

**ESTIMASI PARAMETER DISPERSI BINOMIAL NEGATIF
PADA DATA KECELAKAAN LALU LINTAS DENGAN
METODE *BOOTSTRAP MAXIMUM LIKELIHOOD*
(Studi Kasus: Kota Malang)**

SKRIPSI

Oleh:
AYATUL LUTFI
0710940011-94



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2012**

**ESTIMASI PARAMETER DISPERSI BINOMIAL NEGATIF
PADA DATA KECELAKAAN LALU LINTAS DENGAN
METODE *BOOTSTRAP MAXIMUM LIKELIHOOD*
(Studi Kasus: Kota Malang)**

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Matematika

Oleh:

**AYATUL LUTFI
0710940011-94**



**PROGRAM STUDI MATEMATIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2012**

UNIVERSITAS BRAWIJAYA



LEMBAR PENGESAHAN SKRIPSI

**ESTIMASI PARAMETER DISPERSI BINOMIAL NEGATIF
PADA DATA KECELAKAAN LALU LINTAS DENGAN
METODE *BOOTSTRAP MAXIMUM LIKELIHOOD*
(Studi Kasus: Kota Malang)**

Oleh:
AYATUL LUTFI
0710940011-94

Setelah dipertahankan di depan majelis penguji
pada tanggal 9 Februari 2012
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam Bidang Matematika

Pembimbing I

Pembimbing II

Dra. Endang Wahyu H., M.Si
NIP. 196611121991032001

Prof. Dr. Agus Widodo, M.Kes
NIP. 195305231983031002

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr. Abdul Rouf Alghofari, M.Sc
NIP. 196709071992031001

UNIVERSITAS BRAWIJAYA



LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Ayatul Lutfi
NIM : 0710940011-94
Jurusan : Matematika
Penulis Skripsi berjudul :

**Estimasi Parameter Dispersi Binomial Negatif pada Data
Kecelakaan Lalu Lintas dengan Metode *Bootstrap Maximum
Likelihood*
(Studi Kasus: Kota Malang)**

Dengan ini menyatakan bahwa :

1. Isi skripsi yang saya buat benar-benar karya sendiri dan tidak menjiplak karya orang lain. Karya-karya yang tercantum dalam Daftar Pustaka skripsi ini, semata-mata digunakan sebagai acuan atau referensi.
2. Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya bersedia menanggung segala resiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 9 Februari 2012
Yang menyatakan,

(Ayatul Lutfi)
NIM 0710940011-94

UNIVERSITAS BRAWIJAYA



**ESTIMASI PARAMETER DISPERSI BINOMIAL NEGATIF
PADA DATA KECELAKAAN LALU LINTAS DENGAN
METODE *BOOTSTRAP MAXIMUM LIKELIHOOD*
(Studi Kasus: Kota Malang)**

ABSTRAK

Pada skripsi ini dibahas estimasi parameter dispersi dari distribusi Binomial Negatif untuk pemodelan kecelakaan lalu lintas. Data kecelakaan lalu lintas memiliki karakteristik diskrit, non-negatif, sporadis, nilai rata-rata sampel rendah dan ukuran sampel kecil. Selama ini metode yang digunakan untuk mengestimasi parameter dispersi yaitu *Method of Moment Estimate* (MME) dan *Maximum Likelihood Estimate* (MLE). Untuk mengestimasi parameter dispersi dengan data yang memiliki karakteristik rata-rata sampel rendah dan ukuran sampel kecil, kedua metode tersebut tidak akurat dan tidak stabil dengan nilai bias dan *Mean square Error* (MSE) yang relatif besar. Sebuah metode *Bootstrap Maximum Likelihood Estimate* (BMLE) diusulkan untuk memperoleh hasil estimasi yang lebih baik dari parameter dispersi. Ide dari metode tersebut yaitu menggabungkan teknik resampling Bootstrap dengan MLE. Kinerja dari BMLE dibandingkan dengan MME dan MLE melalui simulasi Monte Carlo. Sebagai aplikasinya, metode tersebut diterapkan pada data lapangan yang dikumpulkan di jalan raya rawan kecelakaan Kota Malang. Secara keseluruhan, hasilnya menunjukkan bahwa metode BMLE menghasilkan bias yang lebih kecil dan estimasi yang lebih stabil daripada MME dan MLE. Hasil tersebut lebih terlihat pada saat nilai rata-rata sampel rendah dan ukuran sampel kecil.

Kata kunci: Parameter Dispersi, *Bootstrap Maximum Likelihood Estimate*

UNIVERSITAS BRAWIJAYA



**ESTIMATING THE DISPERSION PARAMETER OF THE
NEGATIVE BINOMIAL DISTRIBUTION
FOR TRAFFIC CRASH DATA USING A BOOTSTRAPPED
MAXIMUM LIKELIHOOD METHOD
(Field Research: Malang City)**

ABSTRACT

This final project discuss the estimation of the dispersion parameter of The Negative Binomial distribution for modeling motor vehicle collisions. Traffic crash data often exhibit discrete, non-negative, sporadic, low sample mean values and small sample sizes. The previously used methods for estimating the dispersion parameter are the Method of Moment Estimate (MME) and the Maximum Likelihood Estimate (MLE), they may become inaccurate and unstable under such situation. A Bootstrapped Maximum Likelihood Estimate (BMLE) is proposed to obtain better estimates of the dispersion parameter. The proposed method combines the technique of Bootstrap resampling with the MLE. The performance of the BMLE is compared with MME and MLE through Monte Carlo simulations. To validate the simulation results, the methods are applied to observed data collected at accident-prone main streets in Malang. Overall, the results show that the proposed BMLE produces smaller biases and more stable estimates than MME and MLE. The improvements are more pronounced with small samples and low sample means.

Keywords: Dispersion Parameter, Bootstrap Maximum Likelihood Estimate

UNIVERSITAS BRAWIJAYA



KATA PENGANTAR

Segala puji dan syukur ke hadirat Allah SWT yang telah memberikan bimbingan dan petunjuk-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **Estimasi Parameter Dispersi Binomial Negatif pada Data Kecelakaan Lalu Lintas dengan Metode *Bootstrap Maximum Likelihood* (Studi Kasus: Kota Malang)** dengan baik dan lancar. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW sebagai suri tauladan bagi penulis.

Skripsi ini tidak dapat diselesaikan dengan baik tanpa bantuan, bimbingan serta motivasi dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Dra. Endang Wahyu H., M.Si, selaku pembimbing I atas segala bimbingan, motivasi serta kesabaran yang telah diberikan selama penulisan skripsi ini.
2. Prof. Dr. Agus Widodo, M.Kes, selaku pembimbing II atas segala bimbingan, motivasi serta kesabaran yang telah diberikan.
3. Kwardiniya A, S.Si, M.Si selaku dosen penasehat akademik sekaligus dosen penguji atas segala nasehat dan saran yang diberikan selama ini.
4. Drs. Marsudi, MS dan Dr. Sobri Abusini, MT, selaku dosen penguji atas segala saran yang diberikan untuk perbaikan skripsi ini.
5. Seluruh bapak/ibu dosen Jurusan Matematika yang telah memberikan ilmunya kepada penulis selama melaksanakan studi.
6. Segenap staf dan karyawan TU Jurusan Matematika atas segala bantuannya.
7. Bapak Gusti selaku Kepala satuan polisi lalu lintas Polresta Malang atas ijin pengambilan data serta informasi yang telah diberikan.
8. Bapak, Ibu, dan semua anggota keluarga besarku atas segala yang telah diberikan berupa dukungan spiritual dan material.
9. Chendolku yang elek dan cupu atas segala kebaikan, kesabaran, bantuan, doa, dukungan, serta rasa sayang yang diberikan selama ini. Penulis hanya bisa menyampaikan "I love you so much".

10. Andri, Fafan, Anggun, Gyndra, Yogie, Bayu dan anggota Himpunan Mahasiswa Kosong Blong (HMKB) yang lain atas kekosongannya selama ini.

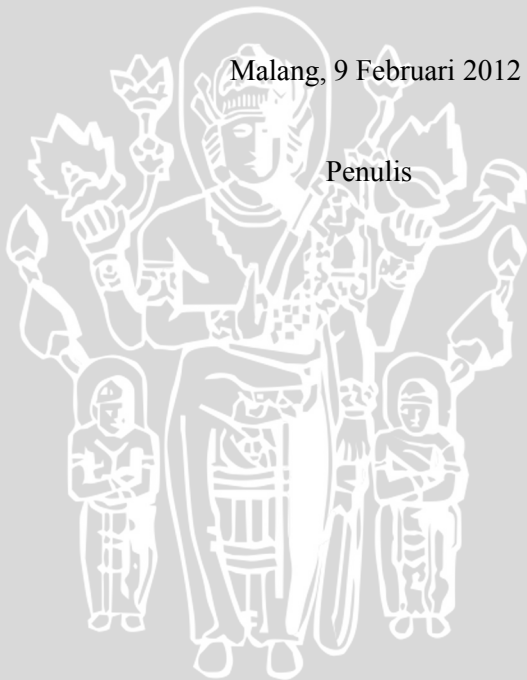
11. Teman-teman yang ada di Jurusan Matematika khususnya Matematika 2007 atas bantuan yang telah diberikan selama ini.

12. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat banyak kekurangan. Untuk itu penulis sangat mengharapkan kritik dan saran melalui email penulis Ayatullutfi@yahoo.com. Akhir kata, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Malang, 9 Februari 2012

Penulis



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Fungsi Kepadatan Peluang.....	5
2.2 Momen.....	5
2.2.1 Momen Sampel.....	6
2.3 Fungsi Pembangkit Momen Faktorial.....	6
2.4 Distribusi Poisson.....	8
2.5 Distribusi Gamma.....	8
2.6 Koefisien Binomial.....	8
2.7 Distribusi Binomial Negatif.....	9
2.8 Estimasi (Pendugaan Parameter).....	13
2.8.1 Sifat-Sifat Estimator.....	14
2.8.1.1 Ketidakbiasan.....	14
2.8.1.2 Rata-Rata Galat Kuadrat (<i>Mean Squared Error</i>).....	16
2.9 Estimasi Parameter Titik Menggunakan <i>Method of Moment Estimate</i>	18

2.10 Estimasi Parameter Titik Menggunakan Metode <i>Maximum Likelihood Estimate</i>	19
2.11 Metode Newton-Raphson.....	21
2.12 Overdispersi	23
2.13 Metode Bootstrap.....	23
2.14 Simulasi Monte Carlo	24
BAB III METODOLOGI PENELITIAN	29
3.1 Deskripsi umum Daerah Studi	29
3.2 Jenis dan Sumber Data	29
3.3 Waktu dan Tempat Penelitian	29
3.4 Metode Pengumpulan Data	29
3.5 Diagram Alir Penelitian	30
3.6 Analisis Data	31
BAB IV HASIL DAN PEMBAHASAN	35
4.1 Estmasi Parameter Binomial Negatif Menggunakan <i>Method of Moment Estimate</i>	35
4.1.1 Fungsi Pembangkit Momen Faktorial Distribusi Binomial Negatif	35
4.2 Estimasi Parameter Binomial Negatif Menggunakan <i>Maximum Likelihood Estimate</i>	40
4.3 Metode <i>Bootstrap Maximum Likelihood</i> Sebagai Estimator k	44
4.4 Simulasi	44
4.5 Hasil Analisis Data Simulasi.....	46
4.6 Aplikasi Menggunakan Data Kecelakaan dari Unit Lakalantas Polresta Malang.....	49
BAB V KESIMPULAN DAN SARAN	53
5.1 Kesimpulan.....	53
5.2 Saran.....	53
DAFTAR PUSTAKA	55
LAMPIRAN	57

DAFTAR TABEL

	Halaman
Tabel 2.1 Beberapa penduga (statistik) dengan parameter targetnya	13
Tabel 2.2 Frekuensi jumlah kecelakaan selama 200 hari	25
Tabel 2.3 Distribusi peluang jumlah kecelakaan	26
Tabel 2.4 Distribusi peluang kumulatif jumlah kecelakaan	26
Tabel 2.5 Penentuan interval angka acak	27
Tabel 2.6 Hasil simulasi	27



UNIVERSITAS BRAWIJAYA



DAFTAR GAMBAR

	Halaman
Gambar 2.1	Ilustrasi metode Newton-Raphson..... 22
Gambar 2.2	Grafik peluang kumulatif 26
Gambar 3.1	Diagram alir penelitian 31
Gambar 3.2	Diagram alir analisis data hasil simulasi 32
Gambar 3.3	Diagram alir analisis data kecelakaan dari Unit Lakalantas Polresta Malang..... 33
Gambar 4.1	Grafik perbandingan bias untuk $k = 1$ dan $\mu = 1$ 46
Gambar 4.2	Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 1$.. 47
Gambar 4.3	Grafik perbandingan bias untuk $k = 4$ dan $\mu = 1$... 48
Gambar 4.4	Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 1$... 49
Gambar 4.5	Grafik perbandingan nilai bias hasil estimasi pada data kecelakaan..... 51
Gambar 4.6	Grafik perbandingan nilai MSE hasil estimasi pada data kecelakaan..... 52



UNIVERSITAS BRAWIJAYA



DAFTAR LAMPIRAN

	Halaman
Lampiran 1	Diagram alir Metode <i>Bootstrap Maximum Likelihood Estimate</i> 57
Lampiran 2	Diagram alir simulasi estimasi parameter dispersi k 58
Lampiran 3	Karakteristik data kecelakaan..... 60
Lampiran 4	Tabel hasil analisis data simulasi..... 61
Lampiran 5	Grafik perbandingan bias dan MSE..... 65
Lampiran 6	Tabel hasil analisis data kecelakaan 81
Lampiran 7	Listing program 82



UNIVERSITAS BRAWIJAYA



BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk menganalisis data kecelakaan dalam analisis keselamatan lalu lintas, distribusi statistik sering digunakan untuk memodelkan data. Poch dan Fred (1996) mengatakan bahwa dalam menganalisis data kecelakaan lalu lintas, sering diasumsikan distribusi jumlah kecelakaan mengikuti distribusi Poisson, karena data kecelakaan memiliki sifat diskrit, non-negatif, sporadis, nilai rata-rata sampel rendah dan ukuran sampel kecil. Distribusi Poisson hanya memiliki satu parameter serta memiliki rata-rata dan varian yang sama. Pada kenyataannya, data kecelakaan lalu lintas memiliki varian yang besar dan rata-rata yang kecil, dan menunjukkan overdispersi.

Distribusi Binomial Negatif, yang juga dikenal sebagai distribusi Poisson-Gamma, telah menjadi distribusi peluang yang kemudian digunakan untuk pemodelan kecelakaan lalu lintas. Distribusi Binomial Negatif dapat menangani overdispersi lebih baik daripada distribusi yang lain karena distribusi Binomial Negatif memuat parameter dispersi di dalamnya. Distribusi Binomial Negatif memiliki dua parameter, yaitu rata-rata μ dan parameter dispersi k .

Data kecelakaan lalu lintas memiliki karakteristik rata-rata sampel yang rendah, atau disebut juga *Low Mean Problem* (LMP). LMP dalam distribusi Binomial Negatif sangat mempengaruhi estimasi parameter dispersi, terutama jika dikombinasikan dengan ukuran sampel yang kecil. Beberapa peneliti dan praktisi data kecelakaan lalu lintas biasanya tidak dapat mengumpulkan data dengan jumlah yang besar. Hal ini disebabkan besarnya biaya dan terbatasnya kondisi di lapangan. Dalam salah satu penelitian (Willson dan Young, 1984), dikatakan bahwa *Method of Moments Estimate* (MME) dan *Maximum Likelihood Estimate* (MLE) memiliki kesulitan dalam mengestimasi parameter dispersi ketika μ kecil dan k besar, karena hasil estimasi dari MME dan MLE cenderung *overestimate* dan tidak stabil.

Metode bootstrap adalah suatu metode resampling, yaitu teknik pengambilan sampel berulang secara acak dari data asli. Metode tersebut akan menjadi dasar dalam memperkirakan karakteristik dari data asli.

Pada skripsi ini akan dibahas tentang estimasi *Bootstrap Maximum Likelihood Estimate* (BMLE) pada data kecelakaan di jalan raya rawan kecelakaan Kota Malang. Ide dari metode BMLE adalah menggabungkan teknik resampling bootstrap dengan metode MLE. Metode BMLE ini akan dibandingkan dengan MME dan MLE melalui simulasi numerik. Diharapkan metode BMLE bisa memberikan hasil estimasi yang lebih baik dari estimator lain pada data yang memiliki karakteristik rata-rata sampel rendah dan ukuran sampel kecil.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya dapat ditarik beberapa rumusan masalah sebagai berikut:

1. Bagaimana algoritma estimasi parameter dispersi Binomial Negatif pada data kecelakaan menggunakan *Bootstrap Maximum Likelihood Estimate*?
2. Bagaimana hasil dan perbandingan estimasi parameter dispersi Binomial Negatif pada data simulasi dan data kecelakaan lalu lintas menggunakan *Method of Moment Estimate*, *Maximum Likelihood estimate*, dan *Bootstrap Maximum Likelihood Estimate*?

1.3 Batasan Masalah

Batasan-batasan masalah dari skripsi ini yaitu:

1. Kinerja dari *Method of Moment Estimate*, *Maximum Likelihood Estimate*, dan *Bootstrap Maximum Likelihood Estimate* dibandingkan melalui simulasi Monte Carlo.
2. Sebagai aplikasinya *Method of Moment Estimate*, *Maximum Likelihood Estimate*, dan *Bootstrap Maximum Likelihood Estimate* diterapkan pada data kecelakaan lalu lintas di jalan raya rawan kecelakaan Kota Malang, yang meliputi Jl. Kolonel Sugiono, Jl. Supriadi, Jl. MT. Haryono dan Jl. Basuki Rachmad pada tahun 2006 hingga 2010.
3. Semua data kecelakaan diasumsikan mengikuti distribusi Binomial Negatif dengan mempertimbangkan besarnya rata-rata dan varian.

1.4 Tujuan penelitian

Tujuan dari skripsi ini adalah:

1. Merumuskan algoritma estimasi parameter dispersi Binomial Negatif pada data kecelakaan menggunakan *Bootstrap Maximum Likelihood Estimate*
2. Menunjukkan hasil dan perbandingan estimasi parameter dispersi Binomial Negatif pada data simulasi dan data kecelakaan lalu lintas menggunakan *Method of Moment Estimate*, *Maximum Likelihood Estimate*, dan *Bootstrap Maximum Likelihood Estimate*.

1.5 Manfaat Penelitian

Adapun manfaat dari skripsi ini yaitu bisa lebih mengerti tentang penerapan distribusi statistik, khususnya distribusi Binomial Negatif dan memberi alternatif dengan memperkenalkan metode lain dalam mengestimasi data yang memiliki karakteristik rata-rata sampel rendah dan ukuran sampel kecil tanpa meningkatkan jumlah sampel data, sehingga mendapatkan hasil yang lebih akurat dan stabil.



UNIVERSITAS BRAWIJAYA



BAB II TINJAUAN PUSTAKA

2.1 Fungsi Kepadatan Peluang

Menurut Walpole dan Myers (1995), suatu variabel acak disebut variabel acak diskrit bila himpunan kemungkinan hasilnya terhitung.

Definisi 2.1

Himpunan pasangan terurut $(x, f(x))$ merupakan suatu fungsi kepadatan peluang variabel acak diskrit X jika untuk setiap kemungkinan hasil x

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x)$.

Variabel acak yang dapat memperoleh semua nilai pada skala kontinu disebut variabel acak kontinu.

Definisi 2.2

Fungsi $f(x)$ adalah fungsi kepadatan peluang variabel acak kontinu X yang didefinisikan pada himpunan semua bilangan *real* \mathbb{R} , jika

1. $f(x) \geq 0$ untuk semua $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $P(a < X < b) = \int_a^b f(x) dx$

2.2 Momen

Momen ke- k dari suatu variabel acak adalah nilai harapan (rata-rata) dari pangkat ke- k variabel acak tersebut. Momen ke- k dinotasikan dengan $(E(X^k))$ atau μ'_k , sedangkan momen pertama disebut rata-rata dan disimbolkan dengan μ .

Rumus untuk momen ke- k adalah

$$\mu'_k = E(X^k) = \begin{cases} \int_{-\infty}^{\infty} x^k f(x) dx & \text{jika } X \text{ variabel acak kontinu} \\ \sum_j x_j^k P(x_j) & \text{jika } X \text{ variabel acak diskrit.} \end{cases} \quad (2.1)$$

(Klugman dkk, 2004)

2.2.1 Momen Sampel

Misalkan X_1, X_2, \dots, X_n adalah barisan n sampel acak, momen sampel ke- k ialah

$$a'_k = \frac{x_1^k + x_2^k + \dots + x_n^k}{n} = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (2.2)$$

dengan k adalah bilangan bulat positif.

Contoh 2.1

Momen sampel pertama yakni $k = 1$ adalah

$$a'_1 = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{X}_n$$

disebut juga sebagai rata-rata sampel (Mishra dan Dudewiez, 1988).

2.3 Fungsi Pembangkit Momen Faktorial

Fungsi pembangkit momen faktorial dari suatu variabel acak X didefinisikan sebagai

$$\eta_X(t) = E(t^X), \quad -\infty < t < \infty. \quad (2.3)$$

Momen – momen faktorial dari variabel acak X dapat diperoleh dengan menurunkan $\eta_X(t)$ terhadap t dan memasukkan $t = 1$.

