 http://dx.doi.org/10.29313/.v0i0.28783

Aplikasi Distribusi Binomial Negatif sebagai Distribusi Campuran pada Data Asuransi Kendaraan Bermotor di Indonesia

Andrianto Wibowo[*], Aceng Komarudin Mutaqin

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

[*]ndrntwbw@gmail.com, aceng.k.mutaqin@gmail.com

Abstract. Data count is data from a random experiment that describes an event that occurs in a certain period of time. A data count whose values can only be positive integers. The distribution that is commonly used to model the count data is the Poisson distribution. The Poisson distribution has the assumption that it must have the same mean and variance (equidispersion). When overdispersion occurs, the assumption of equality of mean and variance in the Poisson distribution is violated, so that another distribution is needed to overcome the count data when overdispersion occurs. One of the distributions that can be used to overcome the overdispersion problem is the negative binomial distribution which is a mixed distribution of the Poisson distribution and the gamma distribution with parameters $\alpha_1 > 0$ and $\beta_1 > 0$. Count data modeling is widely used in various sciences, one of which is actuarial science, especially in the insurance sector to model motor vehicle claim frequency data. The chi-square distribution fit test was carried out to test the suitability of the claim frequency data. The data used in this study is data on the frequency of claims from 24,874 policy holders of comprehensive motor vehicle insurance products PT. X in 2013. The results show that the negative binomial distribution as a mixed distribution of the Poisson distribution and the gamma distribution with the estimated parameters are $\widehat{\alpha_1} = 1{,}6095$ and $\widehat{\beta_1} = 4{,}3996$ is suitable for modeling data on the frequency of claims for vehicle insurance products. motor comprehensive PT. X year 2013.

Keywords: negative binomial distribution, Poisson distribution, gamma distribution, chi-kuadrat.

Abstrak. Data cacahan adalah data hasil percobaan acak yang menggambarkan suatu kejadian yang terjadi pada suatu kurun waktu tertentu. Suatu data cacahan nilai-nilainya hanya dapat berupa bilangan bulat positif. Distribusi yang biasa digunakan untuk memodelkan data cacahan adalah distribusi Poisson. Distribusi Poisson memiliki asumsi yang harus dipenuhi yaitu memiliki nilai *mean* dan *varians* yang sama (*equidispersion*). Saat terjadi overdispersi, asumsi kesamaan *mean* dan *varians* pada distribusi Poisson dilanggar, sehingga diperlukan distribusi lain untuk mengatasi data cacahan saat terjadi overdispersi. Salah satu distribusi yang dapat digunakan untuk mengatasi masalah overdispersi adalah distribusi binomial negatif yang merupakan distribusi campuran dari distribusi Poisson dan distribusi gamma dengan parameter $\alpha_1 > 0$ dan $\beta_1 > 0$. Pemodelan data cacahan banyak digunakan dalam berbagai ilmu salah satunya yaitu ilmu aktuaria khususnya bidang asuransi untuk memodelkan data frekuensi klaim kendaraan bermotor. Pengujian kecocokan distribusi chi-kuadrat dilakukan untuk menguji kecocokan pada data frekuensi klaim. Data yang digunakan dalam penelitian ini adalah data frekuensi klaim dari 24.874 pemegang polis produk asuransi kendaraan bermotor *comprehensive* PT. X tahun 2013. Hasil menunjukan bahwa distribusi binomial negatif sebagai distribusi campuran dari distribusi Poisson dan distribusi Gamma dengan taksiran parameternya adalah $\widehat{\alpha_1} = 1{,}6095$ dan $\widehat{\beta_1} = 4{,}3996$ cocok digunakan untuk memodelkan data frekuensi klaim produk asuransi kendaraan bermotor *comprehensive* PT. X tahun 2013.

Kata Kunci: distribusi binomial negatif, distribusi Poisson, distribusi gamma, chi-kuadrat.

