

## ESTIMASI PARAMETER DISTRIBUSI BINOMIAL NEGATIF-*GENERALIZED* EKSPONENSIAL (BN-GE) PADA DATA OVERDISPERSI

Annisa Ulfiyah<sup>1)</sup>, Rini Cahyandari<sup>2)</sup>, dan Asep Solih Awalluddin<sup>3)</sup>

Jurusan Matematika, Fakultas Sains dan Teknologi, UIN Sunan Gunung Djati

Jl. A. H. Nasution No. 105 Bandung

<sup>1)</sup> [ulfiya.annisa@gmail.com](mailto:ulfiya.annisa@gmail.com), <sup>2)</sup> [rcahyandari@yahoo.com](mailto:rcahyandari@yahoo.com), <sup>3)</sup> [aasolih@gmail.com](mailto:aasolih@gmail.com)

**Abstract:** Poisson distribution is commonly used to modelling count data that is assumed of similarity the average and variance value. But in fact, often found the variance value exceeds the average or better known as overdispersion. Count data that contain excess zeros became one of the cause overdispersion, but the converse does not hold. Poisson distribution is applied to the data that contains over dispersion with excess zeros will generate the calculations becomes underestimate. One of the distribution that can be used is a Negative Binomial distribution. While the Negative Binomial distribution is usually used as an alternative of Poisson distribution will generate the calculations becomes overestimate when applied to the data that contains overdispersion with excess zeros, so it is cannot be used efficiently. Therefore, it is necessary to find another distribution that can be applied for such data analyzing. The Negative Binomial-Generalized Exponential distribution which is a mixed of the Negative Binomial distribution with a Generalized Exponential distribution can be used as an alternative to Poisson and Negative Binomial distribution. In this final task will be explained about estimating parameter of the Negative Binomial-Generalized Exponential distribution by maximum likelihood method which the maximum likelihood estimates may be solved by numerical optimization using statistical software R and the result of parameter estimation used in fitting data which applied at data sample that contains overdispersion. Through the goodness of fit tests using the chi-square test and based on the  $p$ -value with a significance level of 0,05 was obtained that data sample that contains overdispersion spread out followed the Negative Binomial-Generalized Exponential distribution.

**Keywords:** count data with excess zeros, Poisson distribution, Negative Binomial distribution, Negative Binomial-Generalized Exponential distribution, overdispersion, parameter estimation.

**Abstrak:** Distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) merupakan distribusi peluang gabungan antara distribusi Binomial Negatif (BN) dengan distribusi *Generalized* Eksponensial (GE). Distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) ini cocok digunakan untuk memodelkan data cacah dengan nol berlebih, dimana data cacah yang memiliki nilai nol berlebih menjadi salah satu penyebab terjadinya overdispersi, tetapi tidak sebaliknya. Overdispersi merupakan keadaan yang timbul ketika varians data lebih besar dari mean dan umumnya sering terjadi dalam pengamatan suatu data cacah. Tulisan ini membahas tentang estimasi parameter distribusi Binomial Negatif-*Generalized* Eksponensial dengan metode maksimum *likelihood* dimana solusi dari fungsi *likelihood*-nya diselesaikan dengan optimisasi numerik menggunakan *software* statistika R dan hasil estimasi parameter digunakan dalam pencocokan (*fitting*) data yang diterapkan pada sampel data yang mengalami overdispersi, dalam hal ini data banyaknya penghargaan yang diterima oleh mahasiswa. Melalui pengujian hipotesis *goodness of fit* menggunakan uji *chi-square* dan berdasarkan pada  $p$ -value dengan taraf signifikansi sebesar 0,05 diperoleh bahwa

sampel data overdispersi tersebut menyebar mengikuti distribusi Binomial Negatif-*Generalized* Eksponensial.

**Kata kunci:** data cacah dengan nol berlebih, overdispersi, distribusi Binomial Negatif-*Generalized* Eksponensial, estimasi parameter, optimisasi numerik.

