ESTIMASI PARAMETER DISTRIBUSI T MENGGUNAKAN ALGORITMA *EXPECTATION-MAXIMIZATION*

SKRIPSI

AHMAD YANI

H 121 12 011

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

MEI 2019

ESTIMASI PARAMETER DISTRIBUSI T MENGGUNAKAN ALGORITMA *EXPECTATION-MAXIMIZATION*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

AHMAD YANI

H 121 10 011

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

MEI 2019

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

ESTIMASI PARAMETER DISTRIBUSI T MENGGUNAKAN ALGORITMA *EXPECTATION-MAXIMIZATION*

adalah benar hasil kerja saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 20 Mei 2019

Ahmad Yani

NIM : H 121 12 011

ESTIMASI PARAMETER DISTRIBUSI T MENGGUNAKAN ALGORITMA *EXPECTATION-MAXIMIZATION*

Disetujui oleh:

Pembimbing Utama	**Pembimbing Pertama**
<u>Andi Kresna Jaya, S.Si., M.Si.</u>	**<u>Dr. Nurtiti Sunusi, S.Si., M.Si.</u>**
NIP. 19731228 200003 1001	**NIP. 19720117 199703 2002**

Pada 20 Mei 2019

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Ahmad Yani

NIM : H 121 12 011

Program Studi : STATISTIKA

Judul Skripsi : Estimasi Parameter Distribusi T Menggunakan Algoritma *Expectation-Maximization*

Telah berhasil dipertahankan dihadapan dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
1. Ketua	: Andi Kresna Jaya, S.Si., M.Si.	(.......................................)
2. Sekretaris	: Dr. Nurtiti Sunusi, S.Si., M.Si.	(.......................................)
3. Anggota	: Dr. La Podje Talangko, M.Si.	(.......................................)
4. Anggota	: Dra. Nasrah Sirajang., M.Si.	(.......................................)

Ditetapkan di : Makassar

: 20 Mei 2019

KATA PENGANTAR

Segala puji penulis panjatkan kehadirat **Allah *Subhanahu Wata'ala*** yang senantiasa melimpahkan rahmat dan karuniaNya. Shalawat dan salam senantiasa penulis kirimkan kepada Baginda **Nabi Muhammad *Shallallahu'alaihi Wasallam***, Sang Pejuang Kebenaran, yang telah mengajarkan kebenaran dan membimbing umat-umatnya ke arah yang benar. Rasa syukur yang tak terkira atas segala nikmat yang telah diberikan terutama nikmat kesehatan, kesempatan dan kemudahan yang dikaruniakan kepada penulis dalam menyelesaikan tugas akhir dengan judul **"Estimasi Parameter Distribusi T Menggunakan Algoritma *Expectation-Maximization*"**.

Penyusunan tugas akhir ini tentunya tidak lepas dari bantuan berbagai pihak baik moril maupun materil. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus dan penghargaan yang tak terhingga kepada orang tua tercinta **Ayahanda Drs. Kamba** dan **Ibunda Rahmatiah, B.Sc.** atas segala dukungan, doa, restu, nasehat, dan motivasi yang tak henti-hentinya mereka berikan yang telah menjadi penyemangat kepada penulis untuk menggapai cita-cita. Untuk kakak penulis **Muhammad Erwin, Eka Fitrilia,** dan **Juliana Kadri** terima kasih banyak atas ajaran dan motivasi yang telah diberikan kepada penulis. Serta untuk adik **Annisa** yang selalu memberikan keceriaan kepada penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada :

1. Ibu **Prof. Dwia Aries Tina Pulubuhu, M.A** selaku Rektor Unhas yang telah memberi kesempatan menyelesaikan pendidikan di Universitas Hasanudddin.
2. Bapak **Prof. Dr. Ir. Muh. Restu, MP** selaku Wakil Rektor Bidang Akademik, Bapak **Prof. Ir. Sumbangan Baja, M.Phil, Ph.D** selaku Wakil Rektor Bidang perencanaan, Keuangan dan Infrastruktur, Bapak **Prof. Dr. drg. A. Arsunan Arsin, M.Kes.** selaku Wakil Rektor Bidang Kemahasiswaan dan Alumni, k **Prof. dr. Muh. Nasrum Massi, Ph.D** selaku Wakil Rektor Bidang , Inovasi dan Kemitraan, Bapak **Prof. Dr. Ir. Nasaruddin Salam, MT**

selaku Sekretaris Universitas dan segenap **staf Rektorat Unhas** yang telah memberi fasilitas dalam masa pendidikan.

3. Bapak **Dr. Eng. Amiruddin** selaku Dekan FMIPA Unhas.
4. Wakil Dekan I FMIPA Unhas Bapak **Prof Muhammad Ivan Azis, M.Sc,** Wakil Dekan II FMIPA Unhas Bapak **Dr. Muhammad Zakir, M.Sc**, Wakil Dekan III FMIPA Unhas Bapak **Dr Andi Ilham Latunra, M.Si** dan segenap **staf Dekanat FMIPA Unhas** yang telah membantu dalam proses administrasi.
5. Ibu **Dr. Nurtiti Sunusi, S.Si, M.Si.** selaku Ketua Departemen Statistika FMIPA Unhas.
6. Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku Sekretaris Departemen Statistika FMIPA Unhas.
7. **Segenap dosen pengajar**, **staf Departemen Statistika**, dan **staf Departemen Matematika**, yang telah membekali ilmu dan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
8. Ibu **Dr. Nurtiti Sunusi, S.Si, M.Si.** selaku Penasehat Akademik. Terima kasih atas segala masukan dan motivasi yang diberikan kepada penulis selama menjalani pendidikan di Departemen Statistika.
9. Bapak **Andi Kresna Jaya, S.Si., M.Si.** selaku pembimbing utama yang telah meluangkan waktunya dengan penuh kesabaran memberikan bimbingan dan membagi ilmu kepada penulis.
10. Ibu **Dr. Nurtiti Sunusi, S.Si, M.Si.** selaku pembimbing pertama dalam penulisan tugas akhir ini. Terima kasih atas segala masukan dan motivasi yang diberikan kepada penulis selama menjalani pendidikan di Departemen Statistika.
11. Bapak **Dr. La Podje Talangko, M.Si.** dan ibu **Dra. Nasrah Sirajang, M.Si.** selaku Anggota Tim Penguji. Terima kasih telah memberikan kritik dan saran yang membangun dalam penyempurnaan penyusunan tugas akhir ini serta a dukungan dan waktu yang telah diberikan kepada penulis.

bat-sahabatku **Tassimbung (Rahman Syarif, Wihdatul Ummah, St Fadillah RM, dan Trisnawati Andini), Evo Tragels, Rekurensi,**

Statistika 2012, Matematika 2012, dan **MIPA 2012** terima kasih telah memberikan motivasi selama perkuliahan hingga penyusunan tugas akhir ini.

13. Teman seperjuanganku **Oktosar Sabri, Andi Fahruddin Akas, Bogi Kurniawan, Muh Arjuna Ansar, Reski Try Putri, Reski Syafruddin, Inti Arlina,** dan **Kanda Nur Rohma Oktaviani** yang senantiasa menemani, membantu dan menguatkan penulis selama masa penyusunan tugas akhir ini. Kalian Luar Biasa!
14. Kanda-kanda dan adik-adik warga **Himatika FMIPA Unhas** serta **Keluarga Mahasiswa Fakultas MIPA Unhas** yang selalu memberikan semangat untuk menyelesaikan tugas akhir ini, demi menjadi insan yang berilmu.
15. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung, dalam penyusunan tugas akhir ini yang tak sempat penulis sebutkan satu per satu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan sehingga kritik dan saran yang membangun akan penulis terima dan harapan agar tulisan ini dapat memberikan manfaat bagi semua pihak.

