METODE BAYESIAN SELF UNTUK ESTIMASI PARAMETER MODEL SURVIVAL DISTRIBUSI RAYLEIGH

Asty Fistia Ningrum, Neva Satyahadewi, Setyo Wira Rizki

INTISARI

Data survival adalah data yang menunjukkan waktu suatu objek atau individu dapat bertahan selama dilakukan penelitian. Penelitian ini membahas mengenai metode Bayesian SELF untuk estimasi parameter model survival distribusi Rayleigh. Proses estimasi parameter metode Bayesian SELF memerlukan informasi dari fungsi likelihood dan distribusi prior yang kemudian akan membentuk distribusi posterior. Setelah diperoleh estimator dari metode tersebut, selanjutnya diterapkan pada data pasien penderita kanker ovarium yang telah diberi perawatan kemoterapi berdistribusi Rayleigh yang diambil dari program R. Berdasarkan hasil estimasi parameter metode Bayesian SELF dengan prior Vague pada studi kasus penderita kanker ovarium menghasilkan hasil estimasi parameter nilai survival yang kurang baik untuk distribusi Rayleigh meskipun hasil estimasi nilai survival penderita kanker ovariumnya meningkat. Untuk hasil estimasi parameter model survival distribusi Rayleigh yang lebih baik, dapat menggunakan prior lainnya seperti prior Uniform.

Kata Kunci: Kanker Ovarium, Vague, Posterior.

PENDAHULUAN

Analisis *survival* adalah sekumpulan prosedur statistika untuk menganalisis data dengan variabel yang diperhatikan waktu sampai terjadinya suatu *event* [1]. Waktu yang diukur pada analisis *survival* disebut waktu *survival*, karena menunjukkan waktu individu dapat bertahan (*survive*) selama dilakukannya penelitian. Tiga faktor yang dibutuhkan dalam menentukan waktu *survival*, yaitu waktu awal pencatatan (*star point*) yang terdefinisi dengan baik, waktu akhir pencatatan (*end point*) yang terdefinisi dengan baik, dan skala pengukuran sebagai bagian dari waktu yang harus jelas misalnya minggu, hari, atau jam [2].

Terdapat dua model yang digunakan dalam analisis *survival*, yaitu model parametrik dan non-parametrik. Model parametrik adalah suatu model *survival* dengan data *survival* yang mengikuti asumsi distribusi tertentu. Beberapa model parametrik terdiri dari distribusi eksponensial, distribusi Weibull distribusi log-normal, distribusi log-logistik, dan distribusi gamma. Jika distribusi yang mendasari data tidak diketahui, maka data tidak mengikuti suatu distribusi tertentu yang sudah ada sehingga digunakan model non-parametrik. Pada model non-parametrik terdapat dua metode yang terkenal, yaitu metode Kaplan-Meier dan metode Nelson-Aalen.

Pada analisis *survival*, terdapat salah satu distribusi yang ada dalam analisis data uji hidup yaitu distribusi Rayleigh. Distribusi Rayleigh adalah salah satu bentuk khusus dari distribusi Weibull yang pertama kali dikenalkan oleh Lord Rayleigh pada tahun 1880 [3]. Terdapat dua pendekatan yang dikenal untuk mengestimasi parameter, yaitu pendekatan klasik dan pendekatan Bayesian. Pendekatan klasik yaitu metode *maximum likelihood estimation* (MLE). Sedangkan pada pendekatan Bayesian terdapat beberapa metode yaitu *linear exponential loss function* (LINEX), *Lindley approximation*, *general entropy loss function* (GELF), dan *squared error loss function* (SELF) [4]. Pendekatan Bayesian dapat digunakan sebagai alat pengambilan keputusan untuk memperbaharui tingkat kepercayaan dari suatu informasi. Pada pendekatan Bayesian melibatkan suatu prior. Dalam menentukan prior merupakan hal yang cukup sulit karena didalam pemilihan prior tidak ada referensi yang mencantumkan prior aktual pada metode yang akan diteliti, sehingga para peneliti memilih prior yang mudah untuk digunakan. Fungsi *likelihood* yang digabungkan dengan distribusi prior akan

menghasilkan suatu distribusi baru yaitu distribusi posterior yang menjadi dasar untuk inferensi didalam metode Bayesian.