Momen faktorial pertama adalah sebagai berikut

$$\begin{aligned}\eta'_X(t) &= \frac{d}{dt} E[t^X] \Big|_{t=1} \\ &= E[Xt^{X-1}] \Big|_{t=1} \\ &= E[X].\end{aligned}$$

Momen faktorial ke dua diperoleh dari turunan ke dua $\eta_X(t)$ dan memasukkan $t = 1$

$$\begin{aligned}\eta''_X(t) &= \frac{d}{dt} \eta'_X(t) \Big|_{t=1} \\ &= \frac{d}{dt} E[Xt^{X-1}] \Big|_{t=1} \\ &= E[X(X-1)t^{X-2}] \Big|_{t=1} \\ &= E[X(X-1)].\end{aligned}$$

Untuk momen faktorial ke- n adalah

$$\begin{aligned}\eta^n_X(t) &= \frac{d}{dt} \eta^{n-1}_X(t) \Big|_{t=1} \\ &= E[X(X-1)(X-2) \dots (X-(n-1))t^{X-n}] \Big|_{t=1} \\ &= E[X(X-1)(X-2) \dots (X-(n-1))].\end{aligned}$$

Contoh 2.2

Jika X adalah distribusi binomial dengan n percobaan dan peluang berhasil p , maka fungsi pembangkit momen faktorial adalah

$$\eta_X(t) = \sum_{x=0}^n t^x p^x \binom{n}{x} q^{n-x} = \sum_{x=0}^n \binom{n}{x} (tp)^x q^{n-x} = (tp + q)^n$$

sedangkan momen faktorial pertama dan ke dua adalah

$$\begin{aligned}\eta'_X(1) &= n(q + tp)^{n-1} p \Big|_{t=1} = np = E(X) \\ \eta''_X(1) &= n(n-1)(q + tp)^{n-2} p^2 \Big|_{t=1} \\ &= n(n-1)p^2 = E(X(X-1)).\end{aligned}$$

(Mishra, 1988)

2.4 Distribusi Poisson

Anggaplah X adalah suatu variabel acak diskrit yang dapat memiliki nilai-nilai $0, 1, 2, \dots$, sehingga fungsi probabilitas dari X ditentukan oleh

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots \quad (2.4)$$

di mana λ adalah suatu konstanta positif yang diketahui. Distribusi ini disebut distribusi Poisson (mengikuti nama S.D. Poisson yang menemukannya pada awal abad kesembilan belas), dan suatu variabel acak yang memiliki distribusi ini dikatakan terdistribusi Poisson (Indriasari, 2004).

2.5 Distribusi Gamma

Suatu variabel acak X dikatakan memiliki distribusi Gamma, atau terdistribusi Gamma, jika fungsi kepadatan peluangnya adalah

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (\alpha, \beta > 0). \quad (2.5)$$

Di mana $\Gamma(\alpha)$ adalah fungsi Gamma, fungsi Gamma dinotasikan dengan simbol $\Gamma(n)$, dan didefinisikan oleh:

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt \quad n > 0. \quad (2.6)$$

Jika n adalah bilangan bulat positif, maka

$$\Gamma(n+1) = n! \quad (2.7)$$

(Indriasari, 2004)

2.6 Koefisien Binomial

koefisien binomial $\binom{n}{k}$ didefinisikan untuk setiap bilangan

bulat non-negatif n dan k . Bila $k < n$ maka:

$$\binom{n}{k} = 0,$$

dan untuk nilai $k = 0$ maka:

$$\binom{n}{k} = n,$$

sedangkan untuk nilai k yang memenuhi $1 \leq k \leq n$ koefisien binomial didefinisikan dengan:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \quad (2.8)$$

Untuk setiap bilangan real n dan setiap bilangan bulat positif k , didefinisikan:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-(k-1))}{k!} \text{ dan } \binom{n}{0} = 1. \quad (2.9)$$

Jika k bilangan bulat negatif dan untuk setiap bilangan real n maka

$$\binom{n}{k} = 0 \text{ (Tucker, 1984).}$$

2.7 Distribusi Binomial Negatif

Misalkan terdapat suatu urutan percobaan Bernoulli yang saling terpisah satu sama lain. Setiap percobaan memiliki dua kemungkinan yaitu sukses dan gagal. Dalam setiap percobaan peluang terjadi sukses adalah p dan gagal adalah $q = 1 - p$. Di mana urutan ini sampai terjadi r kegagalan, dan r sudah ditentukan sebelumnya. Maka variabel acak keberhasilan X akan memiliki distribusi Binomial Negatif $X \sim NB(r, p)$

Fungsi kepadatan peluang dari distribusi Binomial Negatif yaitu:

$$f(x) = \Pr(X = x) = \binom{r+x-1}{x} p^x (1-p)^r \quad (2.10)$$

untuk $x = 0, 1, \dots$

$\binom{r+x-1}{x}$ disebut koefisien binomial, di mana penjabarannya yaitu:

$$\binom{r+x-1}{x} = \frac{(r+x-1)!}{x!(r-1)!} = \frac{(r+x-1)(r+x-2) \dots (r)}{x!}.$$

Persamaan koefisien binomial dapat ditulis dengan cara berikut:

$$\begin{aligned} \frac{(r+x-1) \dots (r)}{x!} &= (-1)^x \frac{(-r)(-r-1) \dots (-r-x+1)}{x!} \\ &= (-1)^x \binom{-r}{x}. \end{aligned} \quad (2.11)$$

Peluang untuk setiap urutan tertentu dari x sukses dan r gagal adalah $p^x(1-p)^r$, karena $r+x$ percobaan tersebut bersifat bebas atau saling terpisah satu sama lain. Karena kegagalan ke- r berada pada urutan terakhir, maka banyaknya percobaan yaitu $r+x-1$, di mana x adalah banyaknya sukses. Koefisien binomial pada persamaan (2.10) memberikan tepatnya jumlah semua rangkaian sepanjang urutan $r+x-1$ karena koefisien tersebut merupakan interpretasi dari kombinatorial.

Koefisien binomial juga dapat ditulis sebagai fungsi gamma sebagai berikut:

$$\begin{aligned} \binom{r+x-1}{x} &= \frac{(r+x-1)(r+x-2) \dots (r)}{x!} \\ &= \frac{\Gamma(r+x)}{x! \Gamma(r)} \end{aligned} \quad (2.12)$$

Dengan deret binomial dan persamaan (2.11), untuk setiap $0 \leq p \leq 1$ maka diperoleh:

$$(1-p)^{-r} = \sum_{x=0}^{\infty} \binom{-r}{x} (-p)^x$$

$$\begin{aligned}
&= \sum_{x=0}^{\infty} \binom{-r}{x} ((-1)^x p)^x \\
&= \sum_{x=0}^{\infty} \binom{-r}{x} (-1)^x (p)^x \\
&= \sum_{x=0}^{\infty} \binom{r+x-1}{x} p^x. \tag{2.13}
\end{aligned}$$

Distribusi Binomial Negatif juga bisa diperoleh dari gabungan antara distribusi Poisson dan distribusi Gamma. Dengan mengasumsikan bahwa variabel respon X merupakan variabel acak yang saling bebas dan identik (*independent and identically distributed*) yang dinotasikan sebagai *iid*, yaitu $X | \lambda \sim^{iid} \text{Poisson}(\lambda)$ dengan fungsi kepadatan peluang $f(x|\lambda)$ seperti persamaan (2.4), $x = 0, 1, 2, \dots$ dan $\lambda > 0$. Kemudian diasumsikan $\lambda \sim \text{Gamma}(\alpha, \beta)$ dengan rata-rata $\alpha\beta$, varian $\alpha\beta^2$ dan fungsi kepadatan peluangnya seperti persamaan (2.5), maka diperoleh fungsi kepadatan sebagai berikut:

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda} \lambda^{\alpha-1} e^{-\lambda/\beta}}{x! \beta^\alpha \Gamma(\alpha)} \quad x = 0, 1, 2, \dots; \lambda > 0. \tag{2.14}$$

Sehingga fungsi kepadatan peluang tidak bersyarat dari X yaitu:

$$\begin{aligned}
f(x) &= \int_0^{\infty} f(x, \lambda) d\lambda \\
&= \int_0^{\infty} \frac{\lambda^x e^{-\lambda} \lambda^{\alpha-1} e^{-\lambda/\beta}}{x! \beta^\alpha \Gamma(\alpha)} d\lambda \\
&= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^{\infty} e^{-\lambda(1+\frac{1}{\beta})} \lambda^{x+\alpha-1} d\lambda.
\end{aligned}$$

Misal:

$$\begin{aligned}
 u &= \lambda \left(1 + \frac{1}{\beta}\right) \Rightarrow u = \left(\frac{\beta + 1}{\beta}\right) \lambda \\
 &\Rightarrow du = \left(\frac{\beta + 1}{\beta}\right) d\lambda \\
 &\Rightarrow \left(\frac{\beta}{\beta + 1}\right) du = d\lambda \quad (2.15)
 \end{aligned}$$

maka:

$$\begin{aligned}
 f(x) &= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-u} \left(\frac{u\beta}{\beta + 1}\right)^{x+\alpha-1} \left(\frac{\beta}{\beta + 1}\right) du \\
 &= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-u} u^{x+\alpha-1} \left(\frac{\beta}{\beta + 1}\right)^{\alpha+x} du \\
 &= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{\beta + 1}\right)^{\alpha+x} \int_0^\infty e^{-u} u^{x+\alpha-1} du \\
 &= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{\beta + 1}\right)^{\alpha+x} \Gamma(\alpha + x) \\
 &= \frac{\Gamma(\alpha + x) \beta^{\alpha+x}}{\Gamma(\alpha) x! \beta^\alpha \left(\beta + 1\right)^{\alpha+x}} \\
 &= \frac{\Gamma(\alpha + x)}{\Gamma(\alpha) x!} \left(\frac{\beta}{\beta + 1}\right)^x \left(\frac{1}{\beta + 1}\right)^\alpha \quad (2.16)
 \end{aligned}$$

Distribusi Binomial Negatif dengan fungsi kepadatan peluang (2.16) ini mempunyai rata-rata

$$E(X) = E[E(X|\lambda)] = E(\lambda) = \alpha\beta \quad (2.17)$$

dan varian

$$\begin{aligned}
 Var(X) &= E[Var(X|\lambda)] + Var[E(X|\lambda)] \\
 &= Var(\lambda) + E(\lambda) \\
 &= \alpha\beta + \alpha\beta^2. \quad (2.18)
 \end{aligned}$$

Selanjutnya diasumsikan bahwa $\mu = \alpha\beta$ dan $k = 1/\alpha$, sehingga $E(X) = \mu$ dan $Var(X) = \mu + k\mu^2$, varian ini merupakan fungsi

kuadratik yang mengakomodasi parameter overdispersi $k > 0$ sehingga distribusi X menjadi:

$$f(x) = \frac{\Gamma(k^{-1} + x)}{\Gamma(k^{-1})x!} \left(\frac{k\mu}{1 + k\mu}\right)^x \left(\frac{1}{1 + k\mu}\right)^{k^{-1}} \quad (2.19)$$

(Kismiantini, 2008).

2.8 Estimasi (Pendugaan Parameter)

Parameter merupakan karakteristik dari suatu populasi dari fungsi distribusi peluang. Nilai parameter secara eksak dapat diketahui pada penelitian yang mengamati keseluruhan anggota populasi, kegiatan ini dinamakan sensus. Sensus jarang dilakukan karena banyak faktor, antara lain keterbatasan biaya, waktu, tenaga dan tidak dapat dilakukan pada penelitian yang bersifat merusak.

Pendugaan parameter menggunakan hasil observasi yang merupakan sampel acak dari populasi. Misalkan $X = (x_1, \dots, x_n)$ merupakan sampel acak bebas identik dari suatu populasi yang mempunyai distribusi yang tidak diketahui yang dinamakan fungsi distribusi F , maka dari hasil observasi tersebut dapat dibuat suatu penduga $\hat{\theta}$ yang digunakan untuk menduga parameter θ . Pendugaan parameter tidak terlepas dari kesalahan pendugaan (*error of estimation*) dan bias. Kesalahan pendugaan adalah jarak antara penduga dan target parameter $\varepsilon = |\hat{\theta} - \theta|$. Bias adalah selisih antara nilai harapan (*expected value*) penduga dengan parameter yang diduga $Bias = E(\hat{\theta}) - \theta$ (Brain dan Engelhardt, 1992).

Jika memilih $\hat{\theta}$ sebagai penduga maka timbul suatu pertanyaan, seberapa akuratkah penduga tersebut. Timbul persoalan yaitu bagaimana cara menyatakan bahwa $\hat{\theta}$ merupakan penduga yang tepat bagi θ . Untuk itu diperlukan suatu ukuran keakuratan pendugaan yang disebut *standard error* (Rosenfeld, 1992).

Tabel 2.1 Beberapa penduga (statistik) dengan parameter targetnya

Parameter	Simbol	Penduga (Statistik)
Rata-rata	μ	\bar{x}
Varian	σ^2	s^2
Koefisien Korelasi	ρ	r
Koefisien Regresi	β	b

2.8.1 Sifat-Sifat Estimator

Menurut William (1990), keakuratan suatu estimasi parameter tergantung pada ukuran sampel dan metode yang digunakan untuk mengestimasi parameter tersebut. Sifat-sifat dari estimator di antaranya adalah tidak bias dan efisien.

2.8.1.1 Ketidakbiasan

Sifat ketidakbiasan penduga adalah ukuran keakuratan sebagai penduga yang akan dijelaskan pada definisi-definisi berikut.

Definisi 2.3

Sebuah penduga $\hat{\theta}$ dikatakan penduga yang tidak bias bagi θ apabila $E(\hat{\theta}) = \theta$, untuk semua $\theta \in \Omega$. Selainnya dapat dikatakan bahwa $\hat{\theta}$ penduga yang bias bagi θ .

Definisi 2.4

Jika $\hat{\theta}$ adalah penduga yang bias bagi θ , bias bagi penduga tersebut adalah selisih antara nilai harapannya dengan nilai θ yang sebenarnya:

$$\text{Bias} = E(\hat{\theta}) - \theta \quad (2.20)$$

Contoh 2.3

Misalkan X adalah variabel acak dengan rata-rata μ dan varian σ^2 . Misalkan X_1, \dots, X_n adalah sampel acak yang besarnya n dari X . Akan ditunjukkan bahwa rata-rata sampel \bar{X} dan varian sampel s^2 adalah estimator yang tidak bias dari μ dan σ^2 ,

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{\sum_{i=1}^n X_i}{n}\right) \\
 &= \frac{1}{n} E\sum_{i=1}^n X_i \\
 &= \frac{1}{n} \sum_{i=1}^n E(X_i).
 \end{aligned}$$

Diketahui bahwa $E(X_i) = \mu$ untuk semua $i = 1, 2, \dots, n$, maka:

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

Oleh karena itu, rata-rata sampel \bar{X} adalah sebuah estimator yang tidak bias terhadap rata-rata populasi μ .

$$\begin{aligned}
 E(s^2) &= E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] \\
 &= \frac{1}{n-1} E\sum_{i=1}^n (X_i - \bar{X})^2 \\
 &= \frac{1}{n-1} E\sum_{i=1}^n (X_i^2 + \bar{X}^2 - 2\bar{X}X_i) \\
 &= \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) \\
 &= \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right].
 \end{aligned}$$

Dengan mensubstitusikan

$E(X_i^2) = \mu^2 + \sigma^2$ dan $E(\bar{X}^2) = \mu^2 + \sigma^2/n$, diperoleh:

$$\begin{aligned} E(s^2) &= \frac{1}{n-1} \left[\sum_{i=1}^n (\mu^2 + \sigma^2) - n \left(\mu^2 + \frac{\sigma^2}{n} \right) \right] \\ &= \frac{1}{n-1} (n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) \\ &= \sigma^2. \end{aligned}$$

Maka varian sampel s^2 adalah sebuah estimator yang tidak bias terhadap varian populasi σ^2 .

Contoh 2.4

Untuk X yang menyebar secara binomial dengan parameter n dan p , diberikan $\hat{p} = \frac{X}{n}$ sebagai penduga bagi p . Penduga ini adalah penduga tak bias karena:

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{np}{n} = p.$$

2.8.1.2 Rata-Rata Galat Kuadrat (*Mean Squared Error*)

Di dalam mengevaluasi suatu penduga, selain ketidakbiasan, satu hal lagi yang perlu diperhatikan adalah seberapa jauh penyebaran nilai duga tersebut dari nilai parameter sebenarnya (θ). Nilai duga yang sangat menyebar dari θ berarti bahwa nilai $\hat{\theta}$ yang dihasilkan sangat bervariasi dan tidak efisien. Hal ini tentunya tidak diinginkan terjadi pada suatu penduga yang baik, sehingga diharapkan suatu penduga memiliki varian yang sekecil mungkin, yaitu $Var(\hat{\theta})$ minimum. Jika diberikan dua penduga yang tidak bias bagi parameter θ , penduga yang dipilih adalah penduga yang memiliki varian terkecil.

Apabila suatu penduga mempunyai bias tertentu, untuk mengevaluasi penduga tersebut digunakan nilai harapan bagi

$(\hat{\theta} - \theta)^2$, yang didefinisikan sebagai rata-rata galat kuadrat (*Mean Squared Error*) dan dinotasikan dengan MSE.

Definisi 2.5

$\hat{\theta}$ sebagai penduga bagi θ mempunyai rata-rata galat kuadrat:

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 \\ &= \text{Var}(\hat{\theta}) + \text{Bias}^2 \end{aligned} \tag{2.21}$$

yang merupakan fungsi dari varian dan bias bagi $\hat{\theta}$.

Dari definisi tersebut dapat dibuktikan bahwa apabila suatu penduga tidak bias, maka $\text{MSE}(\hat{\theta})$ akan tereduksi menjadi $\text{Var}(\hat{\theta})$, dan bersama besaran ini dapat digunakan sebagai indikator keefisienan suatu penduga.

Contoh 2.5

Diketahui $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ dan $\text{Var}(\hat{\theta}_1) = \sigma_1^2, \text{Var}(\hat{\theta}_2) = \sigma_2^2$. Diberikan penduga baru $\hat{\theta}_3 = a\hat{\theta}_1 + (1 - a)\hat{\theta}_2$. Berapakah nilai a yang meminimumkan varian bagi $\hat{\theta}_3$. Diasumsikan bahwa $\hat{\theta}_1$ dan $\hat{\theta}_2$ saling bebas.

$$\begin{aligned} \text{Var} \hat{\theta}_3 &= \text{Var}(a\hat{\theta}_1 + (1 - a)\hat{\theta}_2) \\ &= a^2\text{Var}(\hat{\theta}_1) + (1 - a)^2\text{Var}(\hat{\theta}_2) \\ &= a^2\sigma_1^2 + (1 - a)^2\sigma_2^2 \end{aligned}$$

agar varian tersebut minimum, maka nilai a yang memenuhi adalah penyelesaian dari turunan pertama varian terhadap a yang disamadengankan nol:

$$\begin{aligned} 0 &= 2a\sigma_1^2 + 2(1 - a)\sigma_2^2 \\ &= a\sigma_1^2 - \sigma_2^2 + a\sigma_2^2 \\ &= a(\sigma_1^2 + \sigma_2^2) - \sigma_2^2 \\ a &= \frac{\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}. \end{aligned}$$

2.9 Estimasi Parameter Titik Menggunakan *Method of Moment Estimate*

Tujuan dari metode momen adalah untuk memperoleh penaksir yang baik ($\hat{\theta}$), distribusi yang mendasari $\hat{\theta}$ harus serupa dengan distribusi dari θ , di mana kesamaan tersebut dibandingkan dengan kesetaraan momen. Namun, momen-momen dari distribusi yang berkorespondensi dengan θ tidak diketahui karena nilai θ juga tidak diketahui. Untuk alasan ini nilai θ akan diperkirakan dengan momen sampel. Momen dihitung dengan menggunakan sampel yang sudah diberikan.

Jika θ adalah vektor dengan k komponen, maka dibutuhkan lebih dari satu persamaan. Lebih jelasnya jika terdapat k parameter yang tidak diketahui, maka dibutuhkan k persamaan untuk menyelesaikannya. Oleh karena itu diperlukan persamaan k momen.

Estimasi parameter menggunakan metode momen diperoleh dengan menyamakan momen sampel (2.2) dengan momen teoritis (2.3), yaitu sebagai berikut:

$$a'_k = E[X^k]. \quad (2.22)$$

Dengan X merupakan variabel acak dari suatu fungsi kepadatan peluang tertentu.

Contoh 2.6

Misal $X_1, X_2, \dots, X_n \sim \text{Gamma}(\alpha, \beta)$, maka $\theta = (\alpha, \beta)$. Diketahui bahwa $E(X) = \alpha\beta$ dan $\text{Var}(X) = \alpha\beta^2$ maka

$$E(X^2) = \text{Var}(X) + E(X)^2 = \alpha\beta^2 + \alpha^2\beta^2$$

Persamaan-persamaan yang harus diselesaikan yaitu:

$$E(X) = a'_1 = \hat{\alpha}\hat{\beta} = \bar{X} \quad (2.23)$$

dan

$$E(X^2) = \hat{\alpha}\hat{\beta}^2 + \hat{\alpha}^2\hat{\beta}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2. \quad (2.24)$$

Dari persamaan (2.23) diperoleh $\hat{\beta} = \frac{\overline{X}}{\hat{\alpha}}$ dan disubstitusikan ke persamaan (2.24) maka diperoleh

$$\hat{\alpha} = \frac{\overline{X}^2}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}^2} \quad \text{dan} \quad \hat{\beta} = \frac{\overline{X}}{\hat{\alpha}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}^2}{\overline{X}}$$

(Hall, 1996).

2.10 Estimasi Parameter Titik Menggunakan *Maximum Likelihood Estimate*

Metode ini digunakan untuk menentukan penduga bagi θ , di mana nilai duga tersebut adalah nilai yang membuat data pengamatan paling mungkin (*the most likely*) terjadi. Berdasarkan prinsip ini, apabila data teramati lebih mungkin (*more likely*) mempunyai nilai $\theta = \theta_1$ daripada mempunyai $\theta = \theta_2$, maka θ_1 akan dipilih sebagai penduga bagi θ .

