ALGORITMA *EXPECTATION-MAXIMIZATION* UNTUK PENDUGAAN PARAMETER DISTRIBUSI *GENERALIZED GAMMA* PADA DATA TERSENSOR KANAN TIPE 1

(Skripsi)

Oleh

Atika Ayu Listianingsih

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

ABSTRACT

EXPECTATION-MAXIMIZATION FOR ESTIMATING THE GENERALIZED GAMMA DISTRIBUTION PARAMETERS WITH TYPE 1 RIGHT CENSORED DATA

By

ATIKA AYU LISTIANINGSIH

Generalized Gamma distribution is the most popular model for analyzing skewed data and various forms of hazard functions, this makes Generalized Gamma distribution often used in survival analysis. Survival analysis aims to predict the chances of survival, recurrence of disease, death, and other events up to a certain period of time. One characteristic of survival data is the possibility of censorship, this study uses right-type censored data type 1. Estimation of Generalized Gamma distribution parameters in survival analysis uses maximum likelihood estimation which is solved by the expectation-maximization algorithm. In this study an analysis of the survival time of patients with gastric cancer will be conducted. Based on the results of the iteration, the estimated value of the parameter is obtained, namely $\alpha = 3.406516$, $\beta = 1.072$, and $\theta = 59.766$. Based on survival function, the chances of survival of patients will continue to decrease until death occurs, while hazard functions have an increasing rate, namely the increasing survival time of patients, the chances of death will increase.

Keywords: Generalized Gamma, Type 1 Right Censored Data, Maximum Likelihood Estimation, Eexpectation-Maximization Algorithm

ABSTRAK

ALGORITMA *EXPECTATION-MAXIMIZATION* UNTUK PENDUGAAN PARAMETER DISTRIBUSI *GENERALIZED GAMMA* PADA DATA TERSENSOR KANAN TIPE 1

Oleh

ATIKA AYU LISTIANINGSIH

Distribusi *Generalized Gamma* adalah model paling popular untuk menganalisis data yang *skewed* dan berbagai bentuk fungsi *hazard*, hal ini membuat distribusi *Generalized Gamma* sering digunakan dalam analisis *survival.* Analisis *survival* bertujuan untuk menduga peluang bertahan hidup, kekambuhan penyakit, kematian, dan peristiwa-peristiwa lainnya sampai pada periode waktu tertentu. Salah satu karakteristik data *survival* adalah kemungkinan adanya sensor, penelitian ini menggunakan data tersensor kanan tipe 1. Pendugaan parameter distribusi *Generalized Gamma* pada analisis *survival* menggunakan *maximum likelihood estimation* yang diselesaikan dengan algoritma *expectation-maximization*. Dalam penelitian ini akan dilakukan analisis waktu bertahan hidup pasien penderita penyakit kanker lambung. Berdasarkan hasil iterasi diperoleh nilai dugaan parameter, yaitu $\alpha = 3.406516$, $\beta = 1.072$, dan $\theta = 59.766$. Berdasarkan fungsi *survival*, peluang bertahan hidup dari pasien akan terus mengalami penurunan sampai terjadinya kematian, sedangkan fungsi *hazard* memiliki laju *increasing* yaitu semakin meningkatnya waktu bertahan hidup dari pasien maka peluang kematiannya akan semakin meningkat.

Kata Kunci: *Generalized Gamma*, Data tersensor kanan tipe 1**,** *Maximum Likelihood Estimation*, Algoritma *Expectation-Maximization*

ALGORITMA *EXPECTATION-MAXIMIZATION* UNTUK PENDUGAAN PARAMETER DISTRIBUSI *GENERALIZED GAMMA* PADA DATA TERSENSOR KANAN TIPE 1

Oleh

Atika Ayu Listianingsih

Skripsi
Sebagai Salah Satu Syarat untuk Mencapai gelar
SARJANA SAINS

Pada
Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2019

Judul Skripsi : **ALGORITMA *EXPECTATION MAXIMIZATION* UNTUK PENDUGAAN PARAMETER DISTRIBUSI *GENERALIZED GAMMA* PADA DATA TERSENSOR KANAN TIPE 1**

Nama Mahasiswa : *Atika Ayu Listianingsih*

No. Pokok Mahasiswa : 1517031019

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dian Kurniasari, S.Si., M.Sc.	**Widiarti, S.Si., M.Si.**
NIP 196903051996032001	NIP 198005022005012003

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP 196311081989022001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dian Kurniasari, S.Si., M.Sc.** ....

Sekretaris : **Widiarti, S.Si., M.Si.** ....

Penguji
Bukan Pembimbing: **Prof. Drs. Mustofa Usman, M.A., Ph.D.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Drs. Suratman, M.Sc.
NIP. 196406041990031002

Tanggal Lulus Ujian Skripsi: **26 Juli 2019**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah in:

Nama	: **Atika Ayu Listianingsih**
Nomor Pokok Mahasiswa	: **1517031019**
Jurusan	: **Matematika**
Judul Skripsi	: **Algoritma *Expectation-Maximization* untyk Pendugaan Parameter Distribusi *Generalized Gamma* pada Data Tersensor Kanan Tipe 1**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menema sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 26 Juli 2019

Yang Menyatakan,

Atika Ayu Listianingsih

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung, pada tanggal 22 April 1997, sebagai anak pertama dari tiga bersaudara, dari pasangan bapak Kasino dan Ibu Lismawati.