Makassar, 20 Mei 2019

Penulis

PERSETUJUAN PUBLIKASI KARYA ILMIAH

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini :

Nama : Ahmad Yani

NIM : H 121 12 011

Program Studi : Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

ESTIMASI PARAMETER DISTRIBUSI T MENGGUNAKAN ALGORITMA *EXPECTATION-MAXIMIZATION*

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, pada tanggal 20 Mei 2019

Yang menyatakan

Yani)

ABSTRAK

Fungsi peluang distribusi t dengan empat parameter merupakan fungsi marjinal dari distribusi normal dan distribusi gamma. Fungsi peluang distribusi normal yang digunakan adalah fungsi peluang distribusi normal dengan mean (Mu) dan variansinya (Seper lambda dan eta). Variabel random Eta pada parameter variansi adalah distribusi gamma dengan parameter alpha dan beta. Untuk mengestimasi parameter mu, lambda, alpha dan beta menggunakan metode Maximum Likelihood melalui algoritma EM. Berdasarkan hasil dari penelitian yang dilakukan maka didapatkan fungsi peluang distribusi t dengan empat parameter (Mu, lambda, alpha dan beta). Dengan menggunakan algoritma EM didapatkan penaksir untuk parameter mu, lambda, alpha dan beta , untuk penaksiran parameter alpha dapat dilanjutkan dengan menggunakan metode numerik dari fungsi penaksir alpha

Kata kunci : *Algoritma EM, Distribusi Gamma, Distribusi Normal, Distribusi T, Maximum Likelihood.*

ABSTRACT

Probability function of t distribution with four parameters is a marginal function of normal and gamma distributions. Probability function of normal distribution used is Probability function of normal distribution with mean (Mu) and variance (One per lambda and eta). Random variable of eta in variance parameter is gamma distribution with alpha and beta parameters. To estimate mu, lambda, alpha and beta parameters use Maximum Likelihood method via EM algorithm. Based on result of research conducted then obtained probability function of t distribution with four parameters (Mu, lambda, alpha dan beta). By using EM algorithm the estimator is obtained for mu, lambda, alpha dan beta, for estimating alpha parameter can be continued using numerical method of estimator function for alpha.

Keywords : *EM Algorithm, Gamma Distribution , Maximum Likelihood, Normal Distribution, T Distribution.*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Latar Belakang

Penelitian merupakan upaya mencari dan membuktikan kebenaran secara ilmiah. Penelitian dikatakan ilmiah apabila dalam cara kerjanya menunjukkan ciri-ciri keilmuan tertentu, yaitu rasional, empiris dan sistematis. Dalam penelitian, analisis data merupakan kegiatan setelah seluruh data terkumpul,dan di kelompokkan berdasarkan variabel dan jenis responden. Bentuk data yang diperoleh bergantung dari cara mendapatkan data tersebut. Setelah data terkumpul, hal yang selanjutnya diperhatikan adalah cara penyajian data atau bagaimana bentuk data yang diinginkan. Terdapat dua jenis pengolahan data statistika, yaitu statistika deskriptif dan statistika inferensial. Statistika deskriptif mengenai tentang bagaimana data dapat digambarkan (dideskripsikan) atau disimpulkan, baik secara numerik (seperti menghitung rata-rata dan standar deviasi) atau secara grafis (dalam bentuk tabel atau grafik) untuk mendapatkan gambaran mengenai data tersebut, sehingga lebih mudah dibaca dan bermakna. Sedangkan statistika inferensial membahas tentang pemodelan data dan proses pengambilan keputusan berdasarkan analisis data, misalnya melakukan pengujian hipotesis, melakukan estimasi pengamatan masa mendatang (prediksi), membuat pemodelan hubungan (korelasi, regresi, ANOVA, deret waktu), dan sebagainya.

Distribusi frekuensi merupakan alat yang dapat digunakan sebagai dasar pembanding dari suatu hasil penelitian dan juga sebagai pengganti distribusi sebenarnya. Distribusi t merupakan salah satu distribusi yang sangat sering digunakan dalam statistika terapan. Selain digunakan untuk menguji suatu hipotesis, distribusi t juga digunakan untuk estimasi interval. Pada umumnya, distribusi t digunakan untuk menguji hipotesis mengenai nilai parameter, paling banyak dari 2 populasi (lebih dari 2 populasi harus menggunakan F), dan dari sampel yang kecil, $n < 100$, bahkan seringkali $n \leq 30$. Untuk n yang cukup besar ($n \geq 100$, gkin cukup $n > 30$) dapat digunakan distribusi normal (Supranto, 2001)

Berbagai metode estimasi statistik telah diusulkan untuk mengestimasi parameter distribusi. Suatu metode estimasi yang dipilih sebaiknya memiliki *error* sampel yang kecil. Perlu dipahami bahwa metode yang digunakan untuk mengestimasi parameter suatu distribusi belum tentu efisien untuk distribusi yang lain. Dari berbagai metode estimasi yang ada, metode Maximum Likelihood merupakan metode estimasi yang digunakan untuk mengestimasi parameter dalam penelitian ini.

Berdasarkan uraian di atas, akan dikaji distribusi t menggunakan metode Maximum Likelihood yaitu algoritma EM dengan judul **“Estimasi Parameter pada Distribusi t Menggunakan Algoritma *Expectation-Maximization*”.**

1.2. Rumusan Masalah

Dalam penelitian ini terdapat beberapa rumusan masalah antara lain:

1. Bagaimana mendapatkan fungsi peluang distribusi t dengan empat parameter menggunakan algoritma EM?

2. Bagaimana model likelihood untuk data yang memenuhi distribusi t menggunakan algoritma EM?

1.3. Batasan Masalah

Dalam penelitian ini, masalah dibatasi pada distribusi t yang terdiri atas empat parameter dengan menggunakan algoritma EM.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mendapatkan penaksir parameter distribusi t dengan empat parameter.

BAB II

TINJAUAN PUSTAKA

2.1. Distribusi Normal

Distribusi normal adalah distribusi peluang kontinu yang terpenting dalam seluruh bidang statistika. Grafiknya berbentuk kurva normal, berbentuk lonceng seperti pada Gambar 2.1, yang menggambarkan dengan cukup baik banyak gejala yang muncul di alam, industri, dan penelitian. Pengukuran fisik di bidang seperti percobaan meteorologi, penelitian curah hujan, dan pengukuran suku cadang yang diproduksi sering dengan baik dapat dijelaskan dengan menggunakan distribusi normal. Selain itu, galat dalam pengukuran ilmiah dapat dihampiri dengan sangat baik oleh distribusi normal. Pada tahun 1733, Abraham DeMoivre menemukan persamaan matematika kurva normal. Ini merupakan dasar bagi banyak teori statistika induktif. Distribusi normal sering pula disebut distribusi Gauss untuk menghormati Karl Friedrich Gauss (1777-1855), yang juga menemukan persamaannya ketika meneliti galat dalam pengukuran yang berulang-ulang mengenai bahan yang sama.

Suatu peubah acak kontinu X yang distribusinya berbentuk lonceng seperti pada Gambar 2.1 disebut peubah acak normal. Persamaan matematika distribusi peluang peubah normal kontinu bergantung pada dua parameter, yaitu rataan (μ) dan simpangan baku (σ). Jadi fungsi padat X akan dinyatakan dengan $n(x; \mu, \sigma)$.

