

## ESTIMASI PARAMETER MODEL *SURVIVAL* DISTRIBUSI PARETO -GAMMA DENGAN METODE BAYESIAN SELF

Bella Amalia, Shantika Martha, Setyo Wira Rizki

### INTISARI

Data *survival* adalah data yang menunjukkan waktu suatu individu atau objek dapat bertahan hidup hingga terjadinya suatu kegagalan atau kejadian tertentu. Data *survival* dikatakan tersensor apabila objek pada penelitian hilang atau sampai akhir penelitian objek tersebut belum mengalami kejadian tertentu. Tujuan dari penelitian ini adalah menentukan estimasi parameter model *survival* distribusi Pareto data tersensor dengan metode Bayesian SELF. Data yang digunakan adalah data sekunder pasien kanker paru-paru. Berdasarkan nilai estimasi metode Bayesian SELF untuk studi kasus penderita kanker paru-paru dapat diketahui bahwa dengan menggunakan metode Bayesian SELF perhitungan peluang hidup pada kasus penderita kanker paru-paru menjadi lebih tinggi. Berdasarkan nilai MAPE yang diperoleh dari fungsi *Survival* distribusi Pareto dengan pendekatan Bayesian SELF adalah sebesar 24,68%. Hal ini berarti bahwa metode Bayesian SELF memiliki kemampuan peramalan yang cukup dalam mengestimasi peluang hidup pasien penderita kanker paru-paru.

**Kata Kunci:** *Distribusi Pareto, Metode Bayesian SELF, MAPE.*

### PENDAHULUAN

Analisis *survival* adalah salah satu teknik statistika yang digunakan untuk menganalisis data yang bertujuan untuk mengetahui hasil dari variabel yang mempengaruhi suatu awal kejadian sampai akhir kejadian[1]. Pada analisis *survival* terdapat konsep penyensoran yaitu pengamatan tersensor dan pengamatan tidak tersensor. Pengamatan dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena objek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian objek tersebut belum mengalami kejadian tertentu sedangkan pengamatan dikatakan tidak tersensor apabila semua objek penelitian atau unit data yang diteliti mati atau gagal.

Terdapat dua model yang digunakan untuk menganalisis data *survival* yaitu model parametrik dan model nonparametrik. 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[2]. Jika distribusi yang mendasari data *survival* tidak diketahui, artinya data tidak mengikuti suatu distribusi tertentu yang sudah ada maka digunakan model non parametrik. Pada model nonparametrik terdapat dua metode yang terkenal yaitu metode Kaplan-Meier dan metode Nelson-Aelen.

Pada analisis *survival*, terdapat salah satu distribusi yang ada dalam analisis data uji hidup adalah distribusi Pareto yang cukup sering digunakan khususnya untuk *reability*. Salah satu metode yang banyak dikembangkan diantaranya adalah metode Bayes. Sebelum menarik sampel dari suatu populasi biasanya diperoleh informasi mengenai parameter yang akan diestimasi. Informasi ini digabungkan dengan informasi dari sampel digunakan untuk mengestimasi parameter populasi. Adapun distribusi prior yang digunakan dalam pengamatan ini adalah prior Gamma. Metode Bayesian merupakan metode estimasi yang menggabungkan distribusi prior dan fungsi *likelihood*[3]. Distribusi prior adalah

distribusi awal yang memberi informasi tentang suatu parameter. Fungsi *likelihood* yang digabung dengan distribusi prior akan menghasilkan suatu distribusi baru yaitu distribusi posterior yang menjadi dasar untuk inferensi di dalam metode Bayesian. Ada beberapa pendekatan dalam metode Bayesian yang dapat digunakan untuk mengestimasi parameter antara lain *General Non-informatif Prior*, *Lindley Aproximation*, *General Entropy Loss Function (GELF)*, dan *Squared Error Loss Function (SELF)*.

Penelitian ini bertujuan untuk menentukan estimasi parameter model *survival* berdistribusi Pareto dengan prior Gamma menggunakan metode Bayesian SELF pada kasus penderita kanker paru-paru. Data yang digunakan dalam penelitian ini adalah data kanker paru-paru yang diambil dari program R versi 3.3.0 dengan jumlah data sebanyak 71 orang. Distribusi yang digunakan untuk model *survival* data tersensor adalah distribusi Pareto. Langkah pertama yang dilakukan pada data kanker paru-paru dimulai dengan menguji kecocokan model untuk mengetahui apakah data yang digunakan berdistribusi Pareto atau tidak dengan uji *Kolmogorov-Smirnov*, dari distribusi pareto ditentukan fungsi kepadatan peluang fungsi distribusi kumulatif, fungsi *survival* dan fungsi *hazard*. Selanjutnya menentukan fungsi *likelihood* dari fungsi kepadatan peluang dan fungsi *survival* kemudian menentukan distribusi prior dari distribusi prior gamma. Setelah memperoleh fungsi *likelihood* dan distribusi prior langkah berikutnya adalah merumuskan distribusi posterior. Setelah mendapatkan nilai parameter dari distribusi posterior, dilakukan pendekatan Bayesian SELF. Langkah terakhir dalam penelitian ini adalah menginterpretasikan data pasien penderita kanker paru-paru dari nilai yang diperoleh.