Definisi 2.7 (Fungsi *likelihood*)

Jika x_1, x_2, \dots, x_n adalah sampel acak berukuran n dari sebuah fungsi kepadatan peluang diskrit atau kontinu, $f_X(x_i|\theta)$, di mana θ merupakan parameter yang tidak diketahui, maka fungsi *likelihood* dinotasikan dengan:

$$L(\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (2.25)$$

Definisi 2.8 (*Maximum Likelihood Estimate*)

Diberikan $L(\theta) = f(x_1, \dots, x_n|\theta)$, $\theta \in \Omega$, dan Ω merupakan himpunan seluruh nilai parameter yang mungkin. Maka *Maximum Likelihood Estimate* (MLE) adalah nilai $\hat{\theta} \in \Omega$ yang memaksimumkan $L(\theta)$. Sehingga $\hat{\theta}$ adalah nilai θ yang memenuhi:

$$f(x_1, \dots, x_n|\hat{\theta}) = \max_{\theta \in \Omega} f(x_1, \dots, x_n|\theta). \quad (2.26)$$

Berdasarkan definisi tersebut $\hat{\theta}$ adalah penyelesaian dari turunan pertama fungsi *likelihood* terhadap θ , yang disamadengankan nol:

$$\frac{d}{d\theta} L(\theta) = \frac{d}{d\theta} \prod_{i=1}^n f(x_i | \theta) = 0. \quad (2.27)$$

Untuk mempermudah perhitungan secara matematis, umumnya digunakan log *likelihood*

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i | \theta). \quad (2.28)$$

Yang juga diturunkan terhadap θ dan disamadengankan nol, untuk memperoleh $\hat{\theta}$:

$$\frac{d}{d\theta} l(\theta) = \frac{d}{d\theta} \sum_{i=1}^n \ln f(x_i | \theta) = 0 \quad (2.29)$$

(Hogg, 2005).

Contoh 2.7

Apabila X_1, X_2, \dots, X_n *iid* dan berdistribusi Poisson, dengan fungsi kepadatan peluang:

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots \text{ dan } \lambda > 0.$$

Fungsi *likelihood* bagi X_1, \dots, X_n adalah perkalian dari fungsi frekuensi masing-masing variabel

$$L(\lambda) = \prod_{i=1}^n f(x_i | \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

Log *likelihood* dari fungsi *likelihood* tersebut yaitu:

$$l(\lambda) = \ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \sum_{i=1}^n \ln x_i !$$

Turunan pertama dari $l(\lambda)$ terhadap λ adalah:

$$\frac{d}{d\lambda} l(\lambda) = -n + \frac{\sum_{i=1}^n x_i}{\lambda}$$

Solusi dari persamaan tersebut jika disamakan dengan nol:

$$\frac{\sum_{i=1}^n x_i}{\hat{\lambda}} = n \Leftrightarrow \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

adalah penduga kemungkinan maksimum (MLE) bagi λ .

2.11 Metode Newton-Raphson

Menurut Salusu (2008), metode yang paling luas dipakai di antara metode-metode pencari akar yang lain yaitu metode Newton-Raphson. Misalkan x_0 merupakan nilai pendekatan dari akar $f(x) = 0$ dan h adalah selisih nilai sebenarnya dengan nilai pendekatan yaitu:

$$x = x_0 + h.$$

Dengan menggunakan deret Taylor di sekitar x_0 diperoleh:

$$f(\alpha) = f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x) + \dots \quad (2.30)$$

Dimana $x = x_0 + \theta h$, $0 < \theta < 1$ yang terletak di antara α dan x_0 . Dengan mengabaikan sisa dan $f(\alpha) = 0$ maka:

$$f(x_0) + hf'(x_0) \approx 0$$

atau

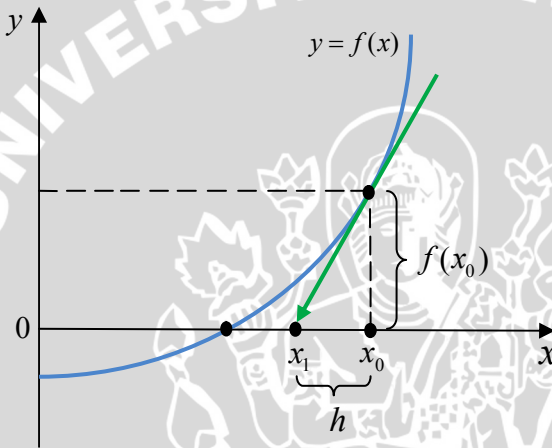
$$h \approx -\frac{f(x_0)}{f'(x_0)}. \quad (2.31)$$

Diperoleh:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

yaitu suatu nilai yang lebih baik dari akar perkiraan semula. Proses diulangi sehingga didapat bentuk umum:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \quad (2.32)$$



Gambar 2.1 Ilustrasi metode Newton-Raphson

Secara geometris, Gambar 2.1 merupakan ilustrasi tentang Metode Newton-Raphson. Diperlihatkan bahwa garis singgung titik awal $(x_0, f(x_0))$ memotong sumbu X di titik $(x_1, 0)$, kemudian dari titik $(x_n, f(x_n))$ dibuat garis singgung lagi. Proses ini diulang sampai diperoleh nilai x_n yang diinginkan. Persamaan yang terbentuk adalah sebagai berikut:

$$y - f(x_n) = f'(x_n)(x - x_n).$$

Karena memotong sumbu X di titik $(x_{n+1}, 0)$:

$$-f(x_n) = f'(x_n)(x_{n+1} - x_n).$$

2.12 Overdispersi

Overdispersi adalah suatu fenomena yang kadang kala terjadi pada data yang dimodelkan menggunakan distribusi Binomial atau distribusi Poisson. Overdispersi pada model Poisson terjadi ketika varian lebih besar daripada rata-rata.

Overdispersi disebabkan oleh varian yang berlebih dari data. Selain itu overdispersi juga bisa terjadi jika terdapat pelanggaran atau pemaksaan dalam asumsi-asumsi distribusi pada data.

Fenomena ini merupakan suatu masalah dalam statistik, karena overdispersi bisa menyebabkan estimasi dari parameter akan menjadi tidak akurat dan tidak stabil.

Terdapat berbagai cara untuk menangani masalah overdispersi dilihat dari penyebab overdispersi itu sendiri. Salah satu cara untuk menangani overdispersi pada data yang dimodelkan dengan distribusi Poisson yaitu dengan menggunakan distribusi Binomial Negatif (Hilbe, 2011).

2.13 Metode Bootstrap

Metode Bootstrap pertama kali diperkenalkan oleh Efron pada tahun 1979. Metode Bootstrap ini menggunakan teknik pengambilan sampel berulang secara acak dari data asli. Misalkan $x = (x_1, \dots, x_n)$ adalah sampel acak dari suatu populasi dengan θ sebagai karakteristik populasinya, maka algoritma dari metode Bootstrap adalah:

1. Diambil sampel Bootstrap yang saling bebas sebanyak B , sehingga didapat B sampel Bootstrap $x^* = (x_1^*, \dots, x_B^*)$, masing-masing terdiri dari n data yang diambil secara berulang dari data asli x . Misalkan $x_1^* = (x_1, \dots, x_n)$. Karena dilakukan pengambilan sampel dengan pengembalian, maka dimungkinkan data pengamatan terambil lebih dari satu kali atau tidak terambil.
2. Dihitung nilai penduga dari masing-masing sampel Bootstrap, $\hat{\theta}^*(b)$, sehingga didapat data baru yakni data Bootstrap $X_{Boot}^* = (\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(b))$, di mana $b = 1, 2, \dots, B$
3. Maka nilai $\hat{\theta}$ dapat didekati oleh $\hat{\theta}^*$ dari sampel hasil pengambilan berulang sebanyak B tersebut

Dengan melihat prosedur di atas maka metode ini perlu bantuan komputer untuk menyelesaikannya (Efron, 1991).

2.14 Simulasi Monte Carlo

Monte Carlo adalah algoritma komputasi yang menggunakan sampling acak yang berulang untuk menghitung hasilnya. Monte Carlo sering digunakan dalam mensimulasikan sistem fisika dan matematika.

Monte Carlo pertama kali diperkenalkan pada tahun 1940-an oleh John von Neumann, Stanislaw Ulam dan Nicholas Metropolis saat mereka bekerja pada proyek senjata nuklir di Laboratorium Nasional Los Alamos. Nama Monte Carlo tersebut diambil dari sebuah nama kasino yang terkenal.

Tidak ada konsensus mengenai bagaimana Monte Carlo harus didefinisikan. Sebagai contoh Ripley (1987) mendefinisikan Monte Carlo sebagai model yang paling mungkin digunakan sebagai simulasi stokastik. Sawilowsky (2003) membedakan antara simulasi, Monte Carlo dan simulasi Monte Carlo. Simulasi adalah representasi fiktif dari realitas. Metode Monte Carlo adalah suatu teknik yang dapat digunakan untuk memecahkan masalah matematika atau statistik. Simulasi Monte Carlo adalah simulasi yang menggunakan sampling berulang untuk menentukan sifat-sifat dari suatu fenomena.

Aplikasi dari metode ini sangat luas, yaitu dalam bidang fisika, matematika, statistika terapan, teknik, biologi, desain dan visual, keuangan dan bisnis, dan telekomunikasi. Menurut Fishman (1995) simulasi Monte Carlo terbagi dalam 5 tahapan, yaitu:

1. Membuat distribusi peluang untuk variabel penting.

Gagasan dasar dari simulasi Monte Carlo adalah membuat nilai dari tiap variabel yang merupakan bagian dari model yang dipelajari. Banyak variabel di dunia nyata yang secara alami mempunyai berbagai peluang yang mungkin ingin disimulasikan. Salah satu cara yang umum dilakukan untuk membuat distribusi peluang untuk suatu variabel adalah memperhitungkan hasil dimasa lalu. peluang atau frekuensi relatif untuk tiap peluang hasil dari tiap variabel ditentukan dengan membagi frekuensi observasi dengan jumlah total observasi.

Contoh 2.8

Frekuensi jumlah kecelakaan pada lokasi tertentu selama 200 hari terlihat di tabel berikut:

Tabel 2.2 Frekuensi jumlah kecelakaan selama 200 hari

Jumlah kecelakaan per hari	Frekuensi
0	10
1	20
2	40
3	60
4	40
5	30
Jumlah	200

Keadaan di atas dapat dirubah menjadi distribusi peluang dengan membagi tiap permintaan dengan total permintaan. Seperti pada tabel berikut:

Tabel 2.3 Distribusi peluang jumlah kecelakaan

Jumlah kecelakaan per hari	Peluang Terjadi
0	$10/200 = 0.05$
1	$20/200 = 0.10$
2	$40/200 = 0.20$
3	$60/200 = 0.30$
4	$40/200 = 0.20$
5	$30/200 = 0.15$
Total Peluang Kumulatif	$200/200 = 1.00$

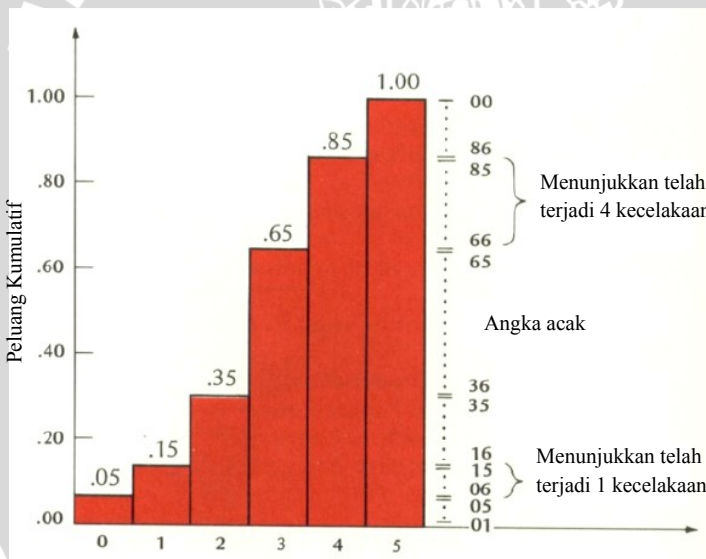
2. Membangun distribusi peluang kumulatif untuk tiap-tiap variabel di tahap pertama.

Konversi dari distribusi peluang, seperti pada kolom kanan Tabel 2.3 menjadi distribusi kumulatif dilakukan dengan menjumlahkan tiap nilai peluang dengan jumlah sebelumnya seperti pada Tabel 2.4 berikut:

Tabel 2.4 Distribusi peluang kumulatif jumlah kecelakaan

Jumlah kecelakaan per hari	Peluang	Peluang Kumulatif
0	0.05	0.05
1	0.10	0.15
2	0.20	0.35
3	0.30	0.65
4	0.20	0.85
5	0.15	1.00

peluang kumulatif terlihat pada gambar di bawah, digunakan pada tahap ke-3 untuk membantu menempatkan nilai acak



Gambar 2.2 Grafik peluang kumulatif

- Menentukan Interval angka acak untuk tiap variabel. Setelah ditentukan peluang kumulatif untuk tiap variabel yang termasuk dalam simulasi, harus ditentukan pula batas angka yang mewakili tiap peluang hasil. Hal tersebut ditunjukkan pada interval angka acak. Penentuan interval didasari oleh peluang kumulatif.

Tabel 2.5 Penentuan interval angka acak

Jumlah kecelakaan per hari	Peluang	Peluang Kumulatif	Interval Angka Acak
0	0.05	0.05	1-5
1	0.10	0.15	6-15
2	0.20	0.35	16-35
3	0.30	0.65	36-65
4	0.20	0.85	66-85
5	0.15	1.00	86-100

4. Membangkitkan angka acak.
Untuk membangkitkan angka acak bisa digunakan *software*, seperti Microsoft Excel, Matlab, Delphi, Minitab dan lain sebagainya.
5. Membuat simulasi dari rangkaian percobaan.
Simulasi bisa dibuat dari sebuah eksperimen dengan mengambil angka acak yang sudah dibangkitkan sebelumnya (langkah 4). Berikut adalah tabel hasil simulasi dari Contoh 2.8

Tabel 2.6 Hasil simulasi

Hari	Angka Acak	Jumlah kecelakaan (Simulasi)
1	28	2
2	50	3
3	78	4
4	8	1
5	16	2
6	61	3
7	98	5
8	51	3
9	45	3
10	21	2
Jumlah		28

Total jumlah kecelakaan untuk 10 hari adalah 28 kecelakaan, rata-rata jumlah kecelakaan per hari adalah 2.8 kecelakaan.

UNIVERSITAS BRAWIJAYA



BAB III

METODOLOGI PENELITIAN

3.1 Deskripsi Umum Daerah Studi

Polresta Malang Kota merupakan sebuah instansi pemerintah yang bertugas melayani masyarakat. Polresta Malang Kota dibagi menjadi beberapa unit yang salah satunya adalah unit kecelakaan lalu lintas (Lakalantas). Salah satu tugas dari unit ini adalah menerima laporan terjadinya kecelakaan lalu lintas dan melakukan penyidikan di tempat kejadian perkara. Selain itu tugas dari unit ini adalah menghimpun dan mencatat data-data yang berkaitan dengan kecelakaan lalu lintas.

3.2 Jenis dan Sumber Data

Data yang digunakan berupa data sekunder. Data sekunder merupakan data yang diperoleh atau dikumpulkan oleh orang yang melakukan penelitian dari sumber-sumber yang telah ada, misalnya dalam bentuk tabel atau diagram. Data sekunder ini yang kemudian digunakan sebagai dasar penentuan model (Azwar, 2001).

Data yang diambil adalah data jumlah kecelakaan lalu lintas di jalan raya rawan kecelakaan. Data tersebut diperoleh dari unit kecelakaan lalu lintas Polresta Malang.

3.3 Waktu dan Tempat Penelitian

Penelitian dalam skripsi ini dilaksanakan di jalan raya rawan kecelakaan Kota Malang menurut Unit Lakalantas Polresta Malang. Jalan yang dimaksud meliputi Jl. Kolonel Sugiono, Jl. Supriadi, Jl. MT. Haryono, dan Jl. Basuki Rachmad pada tahun 2006 hingga 2010.

3.4 Metode Pengumpulan Data

Untuk memperoleh data-data pendukung penelitian ini, maka dilakukan pengumpulan data melalui dua tahapan, yaitu :

1. Penelitian langsung ke Unit Lakalantas Polresta Kota Malang (*Field research*).

Metode ini bertujuan untuk memperoleh data-data pendukung penelitian yang langsung didapatkan di tempat penelitian. Adapun untuk mendapatkannya dengan menggunakan cara-cara berikut (Azwar, 2001):

a. Dokumentasi

Merupakan suatu metode pengumpulan data dengan melihat dan menggunakan data-data berupa arsip-arsip atau catatan yang berhubungan dengan obyek penelitian yang terdapat di polresta. Data-data ini merupakan data sekunder.

b. Wawancara

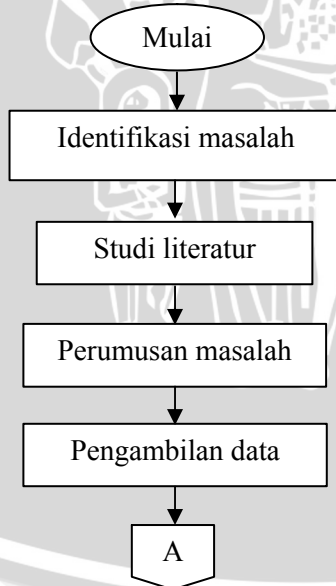
Merupakan suatu metode pengumpulan data dengan melakukan komunikasi atau wawancara mengenai hal-hal yang berhubungan dengan obyek penelitian (Azwar, 2001), yang dalam hal ini dilakukan melalui kunjungan ke polresta dan bertemu Unit Lakalantas.

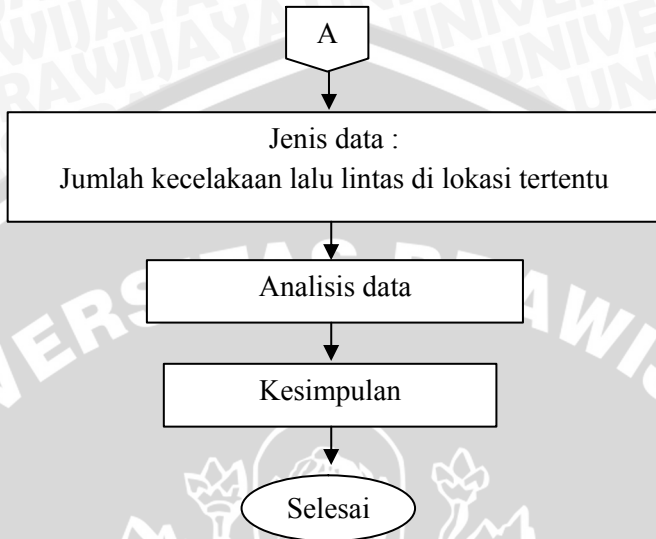
2. Studi literatur.

Metode ini dilakukan dengan tujuan untuk memecahkan permasalahan-permasalahan yang ada dengan menggunakan teori yang telah didapat di perkuliahan

3.5 Diagram Alir Penelitian

Langkah-langkah yang dilakukan dalam skripsi ini digambarkan pada Gambar 3.1 berikut.



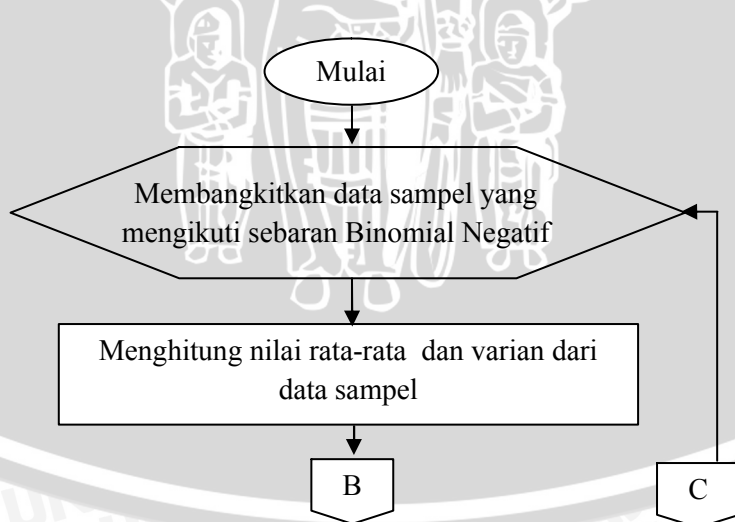


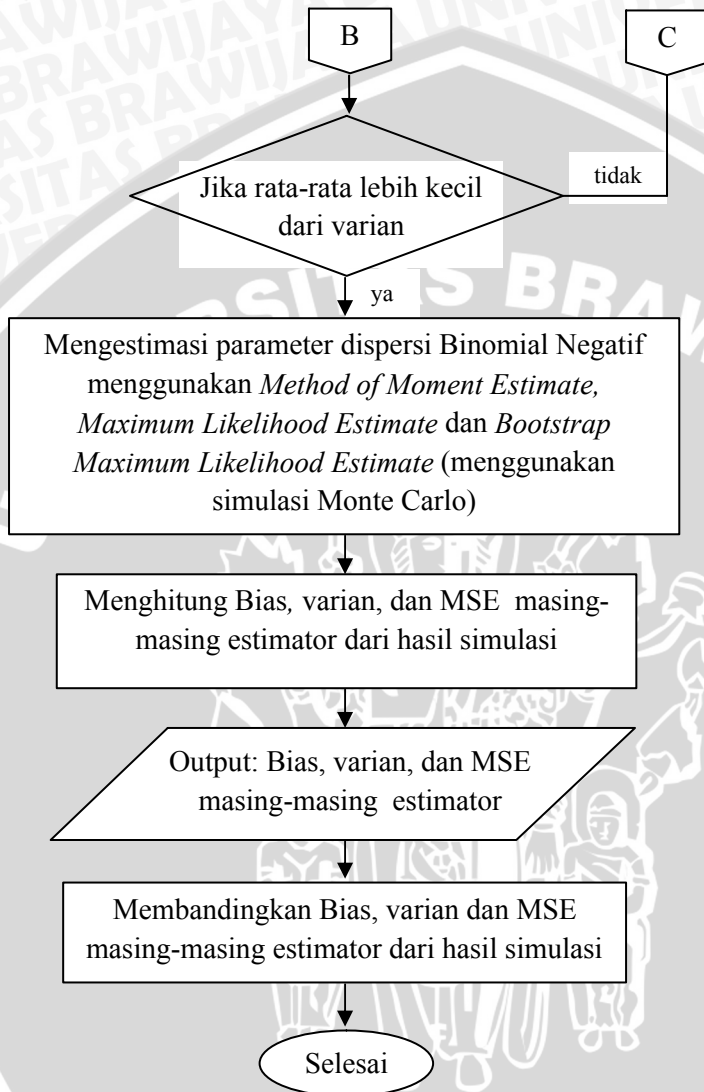
Gambar 3.1 : Diagram alir penelitian

3.6 Analisis Data

Analisis data kecelakaan lalu lintas yang dilakukan dalam penulisan skripsi ini dijelaskan pada diagram alir berikut.