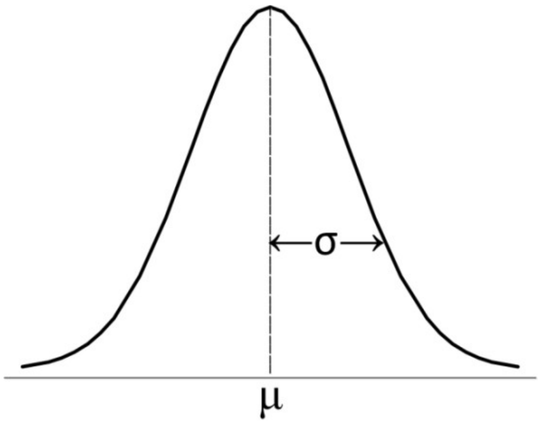
Fungsi padat peubah acak normal X, dengan rataan μ dan variansi σ^2 adalah

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad -\infty < x < \infty \qquad (2.1)$$

dengan $\pi = 3{,}14159\ldots$ dan $e = 2{,}71828\ldots.$

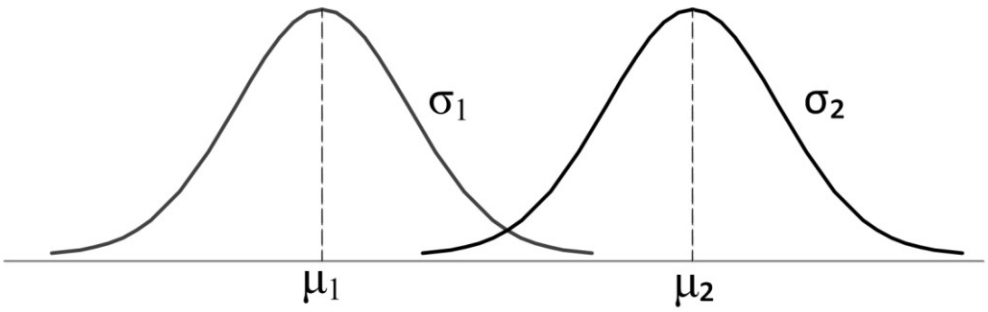
(Walpole & Myers, 1995).

Apabila nilai-nilai μ dan σ diketahui, maka bentuk kurva distribusi normal dapat digambarkan secara jelas.

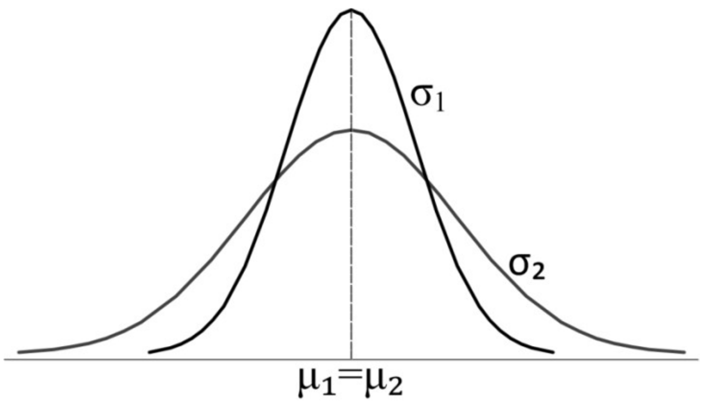


Gambar 2.1 Kurva normal

Bentuk dan ketinggian dari kurva normal sangat bergantung pada μ dan σ. Gambar 2.1 di atas merupakan sebuah kurva normal yang bergantung pada μ dan σ.



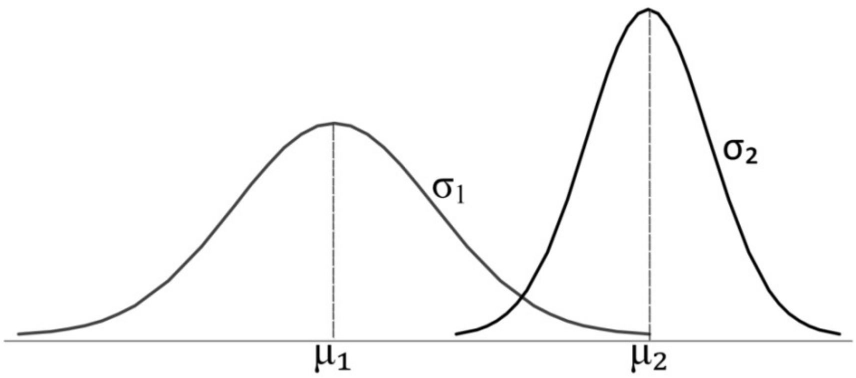
a. Kurva normal dengan $\mu_1 < \mu_2$ dan $\sigma_1 = \sigma_2$



b. Kurva normal dengan $\mu_1 = \mu_2$ dan $\sigma_1 < \sigma_2$

Gambar 2.2 Berbagai kurva normal bergantung pada μ dan σ.

Pada Gambar 2.2a diberikan dua kurva normal dengan rataan yang berbeda tapi simpangan baku yang sama. Hal ini dapat telihat dari titik tengah pada kedua kurva yang letaknya berbeda pada sumbu x, tapi bentuk kedua kurva tersebut memiliki ukuran yang sama. Pada Gambar 2.2b diberikan dua kurva normal dengan rataan yang sama tapi simpangan bakunya berbeda. Terlihat kedua kurva memiliki titik tengah yang sama pada sumbu x, tapi kurva dengan simpangan baku yang lebih besar tampak lebih rendah dan lebih melebar. Perhatikan bahwa luas di bawah kurva peluang harus sama dengan 1 sehingga bila kumpulan data makin berbeda maka makin rendah dan melebar pula kurvanya.



Gambar 2.3 Kurva normal dengan $\mu_1 < \mu_2$ dan $\sigma_1 > \sigma_2$

Gambar 2.3 menunjukkan dua kurva normal yang rataan maupun simpangan bakunya berbeda. Jelas kedua kedua kurva mempunyai letak titik tengah yang berlainan pada sumbu x dan bentuk kedua kurva mencerminkan dua nilai σ yang berbeda.

Dengan mengamati Gambar 2.1 sampai 2.3 serta memeriksa turunan pertama dan kedua dari $n(x;\mu,\sigma)$ dapat diperoleh lima sifat kurva normal berikut:

1. Modus, titik pada sumbu datar yang memberikan maksimum kurva, terdapat pada μ;
2. Kurva setangkup terhadap garis $x = \mu$;
3. Kurva mempunyai titik belok pada $x = \mu \pm \sigma$, cekung dari bawah bila $- \sigma < X < \mu + \sigma$, dan cekung dari atas untuk nilai x lainnya;

edua ujung kurva normal mendekati asimtot sumbu datar bila nilai x ergerak menjauhi μ baik ke kiri maupun ke kanan;

5. Seluruh luas di bawah kurva dan di atas sumbu datar sama dengan 1.

(Walpole & Myers, 1995).

2.2 Distribusi Gamma

Fungsi gamma dapat digambarkan sebagai generalisasi dari $n!$ (n-faktorial), dimana n adalah bilangan bulat positif. Fungsi gamma didefinisikan dari integral tak wajar

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{2.2}$$

untuk $\alpha > 0$ dan $x \in R$.

(Wrede & Murray, 2010)

Peubah acak kontinu X berdistribusi gamma, dengan parameter α dan β, bila fungsi kepadatan peluang berbentuk

$$f(x) = \frac{1}{\beta^\alpha \Gamma\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \qquad x > 0 \tag{2.3}$$

dengan $\alpha > 0$ dan $\beta > 0$

(Walpole & Myers, 1995).

2.3 Distribusi T

Distribusi t pertama kali diterbitkan pada tahun 1908 dalam suatu makalah oleh W.S.Gosset. Pada waktu itu, Gosset bekerja pada perusahaan bir di Irlandia yang melarang penerbitan penelitian oleh karyawannya. Untuk menghindari larangan ini Gossel menerbitkan karyanya secara rahasia dibawah nama student. Karena itulah distribusi t biasanya disebut Distribusi Student. Ciri-ciri distribusi t adalah:

1. Distribusi t seperti distribusi normal yang merupakan sebuah distribusi kontinu, dimana nilainya dapat menempati semua titik pengamatan.
2. Distribusi t seperti distribusi normal berbentuk genta atau lonceng dan simetris dengan nilai rata-rata sama dengan nol.
3. Distribusi t mempunyai rata-rata hitung sama dengan nol, tetapi dengan standar deviasi yang berbeda-beda, sesuai dengan besarnya sampel. Apabila sampel semakin besar, maka distribusi t mendekati normal (Syahza, 2015).