## PENDAHULUAN

Data cacah adalah data hasil percobaan acak yang nilai-nilainya berupa bilangan bulat non-negatif. Dan distribusi yang biasa digunakan untuk memodelkan data cacah adalah distribusi Poisson, dimana pada distribusi Poisson terdapat asumsi yang harus dipenuhi yaitu asumsi kesamaan nilai mean dan varians [8]. Namun pada kenyataannya, pengamatan dalam suatu data cacah biasanya menampilkan keadaan yang dikenal dengan overdispersi, yakni keadaan yang timbul ketika nilai varians dalam data lebih besar dari mean [2]. Dan saat terjadi overdispersi, asumsi kesamaan nilai mean dan varians pada distribusi Poisson dilanggar, sehingga perlu dicari distribusi lain yang dapat digunakan untuk menganalisis data tersebut. Dalam data jumlah yang berkaitan dengan data cacah, *Poisson Mixture Model* seperti Binomial Negatif dapat digunakan sebagai model alternatif [2]. Distribusi Binomial Negatif juga dikenal sebagai distribusi campuran Poisson-Gamma [5]. Tetapi distribusi Binomial Negatif ini lebih baik digunakan untuk data cacah overdispersi yang belum tentu datanya memiliki nilai nol yang berlebih (*excess zeros*). Sedangkan *excess zeros* pada data cacah mengimplikasikan terjadinya overdispersi, tetapi tidak terjadi sebaliknya [8]. Distribusi Poisson cenderung menghasilkan nilai nol yang diberikan pada data menjadi *underestimate*, sedangkan distribusi Binomial Negatif akan menghasilkan nilai nol yang menjadi *overestimate* dalam pengamatan [1]. Akibatnya, distribusi Poisson dan distribusi Binomial Negatif tidak dapat digunakan secara efisien.

Distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) merupakan distribusi peluang gabungan antara distribusi Binomial Negatif (BN) dan distribusi *Generalized* Eksponensial (GE). Dalam hal ini distribusi BN-GE akan digunakan sebagai alternatif untuk distribusi Poisson dan distribusi Binomial Negatif, terutama dalam menganalisis data cacah yang mengalami overdispersi dengan nilai nol yang berlebih (*excess zeros*). Distribusi BN-GE ini diperkenalkan oleh Sirinapa Aryuyuen pada tahun 2013 [1].

Pada tulisan ini akan dibahas estimasi parameter pada distribusi BN-GE menggunakan metode maksimum *likelihood*. Dan dalam menganalisis data yang mengalami overdispersi, hasil estimasi parameter selanjutnya akan digunakan untuk pencocokan (*fitting*) data. Pencocokan (*fitting*) data merupakan cara untuk mengetahui apakah suatu distribusi peluang tertentu sesuai dengan data [2]. Penerapan distribusi BN-GE terhadap data yang mengalami overdispersi sebagai contoh disertakan pada pencocokan (*fitting*) data banyaknya penghargaan yang diterima oleh mahasiswa.

## TINJAUAN PUSTAKA

### Distribusi Binomial Negatif (BN)

Distribusi Binomial Negatif digunakan untuk menghitung jumlah kegagalan ( $x$ ) sebelum terjadi  $r$  buah sukses. Jika  $X$  dinotasikan sebagai variabel acak berdistribusi Binomial Negatif (BN) dengan parameter  $r$  dan  $p$ , maka fungsi massa peluangnya adalah [7]:

$$P(X = x|r, p) = f(x; r, p) = \binom{r+x-1}{x} p^r (1-p)^x, \quad (1)$$

dimana  $x = 0, 1, 2, \dots$  untuk  $r > 0$  dan  $0 < p < 1$ . Jika  $X$  variabel acak berdistribusi Binomial Negatif,  $X \sim \text{BN}(r, p)$ , maka momen faktorial dari  $X$  adalah [1]:

$$\mu_{[k]}(X) = E[X(X-1) \dots (X-k+1)] = \frac{\Gamma(r+k)}{\Gamma(r)} \frac{(1-p)^k}{p^k}, k = 1, 2, \dots$$