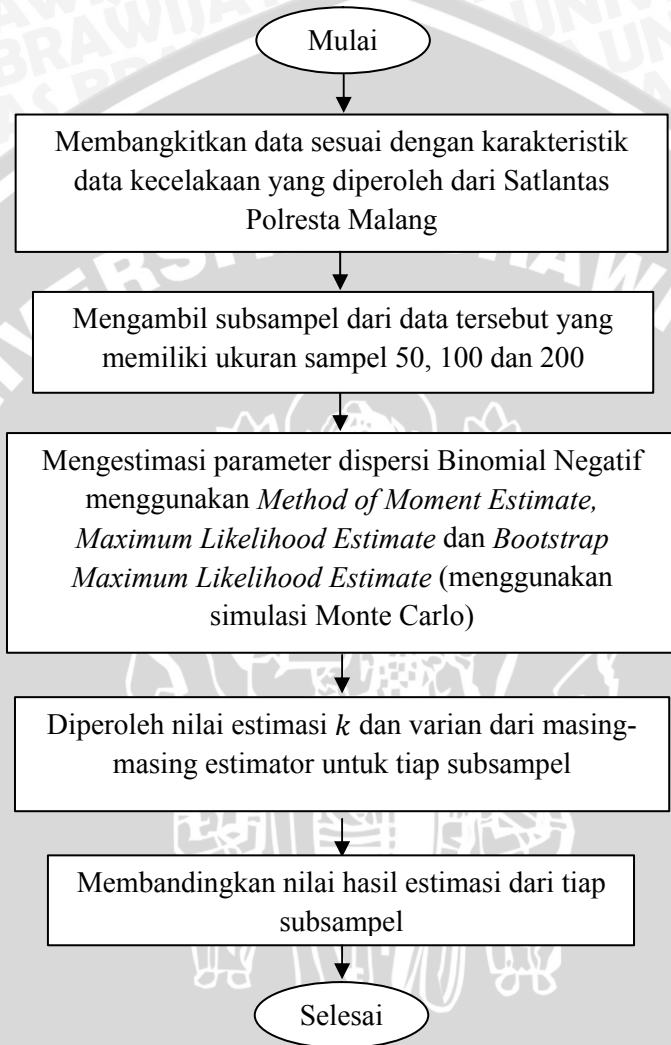
a. Untuk data hasil simulasi.





Gambar 3.2 : Diagram alir analisis data hasil simulasi

b. Untuk data sekunder dari Unit Lakalantas Polresta Malang



Gambar 3.3 : Diagram alir analisis data kecelakaan dari Unit Lakalantas Polresta Malang

UNIVERSITAS BRAWIJAYA



BAB IV HASIL DAN PEMBAHASAN

Dalam pembahasan ini akan dijelaskan bagaimana algoritma dari metode *Bootstrap Maximum Likelihood Estimate* (BMLE) sebagai estimator dari k , dan bagaimana hasil estimasi Binomial Negatif dari metode BMLE dibandingkan dengan *Method of Moment Estimate* (MME) dan *Maximum Likelihood Estimate* (MLE) menggunakan simulasi Monte Carlo pada data kecelakaan.

4.1 Estimasi Parameter Binomial Negatif Menggunakan *Method of Moment Estimate*

Estimasi parameter menggunakan metode momen diperoleh dengan menyamakan momen sampel dengan momen teoritis. Karena distribusi Binomial Negatif mempunyai dua parameter, maka diperlukan momen pertama dan ke dua dari fungsi kepadatan peluangnya.

4.1.1 Fungsi Pembangkit Momen Faktorial Distribusi Binomial Negatif

Fungsi kepadatan peluang untuk distribusi Binomial Negatif berdasarkan persamaan (2.19) adalah

$$f(x) = \frac{\Gamma(k^{-1} + x)}{\Gamma(k^{-1})x!} \left(\frac{k\mu}{1 + k\mu} \right)^x \left(\frac{1}{1 + k\mu} \right)^{k^{-1}}$$

dan persamaan di atas dapat ditulis sebagai:

$$f(x) = \binom{k^{-1} + x - 1}{x} p^x (1 - p)^{k^{-1}} \quad (4.1)$$

dengan $p = \left(\frac{k\mu}{1 + k\mu} \right)$.

Fungsi pembangkit momen faktorial diperoleh dari persamaan (2.3), sehingga fungsi pembangkit momen faktorial untuk distribusi Binomial Negatif adalah

$$\eta_x(t) = E(t^x)$$

$$\begin{aligned}
 \eta_x(t) &= \sum_{x=0}^{\infty} (t^x) \binom{k^{-1} + x - 1}{x} p^x (1-p)^{k-1} \\
 &= (1-p)^{k-1} \sum_{x=0}^{\infty} \binom{k^{-1} + x - 1}{x} (t^x) p^x \\
 &= (1-p)^{k-1} \sum_{x=0}^{\infty} \binom{k^{-1} + x - 1}{x} (tp)^x. \tag{4.2}
 \end{aligned}$$

Penjabaran dari $\binom{k^{-1} + x - 1}{x}$ adalah sebagai berikut

$$\begin{aligned}
 \binom{k^{-1} + x - 1}{x} &= \frac{(k^{-1} + x - 1)!}{(k^{-1} + x - 1 - x)! x!} \\
 &= \frac{(k^{-1} + x - 1)(k^{-1} + x - 2) \dots (k^{-1})(k^{-1} - 1)!}{(k^{-1} - 1)! x!} \\
 &= \frac{(k^{-1} + x - 1)(k^{-1} + x - 2) \dots (k^{-1})}{x!} \\
 &= \frac{k^{-1}(k^{-1} + 1)(k^{-1} + 2) \dots (k^{-1} + x - 2)(k^{-1} + x - 1)}{x!},
 \end{aligned}$$

sedangkan penjabaran dari $\binom{-k^{-1}}{x}$ adalah

$$\begin{aligned}
 \binom{-k^{-1}}{x} &= \frac{-k^{-1}!}{(-k^{-1} - x)! x!} \\
 &= \frac{-k^{-1}(-k^{-1} - 1) \dots (-k^{-1} - x + 1)(-k^{-1} - x)!}{(-k^{-1} - x)! x!} \\
 &= \frac{(-1)^x \{k^{-1}(k^{-1} + 1) \dots (k^{-1} + x - 1)\}}{x!} \\
 &= (-1)^x \binom{k^{-1} + x - 1}{x}.
 \end{aligned}$$

Hubungan antara $\binom{k^{-1} + x - 1}{x}$ dengan $\binom{-k^{-1}}{x}$ adalah

$$\binom{k^{-1} + x - 1}{x} = (-1)^x \binom{-k^{-1}}{x}.$$

Persamaan (4.2) menjadi

$$\eta_X(t) = (1-p)^{k-1} \sum_{x=0}^{\infty} \binom{k-1}{x} (-tp)^x.$$

Mengingat bahwa $(1+x)^p = \sum_{n=0}^{\infty} \binom{p}{n} x^n 1^{p-n}$, dan mengacu pada persamaan (2.11), maka bentuk $\sum_{x=0}^{\infty} \binom{k-1}{x} (-tp)^x$ akan sama dengan $(1-tp)^{-k-1}$.

Oleh karena itu fungsi pembangkit momen faktorial distribusi Binomial Negatif adalah

$$\eta_X(t) = (1-p)^{k-1} (1-tp)^{-k-1}. \quad (4.3)$$

Momen distribusi Binomial Negatif dapat dihitung dengan cara mencari turunan ke- n dari fungsi pembangkit momen faktorial terhadap t pada $t = 1$. Momen pertama dan ke dua dari distribusi Binomial Negatif adalah sebagai berikut:

Momen pertama

$$\begin{aligned} \eta'(t) &= (1-p)^{k-1} (-k^{-1})(1-tp)^{-k-1-1} (-p) \\ &= (1-p)^{k-1} k^{-1} p (1-tp)^{-k-1-1}. \end{aligned}$$

Untuk nilai $t = 1$

$$\begin{aligned} \eta'(1) &= (1-p)^{k-1} k^{-1} p (1-p)^{-k-1-1} \\ &= \frac{k^{-1} p}{(1-p)} \end{aligned}$$

$$\eta'(1) = E(X) = \frac{k^{-1} p}{(1-p)}. \quad (4.4)$$

Momen ke dua

$$\begin{aligned}\eta''(t) &= (1-p)^{k-1} k^{-1} p (-k^{-1} - 1) (-p) (1-tp)^{-k^{-1}-2} \\ &= (1-p)^{k-1} k^{-1} p (k^{-1}p + p) (1-tp)^{-k^{-1}-2}.\end{aligned}$$

Untuk $t = 1$

$$\eta''(1) = (1-p)^{k-1} k^{-1} p (k^{-1}p + p) (1-p)^{-k^{-1}-2}$$

$$\eta''(1) = E(X(X-1))$$

$$= (1-p)^{k-1} k^{-1} p (k^{-1}p + p) (1-p)^{-k^{-1}-2}$$

$$E(X^2) - E(X) = \frac{k^{-1}p(k^{-1}p + p)}{(1-p)^2}$$

$$E(X^2) = \frac{k^{-1}p(k^{-1}p + p)}{(1-p)^2} + E(X)$$

$$= \frac{k^{-1}p(k^{-1}p + p)}{(1-p)^2} + \frac{k^{-1}p}{(1-p)}$$

$$= \frac{k^{-1}p(k^{-1}p + p) + k^{-1}p(1-p)}{(1-p)^2}$$

$$= \frac{(k^{-1}p)^2 + k^{-1}p^2 + k^{-1}p - k^{-1}p^2}{(1-p)^2}$$

$$= \frac{k^{-1}p(k^{-1}p + 1)}{(1-p)^2}. \quad (4.5)$$

Varian distribusi Binomial Negatif adalah

$$Var(X) = E(X^2) - E(X)^2$$

$$= \frac{k^{-1}p(k^{-1}p + 1)}{(1-p)^2} - \left(\frac{k^{-1}p}{(1-p)} \right)^2$$

$$= \frac{(k^{-1}p)^2 + k^{-1}p - (k^{-1}p)^2}{(1-p)^2}$$

$$= \frac{k^{-1}p}{(1-p)^2}. \quad (4.6)$$

Dengan menggunakan persamaan (2.22), (4.4) dan (4.5) diperoleh:

$$E(X) = \frac{\hat{k}^{-1}\hat{p}}{(1-\hat{p})} = \bar{X} \quad (4.7)$$

dan

$$E(X^2) = \frac{\hat{k}^{-1}\hat{p}(\hat{k}^{-1}\hat{p} + 1)}{(1 - \hat{p})^2} = \frac{1}{n} \sum_{i=1}^n X_i^2, \quad (4.8)$$

sehingga dari persamaan (4.7) diperoleh:

$$\hat{k}^{-1} = \frac{\bar{X}(1 - \hat{p})}{\hat{p}}. \quad (4.9)$$

Kemudian persamaan (4.9) disubstitusikan ke persamaan (4.8), yaitu:

$$\begin{aligned} \frac{\bar{X}(1 - \hat{p})}{\hat{p}} \hat{p} \left[\frac{\bar{X}(1 - \hat{p})}{\hat{p}} \hat{p} + 1 \right] &= \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \frac{\bar{X}(\bar{X}(1 - \hat{p}) + 1)}{1 - \hat{p}} &= \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \bar{X}^2 + \frac{\bar{X}}{(1 - \hat{p})} &= \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \frac{1}{(1 - \hat{p})} &= \frac{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2}{\bar{X}} \\ (1 - \hat{p}) &= \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2} \\ (1 - \hat{p}) &= \frac{\bar{X}}{s^2} \\ \hat{p} &= 1 - \frac{\bar{X}}{s^2}. \end{aligned} \quad (4.10)$$

Dengan mensubstitusikan persamaan (4.10) pada persamaan (4.9) diperoleh:

$$\hat{k}^{-1} = \frac{\bar{X}(1 - \hat{p})}{\hat{p}}$$

$$\begin{aligned}\hat{k}^{-1} &= \frac{\bar{X} \left(\frac{\bar{X}}{s^2} \right)}{\left(1 - \frac{\bar{X}}{s^2} \right)} \\ \hat{k}^{-1} &= \frac{\bar{X}^2}{s^2 - \bar{X}} \\ \hat{k} &= \frac{s^2 - \bar{X}}{\bar{X}}.\end{aligned}\quad (4.11)$$

4.2 Estimasi Parameter Binomial Negatif Menggunakan *Maximum Likelihood Estimate*

Metode lain yang digunakan untuk mengestimasi parameter dari distribusi Binomial Negatif yaitu metode *Maximum Likelihood Estimate*. Fungsi *likelihood* dari distribusi Binomial Negatif yaitu:

$$\begin{aligned}L(\mu, k) &= \prod_{i=1}^n f(x_i | \mu, k) \\ &= \prod_{i=1}^n \binom{k^{-1} + x_i - 1}{x_i} p^{x_i} (1-p)^{k^{-1}} \\ &= \prod_{i=1}^n \frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1}) x_i!} \left(\frac{k\mu}{1+k\mu} \right)^{x_i} \left(\frac{1}{1+k\mu} \right)^{k^{-1}}.\end{aligned}\quad (4.12)$$

Fungsi *log-likelihood* dari persamaan (4.12) yaitu:

$$\begin{aligned}l(\mu, k) &= \sum_{i=1}^n \left[\ln \left(\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1}) x_i!} \right) + x_i \ln \left(\frac{k\mu}{1+k\mu} \right) \right. \\ &\quad \left. + k^{-1} \ln \left(\frac{1}{1+k\mu} \right) \right].\end{aligned}\quad (4.13)$$

Misal $A = \ln \left(\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})x_i!} \right)$ maka persamaan (4.13) dapat ditulis

sebagai berikut:

$$\begin{aligned}
 l(\mu, k) &= \sum_{i=1}^n \left[A + x_i \ln \left(\frac{k\mu}{1+k\mu} \right) + k^{-1} \ln \left(\frac{1}{1+k\mu} \right) \right] \\
 &= \sum_{i=1}^n \left[A + x_i (\ln k\mu - \ln(1+k\mu)) + k^{-1} (\ln 1 - \ln(1+k\mu)) \right] \\
 &= \sum_{i=1}^n \left[A + x_i \ln k\mu - x_i \ln(1+k\mu) - k^{-1} \ln(1+k\mu) \right] \\
 &= \sum_{i=1}^n \left[A + x_i \ln k\mu - (x_i + k^{-1}) \ln(1+k\mu) \right]. \quad (4.14)
 \end{aligned}$$

Dengan mensubstitusikan kembali nilai A , maka persamaan (4.14) dapat ditulis sebagai:

$$\begin{aligned}
 l(\mu, k) &= \sum_{i=1}^n \left[\ln \left(\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})x_i!} \right) + x_i \ln k\mu - (x_i + k^{-1}) \ln(1+k\mu) \right] \\
 &= \sum_{i=1}^n \left[\ln \left(\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})} \right) - \ln x_i! + x_i \ln k\mu \right. \\
 &\quad \left. - (x_i + k^{-1}) \ln(1+k\mu) \right] \\
 &= \sum_{i=1}^n \left[\ln \left(\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})} \right) - \ln x_i! + x_i \ln k + x_i \ln \mu \right. \\
 &\quad \left. - (x_i + k^{-1}) \ln(1+k\mu) \right]. \quad (4.15)
 \end{aligned}$$

Bentuk $\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})}$ dapat dijabarkan sebagai berikut:

$$\frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})} = (k^{-1} + x_i - 1)(k^{-1} + x_i - 2) \dots (k^{-1} + 1)(k^{-1}),$$

sehingga

$$\ln \left\{ \frac{\Gamma(k^{-1} + x_i)}{\Gamma(k^{-1})} \right\} = \sum_{v=0}^{x_i-1} \ln(k^{-1} + v). \quad (4.16)$$

Dengan mensubstitusikan persamaan (4.16) ke persamaan (4.15) diperoleh:

$$l(\mu, k) = \sum_{i=1}^n \left[\left(\sum_{v=0}^{x_i-1} \ln(k^{-1} + v) \right) - \ln x_i + x_i \ln k + x_i \ln \mu - (x_i + k^{-1}) \ln(1 + k\mu) \right]. \quad (4.17)$$

Estimasi untuk μ dan k menggunakan *Maximum Likelihood Estimate, MLE* ($\hat{\mu}, \hat{k}$), adalah solusi dari turunan pertama persamaan (4.17) terhadap tiap parameter yang diestimasi.

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^n \left(\frac{x_i}{\mu} - \frac{k(x_i + k^{-1})}{1 + k\mu} \right) = \sum_{i=1}^n \left(\frac{x_i}{\mu} - \frac{kx_i + 1}{1 + k\mu} \right) \quad (4.18)$$

$$\begin{aligned} \frac{\partial l}{\partial k} &= \sum_{i=1}^n \left[\left(\sum_{v=0}^{x_i-1} -\frac{1}{k(1 + kv)} \right) + \frac{x_i}{k} - \left(-\frac{\ln(1 + k\mu)}{k^2} + \frac{x_i k + 1}{k} \frac{\mu}{1 + k\mu} \right) \right] \\ &= \sum_{i=1}^n \left[\left(\sum_{v=0}^{x_i-1} -\frac{1}{k(1 + kv)} \right) + \frac{x_i}{k} + \frac{\ln(1 + k\mu)}{k^2} - \frac{\mu(x_i k + 1)}{k(1 + k\mu)} \right] \end{aligned} \quad (4.19)$$

Untuk memperoleh hasil maksimum dari tiap-tiap parameter, maka persamaan (4.18) dan (4.19) disamadengkan nol,

$$\begin{aligned}
\frac{\partial l}{\partial \hat{\mu}} &= \sum_{i=1}^n \left(\frac{x_i}{\hat{\mu}} - \frac{kx_i + 1}{1 + k\hat{\mu}} \right) = 0 \\
&\Rightarrow \sum_{i=1}^n \left(\frac{x_i}{\hat{\mu}} \right) - \sum_{i=1}^n \left(\frac{kx_i + 1}{1 + k\hat{\mu}} \right) = 0 \\
&\Rightarrow \frac{\sum_{i=1}^n x_i}{\hat{\mu}} - \frac{k \sum_{i=1}^n x_i + n}{1 + k\hat{\mu}} = 0 \\
&\Rightarrow \frac{\sum_{i=1}^n x_i}{\hat{\mu}} - \frac{k \sum_{i=1}^n x_i}{1 + k\hat{\mu}} - \frac{n}{1 + k\hat{\mu}} = 0 \\
&\Rightarrow \frac{(1 + k\hat{\mu}) \sum_{i=1}^n x_i - k\hat{\mu} \sum_{i=1}^n x_i - n\hat{\mu}}{\hat{\mu}(1 + k\hat{\mu})} = 0 \\
&\Rightarrow \frac{\sum_{i=1}^n x_i + k\hat{\mu} \sum_{i=1}^n x_i - k\hat{\mu} \sum_{i=1}^n x_i - n\hat{\mu}}{\hat{\mu}(1 + k\hat{\mu})} = 0 \\
&\Rightarrow \frac{\sum_{i=1}^n x_i - n\hat{\mu}}{\hat{\mu}(1 + k\hat{\mu})} = 0 \\
&\Rightarrow \sum_{i=1}^n x_i - n\hat{\mu} = 0 \\
&\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \\
&\Rightarrow \hat{\mu} = \bar{X}
\end{aligned} \tag{4.20}$$

$$\frac{\partial l}{\partial k} = \sum_{i=1}^n \left[\left(\sum_{v=0}^{x_i-1} -\frac{1}{k(1+k^v)} \right) + \frac{x_i}{k} + \frac{\ln(1+k\mu)}{k^2} - \frac{\mu(x_i k + 1)}{k(1+k\mu)} \right] = 0. \quad (4.21)$$

Untuk menyelesaikan persamaan (4.21) metode yang bisa digunakan adalah Newton-Raphson. Solusi dari persamaan (4.21) tersebut merupakan *Maximum Likelihood Estimate* dari \hat{k} .

Karena tidak mungkin untuk menyelesaikan secara analitik, maka perhitungan solusi dari persamaan (4.21) akan dikerjakan melalui program Matlab yang dapat dilihat pada Lampiran 7.

4.3 Metode *Bootstrap Maximum Likelihood Estimate* Sebagai Estimator k

Low Mean Problem (LMP) dapat diselesaikan dengan meningkatkan ukuran sampel. Akan tetapi tidak semua data bisa diselesaikan dengan meningkatkan ukuran sampelnya, hal itu disebabkan karena keterbatasan biaya, waktu, tenaga dan tidak dapat dilakukan pada penelitian yang bersifat merusak serta kendala-kendala lainnya.