Distribusi t dapat digambarkan sebagai berikut. Misalkan sampel acak n pengamatan diambil dari populasi normal dengan rataan (μ) dan variansi (σ^2). Tiap pengamatan $X_i, i = 1,2, \dots, n,$ dari sampel acak tersebut akan berdistribusi normal yang sama dengan populasi yang diambil sampelnya. Umumnya distribusi normal cukup baik bila $n \geq 30$, sedangkan untuk $n < 30$ maka digunakan distribusi t. Bila ukuran sampelnya kecil, nilai S^2 berubah cukup besar dari sampel ke sampel lainnya dan distribusi peubah acak $(\overline{X} - \mu)/(\sigma/\sqrt{n}\,)$ menyimpang cukup jauh dari distribusi normal baku. Maka akan dibentuk distribusi yang dinamakan T, dengan

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{2.4}$$

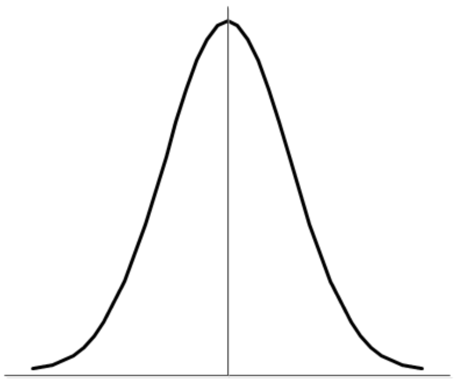
Misalkan Z peubah acak normal baku dan V peubah acak chi-kuadrat dengan derajat kebebasan v. Bila Z dan V bebas, maka bentuk peubah acak T,

$$T = \frac{Z}{\sqrt{V/v}}, \tag{2.5}$$

akan memiliki fungsi peluang:

$$h(t) = \frac{\Gamma[(v+1)/\,2]}{\Gamma(v\,/\,2)\sqrt{\pi v}}\left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}, -\infty < t < \infty. \tag{2.6}$$

Ini dikenal dengan nama distribusi t dengan derajat kebebasan v .



Gambar 2.4 kurva distribusi t

(Walpole & Myers, 1995).

2.4 Metode Maximum Likelihood

Misalkan $Y_1, Y_2,..., Y_n$ adalah peubah acak saling bebas dari populasi yang memiliki fungsi kepadatan peluang $f(y;\theta)$ dengan θ adalah parameter yang tidak diketahui yang merupakan parameter-parameter yang akan diduga, maka fungsi likelihoodnya adalah:

$$L(\theta) = L(\theta; y_1, y_2, \dots, y_n) = \prod_{i=1}^{n} f(y_i;\theta) \tag{2.7}$$

Fungsi log-likelihood dapat ditulis dalam bentuk:

$$l = log\ L(\theta) \tag{2.8}$$

Untuk memaksimumkan fungsi log-likelihood, diperoleh dengan cara menurunkan fungsi log-likelihood terhadap θ kemudian disamakan dengan nol.

$$\frac{\partial}{\partial\theta} log[L(\theta)] = 0 \tag{2.9}$$

(Montgomery,2001)

2.5 Expectation-Maximization

Algoritma EM pertama kali diperkenalkan oleh Dempster dkk pada tahun 1977. Secara garis besar, algoritma EM adalah algoritma untuk menduga suatu parameter dalam suatu fungsi dengan menggunakan MLE, di mana fungsi tersebut mengandung data yang tidak lengkap (Mclachlan & Krishnan, 2008). Algoritma EM merupakan proses yang terbagi atas dua langkah yaitu :

1. Langkah *Expectation* (E-step)

   Pencarian nilai ekspektasi untuk fungsi likelihood berdasarkan variabel yang diamati.

2. Langkah *Maximization* (M-Step)

   Pencarian MLE dari parameter-parameter dengan memaksimumkan ekspektasi likelihood yang dihasilkan dari E-step.

Misalkan kita anggap ada sampel dari n item dimana $n_1$dari item tersebut teramati sementara $n_2 = n - n_1$ item tidak teramati. Item yang teramati dilambangkan dengan $X' = (X_1, X_2, .., X_N)$ dan item yang tidak teramati dilambangkan dengan $Z' = (Z_1, Z_2, .., Z_N)$. Asumsikan X_{is} adalah variable saling bebas dan berdistribusi identic (*independent and identically distribution*) dengan fungsi kepadatan peluang $f(x|\theta)$, dimana $\theta \epsilon \Omega$. Asumsikan X_{is} dan Z_{is} saling bebas. Mari kita lambangkan fungsi kepadatan peluang gabungan dari X dengan $g(x|\theta)$. Kemudian $h(x, z|\theta)$ untuk fungsi kepadatan peluang gabungan untuk data yang teramati dan tidak teramati. Sedangkan $k(z|\theta, x)$ melambangkan notasi fungsi kepadatan peluang bersyarat dari data yang tidak teramati untuk memberikan data yang teramati. Maka dapat kita peroleh

$$k(z|\theta, x) = \frac{h(x, z|\theta)}{g(x|\theta)} \tag{2.10}$$

Fungsi Likelihood data yang teramati yaitu

$$L(\theta|x) = g(x|\theta) \tag{2.11}$$

Kemudian fungsi likelihood untuk data lengkap didefinisikan dengan

$$L^c(\theta|x, z) = h(x, z|\theta) \tag{2.12}$$

Tujuan kita adalah memaksimalkan fungsi likelihood $L(\theta|x)$ dengan menggunakan fungsi likelihood lengkap $L^c(\theta|x, z)$ didalam proses. Gunakan persamaan $h(x, z|\theta)$, kita peroleh

$$\log\ L(\theta|x) = \int \log L(\theta|x).\, k(z|\theta, x) dz$$

$$\log\ L(\theta|x) = \int \log g(x|\theta).\, k(z|\theta, x) dz$$

$$\log\ L(\theta|x) = \int [\log h(x, z|\theta) - \log k(z|\theta, x)]\,.\, k(z|\theta, x) dz$$

$$\log\ L(\theta|x) = \int \log h(x, z|\theta).\, k(z|\theta, x) dz - \int \log k(z|\theta, x).\, k(z|\theta, x) dz$$

$$\log\ L(\theta|x) = E_{\theta_0}[\log L^c(\theta|x,z)|\theta_0,x] - E_{\theta_0}[\log k(z|\theta,x)|\theta_0,x] \tag{2.13}$$

Dimana ekspektasi diambil di bawah fungsi kepadatan peluang bersyarat dari $k(z|\theta,x)$. Kemudian mendefinisikan bagian pertama di sisi kanan pada fungsi di atas

$$Q(\theta|\theta_0,x) = E_{\theta_0}[\log L^c(\theta|x,z)|\theta_0,x] \tag{2.14}$$

Ekspektasi yang didefinisikan fungsi Q dinamakan *E-Step* dari Algoritma EM. Di sini akan dimaksimalkan $\log L(\theta|x)$. Dilambangkan $\hat{\theta}^{(0)}$ inisial estimasi dari θ, berdasarkan pada fungsi likelihood teramati. Kemudian $\hat{\theta}^{(1)}$ menjadi argumen yang memaksimalkan $Q\left(\theta\middle|\hat{\theta}^{(0)},x\right)$. Ini adalah langkah pertama untuk mengestimasi θ kemudian kita definisikan algoritma EM sebagai berikut.

Dilambangkan $\hat{\theta}^{(m)}$ dalam mengestimasi langkah ke-m. Kemudian untuk mengestimasi langkah ke $(m+1)$:

1. Langkah *Expectation* (E-*step*)

$$Q\left(\theta\middle|\hat{\theta}^{(m)},x\right) = E_{\hat{\theta}^{(m)}}\left[\log L^c(\theta|x,z)\middle|\hat{\theta}^{(m)},x\right] \tag{2.15}$$

   Dimana ekspektasi diambil dari fungsi kepadatan peluang bersyarat $k\left(z\middle|\hat{\theta}^{(m)},x\right)$.