1. Pendahuluan

Data cacahan atau *count data* adalah data hasil percobaan acak yang menggambarkan suatu kejadian yang terjadi pada suatu kurun waktu tertentu. Suatu data cacahan nilai-nilainya hanya dapat berupa bilangan bulat positif $\{0, 1, 2, 3, \cdots\}$, karena suatu kejadian tidak mungkin terjadi dalam sejumlah bilangan negatif. Misalkan X merupakan suatu peubah acak yang nilainya berupa data cacahan. Distribusi yang biasa digunakan adalah distribusi Poisson. Distribusi Poisson memiliki asumsi yang harus dipenuhi yaitu memiliki nilai *mean* dan *varians* yang sama (*equidispersion*) (Wang, 2011).

Pada kenyataannya sering terjadi pelanggaran asumsi tersebut dimana varians lebih kecil dari mean (underdispersion) atau varians lebih besar dari mean (overdispersion). Pada kebanyakan data cacahan terkadang ditemukan kasus overdispersi (Consul & Famoye, 1992). Saat terjadi overdispersi, asumsi kesamaan *mean* dan *varians* pada distribusi Poisson dilanggar, sehingga diperlukan distribusi lain untuk mengatasi data cacahan saat terjadi overdispersi. Salah satu distribusi yang dapat digunakan untuk mengatasi masalah overdispersi adalah distribusi binomial negatif yang merupakan distribusi campuran dari distribusi Poisson dan distribusi gamma (*Mixture Poisson-Gamma Distribution*). Distribui campuran atau biasa disebut dengan *mixture distribution* adalah distribusi dari suatu peubah acak yang terbentuk dengan menggabungkan beberapa distribusi sehingga menghasilkan distribusi baru. Prinsip dari pencampuran distribusi adalah reparameterisasi, misalkan X merupakan suatu peubah acak dengan parameter θ, maka parameter θ merupakan suatu peubah acak yang mengikuti distribusi tertentu. Beberapa penelitian telah membahas tentang distribusi binomial negatif sebagai distribusi campuran diantaranya yaitu Panger dan Willmot (Panjer & Willmot, 1981).

Pemodelan data cacahan banyak digunakan dalam berbagai ilmu salah satunya yaitu ilmu aktuaria khususnya bidang asuransi yaitu untuk memodelkan data frekuensi klaim kendaraan bermotor. Frekuensi klaim merupakan jumlah klaim dalam satu blok polis asuransi selama periode waktu tertentu atau banyaknya klaim yang dilakukan oleh pemegang polis selama masa asuransinya biasanya dalam satu tahun.

Dalam penelitian ini distribusi binomial negatif sebagai distribusi campuran dari distribusi Poisson dan distribusi gamma akan diterapkan untuk memodelkan data frekuensi klaim asuransi kendaraan bermotor di Indonesia.

2. Metodologi

Distribusi Poisson

Suatu peubah acak X dikatakan berdistribusi Poisson dengan parameter θ apabila memiliki fungsi massa peluang sebagai berikut:

$$f(x|\theta) = \frac{e^{-\theta}\theta^{x}}{x!}, \text{ untuk } x = 0, 1, 2, \ldots \tag{1}$$

Dengan nilai ekspektasi dan variansnya masing masing adalah:

$$E(X|\theta) = \theta \tag{2}$$

$$V(X|\theta) = \theta \tag{3}$$

Distribusi Poisson mempunyai suatu karakteristik yang khusus yaitu nilai rata-rata dan variansi nya sama. Dalam penerapannya distribusi Poisson sering digunakan untuk mendeskripsikan peluang kejadian yang jarang terjadi dalam selang waktu tertentu, termasuk kejadian acak dalam klaim asuransi kendaraan bermotor (Denuit, Marechal, Pitrebois, &

Walhin, 2009).

Distribusi Gamma

Suatu peubah acak Θ dikatakan berdistribusi gamma dengan parameter bentuk $\alpha_1 > 0$ dan parameter skala $\beta_1 > 0$, jika mempunyai fungsi densitas peluang sebagai berikut:

$$\pi_1(\theta) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\,\theta^{\alpha_1 - 1}\,e^{-\beta_1\theta}, \text{ untuk } \theta > 0 \tag{4}$$

Nilai ekspektasi dan varians dari distribusi gamma masing-masing adalah:

$$E(\Theta) = \frac{\alpha_1}{\beta_1} \tag{5}$$

$$V(\Theta) = \frac{\alpha_1}{\beta_1^2} \tag{6}$$

Pada penerapannya, distribusi gamma sering digunakan sebagai distribusi untuk memodelkan waktu tunggu (*waiting time*), pendapatan per tahun, dan besarnya klaim pada perusahaan asuransi (Shafira, 2011).