Pendidikan Taman Kanak-kanak (TK) An-Nur Kota Sepang, Bandar Lampung diselesaikan tahun 2003, Sekolah Dasar (SD) diselesaikan di SD N 5 Jatimulyo, Lampung Selatan pada tahun 2009, Sekolah Menengah Pertama (SMP) diselesaikan di SMP N 20 Bandar Lampung pada tahun 2012, dan Sekolah Menengah Atas (SMA) diselesaikan di SMA N 13 Bandar Lampung.

Tahun 2015 penulis terdaftar sebagai mahasiswa Jurusan Matematika FMIPA Unila melalui jalur SNMPTN. Selama menjadi mahasiswa penulis pernah menjadi asisten dosen mata kuliah Aljabar Linear Elementer dan asisten pratikum mata kuliah Eksplorasi Data. Penulis aktif di organisasi Rohani Islam (Rois) FMIPA Unia pada tahun 2015-2016 sebagai Anggota Badan Khusus Pemberdayaan Muslimah (BKPM) dan pada tahun 2017 sebagai Bendahara Badan Semi Otonom Bimbingan Belajar Qur'an (BSO-BBQ), pada tahun 2018 penulis aktif di organisasi Bina Rohani Islam (Birohmah) sebagai Sekertaris Departemen Kemuslimahan, pada tahun 2019 penulis aktif di Dewan Perwakilan Mahasiswa sebagai anggota komisi 4.

PERSEMBAHAN

Alhamdulillaahi Rabbil 'Alamin

Segala puji bagi Allah SWT yang telah melimpahkan keberkahan tak terhingga kepada umat-Nya sehingga dapat merasakan indahnya Islam dan tempat penulis mengabdi sebagai hamba serta menggantungkan segala doa dan harapan. Shalawat dan salam semoga tercurahkan kepada Rasulullah SAW.

Penulis persembahkan karya ini untuk:

Ibu, yang selalu mendoakan dan memberikan semangat tanpa henti.

Bapak, yang selalu berkerja keras memberikan yang terbaik untuk keluarga.

Nenek, yang selalu mendoakan disetiap malam.

Kedua saudara penulis, M. Ariawan Dwi Mulyo dan M. Rahmat Nur Hidayat yang selalu mendoakan dan membantu disetiap kesulitan.

Guru-guru penulis, dengan penuh kesabaran memberikan pengetahuan dalam setiap bidang ilmu.

Teruntuk almamater tercinta Universitas Lampung.

KATA INSPIRASI

“Katakanlah: "Hai hamba-hamba-Ku yang malampaui batas terhadap diri mereka sendiri, janganlah kamu berputus asa dari rahmat Allah. Sesungguhnya Allah mengampuni dosa-dosa semuanya. Sesungguhnya Dialah Yang Maha Pengampun lagi Maha Penyayang.”

QS: Az-Zumar:53

“Sesungguhnya sesudah kesulitan itu ada kemudahan”

QS: Al-Insyirah:6

”Sesungguhnya jika kamu bersyukur, pasti Kami akan menambah (nikmat) kepadamu, dan jika kamu mengingkari (nikmat-Ku), maka sesungguhnya azab-Ku sangat pedih”

QS: Ibrahim:7

“ Sesungguhnya Allah tidak merubah keadaan suatu kaum sampai mereka merubah keadaannya sendiri”

QS: Ar- Ra;du: 11

"Aku tidak menciptakan jin dan manusia, melainkan agar mereka beribadah kepada-Ku."

QS: Adz: Dzariyat:56

SANWACANA

Puji syukur penulis ucapkan kehadirat Allah SWT, karena atas berkah dan hidayah-Nya skripsi ini dapat diselesaikan. Skripsi dengan judul *"Algoritma Expectation-Maximization untuk Pendugaan Parameter Distribusi Generalized Gamma pada Data Tersensor Kanan Tipe 1"* adalah salah satu syarat untuk memperoleh gelar Sarjana Matematika (S.Si.) di Universitas Lampung.

Dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Ibu Dian Kurniasari, S.Si., M.Sc., selaku Dosen Pembimbing I yang telah meluangkan waktu untuk membimbing penulis, memberikan motivasi, serta kritik dan saran kepada penulis selama menyusun skripsi.
2. Ibu Widiarti, S.Si., M.Si., selaku Dosen Pembimbing II yang telah memberi masukan dan arahan kepada penulis selama menyusun skripsi.
3. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D., selaku Dosen Pembahas yang memberi masukan dan evaluasi kepada penulis selama menyusun skripsi.
4. Bapak Warsono, Ph.D., selaku Pembimbing Akademik yang telah mengarahkan penulis dari awal sampai lulus kuliah.
5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika FMIPA Universitas Lampung.
6. Bapak Drs. Suratman, M.Sc., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Dosen, staff, dan karyawan Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan bantuan kepada penulis.
8. Bapak, Ibu, nenek, dan adik yang telah memberikan doa, motivasi, dan material kepada penulis.
9. Saudari penulis, Ai Mila, Azizah, Azzahra, Cynthia, Indah, Lena, Liza, Ribut, Risna, Rizki, Silvi, Sindi, dan Wilda yang telah memberikan doa dan motivasi kepada penulis selama perkuliahan.
10. Yunda Kadek, Bang Yefta, kakak tingkat angkatan 2013, kakak tingkat angkatan 2014, dan teman-teman angkatan 2015 yang telah memberi banyak pembelajaran selama perkuliahan.
11. Keluarga besar Rois FMIPA, Birohmah, dan DPM U KBM Unila yang telah memberikan ruang kepada penulis untuk mengembangkan potensi diri serta memberikan doa dan motivasi selama kepengurusan.