2. Langkah *Maximization* (M-*Step*)

$$\hat{\theta}^{(m+1)} = Argmax\ Q\left(\theta\middle|\hat{\theta}^{(m)},x\right) \tag{2.16}$$

Dimana parameter yang dihasilkan dari M-*step* akan digunakan kembali untuk E-*step* yang berikutnya, dan langkah ini akan diulang terus hingga memberikan nilai yang konvergen. Penetuan bahwa parameter yang didapatkan telah konvergen didapatkan dari persamaa berikut:

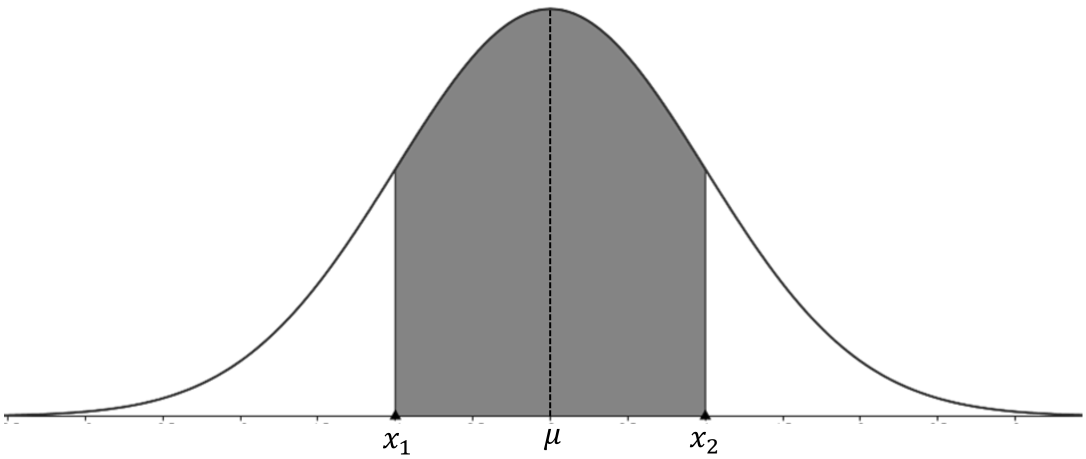
$$L\left(\hat{\theta}^{(m+1)}\middle|x\right) \geq L\left(\hat{\theta}^{(m)}\middle|x\right) \tag{2.17}$$

(Hogg, McKean dan Craig, 2013)

2.6 Luas di Bawah Kurva

Kurva setiap distribusi peluang kontinu atau fungsi padat dibuat sedemikian rupa, sehingga luas di bawah kurva di antara kedua ordinat $x = x_1$ dan $x = x_2$ sama dengan peluang peubah acak X mendapat nilai antara $x = x_1$ dan $x = x_2$ (Walpole & Myers, 1995). Dengan mengintegralkan fungsi distribusi akan didapatkan luas daerah distribusi tersebut. Secara umum dapat jabarkan sebagai berikut:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$



Gambar 2.5 $P(x_1 < X < x_2) =$ luas daerah antara x_1 dan x_2

Dapat dilihat dari Gambar 2.5 bentuk kurva dan daerah yang akan dicari luasnya. Bentuk dan posisi kurva selalu bergantung kepada μ dan σ. Luas daerah yang diarsir merupakan nilai yang akan menjadi patokan untuk mendapatkan nilai x_1 dan x_2.

BAB III

METODE PENELITIAN

3.1 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini, yaitu:

1. Menuliskan bentuk fungsi log likelihood dari data yang teramati dan tidak teramati, persamaan (2.12) $L^c(\theta|x,z) = h(x,z|\theta)$.
2. Menuliskan fungsi data yang tidak teramati, $k(Z|X,\theta)$.
3. Langkah *Expectation*: Menghitunng nilai ekspektasi dari persamaan (2.15):
$$Q\left(\theta\middle|\hat{\theta}^{(m)},x\right) = E_{\hat{\theta}^{(m)}}\left[\log L^c(\theta|x,z)\middle|\hat{\theta}^{(m)},x\right]$$
Dimana ekspektasi diambil dari fungsi kepadatan peluang bersyarat $k\left(z\middle|\hat{\theta}^{(m)},x\right)$.
4. Langkah *Maximization*: Memperbarui nilai dari estimasi $\boldsymbol{\theta}$ dengan memaksimalkan nilai yang didapat pada tahap ekspektasi yang dapat diekspresikan melalui rumus persamaan (2.16) berikut:
$$\hat{\theta}^{(m+1)} = Argmax\ Q\left(\theta\middle|\hat{\theta}^{(m)},x\right)$$
5. Melakukan iterasi hingga terpenuhi syarat persamaan (2.17) berikut:
$$L\left(\hat{\theta}^{(m+1)}\middle|x\right) \geq L\left(\hat{\theta}^{(m)}\middle|x\right)$$

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas tentang cara menaksir parameter pada distribusi t dengan empat parameter menggunakan algoritma EM.

4.1. Distribusi

Beberapa fungsi distribusi yang akan digunakan dalam bab ini, antara lain sebagai berikut :

1. Fungsi distribusi normal

   Persamaan berikut merupakan fungsi distribusi normal yang bergantung pada variabel μ dan σ secara umum dituliskan sebagai berikut :

$$Normal\ (x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty \tag{4.1}$$

   Fungsi distribusi normal yang akan digunakan untuk penurunan rumus ditribusi t dan rumus-rumus yang digunakan dalam algoritma EM adalah fungsi distribusi normal yang bergantung pada variabel mean $= \mu$ dan variansi $= (\lambda\eta)^{-1}$. Distribusi normal ini akan memuat nilai data yang teramati, yang membedakan distribusi ini dengan distribusi normal pada umumnya adalah variansinya berbentuk $(\lambda\eta)^{-1}$. Distribusi normal yang bergantung pada variabel μ dan $(\lambda\eta)^{-1}$ dapat dituliskan sebagai berikut:

$$Normal\ (x|\eta;\mu,\lambda) = \frac{1}{\sqrt{2\pi}\left(\sqrt{\lambda\eta}\right)^{-1}} exp\left[\frac{-(x-\mu)^2}{2(\lambda\eta)^{-1}}\right], -\infty < x < \infty \tag{4.2}$$

2. Fungsi distribusi gamma

   Fungsi distribusi gamma yang bergantung pada variabel α dan β, secara umum dituliskan sebagai berikut :

$$Gamma\ (x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} x^{(\alpha-1)} exp\left[-\frac{x}{\beta}\right], \qquad 0 < x < \infty \tag{4.3}$$

Distribusi gamma yang digunakan dalam penurunan distribusi t dan rumus-rumus yang digunakan dalam algoritma EM menggunakan peubah η. Dimana η memuat data yang tidak teramati. Bentuk distribusi gamma yang digunakan tidak mengalami perubahan, hanya dalam penggunaannya variabel peubahnya memuat data yang tidak teramati. Sehingga fungsi distribusi gamma dengan variabel peubah η sebagai berikut :

$$Gamma(\eta|\alpha,\beta) = \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} \eta^{(\alpha-1)} exp\left[-\frac{\eta}{\beta}\right], 0 < \eta < \infty \tag{4.4}$$

3. Fungsi distribusi t dengan empat parameter.

Misalkan X dinotasikan sebagai sampel acak yang diambil dari populasi normal dengan rataan (μ) dan variansi $(\lambda\eta)^{-1}$; serta η dinotasikan sebagai sampel acak yang diambil dari gamma $(\eta|\alpha,\beta)$; maka X dan η tidak saling bebas. Kemudian fungsi kepadatan peluang bersama dari X dan η adalah $J(x,\eta)$.

$$J(x,\eta) = \frac{1}{\sqrt{2\pi}\left(\sqrt{\lambda\eta}\right)^{-1}} \exp\left[\frac{-(x-\mu)^2}{2(\lambda\eta)^{-1}}\right] \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} \eta^{(\alpha-1)} \exp\left[\frac{-\eta}{\beta}\right]$$

Untuk $-\infty < x < \infty, 0 < \eta < \infty$ dan $J(x,\eta) = 0$ untuk x,η lainnya.

