ESTIMASI PARAMETER DATA BERDISTRIBUSI NORMAL MENGGUNAKAN MAKSIMUM LIKELIHOOD BERDASARKAN NEWTON RAPHSON

ESTIMATION OF NORMAL DISTRIBUTED DATA PARAMETERS USING THE MAXIMUM LIKELIHOOD BASED ON NEWTON RAPHSON

Switamy Angnitha Purba
Program Studi Matematika, Universitas HKBP Nommensen Pematangsiantar, Medan, Indonesia
*email korespondensi: switamyangnithapurba@gmail.com

Abstrak

Estimasi parameter adalah praktik umum dalam statistik. Maximum Likelihood adalah metode estimasi parameter berdasarkan pendekatan distribusi dengan cara memaksimalkan fungsi likelihood. Mean, deviasi standar, proporsi dan lain-lain merupakan perkiraan nilai parameter dengan menggunakan data atau sampel yang dapat diambil dari populasi tersebut. Algoritma Newton Raphson adalah prosedur iteratif yang digunakan untuk menyelesaikan persamaan non-linier. Fokus makalah ini adalah mengestimasi parameter data yang berdistribusi normal menggunakan Maximum Likelihood berdasarkan algoritma iterasi Newton Raphson dengan program matlab.

Kata kunci: estimasi parameter, maximum likelihood, newton raphson

Abstract

The parameter estimation is a common practice in statistics. Maximum Likelihood is a method of estimating parameters based on the distribution approach by maximizing the likelihood function. Mean, standasrt deviation, proportion and others are estimates of parameter value using data or samples that can be taken from the population. Newton Raphson's Algorithm is an iterative procedure used to solve non- linear equations. The focus in this paper is to estimate data parameter that are normally distributed using Maximum Likelihood based on Newton Raphson iteration algorithm with matlab program..

Keywords: parameter estimation, maximum likelihood, newton raphson

Pendahuluan

Estimasi maksimum *likelihood* adalah sebuah metode yang memaksimumkan fungsi *likelihood* untuk memperoleh penaksir parameter dengan kemungkinan maksimum. Dengan begitu diperoleh bentuk implisit dan non linear yang dapat digunakan untuk menyelesaikan algoritma *newton raphson*. *Maksimum Likelihood estimation* adalah metode estimasi parameter pada gugus data yang memiliki sebaran distribusi. Algoritma *Newton Raphson* merupakan prosedur perulangan dalam mencari solusi optimal persamaan nonlinear dengan memanfaatkan vektor turunan pertama dan matriks turunan kedua dari fungsi. Algoritma *Newton Raphson* hampir sama dengan algoritma *Fisher Scoring* akan tetapi yang membedakannya adalah pada *Fisher Scoring* menggunakan nilai ekspektasi dari matriks turunan kedua [4].

Metode *Maksimum Likelihood* merupakan metode yang menggunakan nilai pada ruang parameter Ω untuk mengestimasi nilai parameter yang tidak diketahui. Bain dan engelhard mengatakan bahwa $L(\theta)$ merupakan fungsi peluang dari sampel random dan oleh karena Ω merupakan interval terbuka maka $L(\theta)$ adalah sebuah fungsi yang bisa diturunkan dan diperkirakan dapat maksimum pada Ω [1]. Estimasi parameter perlu dilakukan untuk memperoleh parameter populasi yang tidak diketahui dengan menggunakan data sampelnya. Beberapa riset menggunakan *maksimum likelihood* sudah banyak dilakukan, salah satunya adalah Bolstad yang melakukan estimasi parameter menggunakan data tersensor yang berdistribusi gamma menggunakan algoritma *Expectation Maximation* (EM) [3].

Salah satu kegunaan algoritma Newton Raphson dalam permasalahan statistika adalah bahwa loglikelihood dalam banyak permasalahan mendekati fungsi kuadratik yang menyebar dititik maksimumnya. Hal itulah yang dihubungkan dengan prakiraan kemungkinan maksimumnya yang terdistribusi normal bahwa logaritma dari distribusi normal adalah sebuah fungsi kuadratik. Oleh karena itu, prakiraan dengan *loglikelihood* adalah sebuah pendekatan yang sangat bagus dan mendekati titik maksimumnya [5]. Fokus dalam penulisan ini adalah untuk mengestimasi parameter data yang berdistribusi normal menggunakan *Maximum Likelihood* berdasarkan algoritma iterasi *Newton Raphson* dengan bantuan program Matlab.