Distribusi Binomial Negatif Sebagai Distribusi Campuran

Distribusi binomial negatif merupakan distribusi yang memiliki banyak cara dalam penurunannya. Bosswell dan Patil (1970) menunjukan bahwa ada dua belas cara untuk mendapatkan distribusi binomial negatif. Salah satunya dapat diturunkan sebagai distribusi campuran dari distribusi Poisson dan distribusi gamma. Dengan distribusi Poisson yang memiliki fungsi massa peluang pada Persamaan (1), dimana parameter θ berdistribusi gamma dengan parameter $\alpha_1 > 0$ dan $\beta_1 > 0$, dengan fungsi densitas peluang pada Persamaan (4).

Selanjutnya distribusi marjinal dari X dapat dihitung dengan menggunakan persamaan $P(X = x) = \int_0^\infty P(X = x|\theta)U(\theta)d\theta$ yang membentuk fungsi massa peluang untuk distribusi binomial negatif dengan parameter α_1 dan β_1 sebagai berikut:

$$p_x = \binom{x + \alpha_1 - 1}{x}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{x} \tag{7}$$

Jika $x = 0$ maka:

$$p_0 = \binom{0 + \alpha_1 - 1}{0}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{0}$$

$$= \binom{\alpha_1 - 1}{0}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}$$

$$= \left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1} \tag{8}$$

Sedangkan jika $x = x + 1$, maka:

$$p_{x+1} = \binom{x + 1 + \alpha_1 - 1}{x + 1}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{x+1}$$

$$\begin{aligned}
p_{x+1} &= \binom{x + \alpha_1}{x + 1}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{x}\left(\frac{1}{1 + \beta_1}\right) \\
&= \frac{x + \alpha_1}{x + 1}\,\frac{x + \alpha_1 - 1}{x!\,(\alpha_1 - 1)!}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{x}\left(\frac{1}{1 + \beta_1}\right) \\
&= \frac{x + \alpha_1}{x + 1}\left(\frac{1}{1 + \beta_1}\right)\binom{x + \alpha_1 - 1}{x}\left(\frac{\beta_1}{1 + \beta_1}\right)^{\alpha_1}\left(\frac{1}{1 + \beta_1}\right)^{x} \\
&= \frac{x + \alpha_1}{(x + 1)(1 + \beta_1)}p_x
\end{aligned} \tag{9}$$

Ekspektasi dan variansnya masing-masing adalah:

$$E(X) = \frac{\alpha_1}{\beta_1}$$

$$V(X) = \frac{\alpha_1}{\beta_1}\left(1 + \frac{1}{\beta_1}\right)$$

Misalkan $X_1, X_2, \dots, X_t$ adalah suatu sampel acak berukuran t dari distribusi binomial negatif dengan parameter α_1 dan β_1, dengan nilai sampel acak tersebut adalah Misalkan $x_1, x_2, \dots, x_t$. Penaksir kemungkinan maksimum untuk α_1 dan β_1 dari distribusi binomial negatif tidak dapat diselesaikan secara ekspilit, namun hanya dapat ditaksir dengan menggunakan numerik. Salah satu metode numerik tersebut adalah metode Newton-Raphson. Taksiran parameter α_1 dan β_1 adalah solusi dari 2 persamaan berikut:

$$t \ \ln\left(\frac{1+\beta_1}{\beta_1}\right) = \sum_{i=1}^{t}\sum_{j=0}^{x_i}\frac{1}{\alpha_1 + j} \tag{10}$$

$$\beta_1 = \frac{\alpha_1}{\bar{x}} \tag{11}$$

Perangkat lunak Matlab menyediakan fungsi untuk menghitung taksiran kemungkinan maksimum dari parameter distribusi binomial negatif.