Semoga seluruh kebaikan yang telah diberikan mendapatkan balasan dari Allah SWT dengan limpahan keberkahan. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca sebagai acuan di penelitian selanjutnya.

Bandar Lampung, 26 Juli 2019

Atika Ayu Listianingsih

DAFTAR ISI

Halaman

DAFTAR TABEL .... **xv**

DAFTAR GAMBAR .... **xvi**

I. PENDAHULUAN
1.1 Latar Belakang 1
1.2 Tujuan Penelitian 3
1.3 Manfaat Penelitian 4

II. TINJAUAN PUSTAKA
2.1 Analisis *Survival* 5
2.1.1 Fungsi *Survival* (bertahan hidup) 5
2.1.2 Fungsi *Hazard* (Resiko) 6
2.2 Distribusi *Generalized Gamma* 7
2.3 *Maximum Likelihood Estimation* 7
2.4 Algoritma *Expectation-Maximization* 8
2.5 Data Tersensor 10

III. METODE PENELITIAN
3.1 Waktu dan Tempat 12
3.2 Data Penelitian 12
3.3 Metode Penelitian 12

IV. HASIL DAN PEMBAHASAN
4.1 Data Waktu Bertahan Hidup Pasien Penderita Penyakit Kanker Lambung 15
4.2 Fungsi *Survival* (Bertahan Hidup) dan Fungsi *Hazard* (Resiko) dari distribusi *Generalized Gamma* 16
4.3 Fungsi *Likelihood* Data Lengkap dari Distribusi *Generalized Gamma* 21
4.4 *E- Step* 22
4.5 *M-Step* 26
4.5.1 Persamaan Parameter Baru $\alpha^{(m+1)}, \beta^{(m+1)}$, dan $\theta^{(m+1)}$ 30
4.6 Membentuk Fungsi *Survival* (Bertahan Hidup) dan Fungsi *Hazard* (Resiko) dari Data Bertahan Hidup Pasien Penderita Penyakit Kanker Lambung 31

4.6.1 Membentuk Kurva *Probability Density Function* dan *Cumulative Density Function* 32
4.6.2 Menghitung Nilai Fungsi *Survival* (Bertahan Hidup) dan Fungsi *Hazard* (Resiko) 34
4.6.3 Kurva Fungsi *Survival* (Bertahan Hidup) dan Kurva Fungsi *Hazard* (Resiko) Pasien Penderita Penyakit Kanker Lambung 36
4.7 Simulasi 37

V. KESIMPULAN DAN SARAN
5.1 Kesimpulan 40
5.2 Saran 41

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel Halaman

1. Data Waktu Bertahan Hidup Pasien Penderita Penyakit Kanker Lambung 16

2. Nilai Dugaan Parameter Data Waktu Bertahan Hidup Pasien Penderita Penyakit Kanker Lambung 31

3. Hasil Dugaan Fungsi *Survival* (Bertahan Hidup) dan Fungsi *Hazard (Resiko)* 35

4. Hasil Simulasi 38

DAFTAR GAMBAR

Gambar Halaman

1. Ilustrasi Sensor Kanan Tipe 10

2. Kurva *Probability Density Function* Pasien Penderita Penyakit Kanker Lambung 33

3. Kurva *Cumulative Density Function* Pasien Penderita Penyakit Kanker Lambung 34

4. Kurva Fungsi *Survival* (Bertahan Hidup) 36

5. Kurva Fungsi *Hazard* (Resiko) 37

I. PENDAHULUAN

1.1 Latar Belakang

Distribusi *Generalized Gamma* adalah model paling populer untuk menganalisis data yang *skewed* (miring). Berdasarkan penelitian oleh Stacy (1962) dengan judul, "*A generalization of the Gamma Distribution, Annals of Mathematical Statistic*s", distribusi *Generalized Gamma* termasuk sebagai sub-model khusus yaitu, distribusi *Exponensial*, *Weibull*, *Gamma*, dan *Rayleigh*. Distribusi *Generalized Gamma* sangat cocok untuk memodelkan data dengan berbagai bentuk fungsi *hazard* (resiko), hal ini membuat distribusi *Generalized Gamma* sering digunakan dalam analisis *survival*.