Metode Penelitian

Fokus pada penelitian ini adalah melakukan estimasi parameter pada suatu data berdistribusi normal menggunakan *Maksimum Likelihood* berdasarkan *Newton Raphson*. Oleh karena itu, pada penulisan ini akan mengambil sebuah contoh data berdistribusi normal. Selanjutnya, (a) menemukan fungsi *likelihood* dari data berdistribusi normal. (b) Menemukan fungsi *log-likelihood* dari data berdistribusi normal. (c) Melakukan estimasi parameter dengan menggunakan *Maksimum Likelihood* berdasarkan algoritma *Newton Raphson* $\theta^{(k+1)} = \theta^{(k)} - H\left(\theta^{(k)}\right)^{-1} l'\left(\theta^{(k)}\right)$. $H\left(\theta^{(k)}\right)$ adalah matriks Hessian $s \times s$ dari turunan kedua $l(\theta)$ dievaluasi pada saat $\theta^{(k)}$ dan $l'\left(\theta^{(k)}\right)$ adalah turunan pertama dari *loglikelihood*. (d) Membangun Algoritma *Newton Raphson* pada program Matlab. (e) Mengaplikasikan contoh data berdistribusi normal yang akan digunakan untuk mencari estimasi parameter data dari algoritma *Newton Raphson* pada program Matlab. (f) Membandingkan hasil estimasi parameter data berdistribusi Normal berdasarkan *Newton Raphson* dengan Matlab dan SPSS.

Hasil dan Diskusi

Data Berdistribusi Normal

Pada penulisan ini akan dilakukan estimasi parameter pada sebuah data berdistribusi normal.

Tabel 1. Data Berdistribusi Normal

No	Tahun	Jumlah
1	2005	7
2	2006	10
3	2007	20
4	2008	30
5	2009	50
6	2010	67
7	2011	75
8	2012	60
9	2013	48
10	2014	34
11	2015	22
12	2016	14
13	2017	9

Estimasi Parameter Data Berdistribusi Normal Menggunakan Maksimum Likelihood

Sebuah data berdistribusi normal memiliki fungsi kepadatan peluang bersama (pdf) parameter μ dan σ^2 sebagai berikut :

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Fungsi *likelihood* distribusi normal adalah

$$L(\theta | y_1, y_2, \dots, y_N) = \prod_{i=1}^{N} L(\theta | y_i)$$

$$= \prod_{i=1}^{N} f(y_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} exp\left(-\sum_{i=1}^{N} \frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

Fungsi *log likelihood* distribusi normal adalah

$$l(\theta | y_1, y_2, \dots, y_N) = \log L(\theta | y_1, y_2, \dots, y_N)$$

$$= log\left[\prod_{i=1}^{N} L(\theta | y_i)\right]$$

$$= ln\left[(2\pi\sigma^2)^{-\frac{N}{2}} exp\left(-\sum_{i=1}^{N} \frac{(y_i - \mu)^2}{2\sigma^2}\right)\right]$$

$$= -\frac{N}{2} log(2\pi\sigma^2) - \sum_{i=1}^{N} \frac{(y_i - \mu)^2}{2\sigma^2}$$

Dalam penelitian ini dilakukan estimasi parameter μ dan σ^2 dengan algoritma *Newton Raphson.*

Maksimum Likelihood Berdasarkan Algoritma Newton Raphson

Pada algoritma *Newton Raphson* perlu turunan pertama fungsi *log likelihood* $l'(\theta)$ dan matriks Hessian $H\left(\theta^{(k)}\right)$. Untuk diaplikasikan pada persamaan algoritma *newton raphson*:

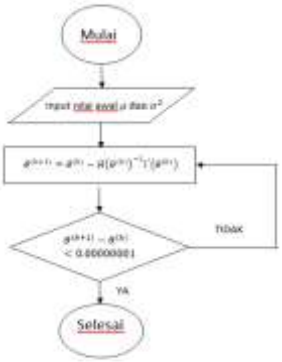
$$\theta^{(k+1)} = \theta^{(k)} - H\left(\theta^{(k)}\right)^{-1} l'\left(\theta^{(k)}\right)$$

