan metode tersebut, selanjutnya diterapkan pada data pasien penderita kanker paru-paru yang diambil dari program R versi 3.3.0 untuk mengetahui peluang individu dapat bertahan hidup. Nilai MSE yang diperoleh untuk fungsi survival dan fungsi hazard dari metode Bayesian SELF ialah $0,293363538$ dan $6,66149 \times 10^{-5}$, sedangkan dari metode Bayesian GELF ialah $3,60504 \times 10^{-6}$ dan $3,60158 \times 10^{-9}$. Berdasarkan nilai MSE dari estimator pada penelitian ini, diperoleh metode Bayesian GELF lebih baik dari pada metode Bayesian SELF. Hasil olah data dari metode Bayesian GELF prior Jeffreys diperoleh peluang seorang pasien untuk bertahan hidup jika mengidap penyakit kanker paru-paru selama 7 hari adalah $0,94447$. Sedangkan selama 63 hari adalah $0,59573$, selama 95 hari adalah $0,46053$, selama 162 hari adalah $0,26655$, dan selama 999 hari adalah $0,0003$.

Kata Kunci : Distribusi Eksponensial, Bayesian SELF, Bayesian GELF.

PENDAHULUAN

Analisis statistika yang digunakan untuk menganalisis data waktu hidup disebut analisis tahan hidup (*survival*) [1]. Analisis *survival* adalah prosedur statistika untuk menganalisis data dengan waktu sampai terjadinya suatu peristiwa tertentu sebagai variabel respon. Terdapat dua model yang digunakan untuk menganalisis data *survival* yaitu model parametrik dan model nonparametrik. Model parametrik adalah model dengan data *survival* yang mengikuti asumsi pada distribusi tertentu. Beberapa model parametrik mengikuti distribusi tertentu diantaranya yaitu distribusi Eksponensial, distribusi Weibull, distribusi Log-Normal, distribusi Log-Logistik, dan distribusi Gamma. Jika distribusi yang mendasari data *survival* tidak diketahui, artinya data tidak mengikuti suatu distribusi tertentu yang sudah ada maka digunakan model nonparametrik. Pada model nonparametrik terdapat dua metode, yaitu metode Kaplan-Meier dan metode Nelson-Aalen.

Distribusi Eksponensial merupakan suatu distribusi yang berguna untuk mencari selisih waktu yang terjadi dalam suatu peluang tertentu. Saat ini dikenal dua metode untuk mengestimasi parameter yaitu metode klasik dan metode Bayesian. Metode Bayesian merupakan metode estimasi yang menggabungkan distribusi prior dan fungsi *likelihood*. Distribusi prior adalah distribusi awal yang memberikan informasi tentang suatu parameter. Informasi dalam distribusi prior dikombinasikan dengan informasi sampel yang dinyatakan dengan fungsi *likelihood* dan hasilnya dinyatakan dalam bentuk distribusi yang disebut distribusi posterior yang selanjutnya menjadi dasar untuk inferensi di dalam metode Bayesian [2]. Terdapat beberapa pendekatan dari metode Bayesian yang dapat digunakan untuk mengestimasi parameter distribusi diantaranya yaitu *Generalized Non-informative prior*, *Linear Exponential Loss Function* (Linex), *Lindley Aproximation*, *General Entropy Loss Function* (GELF), dan *Squared Error Loss Function* (SELF). Penelitian ini membahas tentang Bayesian SELF dan Bayesian GELF karena metode ini digunakan untuk menentukan tingkat estimasi yg terlalu tinggi dan perkiraan parameter yang terlalu rendah [3].