Metode *Bootstrap Maximum Likelihood Estimate* diusulkan untuk mengestimasi nilai k yang telah ditetapkan sebelumnya dari distribusi Binomial Negatif di bawah kondisi rata-rata sampel yang rendah dan ukuran sampel yang kecil. Algoritma dari Metode *Bootstrap Maximum Likelihood Estimate* adalah sebagai berikut:

1. Mengkoleksi data pada $\chi = \{X_1, X_2, \dots, X_n\}$, kemudian menempatkan probabilitas sebesar $1/n$ untuk setiap observasi: $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$.
2. Mengambil sampel acak berukuran n dengan pengembalian dari χ : $\chi^* = \{X_1^*, X_2^*, \dots, X_n^*\}$. Kemudian menghitung varian dan rata-rata dari χ^* . Jika nilai varian kurang dari atau sama dengan rata-rata, maka ulangi kembali langkah ke-2.
3. Menghitung estimasi Bootstrap k^* dari χ^* menggunakan metode *Maximum Likelihood Estimate*.
4. Mengulangi langkah 2 dan 3 sebanyak B kali untuk memperoleh b estimasi dari k^* : $k_{1,b}^*, k_{2,b}^*, \dots, k_{b,b}^*$.

5. Mengurutkan b estimasi dari k^* , maka k bisa didekati dengan nilai median dari $k_{1,b}^*, k_{2,b}^*, \dots, k_{b,b}^*$
diagram alir dari metode Bootstrap dapat dilihat pada Lampiran 1

4.4 Simulasi

Simulasi Monte Carlo digunakan untuk menunjukkan kinerja dan kegunaan dari metode *Bootstrap Maximum Likelihood Estimate* dalam keadaan rata-rata sampel yang rendah dan ukuran sampel yang kecil. Selain itu simulasi Monte Carlo juga digunakan untuk membandingkan hasil estimasi dari *Method of Moment Estimate*, *Maximum Likelihood Estimate*, dan *Bootstrap Maximum Likelihood Estimate*. Algoritma dari simulasi tersebut adalah sebagai berikut:

1. Membangkitkan sebuah sampel $\chi = \{X_1, X_2, \dots, X_n\}$ yang berukuran n dari distribusi Binomial Negatif $X \sim NB(\mu, k)$. Kemudian menghitung varian dan rata-rata dari sampel tersebut, jika nilai varian lebih kecil atau sama dengan rata-rata, maka dibangkitkan kembali χ .
2. Menghitung estimasi dari \hat{k} menggunakan *Maximum Likelihood Estimate*, *Method of Moment Estimate* dan *Bootstrap Maximum Likelihood Estimate*.
3. Mengulangi langkah 1 sampai 2 hingga diperoleh r estimasi dari k .
4. Menghitung rata-rata dari r estimasi dari k tersebut.
5. Menghitung bias serta *Mean Square Error* dari rata-rata estimasi k .

Untuk mengetahui perbedaan kinerja dari masing-masing estimator, maka pada simulasi ini diambil nilai-nilai parameter sebagai berikut:

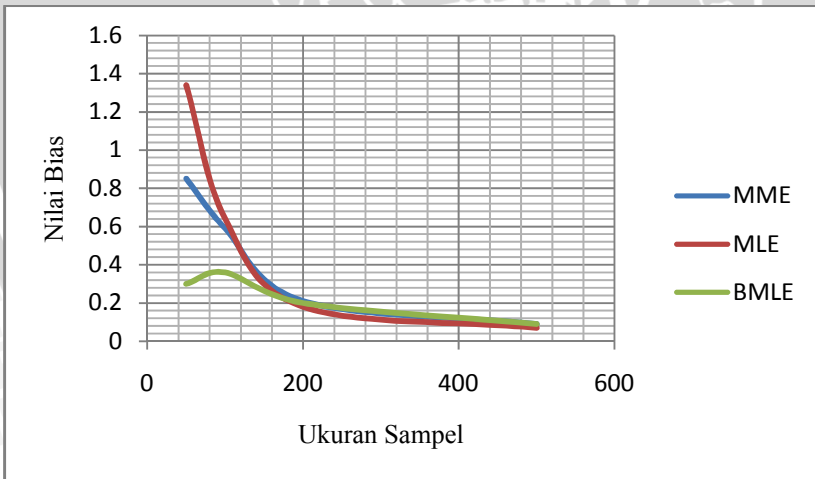
$$\begin{aligned}\mu &= 0.75, 1, 1.5, \text{ dan } 2 \\ k &= 1, 2, 3, \text{ dan } 4 \text{ untuk setiap } \mu \\ n &= 50, 100, 200, \text{ dan } 500 \\ r &= 500.\end{aligned}$$

Dengan μ adalah rata-rata dengan ukuran sampel n , k adalah parameter dispersi, dan r adalah banyaknya estimasi. Menurut Willson (1984), selama simulasi, sebuah sampel tidak digunakan jika estimasi pada sampel tersebut tidak konvergen.

Simulasi ini dijalankan melalui *software* Matlab. Bias, varian, dan *Mean Square Error* (MSE) dari sebanyak r simulasi yang dijalankan, digunakan untuk membandingkan kinerja dari ketiga *estimator* tersebut. Bias merupakan selisih antara nilai sebenarnya dengan nilai rata-rata (persamaan (2.20)) dari semua estimasi; varian adalah ukuran penyebaran dari estimasi disekitar rata-rata untuk setiap estimasi ($\sigma_k^2 = E[(\hat{k} - E[\hat{k}])^2]$); dan MSE dapat dihitung dengan persamaan (2.21). agar lebih jelas dapat dilihat diagram alir untuk simulasi estimasi parameter dispersi k pada Lampiran 2.

4.5 Hasil Analisis Data Simulasi

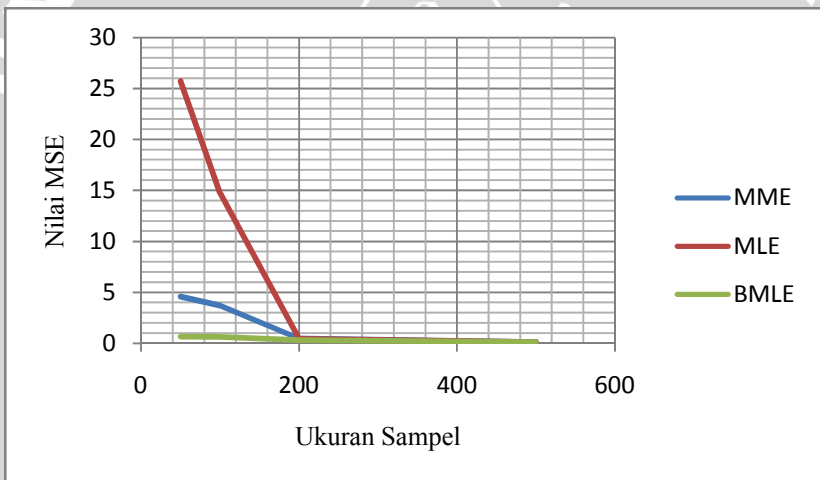
Dengan membandingkan Tabel 1 hingga Tabel 4 pada Lampiran 4 dapat dilihat bahwa nilai bias, varian dan *Mean Square Error* (MSE) meningkat seiring dengan meningkatnya nilai k , terutama untuk *Method of Moment Estimator* (MME) dan *Maximum Likelihood Estimator* (MLE). Tabel-tabel tersebut berisi hasil simulasi dari tiap perlakuan yang berbeda, yaitu ukuran sampel, nilai rata-rata (μ) dan nilai parameter dispersi k yang sudah ditentukan sebelumnya.



Sumber: Analisis data

Gambar 4.1 Grafik perbandingan bias untuk $k = 1$ dan $\mu = 1$

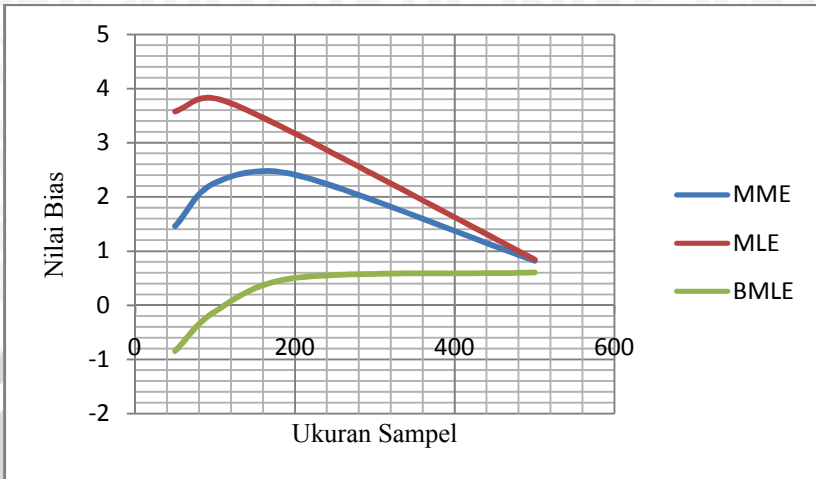
Ketika nilai k rendah, sebagai contoh $k = 1$ dan $\mu = 1$, MME dan MLE relatif akurat yang ditunjukkan dengan nilai bias yang kecil ketika ukuran sampel sebesar 100, yaitu memiliki bias sebesar 0.59 untuk MME dan 0.64 untuk MLE. Akan tetapi nilai MSE dari masing-masing estimator masih relatif besar jika dibandingkan dengan BMLE yang memiliki nilai MSE 0.63, yaitu 3.71 untuk MME dan 14.74 untuk MLE, yang artinya hasil estimasi dari MME dan MLE ini masih kurang efisien. Grafik perbandingan bias dan MSE ketiga estimator pada kondisi $k = 1$ dan $\mu = 1$ untuk tiap ukuran sampel yang berbeda dapat dilihat pada Gambar 4.1 dan Gambar 4.2.



Sumber: Analisis data

Gambar 4.2 Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 1$

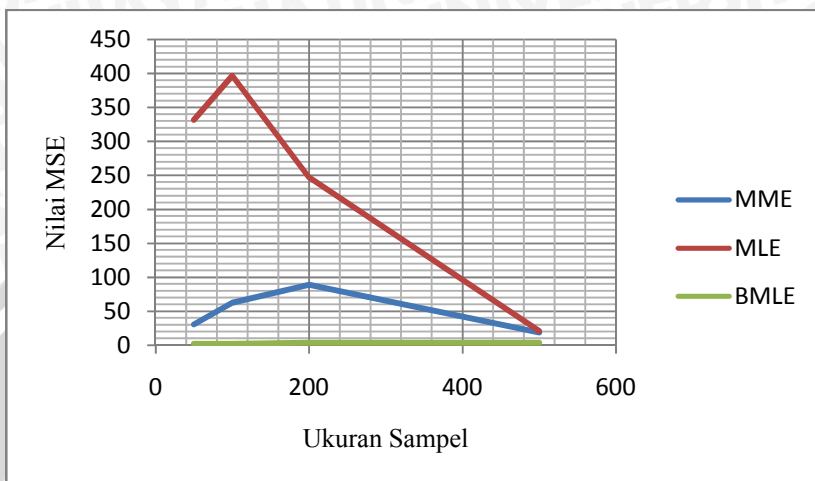
Namun untuk $k = 4$ ($\mu = 1$), ukuran sampel harus lebih dari 500 untuk memperoleh hasil estimasi yang akurat dan efisien. Meskipun pada saat ukuran sampel 500 nilai MSE dari MME dan MLE masih relatif besar, yaitu 18.80 untuk MME dan 20.71 untuk MLE. Dilain pihak, dengan ukuran sampel yang sama nilai MSE dari BMLE relatif rendah meskipun ketika nilai $k = 4$ dan $\mu = 1$ yaitu sebesar 3.82.



Sumber: Analisis data

Gambar 4.3 Grafik perbandingan bias untuk $k = 4$ dan $\mu = 1$

Selain itu BMLE juga lebih akurat dan efisien untuk ukuran sampel sebesar 100 dengan nilai bias -0.12 dan nilai MSE 2.26, jika dibandingkan dengan MME dan MLE yang memiliki nilai bias 2.26 dan 3.82, serta nilai MSE 62.46 dan 396.76 untuk masing-masing estimator dengan kondisi yang sama. Dari hasil estimasi juga dapat dilihat bahwa BMLE lebih stabil dibandingkan dengan MME dan MLE, yang ditunjukkan dengan tidak adanya perbedaan nilai bias dan MSE yang signifikan untuk tiap ukuran sampel dan rata-rata yang berbeda. Untuk lebih memperjelas perbedaan kinerja dari ketiga metode tersebut dapat dilihat pada Gambar 4.3 dan Gambar 4.4.



Sumber: Analisis data

Gambar 4.4 Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 1$

Untuk gambar yang lain bisa dilihat pada Lampiran 5 yang berisikan grafik perbandingan bias dan MSE dari ketiga estimator untuk setiap hasil simulasi.

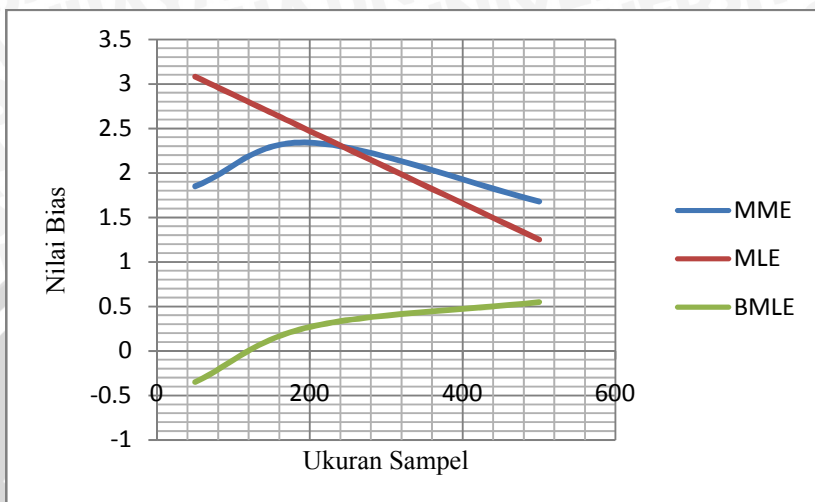
Keempat tabel tersebut juga menunjukkan bahwa secara umum MME memiliki bias yang lebih kecil dari MLE untuk sampel yang relatif kecil, hal ini sesuai seperti yang disampaikan oleh Pieters dkk (1977). MLE memberikan hasil estimasi yang lebih baik daripada MME hanya pada ukuran sampel yang relatif besar, akan tetapi kurang stabil dan efisien pada saat rata-rata sampel rendah dan nilai k tinggi. BMLE lebih baik daripada MME dalam segala kondisi, begitu juga jika dibandingkan dengan MLE, BMLE juga hampir unggul disetiap kondisi kecuali pada saat ukuran sampel sangat besar, yaitu MLE lebih akurat dibandingkan BMLE. Dari perbandingan kinerja ketiga estimator tersebut jelas terlihat bahwa BMLE memiliki lebih banyak keunggulan daripada MME dan MLE untuk mengestimasi nilai parameter dispersi k dengan rata-rata sampel yang rendah dan ukuran sampel yang kecil. BMLE memang memiliki waktu komputasi yang lebih lama daripada yang lain, akan tetapi dengan kemajuan teknologi saat ini hal tersebut dapat diatasi.

4.6 Aplikasi Menggunakan Data Kecelakaan dari Unit Lakalantas Polresta Malang

Data kecelakaan ini digunakan sebagai contoh untuk menunjukkan perbedaan kinerja dari ketiga estimator. Data kecelakaan diperoleh dari jalan rawan kecelakaan lalu lintas di Kota Malang. Yaitu Jl. Kolonel Sugiono, Jl. Supriadi, Jl. MT. Haryono, dan Jl. Basuki Rachmad pada tahun 2006 hingga 2010. Sampel data ini memiliki rata-rata 1.00 dan variannya adalah 4.30. Karakteristik dari data tersebut dapat dilihat pada Lampiran 3.

Pertama-tama ketiga estimator tersebut digunakan untuk mengestimasi parameter k pada data keseluruhan, kemudian diambil 500 subsampel dari data tersebut yang memiliki ukuran sampel 50, 100 dan 200. Hasil dari estimasi dapat dilihat pada Lampiran 6, di mana nilai estimasi k merupakan nilai rata-rata dari 500 subsampel tersebut.

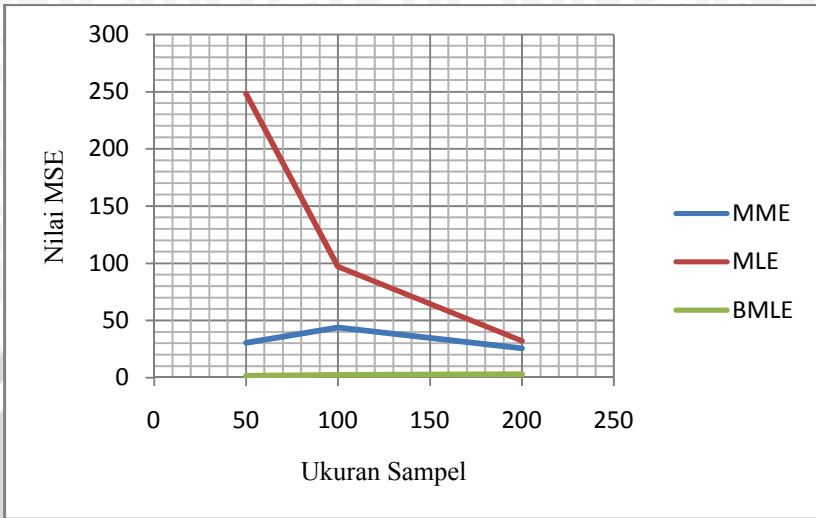
Hasil estimasi nilai k dari ketiga estimator berkisar antara 3 dan 4 untuk data keseluruhan. Estimasi nilai k menggunakan MLE dan BMLE lebih mendekati nilai sesungguhnya, yaitu berkisar pada 3.3, hal itu sesuai dengan hasil simulasi yang menunjukkan bahwa hasil estimasi dari MLE dan BMLE lebih akurat untuk ukuran sampel yang relatif besar. Untuk subsampel yang ukuran sampelnya lebih kecil dapat dilihat bahwa MME dan MLE memberikan hasil estimasi yang tidak stabil dengan varian yang relatif besar. Lain halnya dengan BMLE yang memiliki nilai varian yang jauh lebih kecil dan lebih stabil.



Sumber: Analisis data

Gambar 4.5 Grafik perbandingan nilai bias hasil estimasi pada data kecelakaan

Nilai bias hasil estimasi nilai k dari MME yaitu 1.85, 2.34 dan 1.68 untuk masing-masing ukuran subsampel, dan untuk MLE adalah 3.08, 2.47 dan 1.25, nilai-nilai tersebut cukup jauh pada nilai yang diharapkan. Sedangkan untuk BMLE memberikan nilai bias yang lebih baik, yaitu -0.35, 0.27 dan 0.55 untuk masing-masing ukuran subsampel, serta varian yang relatif kecil. Untuk grafik perbandingan nilai bias dan MSE dapat dilihat pada Gambar 4.5 dan Gambar 4.6.



Sumber: Analisis data

Gambar 4.6 Grafik perbandingan nilai MSE hasil estimasi pada data kecelakaan

Dari hasil estimasi nilai k pada data kecelakaan lalu lintas ini menunjukkan bahwa BMLE memberikan estimasi yang lebih akurat dan varian yang kecil ketika ukuran sampel kecil. Hasil estimasi ini juga menunjukkan kesetabilan dari BMLE yang ditunjukkan oleh nilai hasil estimasi dan varian yang tidak jauh berbeda untuk ukuran sampel yang berbeda.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari pembahasan dapat disimpulkan bahwa:

1. Inti dari metode *Bootstrap Maximum Likelihood Estimate* (BMLE) adalah menggabungkan metode *Maximum Likelihood Estimate* (MLE) dengan teknik resampling Bootstrap. Algoritma dari metode tersebut yaitu mengkolleksi data pada χ , melakukan resampling berukuran n sebanyak b dari χ , mengestimasi parameter dispersi k dari setiap hasil resampling, dan diperoleh hasil estimasi BMLE dengan menghitung median dari b estimasi parameter dispersi k .
2. Berdasarkan hasil analisis data simulasi dan data kecelakaan, dapat dilihat bahwa nilai bias, varian dan *Mean Square Error* (MSE) meningkat seiring dengan meningkatnya nilai k . Nilai bias, varian serta MSE berbanding terbalik dengan nilai rata-rata dan ukuran sampel. Semakin kecil nilai rata-rata dan ukuran sampel, maka nilai bias, varian dan MSE semakin besar, terutama untuk *Method of Moment Estimate* (MME) dan *Maximum Likelihood Estimate* (MLE). Sedangkan untuk *Bootstrap Maximum Likelihood Estimate* (BMLE) nilai rata-rata sampel dan ukuran sampel yang berbeda tidak memberikan perbedaan yang signifikan. Secara umum BMLE lebih akurat dan stabil dibandingkan dengan MME dan MLE dalam segala kondisi, yang ditunjukkan dengan nilai bias dan MSE yang relatif kecil jika dibandingkan dengan MME dan MLE.

5.2 Saran

Untuk selanjutnya dapat digunakan metode *Bias-Corrected Maximum Likelihood* (BCML) untuk mengestimasi nilai parameter dispersi k sebagai pembanding dari *Bootstrap Maximum Likelihood Estimate* (BMLE). Serta penggunaan *boxplot* untuk memperoleh analisis data yang lebih baik.

UNIVERSITAS BRAWIJAYA



DAFTAR PUSTAKA

- Azwar, S. *Metode Penelitian*. Pustaka Pelajar: Yogyakarta. 2001.
- Brain, L and Engelhardt M. *Introduction to Probability and Mathematical Statistic*. PWS-Kent Publishing Company: USA. 1992.
- Efron, B., and R.J. Tibshirani. *An Introduction to the Bootstrap*. New York: Chapman & Hall. 1991.
- Fishman, G. S. *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer. 1995.
- Hall, H. *Notes on Generalized Method of Moments Estimation*. Amsterdam. 1996.
- Hilbe, M. *Negative Binomial Regression, Second Edition*. New York: Cambridge University press, 2011.
- Hogg, V. *Introduction to Mathematical Statistics Sixth Edition*. United States of America, 2005.
- Indriasari, Refina. *Probabilitas dan Statistik*. Jakarta: Erlangga, 2004.
- Kismiantini. *Perbandingan Model Regresi Poisson dan Model Regresi Binomial Negatif*. Yogyakarta: FMIPA Universitas Negeri Yogyakarta. 2008.
- Klugman, S. A., Linda J., and Owen R. G. *From Data to Decision. Second Edition*. United State of America: John Willey & Sons, Inc. 2004.
- Lord, D. Modeling Motor Vehicle Crashes using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. *Accident Analysis & Prevention*, Vol. 38, No. 4, pp. 751-766, 2006.