Uji Kecocokan Distribusi Chi-Kuadrat

Dalam penelitian ini uji kecocokan chi-kuadrat akan digunakan untuk mengetahui apakah $x_1, x_2, \dots, x_n$ merupakan realisasi dari sampel acak $X_1, X_2, \dots, X_n$ yang berasal dari distribusi $F(.)$. Pada uji kecocokan digunakan hipotesis sebagai berikut:

$H_0 = x_1, x_2, \dots, x_n$ merupakan realisasi dari sampel acak yang berdistribusi $F(.)$

$H_1 = x_1, x_2, \dots, x_n$ merupakan realisasi dari sampel acak yang berdistribusi bukan $F(.)$

Statistik uji untuk menguji kecocokan chi-kuadrat, yaitu:

$$\chi^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i} \tag{12}$$

dimana O_i adalah banyaknya observasi pada kategori i, E_i adalah nilai harapan pada kategori i. Untuk nilai E_i dapat dihitung dengan persamaan berikut:

$$E_i = np_x \ \ , x = 0, 1, 2, \dots \tag{13}$$

Nilai kritis dihitung dari distribusi chi-kuadrat dengan derajat bebas $k - p - 1$, dimana k banyaknya kategori dan p banyaknya parameter dalam distribusi. Kriteria pengujiannya adalah tolak hipotesis nol jika statistik uji chi-kuadrat lebih besar dari nilai kuantil dari distribusi chi-kuadrat pada taraf nyata α dan derajat bebas $k - p - 1$ atau $\chi^2 \geq {\chi^2}_{(k-p-1)(1-\alpha)}$. Jika ada kategori yang nilai harapannya (E_i) kurang dari 5, maka kategori-kategori tersebut digabungkan untuk menghasilkan kategori baru yang mempunyai nilai harapan (E_i) yang lebih besar sama dengan 5, hal ini bertujuan untuk menghindari penerimaan hipotesis nol yang bias (Walpole, Myers, & Myers, 2012).

3. Pembahasan dan Diskusi

Data yang digunakan untuk mengaplikasikan metode yang akan dibahas dalam penelitian ini adalah data sekunder yang diperoleh dari PT. X tahun 2013. Data tersebut berisi data frekuensi klaim dan data besar klaim dari setiap klaim yang diajukan oleh 24.874 pemegang polis untuk produk asuransi kendaraan bermotor c*omprehensive* terhadap perusahaan asuransi PT. X tahun 2013.

Uji Kecocokan Distribusi Binomial Negatif Sebagai Distribusi Campuran

Hipotesis untuk pengujian tersebut adalah:

H_0 : Data frekuensi klaim asuransi kendaraan bermotor c*omprehensive* perusahaan PT. X tahun 2013 berasal dari populasi yang berdistribusi binomial negatif sebagai distribusi campuran.

H_1 : Data frekuensi klaim asuransi kendaraan bermotor bermotor c*omprehensive* perusahaan PT. X tahun 2013 bukan berasal dari populasi yang berdistribusi binomial negatif sebagai distribusi campuran.

Langkah selanjutnya adalah menghitung taksiran parameter distribusi binomial negatif sebagai distribusi campuran dengan bantuan perangkat lunak Matlab R2017b hasil taksiran parameter distribusi binomial negatif sebagai distribusi campurannya adalah $\widehat{\alpha_1} = 1{,}6095$ dan $\widehat{\beta_1} = 4{,}3996$. Berdasarkan nilai taksiran parameter distribusi binomial negatif sebagai distribusi campuran yang telah diperoleh, dapat dihitung nilai taksiran peluang untuk setiap frekuensi klaim menggunakan Persamaan (7).

Hasil selengkapnya nilai taksiran peluang untuk setiap frekuensi klaim disajikan dalam Tabel 1 kolom (3), kolom (1) berisikan frekuensi klaim, kolom (2) berisikan banyaknya pemegang polis yang mengajukan klaim. Berdasarkan nilai taksiran peluang frekuensi klaim tersebut, dapat dihitung nilai harapan untuk setiap frekuensi klaim, np_x, untuk setiap x, yaitu dengan mengalikan p_x dengan ukuran sampel.