Data yang digunakan dalam penelitian ini adalah data kasus penderita kanker paru-paru yang diambil dari program R versi 3.3.0 dengan melakukan uji *Kolmogorov-Smirnov* sehingga diketahui data berdistribusi Eksponensial. Langkah pertama adalah menentukan distribusi kumulatif, fungsi *survival*, dan fungsi *hazard* dari distribusi Eksponensial. Langkah kedua menentukan fungsi *likelihood*, distribusi prior, dan distribusi posterior untuk metode Bayesian. Langkah ketiga menentukan estimasi parameter dari metode bayesian SELF dan Bayesian GELF menggunakan prior *Jeffreys*, kemudian langkah keempat melakukan perhitungan fungsi *survival* dan fungsi *hazard* dari hasil estimasi metode Bayesian SELF dan Bayesian GELF. Langkah terakhir adalah menghitung MSE untuk fungsi *survival* dan fungsi *hazard* dari hasil estimasi metode Bayesian SELF dan Bayesian GELF prior *Jeffreys*, dan menerapkan pada kasus penderita kanker paru-paru.

Distribusi Waktu Survival

Distribusi waktu survival dapat dinyatakan dengan beberapa fungsi yaitu, fungsi kepadatan peluang, fungsi distribusi kumulatif, fungsi *survival*, dan fungsi *hazard*. Waktu *survival* (T) merupakan variabel random non-negatif yang mewakili waktu dari individu-individu dalam suatu populasi yang merupakan variabel random kontinu dalam interval $[0, \infty)$ [4]. Fungsi kepadatan peluang atau biasa disebut dengan PDF (*probability density function*) adalah peluang suatu individu mengalami kejadian sesaat dalam interval waktu t sampai $t + \Delta t$. Fungsi kepadatan peluang $f(t)$ dirumuskan sebagai berikut:

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t < T \leq (t + \Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right]$$

Jika T merupakan variabel acak dari waktu hidup suatu individu dalam interval $[0, \infty)$, maka fungsi distribusi kumulatif $F(t)$ untuk distribusi kontinu dengan fungsi densitas peluang $f(t)$ yaitu:

$$F(t) = P(T \leq t) = \int_0^t f(t) dt \quad (1)$$

Fungsi *survival* $S(t)$ didefinisikan sebagai probabilitas seorang individu bertahan lebih besar dari waktu t dengan ($t > 0$) yaitu:

$$S(t) = 1 - P(T \leq t) = 1 - F(t) \quad (2)$$

Fungsi *hazard* $h(t)$ yang juga dikenal sebagai 'laju kegagalan bersyarat' (*conditional failure rate*) atau 'kekuatan mortalitas' (*force of mortality*) adalah probabilitas terjadinya suatu kejadian pada interval waktu antara t dan $t + \Delta t$ dimana waktu *survival* T adalah lebih besar atau sama dengan t . Fungsi *hazard* dinyatakan sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)}, \quad S(t) \neq 0 \quad (3)$$

DISTRIBUSI EKSPONENSIAL

Jika t adalah waktu *survival* dari variabel acak kontinu T yang mengikuti distribusi eksponensial dengan parameter θ , fungsi kepadatan peluang distribusi Eksponensial adalah [5]:

$$f(t) = \begin{cases} \theta e^{-\theta t} & t \geq 0, \theta > 0 \\ 0 & t < 0 \end{cases}$$

Dengan $E(T) = \frac{1}{\theta}$ dan $Var(T) = \frac{1}{\theta^2}$

Fungsi distribusi kumulatif untuk distribusi Eksponensial adalah:

$$F(t; \theta) = 1 - e^{-\theta t}$$

Fungsi *survival* dari distribusi eksponensial adalah:

$$S(t; \theta) = e^{-\theta t}$$

Sehingga fungsi *hazard* adalah:

$$h(t; \theta) = \theta$$

METODE BAYESIAN

Parameter yang digunakan pada metode Bayes merupakan variabel acak yang mempunyai distribusi tertentu (distribusi prior). Pada metode Bayesian sendiri dikenal pula distribusi posterior, dimana distribusi posterior didapat dari fungsi *likelihood* yang digabungkan dengan distribusi prior. Sehingga dalam metode Bayesian ini diperlukan fungsi *likelihood*, distribusi prior dan distribusi posterior untuk mendapatkan parameter. Fungsi *likelihood* merupakan fungsi densitas bersama dari n variabel random X_1, X_2, \dots, X_n dan dinyatakan dalam bentuk $f(x_1, x_2, \dots, x_n; \theta)$. Jika x_1, x_2, \dots, x_n ditetapkan, maka fungsi *likelihood* adalah fungsi dari parameter θ yang dinotasikan dengan $L(\theta)$ [6]. Jika X_1, X_2, \dots, X_n menyatakan suatu sampel random dari $f(x; \theta)$, maka:

$$\begin{aligned} L(\theta) &= f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) \\ &= \prod_{i=1}^n f(x_i; \theta) \end{aligned}$$