Akibatnya, diperoleh:

$$E(X) = \frac{r(1-p)}{p} \text{ dan } (X^2) = \frac{r(1-p)(1+r(1-p))}{p^2}.$$

### Distribusi *Generalized* Eksponensial (GE)

Distribusi *Generalized* Eksponensial (GE) adalah perumuman dari distribusi Eksponensial yang pertama kali diperkenalkan oleh Gupta dan Kundu [3]. Distribusi GE adalah fungsi khusus dari distribusi Gompertz-Verhulst dan distribusi Eksponensial Weibull. Untuk membandingkan tabel kematian dan menghasilkan laju pertumbuhan penduduk, fungsi distribusi tertentu digunakan oleh Gompertz-Verhulst yang didefinisikan sebagai berikut:

$$G(t) = (1 - pe^{-t\beta})^\alpha; t > \frac{1}{\beta} \ln p, p > 0, \alpha > 0, \beta > 0,$$

dari distribusi Gompertz-Verhulst ini, salah satu dari tiga parameternya distandarisasi menjadi satu yaitu  $p = 1$ , sehingga terbentuk distribusi *Generalized* Eksponensial (GE) dengan fungsi distribusi kumulatifnya adalah sebagai berikut:

$$F(x; \alpha, \beta) = (1 - e^{-\beta x})^\alpha$$

dan dari turunan fungsi distribusi kumulatif ini diperoleh fungsi padat peluang distribusi *Generalized* Eksponensial (GE) yang didefinisikan sebagai berikut:

Misalkan  $X$  adalah variabel acak dari distribusi *Generalized* Eksponensial (GE) dengan dua parameter  $(\alpha, \beta)$ , maka fungsi padat peluang dari variabel acak tersebut adalah [3]

$$f(x; \alpha, \beta) = \alpha\beta(1 - e^{-\beta x})^{\alpha-1} e^{-\beta x}; x > 0, \alpha > 0, \beta > 0, \quad (2)$$

dengan  $\alpha =$  parameter bentuk,  $\beta =$  parameter skala.

Fungsi pembangkit momen dari distribusi *Generalized* Eksponensial (GE), yaitu [3]:

$$M_X(t) = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{t}{\beta})}{\Gamma(\alpha-\frac{t}{\beta}+1)}. \quad (3)$$

### Distribusi Campuran

Misalkan  $X$  adalah variabel acak yang bergantung pada parameter  $\lambda$  dengan fungsi peluang bersyarat  $f_{X|\Lambda}(x|\lambda)$ , dimana  $\lambda$  merupakan nilai dari suatu variabel acak  $\Lambda$  dengan

fungsi peluang  $u_{\Lambda}(\lambda)$ , maka distribusi campuran (*mixture distribution*) didefinisikan dengan fungsi peluang sebagai berikut [5]:

$$f_X(x) = \int_{\lambda \in \Lambda} f_{X|\Lambda}(x|\lambda) \cdot u_{\Lambda}(\lambda) d\lambda, \quad (4)$$

dimana distribusi dari  $\Lambda$  disebut sebagai distribusi pencampur atau *mixing distribution*. Distribusi campuran ini akan digunakan untuk menggabungkan distribusi Binomial Negatif (BN) dengan distribusi *Generalized* Eksponensial (GE).

### Estimasi Maksimum *Likelihood*

Misalkan  $x_1, x_2, \dots, x_n$  menyatakan sampel acak berukuran  $n$  dengan fungsi kepadatan peluangnya dinyatakan oleh  $f(x_1, x_2, \dots, x_n|\theta)$  dengan  $\theta \in \square$  adalah nilai parameter yang akan ditaksir dan  $\square$  merupakan himpunan seluruh nilai parameter yang mungkin atau ruang sampel. Maka fungsi *likelihood* dinotasikan dengan [4]

$$L(\theta) = f(x_1|\theta) \dots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta), \quad (5)$$

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

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(x_i|\theta). \quad (6)$$

### Optimisasi Numerik Menggunakan R

*Software* statistika R cocok untuk pemrograman maksimum *likelihood* yang biasa digunakan untuk mencari estimasi parameter. Ada beberapa prosedur untuk mengoptimalkan fungsi *likelihood* dalam R. Namun dalam penelitian ini digunakan perintah `optim()` untuk mendapatkan solusi dari fungsi *likelihood*. Optimisasi melalui `optim()` relatif mudah, karena biasanya tidak diperlukan analisis turunan pertama dan kedua [6].