Penelitian ini bertujuan untuk menentukan estimasi parameter model *survival* berdistribusi Rayleigh dengan prior *Vague* menggunakan metode Bayesian SELF pada kasus penderita kanker ovarium. Data yang digunakan pada kasus penderita kanker ovarium telah diberikan perawatan yang diambil dari program R. Pada data pasien kanker ovarium variabel yang digunakan yaitu variabel waktu pasien yang masih hidup dengan lamanya waktu penelitian selama 1227 hari. Data yang sudah diperoleh kemudian diuji menggunakan *Kolmogorov-Smirnov* dengan bantuan *software easyfit* untuk mengetahui apakah data tersebut berdistribusi Rayleigh. Langkah selanjutnya adalah menentukan fungsi *likelihood*, yang kemudian dikombinasikan dengan distribusi prior untuk mendapatkan distribusi posterior. Distribusi posterior yang diperoleh digunakan untuk menentukan estimasi parameter metode Bayesian SELF, yang selanjutnya diterapkan pada data pasien kasus penderita kanker ovarium.

DISTRIBUSI WAKTU SURVIVAL

Distribusi waktu *survival* dinyatakan dengan tiga fungsi yaitu, fungsi kepadatan peluang, fungsi *survival*, dan fungsi *hazard*.

1. Fungsi kepadatan peluang (*probability density function*)

Probability density function (pdf) adalah peluang suatu individu meninggal atau mengalami kejadian sesaat dalam interval waktu t sampai $(t + \Delta t)$ yang dinotasikan dengan $f(t)$. Pdf dari $f(t)$ yaitu:

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

Jika T merupakan variabel random kontinu non-negatif dalam interval $[0, \infty)$, maka $F(t)$ merupakan fungsi distribusi kumulatif koninu dari T . Sehingga dapat didefinisikan sebagai berikut:

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

2. Fungsi *survival*

Fungsi *survival* $s(t)$ didefinisikan sebagai suatu peluang yang menyatakan bahwa suatu individu dapat bertahan hidup sampai dengan waktu t dengan $t \geq 0$ yaitu:

$$s(t) = 1 - F(t); t \geq 0 \quad (2)$$

3. Fungsi *hazard*

Fungsi *hazard* dari waktu *survival* T memberikan suatu tingkat kegagalan yang dinotasikan dengan $h(t)$. Fungsi *hazard* dinyatakan sebagai berikut:

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

DISTRIBUSI RAYLEIGH

Distribusi Rayleigh adalah salah satu bentuk khusus dari distribusi Weibull yang pertama kali dikenalkan oleh Lord Rayleigh pada tahun 1880 [3]. Fungsi kepadatan peluang distribusi Rayleigh satu parameter dinyatakan sebagai berikut:

$$f(t; \theta) = \frac{t}{\theta^2} \exp\left(-\frac{t^2}{2\theta^2}\right); t \geq 0, \theta > 0 \quad (4)$$

dimana θ merupakan parameter.

Berdasarkan fungsi kepadatan peluang dari distribusi Rayleigh, maka fungsi distribusi kumulatif untuk distribusi Rayleigh ialah:

$$F(t; \theta) = 1 - \exp\left(-\frac{t^2}{2\theta^2}\right); t \geq 0, \theta > 0$$

dengan menggunakan Persamaan (2) diperoleh fungsi *survival* dari distribusi Rayleigh sebagai berikut:

$$s(t; \theta) = \exp\left(-\frac{t^2}{2\theta^2}\right); t \geq 0, \theta > 0 \quad (5)$$

dari persamaan (4) dan (5) diperoleh fungsi *hazard* untuk distribusi Rayleigh sebagai berikut:

$$h(t; \theta) = \frac{t}{\theta^2}; t \geq 0, \theta > 0 \quad (6)$$

PEMBENTUKAN FUNGSI LIKELIHOOD PADA DISTRIBUSI RAYLEIGH

Fungsi *likelihood* $L(t_i; \theta)$ digunakan untuk mengestimasi parameter dengan mrnggabungkan fungsi *likelihood* yang didapat dengan distribusi prior untuk menghasilkan distribusi posterior. Fungsi *likelihood* dari distribusi Rayleigh dengan $\theta > 0$ memiliki bentuk sebagai berikut:

$$\begin{aligned} L(t_i; \theta) &= \prod_{i=1}^n [f(t_i; \theta)] \\ &= \prod_{i=1}^n \left[\frac{t_i}{\theta^2} e^{-\left(\frac{t_i^2}{2\theta^2}\right)} \right] \\ &= \left(\frac{\prod_{i=1}^n t_i}{\theta^{2n}} \right) \exp\left(-\left(\frac{1}{2\theta^2} \sum_{i=1}^n t_i^2\right)\right) \end{aligned} \quad (7)$$