Analisis *survival* (bertahan hidup) bertujuan menduga peluang bertahan hidup, kekambuhan penyakit, kematian, dan peristiwa-peristiwa lainnya sampai pada periode waktu tertentu. Salah satu karakteristik data *survival* adalah kemungkinan adanya sensor. Misalkan X masa hidup seseorang yang diteliti dan waktu penyensoran kanan C_r, X disumsikan saling bebas dengan fungsi kepekatan peluang $f(x)$, fungsi *survival* $S(x)$, dan fungsi *hazard* *h(x)*. Masa hidup X seseorang akan diketahui jika X kurang dari atau sama dengan C_r. Jika X lebih besar dari C_r , X individu tersebut *survive* (bertahan) atau sensor dikanan (Klein and Moeschberger, 1997).

Pendugaan parameter dari distribusi *Genarilized Gamma* pada analisis *survival* menggunakan *Maximum Likelihood Estimation* (MLE) merupakan salah satu metode pendugaan yang paling banyak digunakan. Menurut Millar (2011), *Maximum Likelihood Estimation* sebagai alat untuk inferensi, termasuk evaluasi signifikansi statistik, perhitungan interval kepercayaan, penilaian model, dan pendugaan. *Maximum Likelihood Estimation* memiliki sifat optimal untuk ukuran sampel yang cukup besar. Beberapa kasus sering ditemukan dalam fungsi *log likelihood* tidak dapat dimaksimumkan secara analitik, sehingga diperlukan perhitungan iterasi MLE seperti, algoritma *Expectation-Maximization* (Mclachlan and Krishnan, 2008).

Algoritma *Expectation-Maximization* adalah pendekatan untuk menghitung *Maximation Likelihood Estimation*, berguna dalam keanekaragaman permasalahan data tidak lengkap. Prosedur dalam algoritma *Expectation-Maximization* terdapat dua langkah yaitu, langkah *Expectation* (*E-Step*) dan langkah *Maximation* (*M-Step*). Dasar ide dari algoritma *Expectation-Maximization* adalah mengaitkan masalah data tidak lengkap dengan masalah data lengkap untuk mendapatkan hasil akhir dari *Maximation Likelihood Estimation*. *E-Step* fokus untuk menghasilkan data lengkap, dengan menggunakan kumpulan dari data yang teramati, sehingga langkah sederhana dari *M-Step* memaksimumkan data lengkap (Mclachlan and Krishnan, 2008).

Peneitian sebelumnya oleh Ruhi, *et.al* (2015) dengan judul, *"Mixture Models for Analyzing Product Reliability Data: a Case Study"* menggunakan model *Weibull Mixture* untuk menganalisis data keterandalan produk atau realibilitas dengan pengamatan yang disensor kanan. Penelitian tersebut mendapatkan kesimpulan bahwa algoritma *Expectation-Maximization* pendugaan yang baik dibandingkan metode *Weibull Probability Paper Plots*.

Penelitian oleh Wang and Cheng (2009), yang berjudul, *"EM Algorithm for Estimating the Burr XII Parameters with Multiple Censored Data"* dalam penelitian tersebut *Maximum Likelihood Estimation* melalui algoritma *Expectation-Maximization* diturunkan dan digunakan untuk menemukan parameter distribusi *Burr XII* berdasarkan beberapa data yang disensor. Hasil simulasi menunjukkan bahwa *Maximum Likelihood Estimation* melalui algoritma *Expectation-Maximization* lebih efisien daripada metode *Newton Raphson* untuk parameter *Burr XII* dengan banyak data yang disensor dalam hal bias dan *RMSE* dalam banyak kasus. Terkait hal tersebut, penelitian ini akan membahas algoritma *Expectation-Maximization* untuk pendugaan parameter distribusi *Generalized Gamma* pada data tersensor kanan tipe 1.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut

1. Pendugaan parameter distribusi *Generalized Gamma* pada data tersensor kanan tipe 1 dengan menggunakan algoritma *Expectation-Maximization*.
2. Menerapkan pendugaan parameter distribusi *Generalized Gamma* pada data waktu bertahan pasien penderita penyakit kanker lambung untuk mendapatkan nilai peluang bertahan hidup seseorang dan tingkat resiko *failure* (kegagalan).

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah menambah referensi terhadap ilmu statistika inferensia mengenai analisis *survival* (bertahan hidup) pada data tersensor kanan tipe 1 dengan distribusi *Generalized Gamma* dengan metode algoritma *Expectation-Maximization*.

II. TINJAUAN PUSTAKA

2.1 Analisis *Survival*

Waktu *survival* (bertahan hidup) adalah waktu bertahan hidup individu sampai ke peristiwa tertentu, seperti kegagalan, kematian, kambuh, atau perkembangan penyakit tertentu. Distribusi waktu *survival* (bertahan hidup) biasanya digambarkan atau oleh tiga fungsi, *S(x)* fungsi *survival* (bertahan hidup), *f(x)* fungsi kepekatan peluang, dan *h(x)* fungsi *hazard* (hazard). Tiga fungsi tersebut digunakan untuk menggambarkan berbagai aspek data. Masalah yang mendasar dalam analisis data *survival* adalah pendugaan dari data sampel dari tiga fungsi tersebut untuk menarik kesimpulan tentang pola *survival* dalam populasi (Lee and Wang, 2003).