Estimasi maksimum *likelihood* menggunakan perintah `optim()` memerlukan dua langkah proses, yaitu [6]:

#### 1. Mendeklarasikan fungsi negatif *log-likelihood*

Misal, dinyatakan dalam sintaks berikut:

```
logl <- function(pars, data){
  ll <- -(loglikelihood function)}
```

dengan `logl` adalah nama dari fungsi negatif *log-likelihood*, `pars` adalah nama dari vektor parameter dan `data` adalah nama dari vektor data.

#### 2. Mengoptimalkan fungsi negatif *log-likelihood*

Pernyataan sederhana dari perintah ini adalah seperti dinyatakan dalam sintaks berikut:

```
optim(starting values, logl, data)
```

dengan `starting values` adalah vektor dari nilai awal taksiran parameter, `logl` adalah nama dari fungsi negatif *log-likelihood* yang akan dioptimalkan, dan `data` menyatakan vektor data untuk estimasi.

## HASIL DAN PEMBAHASAN

### Distribusi Binomial Negatif-Generalized Eksponensial (BN-GE)

Distribusi Binomial Negatif-Generalized Eksponensial (BN-GE) merupakan distribusi peluang gabungan antara distribusi Binomial Negatif (BN) dengan distribusi Generalized Eksponensial (GE). Bentuk dari distribusi BN-GE ini pertama-tama diperoleh dengan membentuk *mixed distribution* berdasarkan (2.4), yaitu berupa distribusi peluang gabungan dari distribusi  $BN(r, p)$  dimana  $p = \exp(-\lambda)$ ,  $\lambda > 0$  dan distribusi  $GE(\alpha, \beta)$  dengan menggunakan hukum peluang bersyarat yang di dalamnya terdiri dari tiga parameter, yaitu  $r$ ,  $\alpha$ , dan  $\beta$ .

**Definisi 1:** Misalkan suatu variabel acak  $X$  dikatakan berdistribusi BN-GE( $r, \alpha, \beta$ ),  $X \sim BN-GE(r, \alpha, \beta)$ , apabila distribusi BN memiliki parameter  $r > 0$  dan  $p = \exp(-\lambda)$ , dimana  $\lambda$  berdistribusi GE dengan parameter positif  $\alpha$  dan  $\beta$ ,  $X|\lambda \sim BN(r, p = \exp(-\lambda))$  dan  $\lambda \sim GE(\alpha, \beta)$  [1].

**Teorema 1:** Misalkan  $X \sim BN-GE(r, \alpha, \beta)$ , maka fungsi massa peluang  $X$  adalah [1]

$$f(x; r, \alpha, \beta) = \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \left( \frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)} \right), \quad (7)$$

dimana  $= 0, 1, 2, \dots, r, \alpha$ , dan  $\beta > 0$ .

#### Bukti:

Jika  $X|\lambda \sim BN(r, p = \exp(-\lambda))$  dalam persamaan (1) dan  $\lambda \sim GE(\alpha, \beta)$  dalam persamaan (2), maka fungsi massa peluang dari  $X$  dapat diperoleh dengan *mixing method* menggunakan hukum peluang bersyarat:

$$f(x; r, \alpha, \beta) = \int_0^{\infty} f(x|\lambda) f(\lambda; \alpha, \beta) d\lambda, \quad (8)$$

dimana,  $f(x|\lambda)$  didefinisikan oleh:

$$f(x|\lambda) = \binom{r+x-1}{x} e^{-\lambda r} (1 - e^{-\lambda})^x,$$

dengan menggunakan ekspansi Binomial dari  $(1 - e^{-\lambda})^x$ , maka

$$f(x|\lambda) = \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j e^{-\lambda(r+j)}. \quad (9)$$

Dengan mensubstitusikan persamaan (9) ke persamaan (8), maka diperoleh:

$$\begin{aligned} f(x|\lambda) &= \int_0^{\infty} \left( \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j e^{-\lambda(r+j)} \right) \left( \alpha \beta (1 - e^{-\beta \lambda})^{\alpha-1} e^{-\beta \lambda} \right) d\lambda \\ &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \alpha \beta \int_0^{\infty} (1 - e^{-\beta \lambda})^{\alpha-1} e^{-(\beta+(r+j))\lambda} d\lambda \\ &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j M_{\lambda}(-(r+j)). \end{aligned} \quad (10)$$