ESTIMASI PARAMETER METODE BAYESIAN

Pada metode Bayesian, parameter dipandang sebagai variabel random yang memiliki distribusi yang disebut dengan distribusi prior [4]. Fungsi *likelihood* yang digabungkan dengan distribusi prior akan menghasilkan suatu distribusi baru yaitu distribusi posterior yang menyatakan tingkat keyakinan mengenai suatu parameter setelah sampel diamati. Estimasi parameter metode Bayesian meliputi:

1. Penentuan distribusi prior dan pembentukan distribusi posterior

Distribusi prior $p(\theta)$ merupakan tingkat kepercayaan peneliti terhadap setiap nilai parameter yang mungkin. Dalam Bayesian, parameter θ diperlukan sebagai variabel, sehingga θ mempunyai nilai dalam sebuah domain dengan fungsi densitas $f(\theta)$ dan inilah yang dinamakan dengan distribusi prior θ , dengan adanya informasi prior maka dapat dikombinasikan dengan data sampel yang digunakan untuk membentuk distribusi posterior $\pi^*(\theta|t_i)$. Dalam penelitian ini, prior yang digunakan adalah prior *vague* yang dinyatakan sebagai berikut:

$$p(\theta) = \frac{1}{\theta}, \theta > 0 \quad (8)$$

Sehingga dapat ditentukan distribusi posterior dengan menggunakan Persamaan (8), sebagai berikut:

$$\begin{aligned} \pi^*(\theta|t_i) &= \frac{p(\theta)L(t_i|\theta)}{\int_0^\infty p(\theta)L(t_i|\theta)d\theta} \\ &= \left(\frac{1}{\theta^{(2n+1)}}\right) \prod_{i=1}^n (t_i) \exp\left(-\left(\frac{1}{2\theta^2} \sum_{i=1}^n t_i^2\right)\right) \cdot \left(-\frac{\sum_{i=1}^n t_i^2}{\left(\sum_{i=1}^n t_i\right)(2)^{n-1} \Gamma(n)}\right) \\ &= -\left(\frac{\left(\sum_{i=1}^n t_i^2\right)^n \exp\left(-\left(\frac{1}{2\theta^2} \sum_{i=1}^n t_i^2\right)\right)}{(\theta)^{2n+1} (2)^{n-1} \Gamma(n)}\right) \end{aligned} \quad (9)$$

2. Estimasi parameter metode Bayesian SELF

Pada umumnya, *loss function* digunakan dalam estimasi parameter karena peluang suatu penaksir titik untuk tepat sekali sangat kecil. Estimasi parameter yang digunakan menggunakan *symmetric loss function* yang dikenal sebagai *squared error loss function* (SELF), didefinisikan sebagai berikut:

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

dengan $\hat{\theta}_{rs}$ merupakan estimator Bayesian SELF untuk parameter θ . Estimasi parameter metode Bayesian SELF dapat diperoleh dengan meminimumkan ekspektasi *loss function*, yaitu:

$$\begin{aligned} \frac{\partial [E(\mathcal{L}(\hat{\theta}_{rs}, \theta))]}{\partial \theta} &= 0 \\ \Leftrightarrow \frac{\partial [E((\hat{\theta}_{rs} - \theta)^2)]}{\partial \theta} &= 0 \end{aligned}$$

$$\Leftrightarrow \frac{\partial \left[E \left(\left(\hat{\theta}_{BS}^2 - 2\hat{\theta}_{BS}\theta + \theta^2 \right) \right) \right]}{\partial \theta} = 0$$

$$\Leftrightarrow E(2\theta - 2\hat{\theta}_{BS}) = 0$$

$$\Leftrightarrow E(\theta) - \hat{\theta}_{BS} = 0$$

dengan demikian diperoleh,

$$E(\theta) = \hat{\theta}_{BS} \quad (10)$$

berdasarkan Persamaan (10), dapat ditentukan estimasi parameter dengan metode Bayesian SELF, yaitu:

$$\begin{aligned} \hat{\theta}_{BS} = E(\theta) &= \int_0^{\infty} \theta \pi^*(\theta | t_i) d\theta \\ &= \int_0^{\infty} \theta \cdot \left(\frac{\left(\sum_{i=1}^n t_i^2 \right)^n \exp \left(- \left(\frac{1}{2\theta^2} \sum_{i=1}^n t_i^2 \right) \right)}{(\theta)^{2n+1} (2)^{n-1} \Gamma(n)} \right) d\theta \\ &= \frac{\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma \left(\frac{27}{2} \right)}{(2)^{\frac{1}{2}} \Gamma(14)} \end{aligned} \quad (11)$$

berdasarkan Persamaan (11), diperoleh estimasi fungsi *survival* dan fungsi *hazard* dengan metode Bayesian SELF dari distribusi Rayleigh dengan prior *vague* adalah:

$$\hat{s}_{BS}(t_i; \hat{\theta}_{BS}) = e^{-\frac{t_i^2}{2\hat{\theta}_{BS}^2}} = \exp \left(- \frac{\left((t_i)(2)^{\frac{1}{2}} \Gamma(14) \right)^2}{2 \left(\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma \left(\frac{27}{2} \right) \right)^2} \right) \quad (12)$$

$$\hat{h}_{BS}(t_i; \hat{\theta}_{BS}) = \frac{t_i}{\hat{\theta}_{BS}^2} = \frac{(t_i) \left((2)^{\frac{1}{2}} \Gamma(14) \right)^2}{\left(\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma \left(\frac{27}{2} \right) \right)^2} \quad (13)$$

STUDI KASUS

Studi kasus yang digunakan adalah data waktu *survival* pasien penderita kanker ovarium yang telah diberikan perawatan kemoterapi yang diambil dari program **R**. Pada data pasien kanker ovarium variabel yang digunakan yaitu variabel waktu pasien yang masih hidup dengan lamanya waktu penelitian selama 1227 hari. Agar data dapat digunakan maka data yang diperoleh harus diuji terlebih dahulu. Selanjutnya penerapan hasil estimasi parameter model *survival* berdistribusi Rayleigh menggunakan Bayesian SELF dengan menggunakan prior *vague* pada kasus penderita kanker ovarium, selengkapnya akan dibahas sebagai berikut:

1. Uji kecocokan data penderita kanker ovarium

Uji kecocokan model (*goodness of fit test*) merupakan uji yang digunakan untuk menentukan kesesuaian (kecocokan) antara distribusi sampel dengan distribusi tertentu Uji *Kolmogorov-Smirnov* merupakan salah satu uji yang dapat digunakan untuk memutuskan jika sampel yang digunakan berasal dari populasi distribusi tertentu [5]. Hipotesis untuk distribusi data pada penelitian ini adalah sebagai berikut:

H_0 : Data berdistribusi Rayleigh

H_1 : Data tidak berdistribusi Rayleigh

Taraf signifikansi $\alpha = 5\% = 0,05$. Apabila nilai $p\text{-value} < \alpha$, maka H_0 ditolak. Dari data yang didapat menggunakan *easyfit* dapat diketahui bahwa nilai $p\text{-value}$ untuk distribusi Rayleigh metode Bayesian SELF adalah 0,58271 sehingga dapat disimpulkan bahwa data tersebut berdistribusi Rayleigh.

2. Estimasi parameter metode Bayesian SELF data penderita kanker ovarium

Perhitungan nilai *survival* dan nilai *hazard* dari metode Bayesian SELF menggunakan Microsoft Excel dengan nilai parameter $\theta = 323,9899$ dan nilai estimasi parameter $\hat{\theta}_{BS} = 626,4252$ adalah sebagai berikut:

Tabel 1 Hasil perhitungan fungsi *survival* dan fungsi *hazard* dari metode Bayesian SELF

Waktu (t_i)	<i>Survival Rayleigh</i>	<i>Survival BS</i> ($\hat{s}_{BS}$)	<i>Hazard BS</i> ($\hat{h}_{BS}$)
377	0,5081383	0,83435338	0,00096073
421	0,4298795	0,79784787	0,00107286
448	0,3844231	0,77434833	0,00114166
477	0,3383120	0,74832836	0,00121557
744	0,0716005	0,49395753	0,00189598
769	0,0597955	0,47071537	0,00195969
770	0,0593588	0,46979322	0,00196224
803	0,0463556	0,43972569	0,00204633
855	0,0307432	0,39397845	0,00217885
1040	0,0057879	0,25204302	0,00265029
1106	0,0029480	0,21042508	0,00281849
1129	0,0023077	0,19708406	0,00287710
1206	0,0009800	0,15673222	0,00307332
1227	0,0007683	0,14685372	0,00312684

Dari hasil perhitungan fungsi *survival* dengan menggunakan metode Bayesian SELF perhitungan peluang hidup pada kasus penderita kanker ovarium menjadi lebih tinggi. dimana $\hat{s}_{BS}$ merupakan nilai peluang seseorang pasien penderita kanker ovarium dapat bertahan hidup, sedangkan nilai $\hat{h}_{BS}$

merupakan tingkat laju kematian pasien tersebut. Jika diambil sebarang nilai t_i pada kasus penderita kanker ovarium dengan $t_i=377$ maka diketahui peluang individu dapat bertahan hidup selama 377 hari fungsi *survival* Rayleigh adalah 50,81% dan untuk fungsi *survival* metode Bayesian SELF adalah 83,43%. Dari Tabel 1 dapat dibuat grafik sebagai berikut:

Gambar 1 Grafik fungsi *survival* Bayesian SELF

Berdasarkan Gambar 1 dapat dilihat bahwa nilai *survival* dari data pasien kasus penderita ovarium bergerak turun.