2.1.1 Fungsi *Survival* (Bertahan Hidup)

Menurut Klein and Moeschberger (1997), dasar yang digunakan untuk menggambarkan fenomena waktu-ke-peristiwa adalah fungsi *survival* (bertahan hidup). Peluang seseorang yang bertahan hidup melampaui waktu x (mengalami peristiwa setelah waktu x), didefinisikan sebagai

$$S(x) = \mathrm{P}(X > x) \tag{2.1}$$

Dalam konteks kegagalan peralatan atau barang yang diproduksi, *S(x)* disebut sebagai fungsi keandalan atau *reliability*. Ketika X adalah variabel acak kontinu, fungsi *survival* (bertahan hidup) adalah komplemen dari fungsi distribusi kumulatif, yaitu,

$$S(x) = 1 - F(x) \tag{2.2}$$

Dimana $F(x) = \mathrm{P}(X < x)$, Juga, fungsi *survival* adalah integral dari fungsi kepekatan peluang, $f(x)$ yaitu,

$$S(x) = \mathrm{P}(X > x) = \int_x^\infty f(t)\, dt \tag{2.3}$$

demikian

$$f(x) = -\frac{dS(x)}{dx} \tag{2.4}$$

2.1.2 Fungsi *Hazard* (Resiko)

Menurut Klein and Moeschberger (1997), fungsi *hazard* (resiko) dikenal sebagai tingkat kegagalan kondisional dalam keandalan, kekuatan mortalitas dalam demografi, fungsi intensitas dalam proses stokastik, tingkat kegagalan spesifik usia dalam epidemiologi, dan lain sebagainya. Tingkat *hazard* (resiko) ditentukan oleh,

$$h(x) = \lim_{\Delta x \to 0} \frac{P\,[\,x \le X < x + \Delta x \mid X \ge x]}{\Delta x} \tag{2.5}$$

Jika X adalah variabel acak kontinu, sehingga

$$h(x) = \frac{f(x)}{S(x)} = \frac{d[\ln S(x)]}{dx} \tag{2.6}$$

fungsi *hazard* (resiko) kumulatif *H(x)*, yang didefinisikan

$$H(x) = \int_0^x h(u)\, du = -\ln[S(x)] \tag{2.7}$$

Jadi, untuk masa hidup yang berkelanjutan,

$$S(x) = e^{-H(x)} = e^{-\int_0^x h(u)du} \tag{2.8}$$

2.2 Distribusi *Genaralized Gamma*

Menurut Cordeiro, *et.al* (2012), distribusi *Generalized Gamma* baik untuk memodelkan data dengan berbagai jenis fungsi *hazard rate* (tingkat resiko). Distribusi *Generalized Gamma* telah digunakan di beberapa bidang penelitian seperti teknik, hidrologi, dan analisis *survival*. Fungsi kepekatan Peluang dari distribusi *Generalized Gamma* sebagai berikut,

$$f(x;\alpha,\beta,\theta) = \frac{\beta}{\Gamma(\alpha)}\, \theta^{-\beta\alpha} x^{\beta\alpha-1} \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right] \tag{2.9}$$

Parameter α dan θ dikenal sebagai parameter bentuk dan parameter β dikenal sebagai parameter skala.

2.3 *Maximum Likelihood Estimation*

Menurut Hogg, *et.al* (2005), dimisalkan suatu variabel acak yaitu, $X_1, X_2, \dots, X_n$. Diketahui bahwa $X_1, X_2, \dots, X_n$ adalah saling bebas dengan fungsi kepekatan peluang $f(x;\ \omega)$, $\omega \in \Omega$. Parameter ω tidak diketahui. Dasar dari prosedur pendugaan dari fungsi *likelihood* adalah

$$L(\omega; x) = \prod_{i=1}^{n} f(x_i|\omega),\ \omega \in \Omega \tag{2.10}$$

Dimana x = (x_1, x_2, … , x_n)'. Log dari fungsi *likelihood* umumnya lebih mudah untuk digunakan, yaitu

$$log\ \ \mathrm{L}\ (\omega; x) = log \prod_{i=1}^{n} f(x_i|\omega),\ \omega \in \Omega \qquad (2.11)$$

Dinyatakan bahwa $\widehat{\omega} = \widehat{\omega}(X)$ adalah suatu *maximum likelihood estimator* dari θ jika

$$\widehat{\omega} = Argmax\ log\ \ \mathrm{L}\ (\omega; x) \qquad (2.12)$$

Argmax merupakan cara agar *log* L $(\omega; x)$ mencapai nilai maksimum atas $\widehat{\omega}$. Untuk menentukan *maximum likelihood estimation,* sering sekali menggunakan log dari fungsi *likelihood,* hal tersebut menentukan nilai kritisnya, karena fungsi logaritma monoton naik pada (0, ∞) yang berarti memiliki nilai kritis yang sama. Misalkan *log* L $(\omega; x)$, maka penyelesaian *Maximum Likelihood Estimation* dengan persamaan,