Kemudian dengan mensubstitusikan fungsi pembangkit momen distribusi GE pada persamaan (3) ke persamaan (10), maka diperoleh fungsi massa peluang distribusi BN-GE( $r, \alpha, \beta$ ) sebagai berikut:

$$\begin{aligned} f(x; r, \alpha, \beta) &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \left( \frac{\Gamma(\alpha+1)\Gamma\left(1-\frac{-(r+j)}{\beta}\right)}{\Gamma\left(\alpha-\frac{-(r+j)}{\beta}+1\right)} \right) \\ &= \binom{r+x-1}{x} \sum_{j=0}^x \binom{x}{j} (-1)^j \left( \frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right). \quad \blacksquare \end{aligned}$$

### Estimasi Parameter Distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) Menggunakan Metode Maksimum *Likelihood*

Misalkan  $X_1, X_2, \dots, X_n$  adalah variabel acak dari distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) yang memiliki fungsi peluang  $f(x_i; r, \alpha, \beta)$ , dimana

$$f(x_i; r, \alpha, \beta) = \binom{r+x_i-1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right) \quad x_i = 0, 1, 2, \dots, r, \alpha \text{ dan } \beta > 0,$$

dengan  $r, \alpha$ , dan  $\beta$  merupakan parameter yang nilainya tidak diketahui, sehingga nilai ketiga parameter tersebut akan ditaksir dengan menggunakan metode maksimum *likelihood*. Fungsi *likelihood* dari distribusi BN-GE( $r, \alpha, \beta$ ) yaitu:

$$L(r, \alpha, \beta) = \prod_{i=1}^n \left( \binom{r+x_i-1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left( \frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right) \right) \quad (11)$$

Dan untuk mempermudah perhitungan, dicari fungsi *log-likelihood* dari distribusi BN-GE( $r, \alpha, \beta$ ) yaitu:

$$\begin{aligned} l(r, \alpha, \beta) &= \log L(r, \alpha, \beta) \\ &= \sum_{i=1}^n \log \binom{r+x_i-1}{x_i} + \sum_{i=1}^n \log \left( \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right) \\ &= \sum_{i=1}^n \log(\Gamma(r+x_i) - \Gamma(r) - \Gamma(x_i+1)) + \\ &\quad \sum_{i=1}^n \log \left( \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right). \quad (12) \end{aligned}$$

Selanjutnya untuk mencari solusi dari metode maksimum *likelihood* untuk estimasi parameter  $r, \alpha$ , dan  $\beta$  dilakukan dengan cara optimisasi numerik menggunakan fungsi `optim()` pada R seperti yang telah dijelaskan pada teori optimisasi numerik sebelumnya. Dan berikut ini merupakan rangkuman hasil estimasi parameter  $r, \alpha$ , dan  $\beta$  distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) dengan menggunakan nilai-nilai awal yang berbeda.

**Tabel 1.** Hasil estimasi parameter  $r, \alpha$ , dan  $\beta$  distribusi Binomial Negatif-Generalized Eksponensial (BN-GE)

Nilai Awal ( $r, \alpha, \beta$ )	Hasil Estimasi		
	$r$	$\alpha$	$\beta$
(1,1,1)	14.278601	1.017149	23.910882
(1,2,3)	14.173709	1.017205	23.746633
(2,2,2)	14.149138	1.018569	23.724234
(2,3,4)	14.129862	1.017386	23.686080
(5,1,10)	14.182637	1.017341	23.764893
(10,1,20)	14.110819	1.019015	23.668853
(15,0.5,24)	14.205817	1.018061	23.812423
(20,0.7,30)	14.140773	1.018274	23.712063

Dari Tabel 1 dapat dilihat bahwa dengan nilai awal taksiran parameter yang berbeda-beda, diperoleh nilai estimasi parameter  $r, \alpha$ , dan  $\beta$  masing-masing berada pada kisaran nilai 14, 1 dan 23.