## DISTRIBUSI SURVIVAL

Distribusi *survival* digunakan untuk mengestimasi distribusi waktu *survival*. Distribusi *survival* biasanya ditandai dengan tiga fungsi yaitu: fungsi kepadatan peluang, fungsi *survival*, dan fungsi *hazard*. Hal ini berarti jika salah satu dari persamaan fungsi diketahui maka kedua fungsi lainnya dapat ditentukan[4].

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

Fungsi kepadatan peluang (*probabilty density function*) adalah peluang suatu individu mati atau mengalami kejadian seane dalam interval waktu  $t$  sampai  $(t + \Delta t)$ . Fungsi kepadatan peluang  $f(t)$  dirumuskan sebagai berikut:

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

Jika  $T$  merupakan variabel random non negatif pada interval  $[0, \infty]$ , maka  $F(t)$  merupakan fungsi distribusi kumulatif kontinu dari  $T$ . Didefinisikan sebagai peluang suatu individu mengalami kejadian kurang dari sama dengan waktu  $t$ , yaitu:

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

### 2. Fungsi *Survival*

Fungsi *Survival*  $S(t)$  didefinisikan sebagai peluang suatu individu dapat bertahan hidup waktu *survival* sampai dengan waktu  $t$  dengan  $(t > 0)$  yaitu sebagai berikut:

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

### 3. Fungsi Hazard

Fungsi *hazard* adalah probabilitas kematian selama interval waktu  $((t, t + \Delta t)$  dengan asumsi individu tetap hidup pada interval tersebut dan biasanya dinotasikan dengan  $h(t)$ . Fungsi *hazard* dapat dinyatakan sebagai berikut:

$$h(t) = \lim_{n \rightarrow \infty} \left[ \frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t} \right]$$

Fungsi *hazard* juga dapat dinyatakan dalam distribusi fungsi kumulatif  $F(t)$  dan fungsi kepadatan peluang  $f(t)$  sebagai berikut:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (3)$$

#### DISTRIBUSI PARETO

Distribusi Pareto berasal dari nama seorang ekonom yaitu Vilfredo Pareto yang mengatakan bahwa 80% kekayaan di Milan dimiliki oleh hanya 20% dari penduduknya. Distribusi Pareto disebut juga dengan *power-law*. Jika sebuah kumpulan data memiliki distribusi pareto, maka dikatakan bahwa data-data tersebut tidak sensitif terhadap rata-rata atau standar deviasi dari data tersebut atau dengan kata lain, data itu tidak bersifat acak[5]. Fungsi kepadatan peluang dari Distribusi Pareto dapat dinyatakan sebagai berikut:

$$f(t) = \begin{cases} \frac{\alpha \theta^\alpha}{t^{\alpha+1}}; & t \geq \theta \\ 0; & t < \theta \end{cases} \quad (4)$$

Fungsi distribusi kumulatif untuk distribusi Pareto ialah:

$$F(t) = \begin{cases} 1 - \left(\frac{\theta}{t}\right)^\alpha; & t \geq \theta \\ 0; & t < \theta \end{cases} \quad (5)$$

Fungsi *survival* dari distribusi Pareto sebagai berikut:

$$S(t) = \left(\frac{\theta}{t}\right)^\alpha \quad (6)$$

Sehingga fungsi *hazard* ialah:

$$h(t) = \frac{\alpha}{t} \quad (7)$$

#### ANALISIS METODE BAYESIAN

Pada Metode Bayesian, parameter dipandang sebagai variabel random yang memiliki distribusi yang disebut distribusi prior. Setelah itu dari distribusi prior selanjutnya dapat ditentukan distribusi posteriornya yang merupakan gabungan dari fungsi *likelihood* dan distribusi prior. Hasil yang dinyatakan dalam bentuk posterior yang kemudian menjadi dasar dalam memperoleh estimasi Bayesian. Berikut langkah-langkah yang dilakukan dalam analisis metode Bayesian, yaitu:

### 1. Menentukan fungsi *likelihood* untuk data tersensor

Fungsi *likelihood* pada data tersensor dari data pengamatan  $(t_i, \delta_i)$ ,  $i = 1, 2, \dots, n$  adalah:

$$L(\theta) = \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i} \quad (8)$$

dengan  $\delta_i$  adalah indikator penyensoran, bernilai 1 jika data tersensor dan bernilai 0 jika data tidak tersensor. Nilai  $t_i$  diperoleh dari min  $(T_i, C_i)$   $i = 1, 2, \dots, n$  dengan  $T_i$  adalah waktu hidup individu ke  $i$  dengan  $i = 1, 2, \dots, n$ , dan  $C_i$  adalah waktu penyensoran individu ke  $i$  dengan  $i = 1, 2, \dots, n$ . Sehingga fungsi *Likelihood* dari distribusi Pareto untuk data tersensor memiliki bentuk sebagai berikut:

$$\begin{aligned} L(t_i; \alpha; \theta; \delta_i) &= \prod_{i=1}^n \left[ \frac{\alpha \theta^\alpha}{t^{\alpha+1}} \right]^{\delta_i} \left[ \left( \frac{\theta}{t} \right)^\alpha \right]^{1-\delta_i} \\ &= \frac{\theta^{n\alpha} \alpha^{\sum_{i=1}^n \delta_i}}{\prod_{i=1}^n (t_i)^{\delta_i + \alpha}} \end{aligned} \quad (9)$$

Sehingga diperoleh estimasi fungsi *survival* ( $\hat{s}_{BS}$ ) dan fungsi *hazard* ( $\hat{h}_{BS}$ ) dari distribusi Pareto pada data tersensor adalah:

$$\hat{s}_{BS} = (t_i; \hat{\theta}_{BS}) = \left( \frac{A + n\alpha}{t} \right)^\alpha \quad (10)$$

$$\hat{h}_{BS} = (t_i; \hat{\theta}_{BS}) = \frac{\alpha}{t} \quad (11)$$

### 2. Membentuk distribusi prior

Metode Bayes memberikan metode untuk memilih keyakinan terhadap suatu parameter dari sebuah distribusi jika data diperoleh, ketika populasi mengikuti distribusi tertentu dengan suatu parameter didalamnya ( dalam kasus ini  $\alpha$  ), maka parameter  $\alpha$  mengikuti suatu distribusi peluang yang disebut distribusi prior. Distribusi prior salah satunya adalah Prior sekawan, dalam penelitian ini prior sekawan menggunakan distribusi Gamma sebagai berikut:

$$f(\theta) = \frac{B^A}{\Gamma(A)} \theta^{A-1} e^{-B\theta} \quad (12)$$

### 3. Membentuk distribusi posterior

Dalam estimasi Bayesian, setelah informasi sampel diambil dan prior telah ditentukan maka distribusi posteriornya dicari dengan menggabungkan priornya dengan informasi sampel yang diperoleh dari fungsi *likelihood*, dimana prior ini independen terhadap fungsi *likelihood*[6]. Distribusi posterior dinyatakan sebagai berikut:

$$f(\theta|t_i) = \frac{f(\theta) \cdot f(t_i|\theta)}{\int_0^\infty f(\theta) f(t_i|\theta) d\theta} \quad (13)$$

dengan  $f(\theta|t_i)$  menunjukkan distribusi posterior,  $f(\theta)$  merupakan distribusi prior, dan fungsi *likelihood*. Berdasarkan Persamaan (9) dan (12), maka fungsi kepadatan peluang distribusi posterior sebagai berikut:

$$f(\theta|t_i) = \frac{\theta^{A+n\alpha-1} e^{-B\theta} B^{A+n\alpha}}{\Gamma(A+n\alpha)} \quad (14)$$

### ESTIMASI PARAMETER BAYESIAN SELF

Estimasi parameter yang digunakan menggunakan *Squared Error Loss Function* (SELF) didefinisikan sebagai berikut:

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

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

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

$$E(\theta) = \int_0^{\infty} \theta f(\theta|t_i) d\theta$$

dengan demikian diperoleh,

$$\hat{\theta} = E(\theta) \quad (15)$$

Sehingga berdasarkan Persamaan (15), estimasi parameter dengan metode Bayesian SELF adalah sebagai berikut:

$$\hat{\theta}_{BS} = A + n\alpha \quad (16)$$

### STUDI KASUS

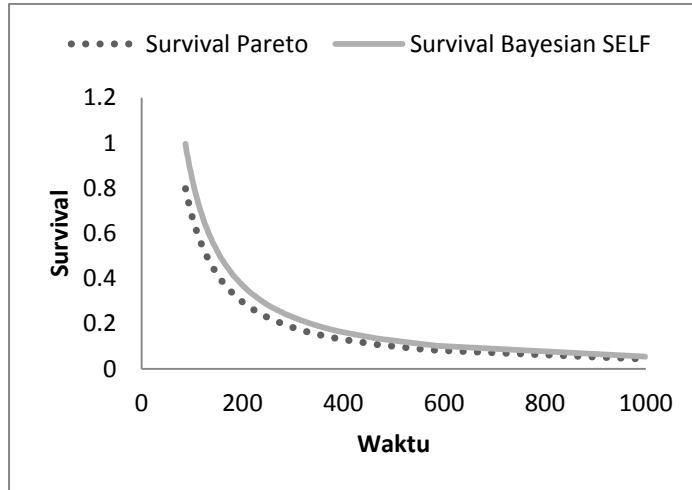
Data yang digunakan adalah data waktu survival pasien penderita kanker paru-paru sebanyak 71. Data tersebut merupakan data sekunder dari program R. Penelitian ini bertujuan untuk menentukan estimasi parameter model *survival* berdistribusi Pareto prior Gamma dengan metode Bayesian SELF dan menerapkan metode Bayesian SELF berdistribusi Pareto dengan menggunakan prior Gamma pada kasus penderita kanker paru-paru. Dari data 71 pasien penderita kanker paru-paru, selanjutnya dilakukan uji distribusi data untuk mengetahui data berdistribusi Pareto atau tidak. Selanjutnya estimasi parameter metode Bayesian SELF untuk penderita kanker paru-paru. Pemaparan selengkapnya adalah sebagai berikut:

#### 1. Uji distribusi data

Uji kecocokan (*Goodness of Fit*) digunakan untuk mengetahui ada atau tidaknya kesesuaian model sebaran yang diasumsikan atau apakah satu variabel bisa didekati dengan menggunakan distribusi atau tidak. Taraf nyata (*Sig*) yang digunakan adalah  $\alpha = 5\% = 0,05$ . Dalam menentukan keputusan akhir untuk menolak atau menerima  $H_0$  didasarkan pada wilayah kritis  $\alpha$  dengan nilai *p-value* yang mendukung suatu uji dalam bentuk peluang, jika nilai *p-value*  $\leq \alpha$  maka  $H_0$  ditolak  $H_1$  diterima. Dari data yang didapat menggunakan EasyFit dapat diketahui bahwa nilai *p-value* untuk distribusi Pareto Bayesian SELF adalah 0,19547 sehingga dapat disimpulkan bahwa data tersebut berdistribusi Pareto.

#### 2. Estimasi parameter metode Bayesian SELF

Dari data kasus pasien penderita kanker paru-paru, dengan menggunakan program R diambil nilai  $\alpha = 1,18765$ ,  $\theta = 72$ ,  $A = 2,37338$ ,  $B = 88,21829$ . Sehingga diperoleh grafik fungsi *survival* dengan metode Bayesian SELF sebagai berikut:



**Gambar 1 Fungsi *survival* Pareto dan Fungsi *survival* dengan Metode Bayesian SELF**

Dari Gambar 1 terlihat bahwa grafik fungsi *survival* Pareto semakin lama semakin turun mendekati 0 dan peluang seseorang untuk bertahan hidup lebih kecil, sedangkan dari grafik fungsi *survival* dengan metode Bayesian SELF juga semakin lama semakin turun mendekati 0 tetapi memiliki perbedaan yaitu dilihat dari hasil estimasi, jika diambil sebarang nilai pada data kasus penderita kanker paru-paru untuk Gambar 1 dengan  $t_1 = 87$  dan  $t_2 = 126$ ,  $t_3 = 200$ ,  $t_4 = 384$   $t_5 = 999$  maka dapat diketahui peluang individu dapat bertahan hidup sebagai berikut:

**Tabel 1 Perhitungan fungsi *Survival* dan fungsi *Survival* dengan Metode Bayesian SELF.**

<i>Survival</i> Pareto	<i>Survival</i> Bayesian SELF
$S(87; 72) = 0,798713268$	$\hat{S}_{BS}(87; 86,69) = 0,995858643$
$S(126; 72) = 0,514465032$	$\hat{S}_{BS}(126; 86,69) = 0,641449778$
$S(200; 72) = 0,297195798$	$\hat{S}_{BS}(200; 86,69) = 0,370552257$
$S(384; 72) = 0,136955652$	$\hat{S}_{BS}(384; 86,69) = 0,170760241$
$S(999; 72) = 0,043997397$	$\hat{S}_{BS}(999; 86,69) = 0,054857219$

Dari hasil perhitungan fungsi *Survival* dan fungsi *Survival* dengan menggunakan metode Bayesian SELF perhitungan peluang hidup pada kasus penderita kanker paru-paru menjadi lebih tinggi.