$$\frac{\partial\ L(\omega)}{\partial \omega} = 0 \qquad (2.13)$$

2.4 Algoritma *Expectation-Maximation*

Menurut Mclachlan and Krishnan (2008), algoritma *Expectation-Maximization* adalah pendekatan yang secara luas untuk perhitungan iterasi *Maximum Likelihood Estimation* yang bermanfaat dalam berbagai ketidaklengkapan data, di mana algoritma seperti metode *Newton-Raphson* mungkin akan menjadi lebih rumit dalam mengatasi ketidaklengkapan data. Pada setiap iterasi algoritma *Expectation-Maximization* ada dua langkah yang disebut, langkah ekspektasi atau *E-Step* dan langkah maksimalisasi atau *M-Step*. Dasar ide dari Algoritma *Expectation-Maximization* adalah mengaitkan masalah data tidak lengkap dengan masalah data lengkap untuk mendapatkan hasil akhir dari *Maximation Likelihood Estimation*. *E-Step* fokus untuk menghasilkan data lengkap, dengan menggunakan kumpulan dari data yang teramati, sehingga langkah sederhana dari *M-Step* memaksimumkan data lengkap.

Dimisalkan **X** menjadi vektor acak yang sesuai dengan data yang diamati **y**, memiliki p.d.f. g(y; ω), di mana $\omega = (\omega_{\mathbf{1}}, \ldots, \omega_{\mathbf{d}})^T$ adalah vektor dari parameter yang tidak diketahui dengan ruang parameter sebuah **Ω**. Algoritma *Expectation-Maximization* adalah algoritma yang diterapkan secara luas yang menyediakan prosedur berulang (iterasi) untuk menghitung *maximation likelihood estimation*. Dalam konteks ini, vektor data yang diamati **x** sebagai data tidak lengkap. Ide 'data tidak lengkap' mencakup pengertian hilang secara konvensional data. Misalkan **y** menunjukkan vektor yang berisi data teramati yang disebut augmented, dan misalkan **z** menunjukkan vektor yang berisi data tambahan, disebut sebagai data yang tidak dapat diobservasi atau hilang.

Misalkan $\omega^{(0)}$menjadi nilai awal untuk ω. Kemudian pada iterasi pertama, *e-step* membutuhkan perhitungan berikut

$$Q(\omega; \omega^{(0)}) = E_{\omega^{(0)}}\{ \log L_c(\omega) \mid \mathbf{x}\} \tag{2.14}$$

m-step selanjutnya memaksimalisasi $Q(\omega; \omega^{(0)})$ sehubungan dengan $\omega^{(0)}$ atas ruang parameter Ω. Yaitu, kami memilih $\omega^{(1)}$ sedemikian rupa

$$Q\left(\omega^{(1)}; \omega^{(0)}\right) \geq Q(\omega; \omega^{(0)}) \tag{2.15}$$

untuk semua $\omega \in \Omega$. Pada iterasi ke *(m + 1)*, *E-Step* dan *M-Step* didefinisikan sebagai berikut, *E-Step* menghitung Q(ω; $\omega^{(m)}$), dimana

$$Q(\omega; \omega^{(m)}) = E_{\omega^{(m)}}\{\log L_c(x; \omega) \mid \mathbf{y}\} \tag{2.16}$$

M-Step memaksimumkan fungsi $Q(\omega; \omega^{(m)})$, sehingga mendapatkan persamaan parameter baru $\omega^{(m+1)}$ dengan syarat iterasi sebagai berikut,

$$Q\left(\omega^{(k+1)};\ \omega^{(k)}\right) \geq\ Q(\omega; \omega^{(k)}) \tag{2.17}$$

untuk semua $\omega \in \Omega$.

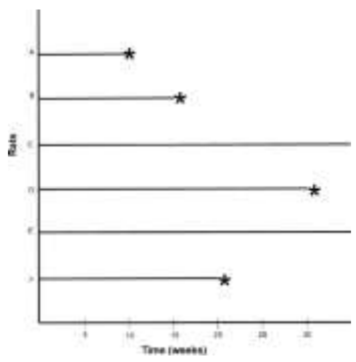
2.5 Data Tersensor

Penyensoran adalah hilangnya pengamatan/informasi pada variabel masa hidup yang diamati dalam penelitian. Dalam data *survival* (bertahan hidup), penyensoran sering terjadi karena berbagai alasan. Dalam uji klinis tentang efektivitas perawatan medis untuk penyakit, misalnya, pasien mungkin meninggalkan rumah sakit karena migrasi atau masalah kesehatan (Liu, 2012).

Menurut Klein and Moeschberger (1997), terdapat tiga tipe penyensoran sebagai berikut:

1. Sensor Kanan

   Misalkan X adalah masa hidup seseorang dan waktu penyensoran C_r Masa hidup X seseorang akan diketahui jika X $\leq$ C_r. Jika X lebih besar dari C_r maka orang tersebut dikatakan *survive* (bertahan) atau sensor kanan.