Sehingga diperoleh bentuk algoritma metode *newton raphson* untuk populasi data kecelakaan yang berdistribusi normal adalah sebagai berikut:

$$= \binom{\mu}{\sigma^2}^k - \begin{bmatrix} -\frac{N}{\sigma^2} & -\frac{\sum_{i=1}^{N} -y_i + N\mu}{\sigma^4} \\ -\frac{\sum_{i=1}^{N} -y_i + N\mu}{\sigma^4} & \frac{N}{2\sigma^4} - \frac{\sum_{i=1}^{N} (y_i - \mu)^2}{\sigma^6} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^{N} (y_i - \mu) \\ -\frac{N}{2\sigma^2} + \sum_{i=1}^{N} \frac{(y_i - \mu)^2}{2\sigma^4} \end{bmatrix}$$

Membangun Algoritma Maksimum Likelihood Berdasarkan Newton Raphson Pada Matlab

Algoritma *newton raphson* yang dibangun adalah algoritma iterasi dengan nilai awal estimasi parameternya adalah sembarang dengan selisih perbedaannya adalah 0.00000001. Jadi, membutuhkan waktu yang lama untuk mendapatkan mengestimasi parameter tersebut. Oleh karena itu, pada penelitian ini menggunakan Program Matlab. Berikut diagram alir algoritma *Maksimum Likelihood* berdasarkan *Newton Raphson*.



Gambar 1. Diagram Alir Newton Raphson

Hasil Algoritma Maksimum Likelihood berdasarkan Newton Raphson

Dari algoritma yang telah dibangun diperoleh hasil iterasi sebagai berikut.

Tabel 2. Hasil $\mu\ dan\ \sigma^2$ menggunakan Matlab

Iterasi	μ	σ^2
1	32.2059568	70.7757296
2	33.3456395	103.1727614
3	33.8822083	148.8026415
4	34.1316820	210.3579703
5	34.2427661	287.9541746
6	34.2881272	374.7276378
7	34.3036555	452.0446314
8	34.3073012	495.8408571
9	34.3076843	505.9381605
10	34.3076923	506.3661386
11	34.3076923	506.3668639

Berikut hasil perhitungan $\mu\ dan\ \sigma^2$ menggunakan SPSS tanpa mengklasifikasikan data tersebut merupakan data berdistribusi normal atau tidak.

Tavel 3. Hasil $\mu\ dan\ \sigma^2$ menggunakan SPSS

Descriptive Statistics

	N	Mean	Variance
VAR00006	13	34.3077	548.564
Valid N (listwise)	13		

Perbandingan hasil estimasi parameter dengan menggunakan metode Newton Raphson dan SPSS.

Tabel 4. Perbandingan hasil Matlab dan SPSS

	μ	σ^2
Matlab	34.3076923	506.3668639
SPSS	34.3077	548,564

Simpulan

Hasil estimasi parameter μ dan σ^2 data berdistribusi normal dengan algoritma *Maksimum Likelihood* berdasarkan *Newton Raphson* sangat baik untuk estimasi parameter μ sedangkan untuk estimasi parameter σ^2 perlu dilakukan studi lebih lanjut lagi karena terdapat perbedaan yang cukup besar. Serta penelitian selanjutnya dapat dilakukan dengan data berdistribusi lainnya dan dengan menggunakan metode yang berbeda.

Ucapan Terima Kasih

Saya berterimakasih kepada Universitas HKBP Nommensen Pematangsiantar. Saya juga berterimakasih kepada kedua orangtua yang telah mendoakan dan memotivasi saya.

Pustaka

[1] Bain, J., dan Engelhardt, M. (1992). *Introduction to probability and mathematical statistics 2nd edition*. Duxbury. USA.

[2] Chan, S. H. (2015). *Expectation maximation algorithm*. University PURDUE.

[3] Bolstad, B.M. (1998). *Comparing some iterative method of parameter estimation for censored gamma data.* The University of Waikato.

[4] Ehlers, R. (2002). *Maximum likelihood estimation prosedures for categorial data.* University of Pretoria, South Africa.

[5] Storvik, G. (2011). *Numerical optimization of likelihoods: additiona literature for STK2120.* University of Oslo.