Mishra S.H. dan Dudewiez E. J. *Modern Mathematical Statistics*. New York: Wiley. 1988

Piegorsch, W.W. Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. *Biometrics* 46, pp. 863-867, 1990.

Pieters, E.G., Gates C.E., Matis J.H., and Sterling W.L. Small Sample Comparisons of Different Estimators of Negative Binomial Parameters. *Biometrics* 33, pp. 718-723, 1977.

Poch, M., and Fred L. M. Negative Binomial Analysis of Intersection Accident Frequency. *Journal of Transportation Engineering*, Vol. 122, No. 2, pp.105-113, 1996.

Ripley, B. D. *Stochastic Simulation*. New York: Wiley & Sons. 1987.

Rosenfeld, R. *The McGraw Hill 36-Hour Business Statistics Course*. McGraw-Hill Inc: Singapore. 1992.

Salusu, A. *Metode Numerik*. Jakarta: Graha Ilmu. 2008.

Sawilowsky, S. "You think you've got trivials?". *Journal of Modern Applied Statistical Methods* 2 (1): 218–225. 2003.

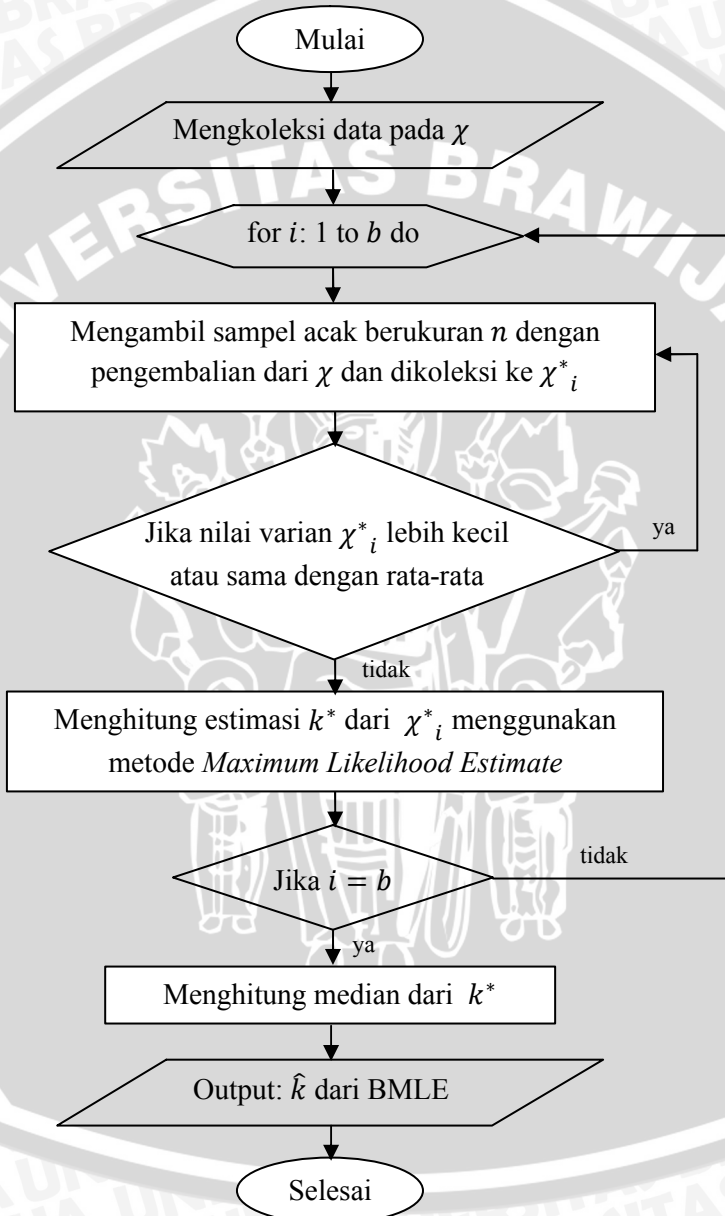
Tucker, Alan. *Applied Combinatorics Second Edition*. Canada: John Wiley & Sons, Inc. 1984.

Walpole, R. E. dan Myers H. R. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Edisi ke-4*. Bandung: ITB Bandung. 1995.

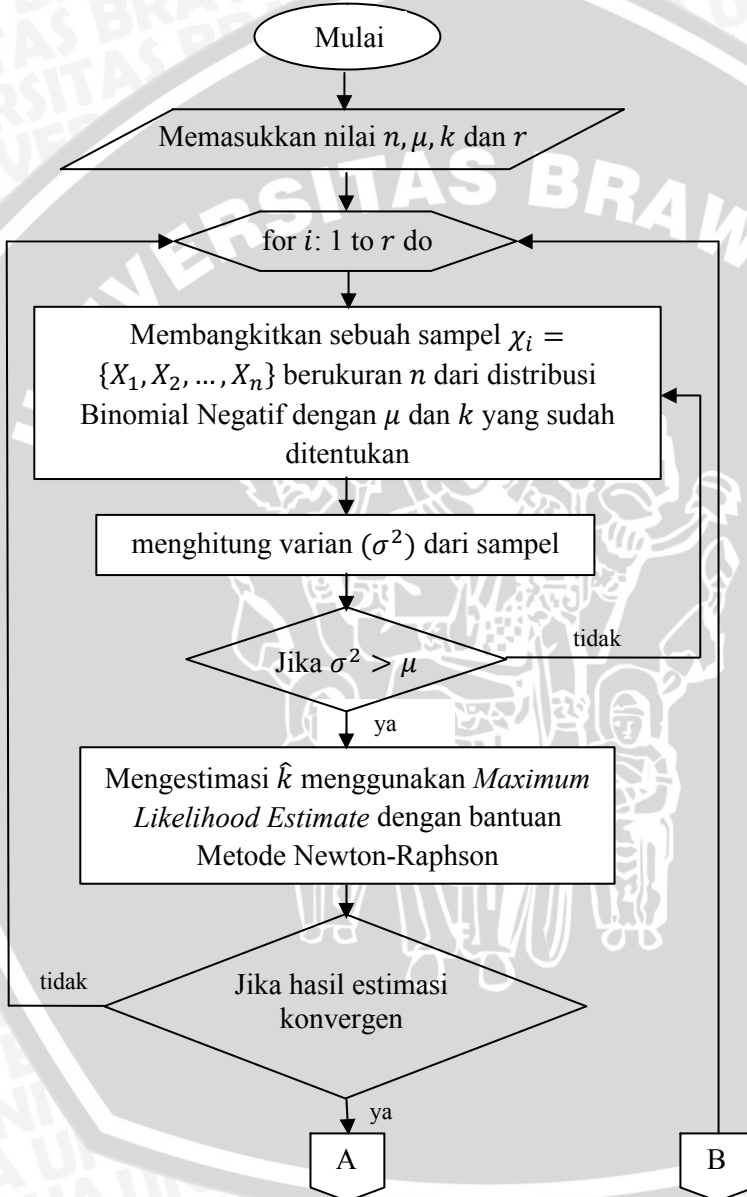
William, W. *Probabilita dan Statistik dalam Ilmu Rekayasa dan Manajemen*. Jakarta: Universitas Indonesia, 1990.

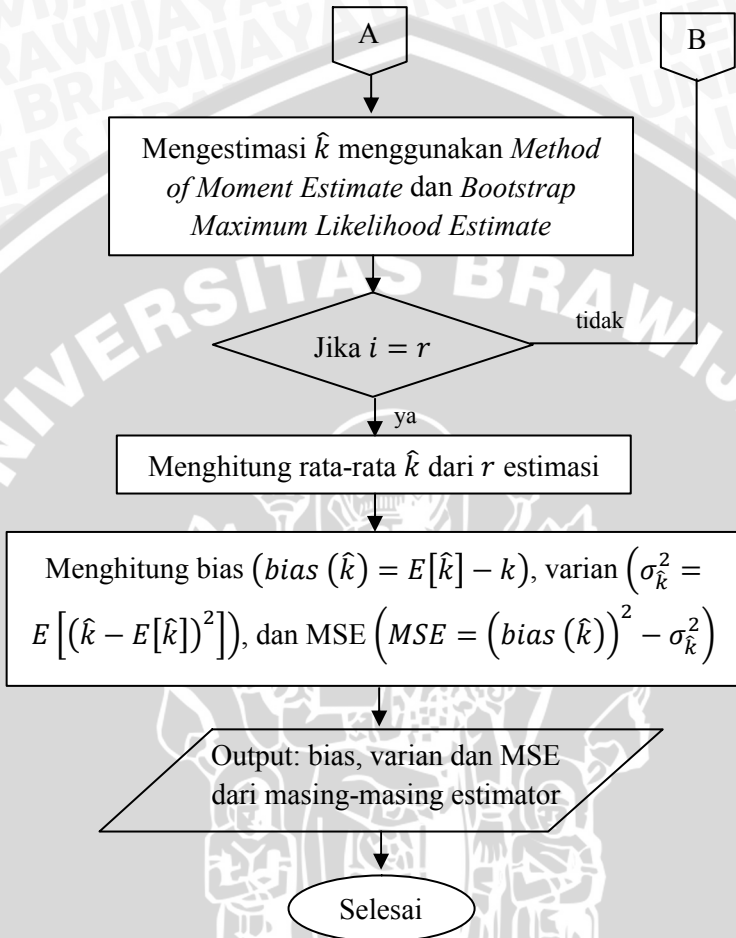
Willson, L.J., J.L. Folks, and Young J.H. Multistage Compared with Fixed-Sample-Size Estimation of the Negative Binomial Parameter k . *Biometrics*, Vol. 109-117, 1984.

Lampiran 1: Diagram alir Metode *Bootstrap Maximum Likelihood Estimate*



Lampiran 2: Diagram alir simulasi estimasi parameter dispersi k





Lampiran 3: Karakteristik data kecelakaan

No	Jenis Kecelakaan	Tahun Kejadian				
		2006	2007	2008	2009	2010
1	Meninggal	76	59	64	42	23
2	Luka Parah	62	65	68	72	57
3	Luka Ringan	530	225	144	241	204
	Total	668	349	276	255	284

Sumber: Satlantas Polresta Malang

Rata-rata	Standar Deviasi	Kepadatan Kendaraan	Max-Min Jumlah Kecelakaan per Hari	Nilai parameter dispersi (k)
1.00	2.07	5,857-53,531	5-0	3.3

Lampiran 4: Tabel hasil analisis data simulasi

Tabel 1 Hasil analisis data simulasi untuk $k = 1$

Mean	Ukuran sampel	MME			MLE			BMLE		
		Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
0.75	50	0.69	2.60	3.09	1.11	17.78	19.01	0.14	0.39	0.41
	100	0.88	7.70	8.47	1.55	110.69	113.1	0.28	0.46	0.54
	200	0.26	0.82	0.89	0.23	0.97	1.02	0.21	0.30	0.34
	500	0.14	0.21	0.23	0.11	0.18	0.19	0.13	0.18	0.20
1	50	0.85	3.86	4.58	1.34	23.91	25.70	0.30	0.56	0.65
	100	0.59	3.36	3.71	0.64	14.33	14.74	0.36	0.51	0.63
	200	0.21	0.43	0.48	0.18	0.39	0.42	0.20	0.26	0.31
	500	0.09	0.12	0.13	0.07	0.09	0.10	0.09	0.10	0.11
1.5	50	0.76	3.53	4.11	1.05	27.81	28.91	0.43	0.68	0.86
	100	0.35	2.35	2.48	0.43	31.99	32.17	0.28	0.43	0.51
	200	0.14	0.16	0.18	0.11	0.13	0.14	0.14	0.14	0.16
	500	0.06	0.07	0.07	0.05	0.05	0.05	0.06	0.05	0.06
2	50	0.69	3.22	3.69	0.76	7.81	8.39	0.52	0.94	1.21
	100	0.22	0.41	0.46	0.17	0.34	0.37	0.22	0.33	0.38
	200	0.12	0.11	0.13	0.09	0.09	0.09	0.11	0.09	0.11
	500	0.05	0.05	0.05	0.03	0.04	0.04	0.04	0.04	0.04

Sumber: Analisis data

Tabel 2 Hasil analisis data simulasi untuk $k = 2$

mean	Ukuran sampel	MME			MLE			BMLE		
		Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
0.75	50	1.39	14.77	16.69	2.87	235.55	243.78	-0.01	0.74	0.74
	100	1.31	19.77	21.00	1.77	66.72	69.87	0.28	1.07	1.14
	200	0.98	33.36	34.31	1.94	841.80	845.57	0.44	1.21	1.41
	500	0.20	0.82	0.86	0.19	0.78	0.82	0.21	0.56	0.61
1	50	1.38	15.32	17.23	2.21	117.46	122.34	0.32	1.12	1.23
	100	1.07	13.40	14.54	1.30	41.69	43.39	0.50	1.33	1.58
	200	0.35	1.62	1.74	0.33	1.57	1.68	0.34	1.04	1.15
	500	0.14	0.31	0.33	0.12	0.28	0.30	0.15	0.30	0.32
1.5	50	1.21	17.90	19.37	1.91	161.23	164.87	0.55	1.45	1.75
	100	0.44	2.07	2.27	0.43	2.23	2.42	0.43	1.17	1.35
	200	0.17	0.48	0.51	0.15	0.43	0.46	0.20	0.46	0.50
	500	0.10	0.18	0.19	0.08	0.15	0.16	0.10	0.16	0.17
2	50	0.83	11.33	12.02	1.19	165.29	166.71	0.61	1.77	2.14
	100	0.31	1.10	1.20	0.30	1.12	1.21	0.36	0.95	1.08
	200	0.16	0.33	0.36	0.14	0.28	0.30	0.18	0.29	0.33
	500	0.06	0.11	0.11	0.05	0.09	0.09	0.06	0.09	0.10

Sumber: Analisis data

Tabel 3 Hasil analisis data simulasi untuk $k = 3$

mean	Ukuran sampel	MME			MLE			BMLE		
		Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
0.75	50	1.30	22.25	23.95	3.69	434.12	447.72	-0.68	0.86	1.32
	100	2.24	51.12	56.12	4.16	387.81	405.11	-0.07	1.34	1.35
	200	1.82	44.85	48.15	2.33	106.92	112.34	0.35	1.89	2.01
	500	0.66	16.94	17.37	0.72	30.20	30.72	0.45	2.05	2.26
1	50	1.61	26.19	28.79	3.48	797.57	809.66	0.44	1.23	1.26
	100	1.95	33.32	37.14	2.77	127.68	135.36	0.54	2.08	2.27
	200	0.93	11.23	12.09	0.98	12.86	13.83	0.39	2.27	2.55
	500	0.37	2.24	2.38	0.38	2.34	2.48	0.17	1.61	1.76
1.5	50	2.12	37.55	42.03	3.77	343.90	358.08	0.56	2.63	2.94
	100	1.29	23.20	24.86	1.54	51.16	53.52	0.74	2.97	3.51
	200	0.49	3.27	3.52	0.50	3.32	3.57	0.51	2.20	2.47
	500	0.19	0.64	0.67	0.18	0.60	0.63	0.22	0.62	0.67
2	50	2.06	46.27	50.50	3.35	275.58	286.80	0.92	3.62	4.47
	100	0.80	7.84	8.47	0.80	10.03	10.68	0.71	2.77	3.28
	200	0.31	1.27	1.37	0.31	1.22	1.31	0.38	1.28	1.43
	500	0.11	0.36	0.38	0.10	0.34	0.35	0.17	0.34	0.36

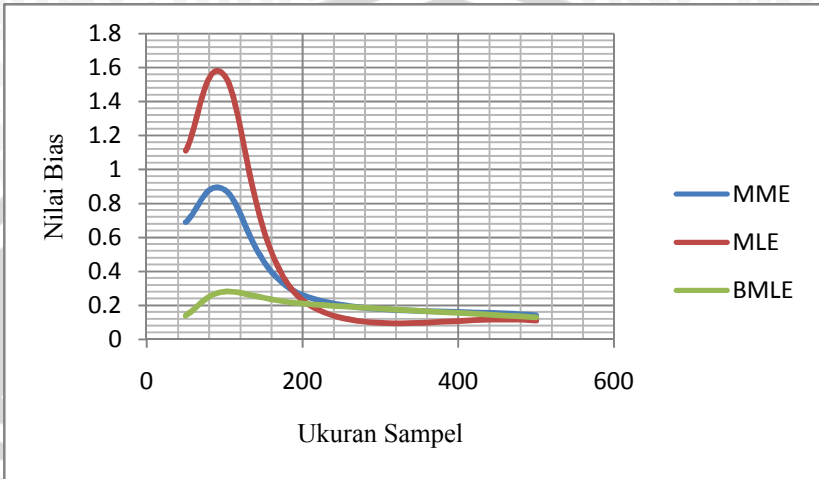
Sumber: Analisis data

Tabel 4 Hasil analisis data simulasi untuk $k = 4$

mean	Ukuran sampel	MME			MLE			BMLE		
		Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
0.75	50	0.80	23.04	23.69	3.45	407.23	419.14	-1.55	0.85	3.25
	100	2.29	58.11	63.37	4.37	334.45	353.48	-0.70	1.40	1.89
	200	2.98	120.02	128.89	5.18	829.51	856.37	0.00	2.25	2.25
	500	1.94	91.61	95.37	2.31	213.92	219.25	0.62	3.63	4.01
1	50	1.46	28.49	30.62	3.58	319.03	331.84	-0.85	1.46	2.18
	100	2.26	57.34	62.46	3.82	382.18	396.76	-0.12	2.25	2.26
	200	2.41	82.98	88.77	3.17	236.82	246.88	0.50	3.30	3.55
	500	0.82	18.14	18.80	0.84	20.00	20.71	0.60	3.46	3.82
1.5	50	2.40	61.00	66.76	7.40	2828.00	2882.76	0.10	2.70	2.71
	100	2.46	71.07	77.13	3.64	264.57	277.79	0.73	4.06	4.59
	200	1.31	55.52	57.23	1.80	425.10	428.36	0.78	4.48	5.08
	500	0.28	1.73	1.81	0.28	1.64	1.71	0.34	1.68	1.80
2	50	2.60	52.54	59.30	4.33	617.93	636.68	0.74	4.17	4.71
	100	1.91	61.97	65.62	2.64	307.03	314.00	0.93	4.96	5.83
	200	0.83	8.16	8.85	0.85	8.66	9.38	0.81	4.06	4.71
	500	0.21	1.11	1.16	0.22	1.08	1.12	0.27	1.12	1.19

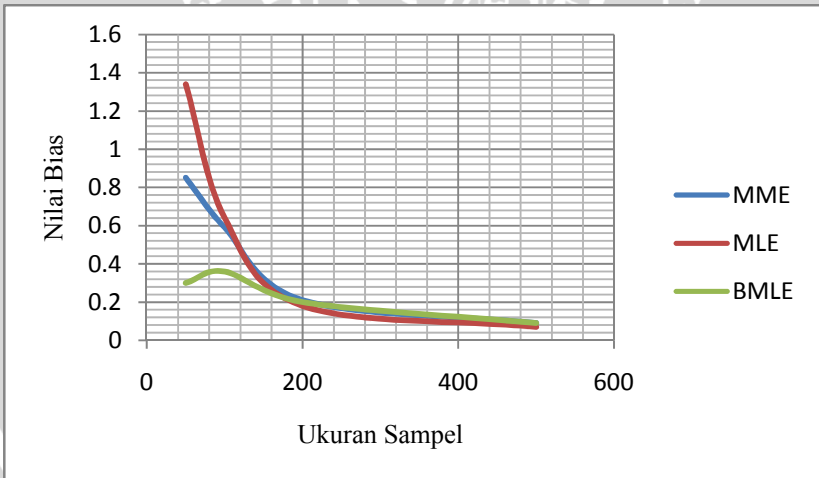
Sumber: Analisis data

Lampiran 5: Grafik perbandingan bias dan MSE
Grafik perbandingan bias untuk $k = 1$ dan $\mu = 0.75$



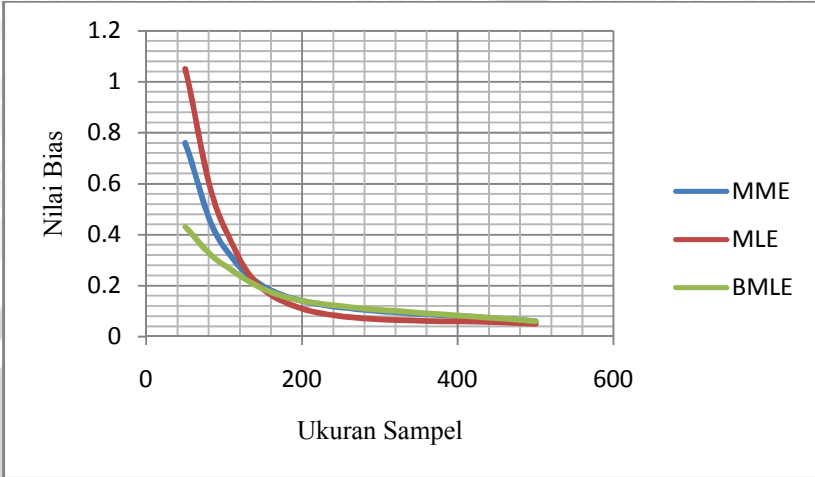
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 1$ dan $\mu = 1$