Setelah diperoleh nilai estimasi parameter, dapat dilakukan pencocokan (*fitting*) data. Berikut ini merupakan pencocokan (*fitting*) data dari distribusi Poisson, Binomial Negatif (BN) dan Binomial Negatif-Generalized Eksponensial (BN-GE).

a. Pencocokan (*Fitting*) Data dengan Distribusi Poisson

Sebelum dilakukan pencocokan (*fitting*) data, terlebih dahulu dicari nilai estimasi parameter  $\lambda$  dari distribusi Poisson menggunakan bantuan *software* R, sehingga diperoleh nilai parameter  $\lambda = 0,63$ . Selanjutnya nilai parameter tersebut disubstitusikan ke persamaan

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

untuk mengetahui peluang dari distribusi Poisson. Hasil pencocokan (*fitting*) data distribusi Poisson dapat dilihat pada Tabel 2 yang menghasilkan pencocokan (*fitting*) data yang jauh berbeda dengan frekuensi sebenarnya pada data observasi. Hal ini menunjukkan bahwa distribusi Poisson mempunyai ketidaksesuaian terhadap data yang mengalami overdispersi dengan nilai *chi-square* hitung ( $\chi^2_{hitung}$ ) yang dihasilkan sebesar 25,559581333 dan *p-value* sebesar 0,000002817.

b. Pencocokan (*Fitting*) Data dengan Distribusi Binomial Negatif (BN)

Sama halnya dengan distribusi Poisson, untuk mencari nilai estimasi parameter distribusi Binomial Negatif (BN) dapat menggunakan bantuan *software* R, sehingga diperoleh nilai  $r = 0,85634685$  dan  $p = 0,5761417$ . Nilai dari estimasi parameter tersebut kemudian disubstitusikan ke persamaan (2.1) untuk mengetahui peluang dari distribusi Binomial Negatif (BN). Selanjutnya nilai peluang yang telah diperoleh digunakan untuk mengetahui hasil pencocokan (*fitting*) data. Hasil pencocokan (*fitting*) data pada Tabel 2

memperlihatkan bahwa distribusi Binomial Negatif (BN) menghasilkan pencocokan (*fitting*) data yang tidak jauh berbeda dengan frekuensi sebenarnya. Sehingga dapat dikatakan bahwa distribusi Binomial Negatif (BN) memiliki kesesuaian yang cukup baik terhadap data observasi. Hal ini dapat dilihat melalui perbedaan antara frekuensi pencocokan (*fitting*) dan frekuensi sebenarnya yang menghasilkan nilai *chi-square* hitung ( $\chi^2_{hitung}$ ) dan *p-value* masing-masing sebesar 3,340907925 dan 0,3419895.

c. Pencocokan (*Fitting*) Data dengan Distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE)

Hasil estimasi parameter distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) yang telah diperoleh pada Tabel 1 akan digunakan untuk mengetahui hasil pencocokan (*fitting*) data. Di sini penulis memilih nilai estimasi parameter dengan nilai awal (1,1,1) yaitu  $r = 14,278601$ ,  $\alpha = 1,017149$  dan  $\beta = 23,910882$ . Selanjutnya nilai estimasi dari masing-masing parameter disubstitusikan ke persamaan (3.1) untuk memperoleh nilai peluang dari distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE). Hasil pencocokan (*fitting*) data pada Tabel 2 menunjukkan bahwa distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) memiliki kesesuaian yang cukup baik terhadap data observasi. Hal ini dapat dilihat melalui perbedaan yang dihasilkan antara frekuensi pencocokan (*fitting*) dan frekuensi sebenarnya yang tidak jauh berbeda, dengan nilai *chi-square* hitung ( $\chi^2_{hitung}$ ) dan *p-value* yang dihasilkan masing-masing sebesar 3,187473896 dan 0,3636137.