Sehingga fungsi *likelihood* untuk distribusi Eksponensial adalah sebagai berikut:

$$\begin{aligned} L(t_i; \theta) &= \prod_{i=1}^n [f(t_i; \theta)] \\ &= \theta e^{-\theta t_1} \cdot \theta e^{-\theta t_2} \dots \theta e^{-\theta t_n} \\ &= \theta^n e^{-\theta \sum_{i=1}^n t_i} \end{aligned} \tag{4}$$

Dalam metode Bayesian, ketika suatu populasi mengikuti distribusi tertentu dengan suatu parameter didalamnya (dalam hal ini θ), maka parameter θ mengikuti suatu distribusi peluang yang disebut distribusi prior. Dalam kasus ini, distribusi prior *Jeffreys* ditetapkan sebagai prior non-informatif, sehingga distribusi prior *Jeffreys* ialah [7]:

$$f(\theta)_J = k \frac{\sqrt{n}}{\theta}, \quad \theta \neq 0 \tag{5}$$

Setelah informasi sampel diambil dan prior telah ditentukan maka distribusi posteriornya dicari dengan menggabungkan prior dengan fungsi *likelihood*. Distribusi posterior dinyatakan sebagai berikut [8]:

$$f(\theta|t_i) = \frac{f(\theta) f(t_i|\theta)}{\int_{-\infty}^{\infty} f(\theta) f(t_i|\theta) d\theta}, \quad \theta > 0, t_i > 0$$

Fungsi kepadatan peluang $f(\theta|t_i)$ dan $f(\theta)$ masing-masing menunjukkan distribusi posterior dan distribusi prior, sedangkan $f(t_i|\theta)$ menunjukkan fungsi *likelihood*. Berdasarkan Persamaan (4) dan

(5) distribusi posterior untuk model *survival* berdistribusi Eksponensial menggunakan prior *Jeffreys* adalah:

$$\begin{aligned}
 f(\theta|t_i) &= \frac{\frac{k\sqrt{n}}{\theta} \theta^n e^{-\theta \sum_{i=1}^n t_i}}{\int_0^{\infty} \frac{k\sqrt{n}}{\theta} \theta^n e^{-\theta \sum_{i=1}^n t_i} d\theta} \\
 &= \frac{k\sqrt{n} \theta^{n-1} e^{-\theta \sum_{i=1}^n t_i}}{k\sqrt{n} \int_0^{\infty} \theta^{n-1} e^{-\theta \sum_{i=1}^n t_i} d\theta} \\
 f(\theta|t_i) &= \frac{\theta^{n-1} \left(\sum_{i=1}^n t_i \right)^n e^{-\theta \sum_{i=1}^n t_i}}{\Gamma(n)} \quad (6)
 \end{aligned}$$

METODE BAYESIAN SELF

Estimasi parameter yang digunakan dalam penelitian ini menggunakan *Symmetric Loss Function* yang dikenal sebagai *Squared Error Loss Function* (SELF) yang didefinisikan sebagai berikut:

$$\mathcal{L}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \quad 0 < \theta < \infty$$

dengan $\hat{\theta}$ merupakan estimator Bayesian SELF untuk parameter θ . Estimasi Bayesian SELF dari θ pada distribusi Eksponensial diperoleh dengan meminimumkan ekspektasi *loss function* yang diperoleh sebagai berikut:

$$\hat{\theta} = E(\theta) \quad (7)$$

Sehingga berdasarkan Persamaan (7) maka estimasi parameter dengan metode Bayesian SELF adalah:

$$\begin{aligned}
 \theta_{BS} &= \int_0^{\infty} \theta f(\theta|t_i) d\theta \\
 &= \int_0^{\infty} \theta \frac{\theta^{n-1} \left(\sum_{i=1}^n t_i \right)^n e^{-\theta \sum_{i=1}^n t_i}}{\Gamma(n)} d\theta \\
 &= \frac{1}{\Gamma(n)} \left(\sum_{i=1}^n t_i \right)^n \int_0^{\infty} \theta^n e^{-\theta \sum_{i=1}^n t_i} d\theta \\
 &= \frac{1}{\Gamma(n)} \left(\sum_{i=1}^n t_i \right)^n \int_0^{\infty} \left(\frac{u}{\sum_{i=1}^n t_i} \right)^{n-1+1} e^{-u} \left(\frac{du}{\sum_{i=1}^n t_i} \right) \\
 &= \frac{1}{\sum_{i=1}^n t_i}
 \end{aligned}$$

Estimasi parameter fungsi *survival* dan fungsi *hazard* dengan metode Bayesian SELF dari distribusi Eksponensial prior *Jeffreys* diperoleh:

$$S_{BS} = e^{-\theta_{BS} t_i} = e^{-\left(\frac{1}{\sum_{i=1}^n t_i}\right) t_i} \tag{8}$$

$$h_{BS} = \theta_{BS} = \frac{1}{\sum_{i=1}^n t_i} \tag{9}$$

METODE BAYESIAN GELF

Estimasi parameter dalam penelitian ini juga menggunakan *General Entropy Loss Function* (GELF) yang didefinisikan sebagai berikut:

$$\mathcal{L}(\theta, \theta) = \left(\frac{\theta}{\theta}\right)^k - k \ln\left(\frac{\theta}{\theta}\right) - 1$$

Untuk $k \neq 0$, $0 < \theta < \infty$ dengan k menunjukkan penyimpangan asimetri dan θ merupakan estimator Bayesian GELF untuk parameter θ .

Estimasi Bayesian GELF dari θ pada distribusi Eksponensial diperoleh dengan meminimumkan ekspetasi *Loss Function* sehingga diperoleh sebagai berikut:

$$\theta = \left[E(\theta^{-k}) \right]^{-\frac{1}{k}} \tag{10}$$

Sehingga berdasarkan Persamaan (10) maka metode estimasi parameter dengan metode Bayesian GELF adalah:

$$\begin{aligned} \theta_{BG} &= \left[E(\theta^{-k}) \right]^{-\frac{1}{k}} \\ E(\theta^{-k}) &= \int_0^\infty \theta^{-k} \frac{\theta^{n-1} \left(\sum_{i=1}^n t_i\right)^n e^{-\theta \sum_{i=1}^n t_i}}{\Gamma(n)} d\theta \\ &= \frac{1}{\Gamma(n)} \left(\sum_{i=1}^n t_i\right)^n \int_0^\infty \theta^{n-k-1} e^{-\theta \sum_{i=1}^n t_i} d\theta \\ &= \frac{1}{\Gamma(n)} \frac{\left(\sum_{i=1}^n t_i\right)^{n-1}}{\left(\sum_{i=1}^n t_i\right)^{n-k-1}} \Gamma(n-k) \\ &= \frac{1}{\Gamma(n)} \left(\sum_{i=1}^n t_i\right)^k \Gamma(n-k) \\ \left[E(\theta^{-k}) \right]^{-\frac{1}{k}} &= \left(\frac{1}{\Gamma(n)} \left(\sum_{i=1}^n t_i\right)^k \Gamma(n-k) \right)^{-\frac{1}{k}} \end{aligned}$$

$$= \frac{(\Gamma(n-k))^{\frac{1}{k}}}{\left(\sum_{i=1}^n t_i\right) \Gamma(n)^{\frac{1}{k}}}$$

Sehingga diperoleh θ_{BG} adalah:

$$\theta_{BG} = \frac{(\Gamma(n-k))^{\frac{1}{k}}}{\left(\sum_{i=1}^n t_i\right) \Gamma(n)^{\frac{1}{k}}} \quad (11)$$