**3. Perhitungan Nilai MAPE**

*Mean Absolute Percentage Error* (MAPE) dihitung dengan menggunakan kesalahan absolut padatiap periode dibagi dengan nilai observasi yang nyata pada periode itu[7]. Nilai MAPE dapat dihitung dengan persamaan berikut:

$$MAPE = \frac{\sum |(S - \hat{S})/S|}{n} \times 100 \tag{17}$$

Dimana S merupakan data awal *survival*,  $\hat{S}$  merupakan data *survival* yang sudah di estimasi  $n$  adalah jumlah data dengan syarat  $t > \theta$ .

Kriteria nilai MAPE jika dilihat dengan tabel adalah sebagai berikut:

**Tabel 2 Kriteria MAPE**

MAPE	Status
< 10%	Sangat Baik
10% – 20%	Baik
20% – 50%	Cukup
> 50%	Buruk

Berdasarkan Persamaan (17) diperoleh nilai MAPE sebagai berikut:

$$MAPE = \frac{|15,797|}{64} \times 100 = 24,68 \%$$

## KESIMPULAN

1. Estimasi parameter model *survival* berdistribusi Pareto pada data tersensor menggunakan metode Bayesian SELF adalah sebagai berikut:

$$\hat{S}_{BS} = (t_i; \hat{\theta}_{BS}) = \left( \frac{A + n\alpha}{t} \right)^\alpha$$

$$\hat{h}_{BS} = (t_i; \hat{\theta}_{BS}) = \frac{\alpha}{t}$$

2. Berdasarkan hasil estimasi metode Bayesian SELF pada kasus penderita kanker paru-paru dapat diketahui peluang seseorang untuk bertahan hidup semakin lama semakin kecil (mendekati nol). Dan berdasarkan grafik fungsi *Survival* Pareto dan grafik fungsi *Survival* dengan Metode Bayesian SELF yaitu perhitungan peluang hidup seorang penderita kanker paru-paru menjadi lebih tinggi.
3. Berdasarkan nilai MAPE yang diperoleh dari fungsi *Survival* distribusi Pareto dengan pendekatan Bayesian SELF adalah sebesar 24,68%. Hal ini berarti bahwa metode Bayesian SELF memiliki kemampuan peramalan yang cukup dalam mengestimasi peluang hidup pasien penderita kanker paru-paru.

## DAFTAR PUSTAKA

- [1]. Kleinbaum, D.G., and Klein, M. *Survival Analysis : A Self-Learning Text Second Edition*, Springer Science Business Media, Inc, New York, 2005.
- [2]. Fitria, S., Helmi, & Rizki, S.W. Estimasi Parameter Model *Survival* distribusi Eksponensial Data Tersensor Dengan Metode Maksimum Likelihood dan Bayesian SELF. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 3: 213-220, 2016.
- [3]. Box, G.E.P and Tiao, G.C. *Bayesian Inference in Statistical Analysis*, Addison Wesley Publishing Company, London, 1973.
- [4]. Lee, E.T. and Wang, J.W. *Statistical Methods for Survival Data Analysis Third Edition*, Canada : A Jhon Wiley & Sons, Inc, 2003.
- [5]. Jamilah, Firdaus & Sugiarto, S. *Penduga Interval Parameter Bentuk Dari Distribusi Pareto Berdasarkan Metode Momen dan Maksimum Likelihood*. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Riau. Pekanbaru, 2014.
- [6]. Bolstad, W.M. *Introduction to Bayesian Statistical Second Edition*, A.Jhon Wiley & Sons, Inc, Amerika, 2007.
- [7]. Pakaja, F., Naba, A., Purwanto. Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan Certainty Factor, *Jurnal EECCIS*, Vol.6, No.1, Juni 2012

BELLA AMALIA :Jurusan Matematika FMIPA UNTAN,Pontianak  
[bellaamalia03@gmail.com](mailto:bellaamalia03@gmail.com)

SHANTIKA MARTHA :Jurusan Matematika FMIPA UNTAN,Pontianak  
[shantika.martha@math.untan.ac.id](mailto:shantika.martha@math.untan.ac.id)

SETYO WIRA RIZKI :Jurusan Matematika FMIPA UNTAN,Pontianak  
[setyo.wirarizki@math.untan.ac.id](mailto:setyo.wirarizki@math.untan.ac.id)

---