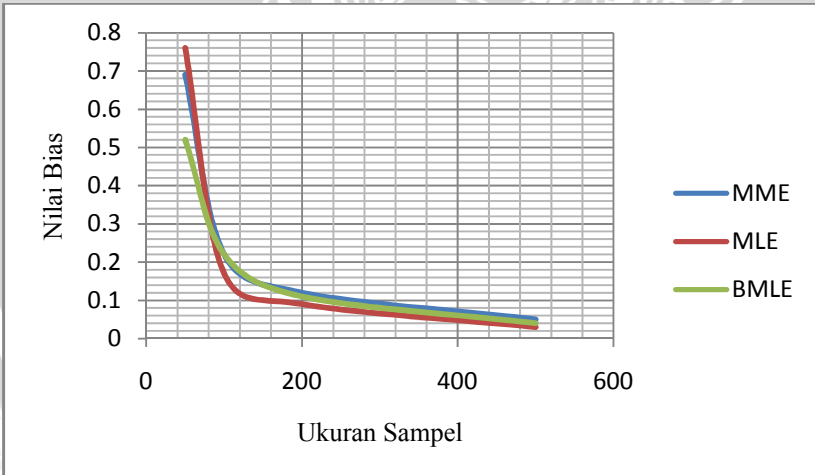
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 1$ dan $\mu = 1.5$



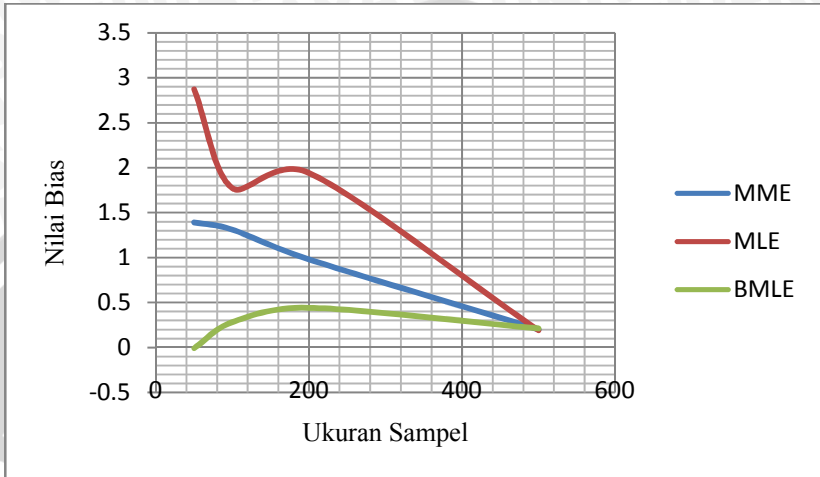
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 1$ dan $\mu = 2$



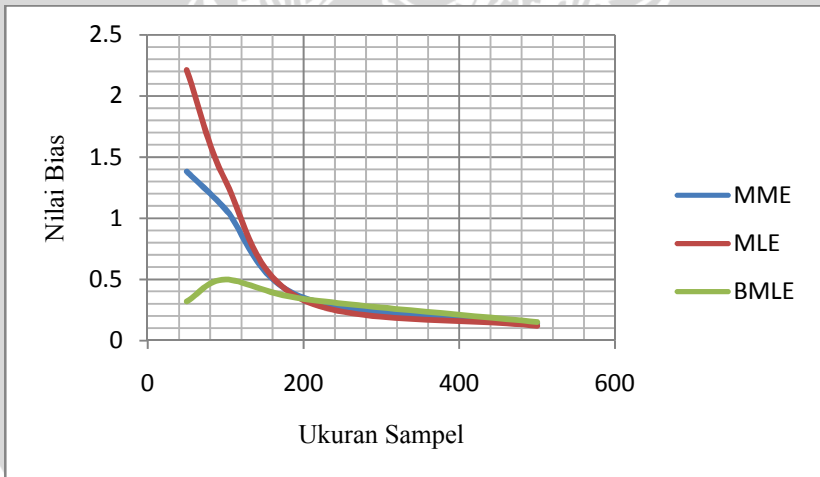
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 2$ dan $\mu = 0.75$



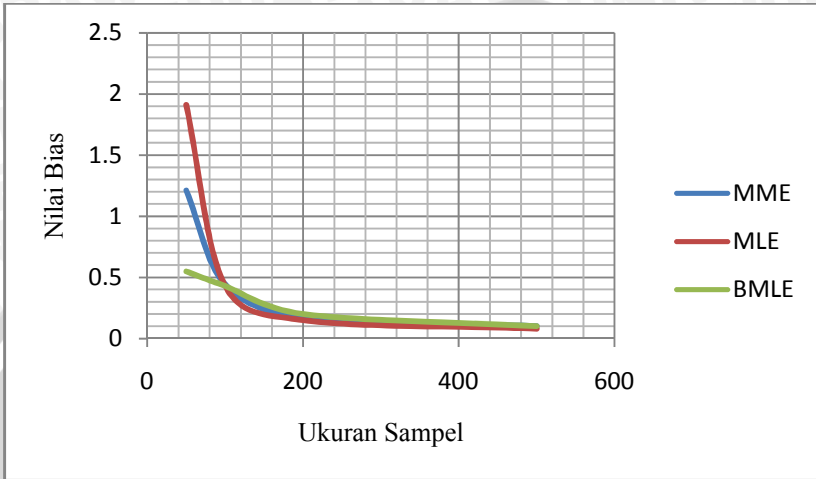
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 2$ dan $\mu = 1$



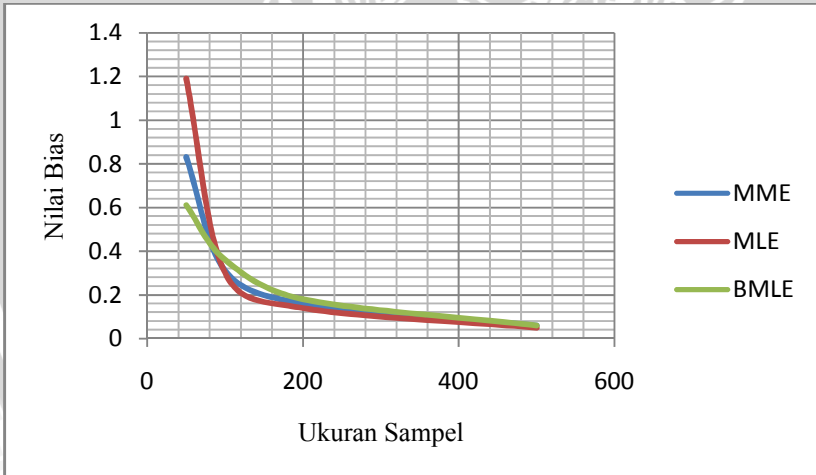
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 2$ dan $\mu = 1.5$



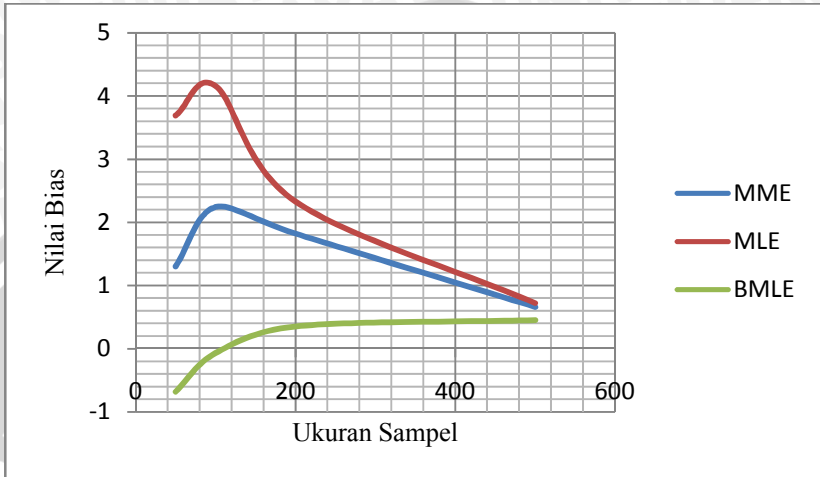
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 2$ dan $\mu = 2$



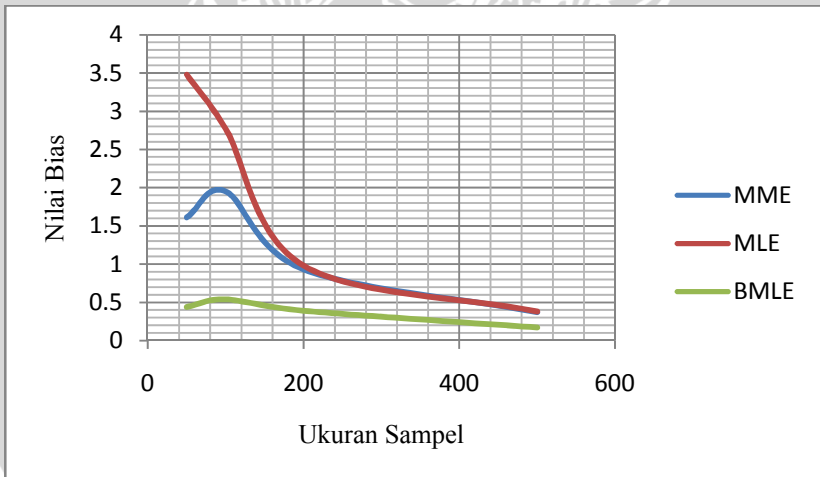
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 3$ dan $\mu = 0.75$



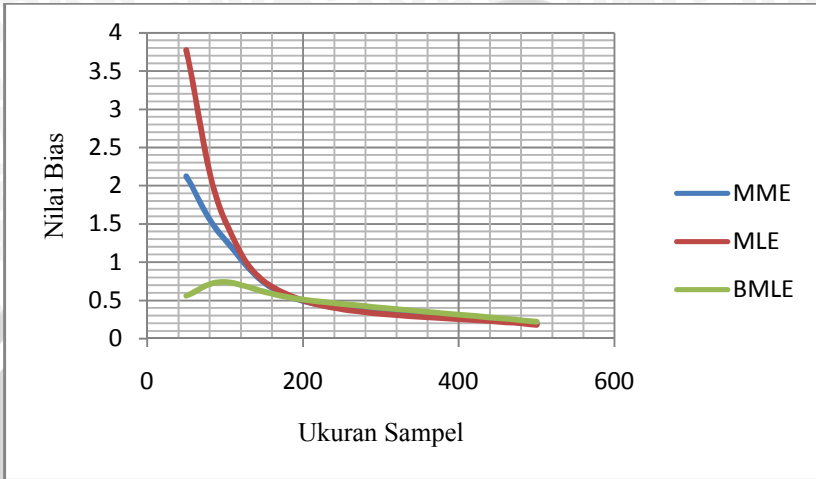
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 3$ dan $\mu = 1$



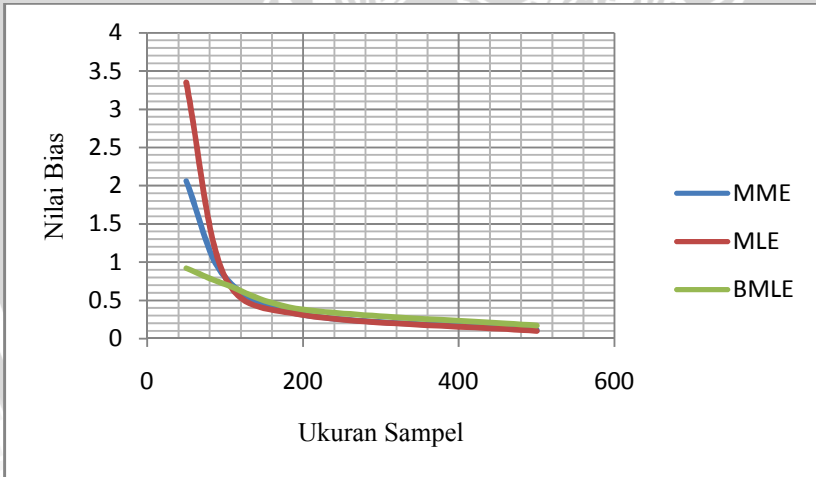
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 3$ dan $\mu = 1.5$



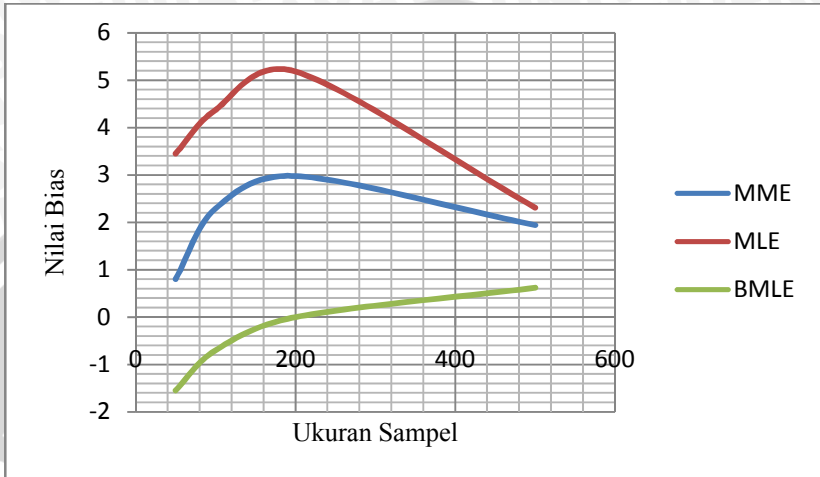
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 3$ dan $\mu = 2$



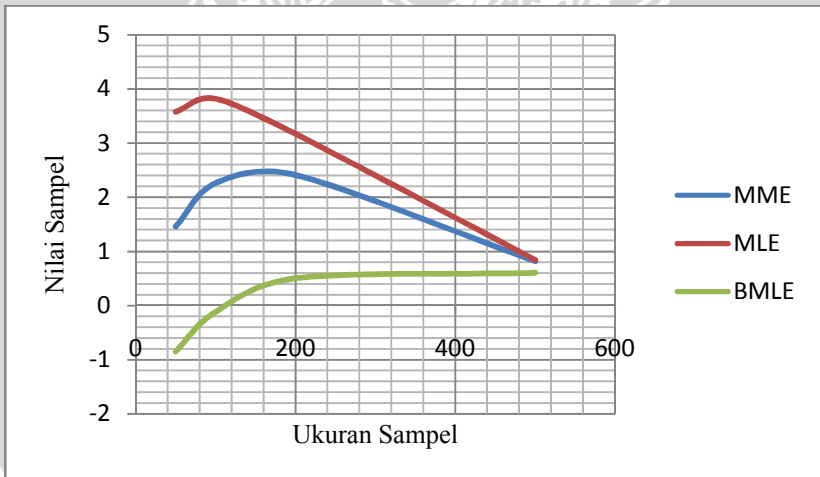
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 4$ dan $\mu = 0.75$



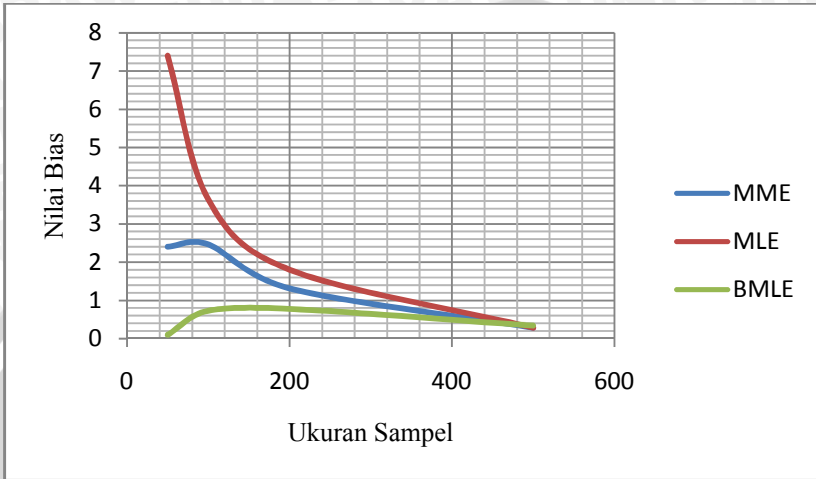
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 4$ dan $\mu = 1$



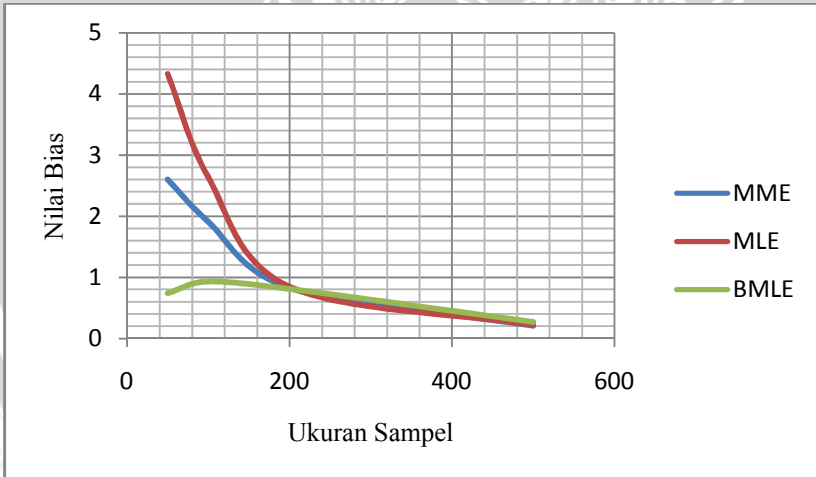
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 4$ dan $\mu = 1.5$



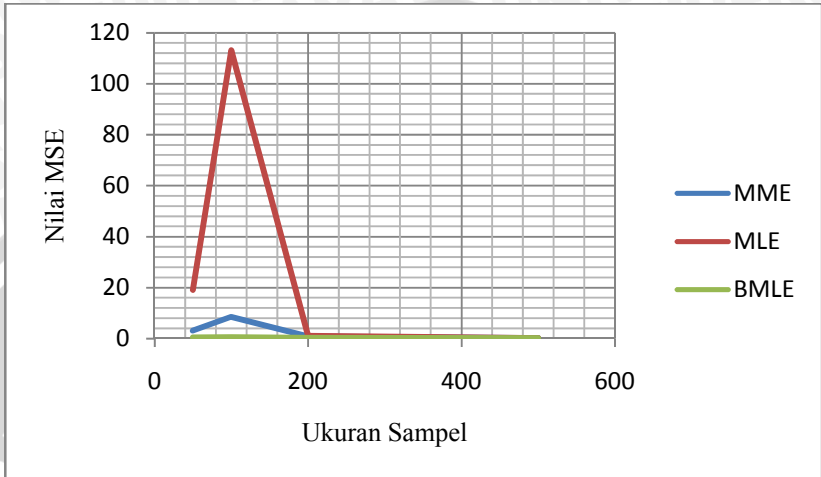
Sumber: Analisis data

Grafik perbandingan bias untuk $k = 4$ dan $\mu = 2$



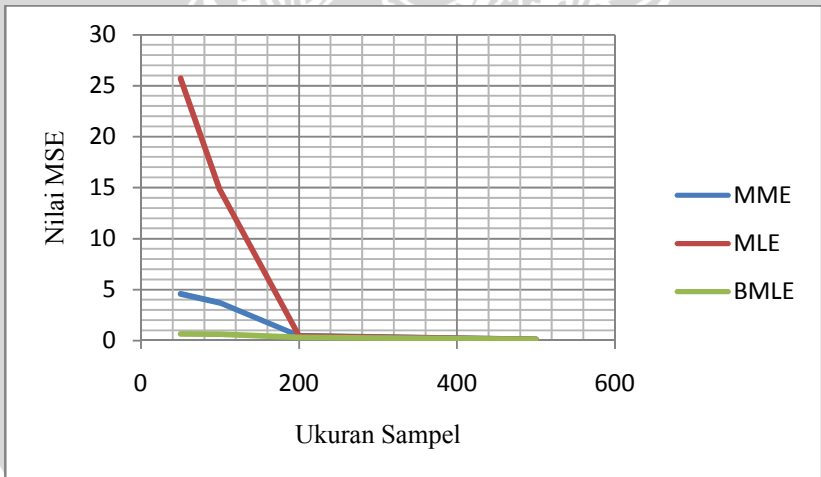
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 0.75$



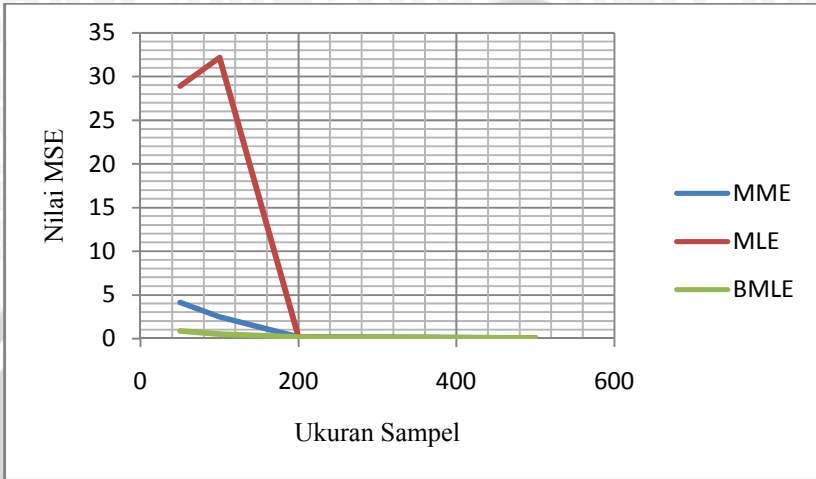
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 1$



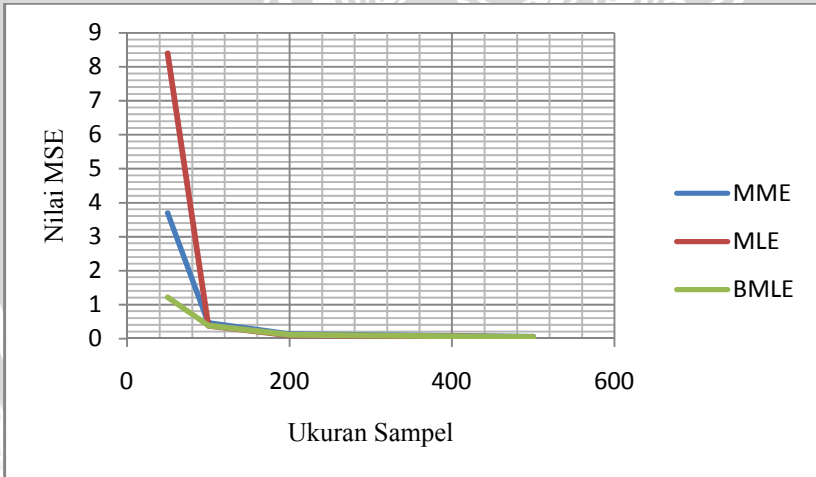
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 1.5$