Estimasi parameter fungsi *survival* dan fungsi *hazard* dengan metode Bayesian GELF dari distribusi Eksponensial prior *Jeffreys* diperoleh:

$$S_{BG} = e^{-\theta_{BG} t_i} = e^{-\left[\frac{(\Gamma(n-k))^{\frac{1}{k}}}{\left(\sum_{i=1}^n t_i\right) \Gamma(n)^{\frac{1}{k}}} \right] t_i} \quad (12)$$

$$h_{BG} = \theta_{BG} = \frac{(\Gamma(n-k))^{\frac{1}{k}}}{\left(\sum_{i=1}^n t_i\right) \Gamma(n)^{\frac{1}{k}}} \quad (13)$$

STUDI KASUS

Estimasi parameter yang diperoleh dari metode Bayesian SELF dan Bayesian GELF prior *Jeffreys* dibandingkan menggunakan nilai *Mean Square Error* (MSE) dengan menggunakan data 137 pasien penderita kanker paru-paru yang diambil dari Program R versi 3.3.0. MSE dari estimator untuk fungsi *survival* dan fungsi *hazard* didefinisikan sebagai berikut:

$$MSE(S) = E\left[(S - \hat{S})^2\right] \quad (14)$$

$$MSE(h) = E\left[(h - \hat{h})^2\right] \quad (15)$$

Dari data kanker paru-paru diperoleh $E(T) = 121,627$, $\sum_{i=1}^n t_i = 16.663$, $n = 137$, $k = 1$, $S = 0,48905$, $\theta_{BS} = 6,00132 \times 10^{-5}$ dan $\theta_{BG} = 0,008161796$, dengan $S_{BS}(t_i; \theta_{BS})$ adalah fungsi *survival* dari metode Bayesian SELF, $h_{BS}(t_i; \theta_{BS})$ adalah fungsi *hazard* dari metode Bayesian SELF, $S_{BG}(t_i; \theta_{BG})$ adalah fungsi *survival* dari metode Bayesian GELF dan $h_{BG}(t_i; \theta_{BG})$ adalah fungsi *hazard* dari metode Bayesian GELF, maka diperoleh:

$$S_{BS}(t_i; \theta_{BS}) = e^{-\left(\frac{1}{16663}\right) t_i}$$

$$h_{BS}(t_i; \theta_{BS}) = \frac{1}{16663} = 6,00132 \times 10^{-5}$$

$$S_{BG}(t_i; \theta_{BG}) = e^{-\theta_{BG} t_i} = e^{-\left[\frac{(\Gamma(137-1))^{\frac{1}{1}}}{(16663) \Gamma(137)^{\frac{1}{1}}} \right] t_i}$$

$$h_{BG}(t_i; \theta_{BG}) = \theta_{BG} = \frac{(\Gamma(137-1))^{\frac{1}{1}}}{(16663)\Gamma(137)^{\frac{1}{1}}} = 0,008161796$$

Selanjutnya nilai MSE yang dihitung menggunakan Persamaan (14) dan (15) diperoleh dalam tabel berikut:

Tabel 1 Hasil perbandingan metode Bayesian SELF dan metode Bayesian GELF

Metode Estimasi	<i>Mean Square Error (MSE)</i>	
	<i>Fungsi Survival</i>	<i>Fungsi Hazard</i>
Bayesian SELF	0,293363538	$6,66149 \times 10^{-5}$
Bayesian GELF	$3,60504 \times 10^{-6}$	$3,60158 \times 10^{-9}$

Dari Tabel 1 terlihat bahwa untuk fungsi *survival* dan fungsi *hazard* metode Bayesian GELF memiliki nilai MSE lebih kecil dari pada metode Bayesian SELF. Berdasarkan nilai MSE hasil estimasi untuk parameter θ diperoleh metode Bayesian GELF lebih baik dari pada metode Bayesian SELF yang digunakan untuk mengestimasi model *survival* berdistribusi Eksponensial.