Fungsi kepadatan peluang marjinal t ialah

Untuk $\theta = \{\mu, \lambda, \alpha, \beta\}$

$$t(x|\theta) = \int_{-\infty}^{\infty} J(x,\eta)\, d\eta$$

$$t(x|\theta) = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}\left(\sqrt{\lambda\eta}\right)^{-1}} \exp\left[\frac{-(x-\mu)^2}{2(\lambda\eta)^{-1}}\right] \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} \eta^{(\alpha-1)} \exp\left[\frac{-\eta}{\beta}\right] d\eta$$

$$t(x|\theta) = \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} \left(\frac{\lambda}{2\pi}\right)^{\left(\frac{1}{2}\right)} \int_{0}^{\infty} \eta^{\left(\alpha+\frac{1}{2}-1\right)} \exp\left[-\eta\left(\frac{\lambda(x-\mu)^2}{2} + \frac{1}{\beta}\right)\right] d\eta$$

Berdasarkan bentuk integral dari fungsi gamma, maka didapatkan persamaan distribusi t dengan empat parameter $t\ (x|\mu, \lambda, \alpha, \beta)$ yaitu :

$$t\ (x|\theta) = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{\lambda\beta}{2\pi}\right)^{\left(\frac{1}{2}\right)} \left[1 + \frac{\lambda\beta(x-\mu)^2}{2}\right]^{-\left(\alpha+\frac{1}{2}\right)}, -\infty < t < \infty \qquad (4.5)$$

4.2. Penurunan Rumus Yang Akan Digunakan Dalam Algoritma EM

Sebelum masuk dalam algoritma EM diberikan beberapa turunan fungsi yang akan digunakan antara lain :

1. Fungsi likelihood untuk data yang teramati

   Fungsi distribusi yang digunakaan untuk data yang teramati adalah fungsi distribusi t dengan empat parameter. Data yang teramati diidentifikasikan sebagai $X = \{x_i\}$. Dalam algoritma EM fungsi distribusi dari data yang teramati merupakan fungsi yang akan dicari nilai-nilai parameternya. Dimana parameter yang akan dicari adalah $\theta = \{\mu, \lambda, \alpha, \beta\}$. Fungsi likelihood dari data yang teramati, yaitu sebagai berikut:

$$L(\theta|X) = g(X|\theta)$$

$$g(X|\theta) = \prod_{i=1}^{n} t\ (x_i|\mu, \lambda, \alpha, \beta)$$

$$g(X|\theta) = \prod_{i=1}^{n} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{\lambda\beta}{2\pi}\right)^{\left(\frac{1}{2}\right)} \left[1 + \frac{\lambda\beta(x_i - \mu)^2}{2}\right]^{-\left(\alpha+\frac{1}{2}\right)} \qquad (4.6)$$

Fungsi $g(X|\theta)$ merupakan fungsi likelihood dari distribusi t dengan empat parameter. Salah satu kegunaan dari fungsi $g(X|\theta)$ yaitu digunakan sebagai syarat berhentinya iterasi dalam algoritma EM.

2. Fungsi likelihood untuk data lengkap

Fungsi likelihood lengkap memuat data yang teramati ($X = \{x_i\}$) dan data tidak teramati ($Z = \{\eta_i\}$). Fungsi ini merupakan fungsi likelihood dari distribusi normal dan distribusi gamma yang dijabarkan sebagai berikut:

$$L^c(\theta|X,Z) = h(X,Z|\theta)$$

$$h(X,Z|\theta) = \prod_{i=1}^{N} Normal(x_i|\mu_i(\lambda\eta_i)^{-1})Gamma(\eta_i|\alpha,\beta)$$

Sehingga didapatkan fungsi likelihood untuk data lengkap $h(X,Z|\theta)$ sebagai berikut:

$$h(X,Z|\theta) = \prod_{i=1}^{N} \frac{\sqrt{\lambda\eta_i}}{\sqrt{2\pi}} exp\left[\frac{-\lambda\eta_i(x_i - \mu)^2}{2}\right] \frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}} \eta_i{}^{(\alpha-1)} exp\left[\frac{-\eta_i}{\beta}\right] \qquad (4.7)$$

3. Fungsi log-Likelihood lengkap

Fungsi log-likelihood lengkap merupakan fungsi $h(X,Z|\theta)$. Fungsi ini akan digunakan dalam langkah *expectation*. Sehingga fungsi ini adalah salah satu fungsi yang sangat penting untuk dijabarkan, yaitu sebagai berikut:

$$g\ h(X,Z|\theta) = \sum_{i=1}^{N} log\ Norma(x_i|\mu_i(\lambda\eta_i)^{-1}) + log\ \ Gamma(\eta_i|\alpha,\beta)$$

ehingga didapatkan fungsi log-likelihood untuk data lengkap $log\ h(X,Z|\theta)$ bagai berikut:

$$log\ h(X,Z|\theta) = \sum_{i=1}^{N}\Big(-\frac{1}{2}log\ 2\pi + \frac{1}{2}log\ \lambda + \frac{1}{2}log\ \eta_i - \frac{\lambda\eta_i}{2}(x_i-\mu)^2$$
$$- log\ \Gamma(\alpha) - \alpha\ log\ \beta + (\alpha-1)\ log\ \eta_i - \frac{\eta_i}{\beta}\Big) \tag{4.8}$$

4. Distribusi dari data yang tidak teramati

Distribusi dari data yang tidak teramati dinotasikan sebagai $k(Z|X,\theta)$. Dimana $k(Z|X,\theta)$ proporsional terhadap $h(X,Z|\theta)$, maka persamaan untuk $k(Z|X,\theta)$ dapat dijabarkan sebagai berikut:

$$h(X,Z|\theta) = \prod_{j=1}^{N} Norma\left(x_j|\mu_j(\lambda\eta_j)^{-1}\right) Gamma(\eta_j|\alpha,\beta)$$

$$k(Z|X,\theta) \propto h(X,Z|\theta)$$

$$k(Z|X,\theta) \propto \prod_{j=1}^{N}\frac{\sqrt{\lambda\eta_j}}{\sqrt{2\pi}} exp\left[\frac{-\lambda\eta_j(x_j-\mu)^2}{2}\right]\frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}}\eta_j{}^{(\alpha-1)} exp\left[\frac{-\eta_j}{\beta}\right]$$

$$k(Z|X,\theta) \propto \frac{\sqrt{\lambda}}{\sqrt{2\pi}}\frac{1}{\Gamma(\alpha)(\beta)^{(\alpha)}}\prod_{j=1}^{N}\eta_j{}^{\left(\alpha+\frac{1}{2}-1\right)} exp\left[-\eta_j\left(\frac{\lambda(x_j-\mu)^2}{2}+\frac{1}{\beta}\right)\right]$$

$$k(Z|X,\theta) \propto \prod_{j=1}^{N} Gamma\left(\eta_j\middle|\alpha+\frac{1}{2},\frac{2\beta}{\lambda\beta(x-\mu)+2}\right)$$

Misal $a_j = \alpha + \frac{1}{2}$ dan $b_j = \frac{2\beta}{\lambda\beta(x_j-\mu)^2+2}$, maka

$$k(Z|X,\theta) = \prod_{i=1}^{N} Gamma\left(\eta_j|\alpha_j,\beta_j\right)$$

Dari hasil penjabaran di atas terlihat fungsi $h(X,Z|\theta)$ yang merupakan fungsi
d dari gabungan distribusi normal dan distribusi gamma proporsional
fungsi $k(Z|X,\theta)$ yang merupakan fungsi likelihood dari distribusi gamma
ı. Jadi, fungsi $k(Z|X,\theta)$ dituliskan sebagai berikut:

$$k(Z|X,\theta) = \prod_{j=1}^{N} \frac{1}{\Gamma(a_j)(b_j)^{(a_j)}} {\eta_j}^{(a_j-1)} exp\left[\frac{-\eta_j}{b_j}\right] \tag{4.9}$$

4.3. Algoritma EM

Algoritma EM terbagi atas dua langkah yaitu:

1. Langkah *Expectation*

Pada langkah ini akan dicari fungsi $Q\left(\theta\middle|\hat{\theta}^{(m)},x\right)$ pada persamaan (2.14). Fungsi yang digunakan dalam langkah *expectation* adalah perkalian dari persamaan (4.9) dan persamaan (4.8). Hasil dari perkalian persamaan tersebut kemudian diintegralkan terhadap η. Berdasarkan bentuk integral dari fungsi gamma dan bentuk ekspektasi, maka fungsi $Q\left(\theta\middle|\hat{\theta}^{(m)},x\right)$ dijabarkan sebagai berikut:

$$Q\left(\theta\middle|\hat{\theta}^{(m)},x\right) = E_{\theta_0}[\log L^c(\theta|x,z)|\theta_0,x]$$

$$= \int_0^{\infty} k(Z|X,\theta_0)\ log\ h\,(X,Z|\theta)dz$$

$$= \int_0^{\infty} \prod_{j=1}^{N} \frac{1}{\Gamma(a_j)(b_j)^{(a_j)}} {\eta_j}^{(a_j-1)} \exp\left[\frac{-\eta_j}{b_j}\right] \sum_{i=1}^{N} \Bigg(-\frac{1}{2}\log 2\pi + \frac{1}{2}\log\lambda$$

$$+ \frac{1}{2}\log\ \eta_i - \frac{\lambda\eta_i}{2}(x_i-\mu)^2 - \log\Gamma(\alpha) - \alpha\log\beta$$

$$+ (\alpha-1)\log\eta_i - \frac{\eta_i}{\beta}\Bigg)\,d\eta_i$$

Sehingga didapatkan fungsi $Q\left(\theta\middle|\hat{\theta}^{(m)},x\right)$ dari langkah *expectation* sebagai berikut:

$$Q\left(\theta\middle|\hat{\theta}^{(m)},x\right) = -\frac{N}{2}\log 2\pi + \frac{N}{2}\log\lambda + \frac{1}{2}\sum_{i=1}^{N}E[\log\eta_i]$$

$$-\frac{\lambda}{2}\sum_{i=1}^{N}(x_i-\mu)^2E[\eta_i] - N\log\Gamma(\alpha) - N\,\alpha\log\beta$$

$$+(\alpha-1)\sum_{i=1}^{N}E[\log\eta_i] - \frac{1}{\beta}\sum_{i=1}^{N}E[\,\eta_i]$$

(4.10)

dimana untuk nilai $E[\eta_i]$ dan $E[\log\eta_i]$ merupakan ekspektasi dari fungsi $k(Z|X,\theta_0)$. Penjabaran $E[\eta_i]$ dan $E[\log\eta_i]$, yaitu sebagai berikut:

$$E[\eta_i] = \int_0^\infty \eta_i \prod_{j=1}^{N} Gamma\left(\eta_j\middle|a_j,b_j\right)d\eta_i$$

$$= \int_0^\infty \eta_i\,Gamma\,(\eta_i|a_i,b_i)\,d\eta_i$$

$$= \int_0^\infty \eta_i\frac{1}{\Gamma(a_i)(b_i)^{(a_i)}}\eta_i^{(a_i-1)}exp\left[\frac{-\eta_i}{b_i}\right]d\eta_i$$

Misalkan $y = \frac{\eta_i}{b_i}$

$$= \int_0^\infty y.b_i\frac{1}{\Gamma(a_i)(b_i)^{(a_i)}}(y.b_i)^{(a_i-1)}exp[-y]\,b_i\,dy$$

$$= \frac{b_i}{\Gamma(a_i)}\int_0^\infty y^{(a_i+1-1)}exp[-y]\,dy$$

$$= \frac{b_i}{\Gamma(a_i)}\Gamma(a_i+1)$$

$$= a_i.b_i$$

$$= \left(\alpha_0 + \frac{1}{2}\right) . \left(\frac{2\beta_0}{\lambda_0\beta_0(x_i - \mu) + 2}\right)$$

Sehingga persamaan untuk $E[\eta_i]$ sebagai berikut.

$$E[\eta_i] = \frac{(2\alpha_0 + 1)\beta_0}{\lambda_0\beta_0(x_i - \mu)^2 + 2} \tag{4.11}$$

$$E[log\ \eta_i] = \int_0^\infty log\ \eta_i \prod_{j=1}^{N} Gamma\ \left(\eta_j \middle| a_j, b_j\right)\ d\eta_i$$

$$= \int_0^\infty log\ \eta_i\ Gamma\ (\eta_i|a_i, b_i)\ d\eta_i$$

$$= \int_0^\infty log\ \eta_i \frac{1}{\Gamma(a_i)(b_i)^{(a_i)}} \eta_i{}^{(a_i-1)} exp\left[\frac{-\eta_i}{b_i}\right]\ d\eta_i$$

Misalkan $y = \frac{\eta_i}{b_i}$

$$= \int_0^\infty (log\ y{+}log\ b_i) \frac{1}{\Gamma(a_i)(b_i)^{(a_i)}} (y.b_i)^{(a_i-1)} exp[-y]\ b_i\ dy$$

$$= \int_0^\infty log\ y \frac{b_i(b_i)^{(a_i-1)}}{\Gamma(a_i)(b_i)^{(a_i)}} (y)^{(a_i-1)} exp[-y]\ dy + \int_0^\infty log\ b_i \frac{b_i(b_i)^{(a_i-1)}}{\Gamma(a_i)(b_i)^{(a_i)}} (y)^{(a_i-1)} exp[-y]\ dy$$

$$= \frac{1}{\Gamma(a_i)} \int_0^\infty log\ y\ (y)^{(a_i-1)} exp[-y]\ dy + log\ b_i \int_0^\infty \frac{1}{\Gamma(a_i)} (y)^{(a_i-1)} exp[-y]\ dy$$

$$= \frac{\Gamma'(a_i)}{\Gamma(a_i)} + log\ b_i$$

Untuk $\frac{\Gamma\prime(a_i)}{\Gamma(a_i)} = \psi(a_i)$

Sehingga persamaan untuk $E[log\ \eta_i]$ sebagai berikut:

$$E[log\ \eta_i] = \psi\left(\alpha_0 + \frac{1}{2}\right) + log\ \left(\frac{2\beta_0}{\lambda_0\beta_0(x_i-\mu)^2+2}\right) \tag{4.12}$$

2. Langkah *Maximization*

Pada langkah ini akan dimaksimumkan persamaan (4.10) untuk mendapatkan penaksir parameter $\theta = \{\mu, \lambda, \alpha, \beta\}$. Metode yang digunakan untuk mencari penaksir parameter $\mu, \lambda, \alpha, \beta$ adalah metode turunan. Persamaan (4.10) akan diturunkan terhadap μ, λ, α,dan β sehingga didapatkanlah fungsi penaksir untuk setiap parameter $\mu, \lambda, \alpha, \beta$.

Langkah *Maximization* untuk mendapatkan masing-masing penaksir parameter $\mu, \lambda, \alpha, \beta$, yaitu sebagai berikut:

1) Penaksir untuk parameter μ sebagai berikut :

$$\frac{\partial Q}{\partial \mu} = 0$$

$$\lambda\sum_{i=1}^{N}(x_i-\mu)E[\eta_i] = 0$$

$$\lambda\left(\sum_{i=1}^{N}x_i\ E[\eta_i] - \sum_{i=1}^{N}\mu\ E[\eta_i]\right) = 0$$

$$\lambda\mu\sum_{i=1}^{N}E[\eta_i] = \lambda\sum_{i=1}^{N}x_i\ E[\eta_i]$$

Maka diperoleh persamaan penaksir untuk mencari nilai μ sebagai berikut :

$$\mu = \frac{\sum_{i=1}^{N} x_i \, E[\eta_i]}{\sum_{i=1}^{N} E[\eta_i]} \tag{4.14}$$

2) Penaksir untuk parameter λ sebagai berikut :

$$\frac{\partial Q}{\partial \lambda} = 0$$

$$\frac{N}{2\lambda} - \frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^2 E[\eta_i] = 0$$

$$\frac{N}{2\lambda} = \frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^2 E[\eta_i]$$

Maka diperoleh persamaan penaksir untuk mencari nilai λ sebagai berikut :

$$\lambda = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 E[\eta_i]\right)^{-1} \tag{4.15}$$

Nilai μ yang digunakan dalam persamaan (4.15) adalah nilai μ yang baru didapatkan dari persamaan (4.14).