Gambar 2 Grafik fungsi *hazard* Bayesian SELF

Berdasarkan Gambar 2 dapat dilihat bahwa nilai *hazard* yang naik.

Berdasarkan Gambar 1 dan Gambar 2 dapat disimpulkan bahwa estimasi parameter distribusi Rayleigh dengan metode Bayesian SELF kurang baik untuk melakukan analisis data kasus pasien penderita kanker ovarium ini. Hal ini ditunjukkan dari grafik nilai *survival* dan nilai *hazard* yang sesuai dengan karakteristik dari fungsi *survival* dan fungsi *hazard*.

KESIMPULAN

1. Hasil estimasi parameter model *survival* berdistribusi Rayleigh menggunakan metode Bayesian SELF adalah sebagai berikut:

$$\hat{\theta}_{BS} = \frac{\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma\left(\frac{27}{2}\right)}{(2)^{\frac{1}{2}} \Gamma(14)}$$

$$\hat{s}_{BS}(t_i; \hat{\theta}_{BS}) = e^{-\frac{t_i^2}{2\hat{\theta}_{BS}^2}} = \exp \left(- \frac{\left((t_i)(2)^{\frac{1}{2}} \Gamma(14) \right)^2}{2 \left(\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma\left(\frac{27}{2}\right) \right)^2} \right)$$

$$\hat{h}_{BS}(t_i; \hat{\theta}_{BS}) = \frac{t_i}{\hat{\theta}_{BS}^2} = \frac{(t_i) \left((2)^{\frac{1}{2}} \Gamma(14) \right)^2}{\left(\left(\sum_{i=1}^n t_i^2 \right)^{\frac{1}{2}} \Gamma\left(\frac{27}{2}\right) \right)^2}$$

2. Berdasarkan nilai parameter $\theta = 323,9899$ dan nilai dari parameter $\hat{\theta}_{BS}$ untuk metode Bayesian SELF adalah 626,4252. Jika diambil sebarang nilai t_i pada kasus penderita kanker ovarium dengan $t_i=377$ maka diketahui peluang individu dapat bertahan hidup selama 377 hari fungsi *survival* Rayleigh adalah 50,81% dan untuk fungsi *survival* Bayesian SELF adalah 83,43%. Berdasarkan hasil estimasi parameter metode Bayesian SELF dengan prior Vague pada studi kasus penderita kanker ovarium menghasilkan hasil estimasi parameter nilai survival yang kurang baik untuk distribusi Rayleigh meskipun hasil estimasi nilai *survival* penderita kanker ovariumnya meningkat. Untuk hasil estimasi parameter model *survival* distribusi Rayleigh yang lebih baik, dapat menggunakan prior lainnya seperti prior Uniform [6].

DAFTAR PUSTAKA

- [1] Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. New York: Springer Science Business Media; 2005.
- [2] Fitria S, Helmi, Rizki SW. Estimasi Parameter Model *Survival* Distribusi Eksponensial Data Tersensor dengan Metode Maksimum *Likelihood* dan Bayesian SELF. *Buletin Ilmiah Math, Stat. dan Terapannya (Bimaster)*. 2016; 3:213-220.
- [3] Dey S, Dey T. Rayleigh Distribution Revisited via Extension of Jeffreys Prior Information and A New Loss Function. *Statistical Journal*. 9(3):213-226.
- [4] Guure CB, Ibrahim NA. Bayesian Analysis Survival Function and Failure Rate of Weibull Distribution with Censored Data. *Journal Mathematical Problem in Engineering*. 2012. Article ID 329489.
- [5] Bain LJ, Engelhardt. *Introduction to Probability and Mathematical Statistics*. California: Duxbury of Wathfor; 1992.
- [6] Wahyuni ER, Rizki SW, Perdana H. Estimasi Parameter Model *Survival* Distribusi Rayleigh Prior Uniform dengan Metode Bayesian SELF. *Buletin Ilmiah Math, Stat. dan Terapannya (Bimaster)*. 2019;8(2):357-362.

ASTY FISTIA NINGRUM :Jurusan Matematika FMIPA UNTAN, Pontianak,
astyfistia@gmail.com

NEVA SATYAHADEWI :Jurusan Matematika FMIPA UNTAN, Pontianak,
neva.satya@math.untan.ac.id

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