Dari hasil estimasi dengan metode Bayesian GELF, kemudian diambil sebarang nilai t_i menggunakan teknik pengambilan sampel acak sederhana pada data penderita kanker paru-paru. Diperoleh $t_1 = 7, t_2 = 63, t_3 = 95, t_4 = 162,$ dan $t_5 = 999$ maka dapat diketahui peluang individu dapat bertahan hidup sebagai berikut:

Tabel 2 Hasil Estimasi Parameter Fungsi Survival Distribusi Eksponensial dengan Metode Bayesian GELF

No	Waktu	<i>Survival BG</i> $\left(\hat{S}_{BG} \right)$
1	7	0,94447
2	63	0,59573
3	95	0,46053
4	162	0,26655
5	999	0,00029

Berdasarkan Tabel 2 diketahui peluang seorang pasien pada kasus ini untuk bertahan hidup jika mengidap penyakit kanker paru-paru selama 7 hari adalah 0,94447, selama 63 hari adalah 0,59573, selama 95 hari adalah 0,46053, selama 162 hari adalah 0,26655, dan selama 999 hari adalah 0,00029.

PENUTUP

Berdasarkan nilai *Mean Square Error (MSE)* yang diperoleh dari metode Bayesian SELF untuk fungsi *survival* dan fungsi *hazard* ialah 0,293363538 dan $6,66149 \times 10^{-5}$. Sedangkan nilai MSE yang diperoleh dari metode Bayesian GELF untuk fungsi *survival* dan fungsi *hazard* ialah $3,60504 \times 10^{-6}$ dan $3,60158 \times 10^{-9}$. Dapat dilihat nilai MSE yang diperoleh dari metode Bayesian GELF lebih kecil dibandingkan dengan nilai MSE yang diperoleh dari metode Bayesian SELF, sehingga untuk estimasi parameter model *survival* distribusi Eksponensial metode Bayesian GELF lebih baik dari pada metode Bayesian SELF. Berdasarkan hasil estimasi metode Bayesian GELF untuk studi kasus penderita kanker paru-paru dapat diketahui peluang seorang pasien untuk bertahan hidup jika mengidap penyakit kanker paru-paru selama 7 hari adalah 0,94447, selama 63 hari adalah 0,59573, selama 95 hari adalah 0,46053, selama 162 hari adalah 0,26655, dan selama 999 hari adalah 0,00029.

DAFTAR PUSTAKA

- [1]. Lawless JF. *Statistical Methods for Survival Data Analysis*. New York: John Wiley & Sons; 1982.
- [2]. Berger JO. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer; 1985.
- [3]. Guure CB & Ibrahim NA. *Bayesian Analysis of the Survival Function and Failure Rate of Weibull Distribution with Censored Data*. Malaysia: Journal Mathematical Problem in Engineering; 2012.
- [4]. Lawless JF. *Statistical Model and Methods for Lifetime Data*. New Jersey: Jhon Wiley and Sons; 2003.
- [5]. Lee ET & Wang JW. *Stastitical Methods for Survival Data Analysis*. Canada: A Jhon Wiley & Sonc, Inc;2003.
- [6]. Bain LJ dan Engelhardt M. *Introduction to Probability and Mathematical Statistics*. California: Duxbury Press; 1992.
- [7]. Al-Kutubi HS & Ibrahim NA. *Bayes Estimator for Exponential Distribution with Extension of Jeffreys Prior Information*. Malaysia: Malaysian Journal of Mathematical Sciences; 2009.
- [8]. Soejati Z & Soebanar. *Inferensi Bayesian*. Jakarta: Karunika Untiversitas Terbuka; 1988.

DEWI PUSPITAWATI : Jurusan Matematika FMIPA UNTAN, Pontianak,
puspitawati567@gmail.com

SETYO WIRA RIZKI : Jurusan Matematika FMIPA UNTAN, Pontianak,
setyo.wirarizki@math.untan.ac.id

NURFITRI IMRO'AH : Jurusan Matematika FMIPA UNTAN, Pontianak,
nurfitriimroah18@gmail.com