3) Penaksir untuk parameter α sebagai berikut :

$$\frac{\partial Q}{\partial \alpha} = 0$$

$$-N\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - N \log \beta + \sum_{i=1}^{N} E[\log \eta_i] = 0$$

Maka diperoleh persamaan penaksir untuk mencari nilai α sebagai berikut :

$$N\psi(\alpha) + N\ log\ \beta - \sum_{i=1}^{N} E[log\ \eta_i] = 0 \quad (4.16)$$

4) Penaksir untuk parameter β sebagai berikut :

$$\frac{\partial Q}{\partial \beta} = 0$$

$$-\frac{N\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{N} E[\ \eta_i] = 0$$

$$\frac{N\alpha}{\beta} = \frac{1}{\beta^2}\sum_{i=1}^{N} E[\ \eta_i]$$

$$\frac{\beta^2}{\beta} = \frac{1}{N\alpha}\sum_{i=1}^{N} E[\ \eta_i]$$

Maka diperoleh persamaan penaksir untuk mencari nilai β sebagai berikut :

$$\beta = \frac{1}{N\alpha}\sum_{i=1}^{N} E[\ \eta_i] \quad (4.17)$$

Karena dalam persamaan (4.16) terdapat parameter β, maka akan dilakukan subtitusi persamaan (4.17) ke dalam persamaan (4.16) untuk mencari parameter α yang dijabarkan sebagai berikut:

$$N\psi(\alpha) + N\ log\ \beta - \sum_{i=1}^{N} E[log\ \eta_i] = 0$$

$$N\psi(\alpha) + N\log\left(\frac{1}{N\alpha}\sum_{i=1}^{N} E[\,\eta_i]\right) - \sum_{i=1}^{N} E[log\,\eta_i] = 0$$

Sehingga didapatkan persamaan:

$$\psi(\alpha) + \ log\ \frac{1}{N}\sum_{i=1}^{N} E[\,\eta_i] + log\ \frac{1}{\alpha} - \frac{1}{N}\sum_{i=1}^{N} E[log\,\eta_i] = 0 \qquad (4.18)$$

Kemudian dari persamaan (4.18) akan didapat nilai parameter α menggunakan metode numerik. Nilai α yang baru akan dimasukkan ke dalam persamaan (4.17) untuk mendapatkan nilai parameter β. Sehingga dari langkah *maximization* ini akan didapatkan parameter $\mu, \lambda, \alpha, \beta$ yang baru atau dilambangkan $\hat{\theta}^{(m+1)}$. Hal tersebut dilakukan hingga terpenuhi syarat berikut $L\left(\hat{\theta}^{(m+1)}\middle|x\right) \geq L\left(\hat{\theta}^{(m)}\middle|x\right)$ persamaan (1.17).

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan di atas maka dapat ditarik kesimpulan sebagai berikut:

1. Fungsi distribusi t dengan empat parameter sebagai berikut:

$$t\ (x|\alpha,\beta,\mu,\lambda) = \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)}\left(\frac{\lambda\beta}{2\pi}\right)^{\left(\frac{1}{2}\right)}\left[1+\frac{\lambda\beta(x-\mu)^2}{2}\right]^{-\left(\alpha+\frac{1}{2}\right)} \quad -\infty < t < \infty$$

2. Model likelihood untuk distribusi t dengan empat parameter sebagai berikut:

$$L(\alpha,\beta,\mu,\lambda) = \prod_{i=1}^{n}\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)}\left(\frac{\lambda\beta}{2\pi}\right)^{\left(\frac{1}{2}\right)}\left[1+\frac{\lambda\beta(x_i-\mu)^2}{2}\right]^{-\left(\alpha+\frac{1}{2}\right)}$$

   Dengan menggunakan algoritma EM, didapatkan penaksir parameter pada distribusi t dengan empat parameter dengan persamaan sebagai berikut:

   A. $\mu = \frac{\sum_{i=1}^{N} x_i\, E[\eta_i]}{\sum_{i=1}^{N} E[\eta_i]}$

   B. $\lambda = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 E[\eta_i]\right)^{-1}$

   C. $\psi(\alpha) + \ log\ \frac{1}{N}\sum_{i=1}^{N} E[\ \eta_i] + log\ \frac{1}{\alpha} - \frac{1}{N}\sum_{i=1}^{N} E[log\ \eta_i] = 0$

      Nilai α dari persamaan pada poin C dapat dicari menggunakan metode numerik.

   D. $\beta = \frac{1}{N\alpha}\sum_{i=1}^{N} E[\ \eta_i]$

5.2 Saran

...am pengaplikasiannya, nilai α dapat dicari dengan menggunakan metode ... seperti bisection, newton raphson dan secant. Namun pada kenyataannya tidak didapatkan nilai α yang sesuai, disebabkan metode numerik yang digunakan atau data yang tidak cocok. Oleh karena itu, pada penelitian selanjutnya disarankan untuk menggunakan metode IRLS (*Iteratively Reweighted Least Squares*) dalam mencari nilai α.

DAFTAR PUSTAKA

Hogg, R. V., McKean J. W., & Craig, A. T. (2013). *Introduction to Mathematical Statistics Seventh edition*. United States of America : Pearson Education.

Mclachlan, Geoffrey J. & Krishnan, Thriyambakam. (2008). *The EM algorithm and extensions.* New Jersey: John Wiley & Sons.

Montgomery, Douglas C., (2001). *Design and Analysis of Experiments*, 5^{th} Edition. New York: John Wiley & Sons.

Supranto, J. (2001). Statistika : *Teori dan Aplikasi Edisi Keenam*. Jakarta: Erlangga.

Syahza, Almasdi. (2015). *Metodologi Penelitian.* Bandung: UPI.

Walpole, Ronald E.& Myers, Raymond H. (1995). *Ilmu Peluang Dan Statistika untuk Insinyur dan Ilmuawan*. Bandung: ITB.

Wrede, Robert & R. Spiegel, Murray. (2010). *Schaum's Outlines Advance Calculus Third Edition*. United States: The McGraw-Hill Companies.

LAMPIRAN

Lampiran 1

Tabel Distribusi T

Tabel berikut ini menampilkan beberapa nilai dari distribusi t dengan empat parameter, yaitu nilai t sedemikian sehingga :

$$P(T \leq t) = \int_{-\infty}^{t} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)} \left(\frac{\lambda\beta}{2\pi}\right)^{\left(\frac{1}{2}\right)} \left[1 + \frac{\lambda\beta(x-\mu)^2}{2}\right]^{-\left(\alpha+\frac{1}{2}\right)} dx$$

Dengan syarat $\beta = 2,\ \mu = 0$ dan $\lambda = \frac{1}{2\alpha}$, maka:

α	$P(T \leq t)$				
	0,005	**0,01**	**0,025**	**0,05**	**0,1**
1	9,925	6,965	4,303	2,920	1,886
2	4,604	3,747	2,776	2,132	1,533
3	3,707	3,143	2,447	1,943	1,440
4	3,355	2,896	2,306	1,860	1,397
5	3,169	2,764	2,228	1,812	1,372
6	3,055	2,681	2,179	1,782	1,356
7	2,977	2,624	2,145	1,761	1,345
8	2,921	2,583	2,120	1,746	1,337
9	2,878	2,552	2,101	1,734	1,330
10	2,845	2,528	2,086	1,725	1,325
11	2,819	2,508	2,074	1,717	1,321
12	2,797	2,492	2,064	1,711	1,318
13	2,779	2,479	2,056	1,706	1,315
14	2,763	2,467	2,048	1,701	1,313
15	2,750	2,457	2,042	1,697	1,310
16	2,738	2,449	2,037	1,694	1,309
17	2,728	2,441	2,032	1,691	1,307
18	2,719	2,434	2,028	1,688	1,306
19	2,712	2,429	2,024	1,686	1,304
20	2,704	2,423	2,021	1,684	1,303
21	2,698	2,418	2,018	1,682	1,302
22	2,692	2,414	2,015	1,680	1,301
23	2,687	2,410	2,013	1,679	1,300
24	2,682	2,407	2,011	1,677	1,299
25	2,678	2,403	2,009	1,676	1,299
26	2,674	2,400	2,007	1,675	1,298
27	2,670	2,397	2,005	1,674	1,297
28	2,667	2,395	2,003	1,673	1,297
29	2,663	2,392	2,002	1,672	1,296