Gambar 1. Ilustrasi Sensor Kanan Tipe 1

Sebagai contoh dari (Lee and Wang, 2003), misalkan ada enam tikus yang terkena karsinogen dengan menyuntikkan sel tumor ke bantalan kaki mereka. Waktu untuk mengembangkan tumor dengan ukuran tertentu diamati. Peneliti memutuskan untuk menghentikan percobaan setelah 30 minggu. Gambar 1. adalah grafik waktu perkembangan tumor. Tikus A, B, dan D mengalami tumor masing-masing setelah 10, 15, dan 25 minggu. Tikus C dan E tidak mengalami tumor pada akhir penelitian, sehingga waktu kemungkinan terkena tumor adalah 30 minggu lebih. Tikus F meninggal tanpa tumor setelah 19 minggu pengamatan. Sehingga data waktu *survival* adalah 10, 15,30+, 25, 30+, dan 19+ minggu. Tanda tambah menunjukkan pengamatan yang disensor.

2. Sensor Kiri

   Misalkan X adalah masa hidup seseorang dalam penelitian. Jika terjadi sensor kiri $C_{l,}$ berarti seseorang tersebut telah mengalami kejadian sebelum diamati dalam penelitian.

3. Sensor Interval

   Masa hidup seseorang terjadi dalam suatu interval waktu.

III. METODE PENELITIAN

3.1 Waktu dan Tempat

Penelitian ini dilakukan pada semester genap tahun akademik 2018/2019, bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data waktu bertahan hidup pasien penderita penyakit kanker lambung yang didiperoleh dari *package asaur* dalam *software R* (Wang, *et al*, 2014).

3.3 Metode Penelitian

Penelitian ini bertujuan mencari pendugaan parameter dari distribusi *Generalized Gamma* dengan metode algoritma *Expectation-Maximization* yang akan diterapkan dalam analisis *survival*. Perhitungan numerik dilakukan dengan software *open source* yaitu *R* 3.6.0. Adapun langkah-langkah penelitian ini, sebagai berikut:

1. Menganalisis data waktu bertahan hidup pasien penderita penyakit kanker lambung secara deskriptif.
2. Mencari fungsi *survival* (bertahan hidup) dan fungsi *hazard* (resiko) dari distribusi *Generalized Gamma*.
   a. Mencari fungsi kumulatif dari distribusi *Generalized Gamma*.
   b. Mencari fungsi *survival* (bertahan hidup) dari distribusi *Generalized Gamma*.
   c. Mencari fungsi *hazard* (resiko) dari distribusi *Generalized Gamma*.
3. Menentukan fungsi *likelihood* data lengkap dari distribusi *Genaralized Gamma*.
   a. Menentukan persamaan fungsi *likelihood* data lengkap dari distribusi *Generalized Gamma*.
   b. Menentukan persamaan fungsi *log likelihood* data lengkap dari distribusi *Generalized Gamma*
4. *E-step* (tahap ekspektasi)
   a. Mempartisi persamaan fungsi *log likelihood* untuk data lengkap dari distribusi *Generalized Gamma*
   b. Mencari ekspektasi dari persamaan fungsi *log likelihood* untuk data lengkap bersyarat data teramati.
5. *M-step* (tahap maksimasi)
   a. Memaksimumkan persamaan fungsi *likelihood* data lengkap $Q(\omega, \omega^{(m)})$ terhadap parameter α.
   b. Memaksimumkan persamaan fungsi *likelihood* data lengkap $Q(\omega, \omega^{(m)})$ terhadap parameter β.

c. Memaksimumkan persamaan fungsi *likelihood* data lengkap $Q\left(\omega, \omega^{(m)}\right)$ terhadap parameter θ.

d. Mendapatkan persamaan parameter baru $\alpha^{(m+1)}, \beta^{(m+1)}$, dan $\theta^{(m+1)}$ dengan bantuan bantuan metode iterasi *Newton Raphson*

6. Menerapkan pendugaan parameter distribusi *generalized gamma* pada data waktu bertahan hidup pasien penderita penyakit kanker lambung ke dalam fungsi *survival* (bertahan hidup) dan fungsi *hazard* (resiko).
   a. Menghitung nilai fungsi *survival* (bertahan hidup) dan fungsi *hazard* (resiko) dari data waktu bertahan hidup pasien penderita penyakit kanker lambung pada waktu bertaha hidup 50, 100, 150, 200, dan 250 minggu.
   b. Membentuk kurva fungsi *survival* dan fungsi *hazard* pada pasien penderita penyakit kanker lambung.
7. Melakukan simulasi dengan ukuran sampel yang berbeda-beda, dengan langkah-langkah sebagai berikut:
   a. Membangkitkan data yang berdistribusi *Generalized Gamma* dengan ukuran sampel 50 dan100.
   b. Membangkitkan data yang berdistribusi *Bernoulli* dengan peluang kejadian 80%, 70%, 60%, dan 50%.
   c. Melakukan pendugaan parameter dengan *Maximum Likelihood Estimation*
   d. Menghitung bias dugaan dari masing-masing parameter dengan cara $bias(\hat{\alpha}) = |\alpha - \hat{\alpha}|$, $bias(\hat{\beta}) = |\beta - \hat{\beta}|$, dan $bias\ (\hat{\theta}) = |\theta - \hat{\theta}|$.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari penelitian yang telah dilakukan, maka diperoleh kesimpulan bahwa fungsi *survival* (bertahan hidup) dan fungsi *hazard* (resiko) distribusi *Generalized Gamma* sebagai berikut :

$$S(x) = \frac{\Gamma\left(\alpha,\ (\frac{x}{\theta})^{\beta}\right)}{\Gamma(\alpha)}$$

$$h(x) = \frac{\frac{\beta\ x^{\beta\alpha-1}}{\theta^{\beta\alpha}}\ e^{-(\frac{x}{\theta})^{\beta}}}{\Gamma\left(\alpha,\ (\frac{x}{\theta})^{\beta}\right)}$$