Hasil pencocokan (*fitting*) data tertera pada tabel berikut:

**Tabel 2.** Banyaknya penghargaan yang diterima oleh mahasiswa dari hasil observasi [9] dan pencocokan (*fitting*) data distribusi

$x$	Data Observasi	Poisson	BN	BN-GE
0	124	106,518360201	124,726870133	124,525447411
1	49	67,106566927	45,272077135	45,944788121
2	13	21,138568582	17,810669408	17,547808111
3	9	4,439099402	7,187711268	6,950808011
4	2	0,699158156	2,937158696	2,850192408
5	2	0,088093928	1,209171207	1,207113639
6	1	0,009249862	0,500246466	0,526853975
<b>Total</b>	200	200	200	200
<b>Estimasi Parameter</b>		$\hat{e} = 0,63$	$r = 0,85634685$ $p = 0,5761417$	$r = 14,278601$ $\hat{a} = 1,017149$ $\hat{a} = 23,910882$
<b>Chi-square</b>		25,559581333	3,340907925	3,187473896
<b>p-value</b>		0,000002817	0,3419895	0,3636137

Dari tabel 2 dapat disimpulkan bahwa distribusi Binomial Negatif-*Generalized* Eksponensial (BN-GE) memberikan hasil pencocokan (*fitting*) data yang paling baik terhadap data



observasi karena memiliki nilai *chi-square* hitung ( $\chi^2_{\text{hitung}}$ ) yang terkecil yaitu 3,187473896 dan memiliki *p-value* yang terbesar yaitu 0,3636137.

## KESIMPULAN

Distribusi Binomial Negatif-Generalized Eksponensial (BN-GE) merupakan distribusi peluang gabungan yang diperoleh dengan cara mixing method yaitu metode penggabungan distribusi peluang antara distribusi Binomial Negatif (BN) dengan distribusi Generalized Eksponensial (GE). Distribusi Binomial Negatif-Generalized Eksponensial (BN-GE) memuat tiga parameter yaitu  $r$ ,  $\hat{a}$ , dan  $\hat{\lambda}$ . Estimasi parameter distribusi Binomial Negatif-Generalized Eksponensial dilakukan dengan metode maksimum likelihood dimana solusi dari fungsi likelihood-nya diselesaikan dengan optimisasi numerik menggunakan software R.

Hasil estimasi parameter dari distribusi Binomial Negatif-Generalized Eksponensial menghasilkan pencocokan (fitting) data yang cukup baik. Melalui pengujian hipotesis menggunakan uji chi-square dan berdasarkan pada *p-value* dengan taraf signifikansi sebesar 0,05 dapat diketahui bahwa data banyaknya penghargaan yang diterima oleh mahasiswa menyebar mengikuti distribusi Binomial Negatif-Generalized Eksponensial.

## REFERENSI

- [1] Aryuyuen, S. dan W. Bodhisuwan, The Negative Binomial-Generalized Exponential (NB-GE) Distribution. *Applied Mathematical Sciences*. **7(22)**, 1093-1105, 2013.
- [2] Fakhri, B. S, *Distribusi Gabungan Beta-Binomial Negatif*, Skripsi, Jurusan Matematika, FMIPA, Universitas Brawijaya, Malang: 2013.
- [3] Gupta, R. D. dan D. Kundu, Generalized Exponential Distributions. *Austral. & New Zealand J. Statist.* **41(2)**, 173-188, 1999.
- [4] Hogg, R. V, *Introduction to Mathematical Statistics, 6th Edition*, Prentice-Hall International, New Jersey: 2005.
- [5] Shafira, *Penaksiran Parameter Distribusi Binomial Negatif pada Kasus Overdispersi*, Skripsi, FMIPA, Universitas Indonesia, Depok: 2011.
- [6] Steenbergen, M. R, *Maximum Likelihood Programming in R*, Department of Political Science, University of North Carolina, Chapel Hill: 2006.
- [7] Walck, C, *Hand-book on Statistical Distributions for Experimentalists*, University of Stockholm, Stockholm: 2007.
- [8] Zhaoliang Wang, One Mixed Negative Binomial Distribution with Application, *Statistical Planning and Inference*, **141**, 1153-1160. 2011.
- [9] R Data Analysis Examples: Poisson Regression, [http://www.ats.ucla.edu/stat/data/poisson\\_sim.csv](http://www.ats.ucla.edu/stat/data/poisson_sim.csv) [diakses 15 Desember 2015].