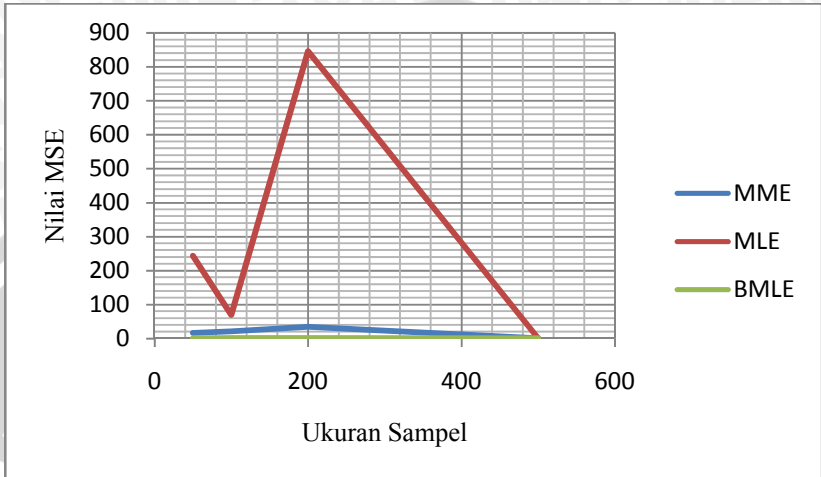
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 1$ dan $\mu = 2$



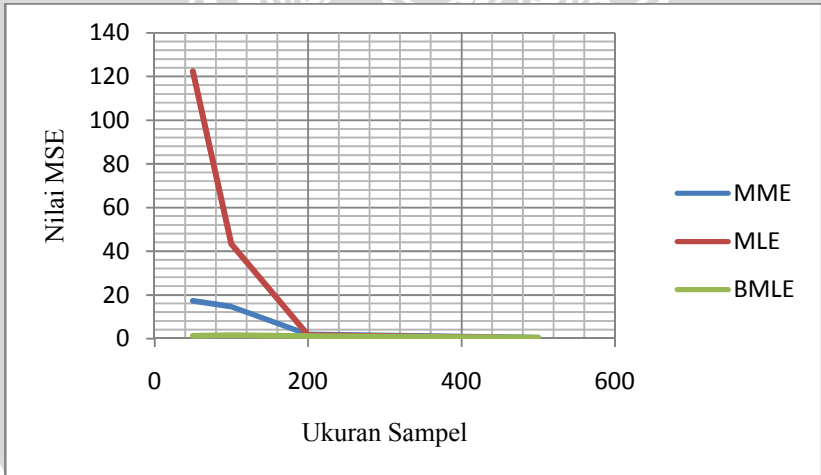
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 2$ dan $\mu = 0.75$



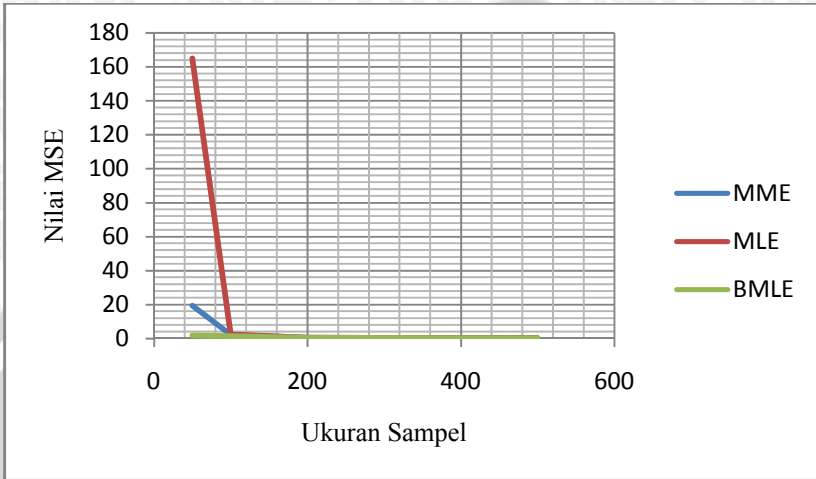
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 2$ dan $\mu = 1$



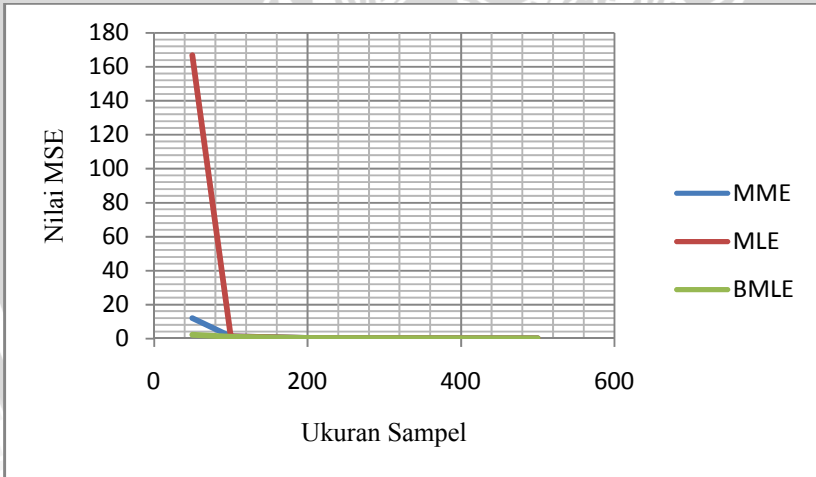
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 2$ dan $\mu = 1.5$



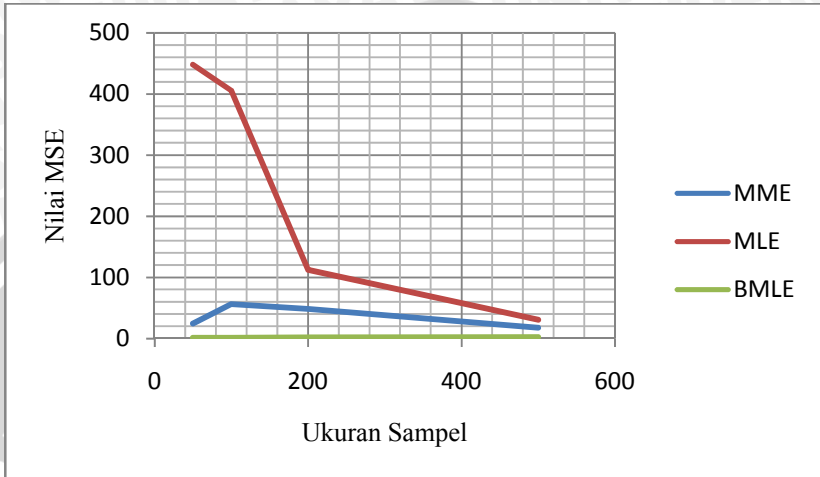
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 2$ dan $\mu = 2$



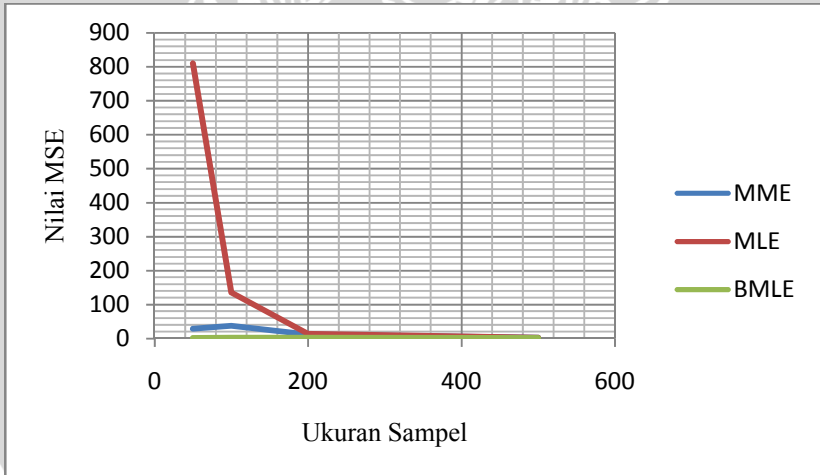
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 3$ dan $\mu = 0.75$



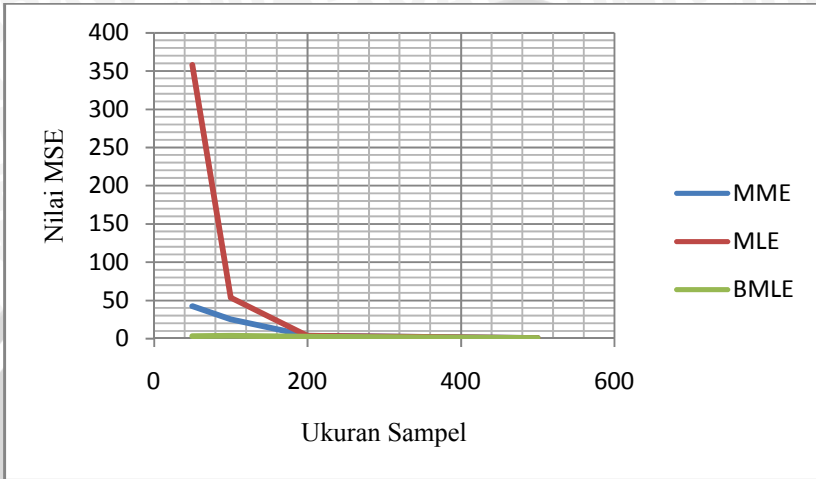
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 3$ dan $\mu = 1$



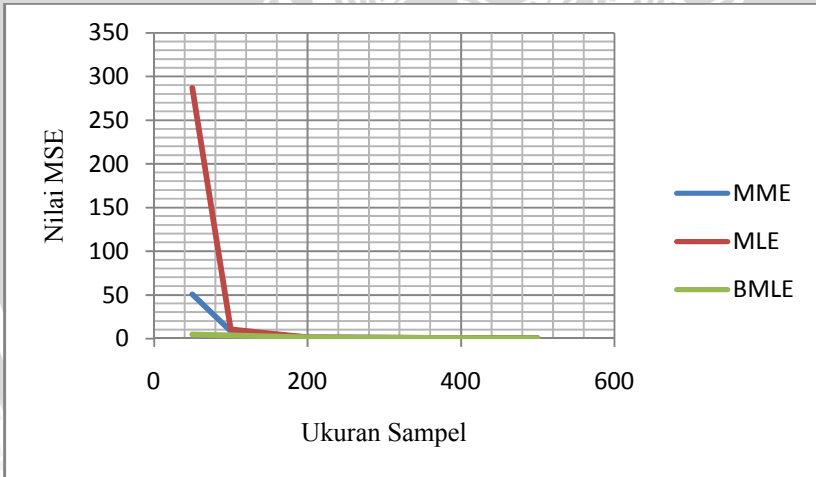
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 3$ dan $\mu = 1.5$



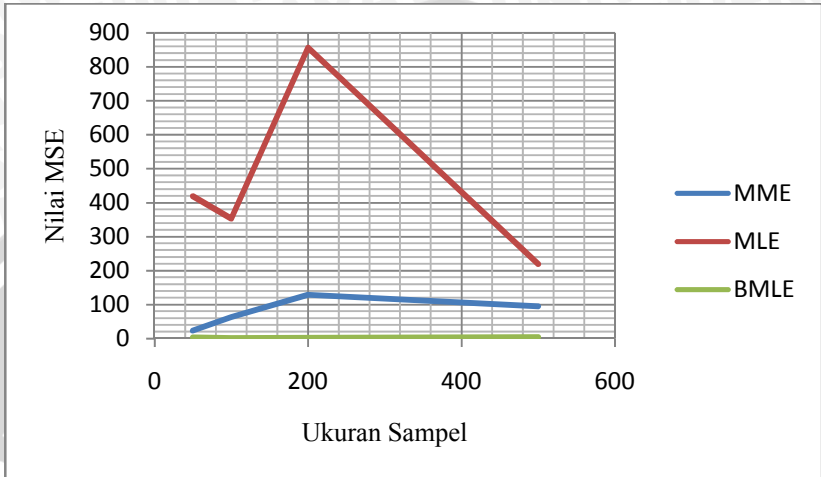
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 3$ dan $\mu = 2$



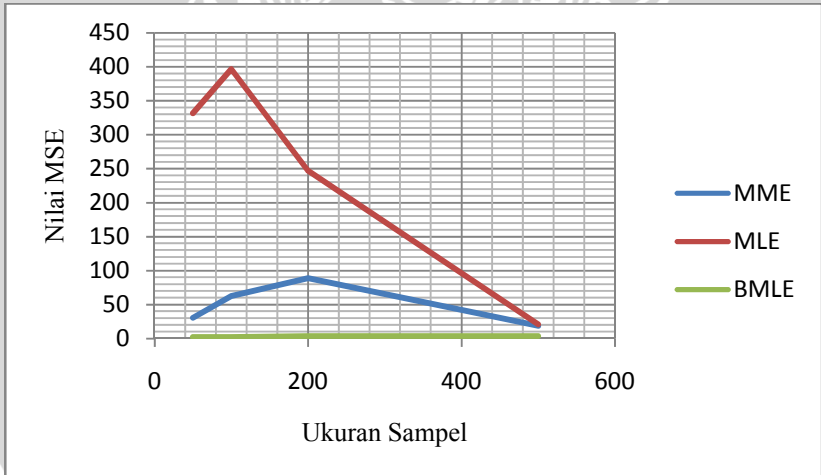
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 0.75$



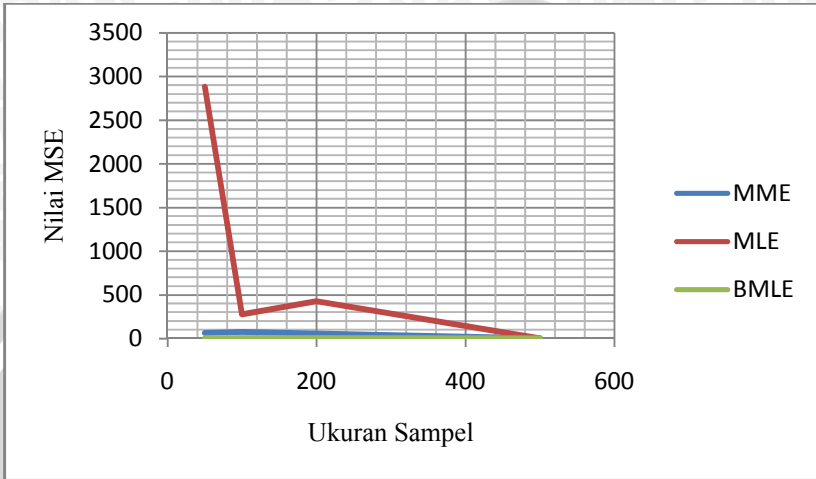
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 1$



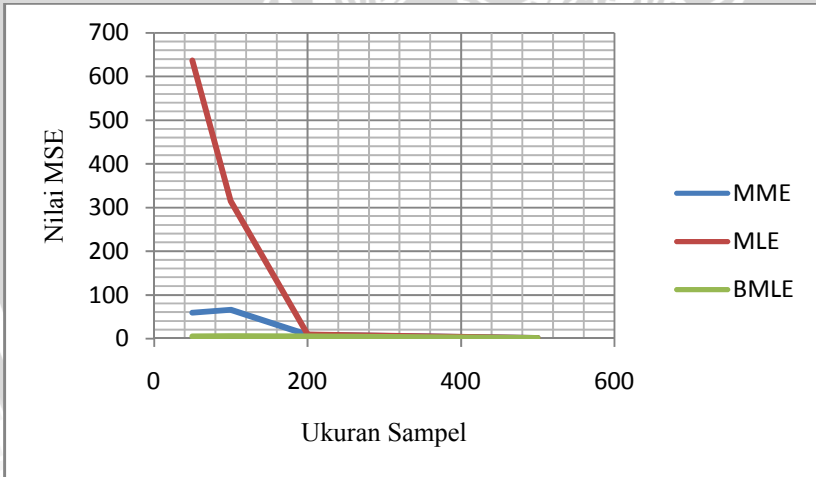
Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 1.5$



Sumber: Analisis data

Grafik perbandingan MSE untuk $k = 4$ dan $\mu = 2$



Sumber: Analisis data

Lampiran 6: Tabel hasil analisis data kecelakaan

Ukuran Sampel	MME			MLE			BMLE		
	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
1825	0.48	-	0.23	-0.04	-	0.00	0.05	-	0.00
50	1.85	26.99	30.41	3.08	238.55	248.04	-0.35	1.49	1.61
100	2.34	38.49	43.97	2.47	90.82	96.92	0.27	2.36	2.43
200	1.68	22.88	25.70	1.25	30.56	32.12	0.55	2.68	2.98

Sumber: Analisis data



Lampiran 7: Listing program

```
clear all;
clc;
format long
mu=[0.75 1 1.5 2];
n=[50 100 200 500];
ulangan=500;
p2=0.5; sumF=0; sumG=0; sumF1=0; sumG1=0;

for k=1:4
fprintf('\n
Tabel Perbandingan\n
k = %d\n',k)
fprintf('=====
=====
');
fprintf('  Mean      n      MME
MLE      BMLE\n');
fprintf('
Bias      Var      MSE      Bias      Var      MSE
Bias      Var      MSE      Bias      Var      MSE\n');
fprintf('=====
=====
');
for i=1:4
fprintf('  %2.2f\n',mu(i));
for j=1:4
r=1; itNR=1; itM=1;
P1=[];
P2=[];
KK2=[];
while r<=500
x =
random('nbin',1/k,mu(i)*k/(1+mu(i)*k),1,n(j));
varX = var(x);
meanX = mean(x);
while varX<=meanX
x =
random('nbin',1/k,mu(i)*k/(1+mu(i)*k),1,n(j));
varX = var(x);
meanX = mean(x);
end;
p1 = (varX-meanX)/meanX^2;
P1 = [P1 p1];
```

```

while itNR<=15
for a=1:n(j)
    f = log(gamma(1/p2+x(a))/gamma(1/p2)) +
    x(a)/p2 + log(1+p2*mu(i))/p2^2 -
    mu(i)*(x(a)*p2+1)/(p2*(1+p2*mu(i)));
    g1=-x(a)/p2^2 +
    mu(i)/(p2^2*(1+p2*mu(i))) - 2*log(1+p2*mu(i))/p2^3
    - mu(i)*x(a)/(p2*(1+p2*mu(i))) +
    (1+p2*x(a))*mu(i)^2/(p2*(1+p2*mu(i))^2) +
    (1+p2*x(a))*mu(i)/(p2^2*(1+p2*mu(i)));
    g = g1 + (Psi(1/p2) -
    Psi(1/p2+x(a)))/p2^2;
    sumF = sumF + f;
    sumG = sumG + g;
end;
p2 = p2 - sumF/sumG;
itNR = itNR + 1;
P2 = [P2 p2];
end;
for kw=1:ulangan,
for ii=1:n(j),
    m=randint(1,1,[1 n(j)]);
    yy(ii)=x(m);
end
varY = var(yy);
meanY = mean(yy);
while varY<=meanY
for ii=1:n(j),
    m=randint(1,1,[1 n(j)]);
    yy(ii)=x(m);
end
varY = var(yy);
meanY = mean(yy);
end;
for ii=1:n(j)
    y(kw,ii)=yy(ii);
end
end

kk2(j)=0.5;
while itM<=15
for kw=1:ulangan,
for a=1:n(j),

```

```

f1 =
log(gamma(1/kk2(j)+y(kw,a))/gamma(1/kk2(j))) +
y(kw,a)/kk2(j) + log(1+kk2(j)*mu(i))/kk2(j)^2 -
mu(i)*(y(kw,a)*kk2(j)+1)/(kk2(j)*(1+kk2(j)*mu(i)));
g11=-y(kw,a)/kk2(j)^2 +
mu(i)/(kk2(j)^2*(1+kk2(j)*mu(i))) -
2*log(1+kk2(j)*mu(i))/kk2(j)^3 -
mu(i)*y(kw,a)/(kk2(j)*(1+kk2(j)*mu(i))) +
(1+kk2(j)*y(kw,a))*mu(i)^2/(kk2(j)*(1+kk2(j)*mu(i))
^2) +
(1+kk2(j)*y(kw,a))*mu(i)/(kk2(j)^2*(1+kk2(j)*mu(i))
);
g1 = g11 + (Psi(1/kk2(j))-
Psi(1/kk2(j)+y(kw,a)))/kk2(j)^2;
sumF1 = sumF1 + f1;
sumG1 = sumG1 + g1;
end;
end;
kk2(j) = kk2(j) - sumF1/sumG1;
itM = itM + 1;
end;
KK2=[KK2 kk2];
kk3= median(KK2);
KK3=[KK3 kk3];
r = r+1;
end;
bias1 = mean(P1)-k;
varian1 = var(P1);
MSE1 = bias1^2 + varian1;
bias2 = mean(P2)-k;
varian2 = var(P2);
MSE2 = bias2^2 + varian2;
bias_ky = mean(KK3)-k;
varian_ky = var(KK3);
MSE_ky = bias_ky^2 + varian_ky;
fprintf(' %11.0f %8.2f %7.2f %7.2f %8.2f
%7.2f %7.2f %9.2f %8.2f
%8.2f\n',n(j),bias1,varian1,MSE1,bias2,varian2,MSE2
,bias_ky,varian_ky,MSE_ky);
end;
end;
end;

```