Dalam pendugaan parameter menggunakan *Maximum Likelihood Estimation* pada data waktu bertahan hidup pasien penderita penyakit kanker lambung yang tersensor kanan tipe 1 distibusi *Generalized Gamma* dengan bantuan algoritma *Expectation-Maximization* dihasilkan nilai dugaan parameter yaitu, $\hat{\alpha} = 3.406516$, $\hat{\beta} = 1.072$, dan $\hat{\theta} = 59.766$.

Penerapan pendugaan parameter distribusi *Generalized Gamma* pada analisis *survival* menghasilkan peluang bertahan hidup dari pasien penyakit penderita kanker lambung akan terus mengalami penurunan sampai terjadinya kematian, sedangkan fungsi *hazard* (resiko) pasien penderita penyakit kanker lambung memiliki laju kegagalan *increasing* yaitu semakin meningkatnya waktu bertahan hidup dari individu maka laju kegagalannya akan semakin meningkat. Berdasarkan hasil simulasi, semakin besar jumlah data maka semakin kecil bias dugaan.

5. 2 Saran

Adapun saran dari penelitian ini yaitu, algoritma *Expectation-Maximizaion* dapat dikembangkan lebih lanjut dengan menggunakan jenis distribusi berbeda, seperti distribusi *Rayleigh, Generalized Weibull, Gamma Mixture* dan sebagainya dengan tipe penyensoran yang berbeda.

DAFTAR PUSTAKA

Abramowitz, M. and Stegun, I.A. 1964. *Handbook of Mathematical Functions with Formulas, Grpahs and Mathematical Tables*. U.S.Government, Washington.

Almkvist, G and Zeilberger, D. 1990. The Method of Differentiating Under the Integral Sign. *J. Symb. Comp.* 10(6): 571-591.

Bain, L.J. and Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics.* Duxbury, United States.

Cardeiro, G.M., Castellares, F., Montenegro, L.C., and Casto, M. 2012. The Beta Generalizatiom Gamma Distribution. *J. of Theoretical and Applied Statistics.* 1-13.

Hogg, V.H., Mckean, J.W., and Craig, A.T. 2013. *Introduction to Mathematical Statistics*. Pearson Education Inc., Boston.

Huang, P.H., and Hwang, T.Y. 2006. On New Moment Estimation of Parameters of the Generalized Gamma Distribution Using It's Characterization. *Taiwanese J. of Mathematics.* 10(4): 1083-1093.

Kurniasari, D., Maninja, D.R., and Warsono. 2007. Sifat Asimtotik Normalitas dan Ketakbiasan Penduga Kemungkinan Maksimum Parameter Distribusi *Generalized Gamma*. *J. Sains MIPA.* 14(1): 41-46.

Klein, J.P. and Moeschberger, M.L. 1997. *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag, New York.

Lange, K. 2014. A Gradient Algorithm Locally Equivalent to the EM Algorithm. *J. of the Royal Statistical Society*. 57(2): 425-437

Lee, E. and Wang, J.W. 2003. *Statistical Method for Survival Data Analysis.* John Wiley & Sons Inc., New York.

Liu, X. 2012. *Survival Analysis : Models and Applications*. John Wiley & Sons, USA.

Mclachlan, G.J. and Krishnan, T. 2008. *The EM Algorithm and Extensions.* John Wiley & Sons Inc., Canada.

Millar, R.B., 2011. *Maximum Likelihood Estimation and Inference: With Examples in R, SAS, and ADMB.* John Wiley & Sons, Ltd., United Kingdom.

Ruhi, S., Sarker, S., and Karim, M.R. 2015. Mixture Model for Analyzing Product Reliability Data: a Case Study. *a Springer Open J.* 4:364.

Stacy, E.W. 1962. A Generalization of The Gamma Distribution. Institute of Mathematical Statistics. *The Annals of Mathematical Statistics.* 33(3):1187-1192.

Varberg, D., Purcell, E.J., and Rigdon, S.E. 2008. *Kalkulus*. Erlangga, Jakarta.

Wang, F.K. and Cheng, Y.F. 2009. EM Algorithm for Estimating The Burr XII Parameters wth Multiple Censored Data. *John Wiley & Sons Inc*. 615-630

Wang, Y., Yu, Y., Li, W., Feng, Y., Hou, J., Ji, Y., Sun, Y., Shen, K., Shen, Z., Qin, X. and Liu, T. 2014 A Phase II Trial of Xeloda and Oxaliplatin (XELOX) Neo-Adjuvant Chemotherapy Followed by Surgery for Advanced Gastric Cancer Patiets with Para-Aortic Lymph Nide Metastasis. *Cancer Chemotherapy and Pharmacology*. 73(6): 1155-1161.