Tabel 1. Taksiran Nilai Peluang dan Nilai Harapan Terjadinya Klaim

Frekuensi Klaim (x)	Banyaknya Tertanggung O_i	Peluang Terjadinya Klaim p_x	Nilai Harapan Terjadinya Klaim E_i
0	17908	0,7192	17888,8815
1	5254	0,2144	5332,2755
2	1372	0,0518	1288,4818
3	276	0,0115	287,1062
4	47	0,0025	61,2738
5	14	0,0005	12,7311
≥6	3	0,0001	3,2501
Jumlah	24874	1	24874

Terlihat dalam tabel diatas ada nilai harapan terjadinya klaim yang kurang dari 5 yaitu untuk frekuensi klaim ≥6. Oleh karena itu frekuensi klaim ≥6 akan digabungkan dengan frekuensi klaim 5 untuk mendapatkan nilai harapan terjadinya klaim yang lebih besar dari 5.

Tabel 2. Nilai-nilai yang Dibutuhkan untuk Perhitungan Statistik Uji

Frekuensi Klaim (x)	Banyaknya Tertanggung O_i	Peluang Terjadinya Klaim p_x	Nilai Harapan Terjadinya Klaim E_i	$\frac{(\boldsymbol{O_i}-\boldsymbol{E_i})^2}{\boldsymbol{E_i}}$
0	17908	0,7192	17888,8815	0,0204
1	5254	0,2144	5332,2755	1,1491
2	1372	0,0518	1288,4818	5,4136
3	276	0,0115	287,1062	0,4296
4	47	0,0025	61,2738	3,3251
≥5	17	0,0006	15,9812	0,0649
Jumlah	24874	1	24874	10,4027

Nilai statistik uji Chi-Kuadrat ada dalam Tabel 2 kolom (5) baris terakhir, yaitu 10,4027. Dengan taraf nyata 1% nilai kuantil distribusi Chi-Kuadrat dengan derajat bebas $(6-2-1)$ adalah 11,3449. Terlihat bahwa nilai statistik ujinya lebih kecil dibandingkan dengan nilai kuantilnya $(10{,}4027 < 11{,}3449)$. Dengan demikian hipotesis nol diterima dan dapat disimpulkan bahwa data frekuensi klaim pemegang polis untuk produk asuransi kendaraan bermotor c*omprehensive* perusahaan PT. X tahun 2013 berasal dari populasi berdistribusi binomial negatif sebagai distribusi campuran.

4. Kesimpulan

Berdasarkan hasil penelitian ini dapat disimpulkan bahwa distribusi binomial negatif sebagai distribusi campuran dari distribusi Poisson dan distribusi gamma dengan taksiran parameternya adalah $\widehat{\alpha_1} = 1{,}6095$ dan $\widehat{\beta_1} = 4{,}3996$ cocok digunakan untuk memodelkan data frekuensi klaim produk asuransi kendaraan bermotor *comprehensive* PT. X tahun 2013.

Acknowledge

With a very deep sense of gratitude and my gratitude to Allah SWT for His grace and guidance alhamdulillah, I was able to complete this research on time and at the right time. I dedicate this thesis to two great people in my life, namely my parents. Don't forget to also say thank you and my love to my brothers, nephews, family, friends, girlfriends and all good people involved who always provide help, prayers and support to the author. I will forever be so grateful to have you all in my life.

Daftar Pustaka

[1] Consul, P. C., & Famoye, F. (1992). Generalized Poisson Regression Model. *Communications in Statistics - Theory and Methods*, 21(1), 89-109.

[2] Denuit, M., Marechal, X., Pitrebois, S., & Walhin, J.-F. (2009). *Actuarial Modelling of Claim Counts Risk Classification, Credibility and Bonus-Malus Systems.* John Willey & Sons.

[3] Panjer, H. H., & Willmot, G. E. (1981). Finite Sum Evaluation of the Negative Binomial-Exponential Model. *ASTIN Bulletin, Cambridge University Press*, 12(2), 133-137.

[4] Shafira. (2011). *Penaksiran Parameter Distribusi Binomial Negatif Pada Kasus Overdispersi.* Depok: Skripsi, Universitas Indonesia.

[5] Subagiyo, D. T., & Salviana, F. M. (2016). *Hukum Asuransi.* Surabaya: PT REVKA PETRA MEDIA.

[6] Walpole, R. E., Myers, R. H., & Myers, S. L. (2012). *Probability and Statistics for Engineers and Scientists Ninth Edition.* United States of America: Pearson Education, Inc.

[7] Wang, Z. (2011). One Mixed Negative Binomial Distribution with Application. *Journal of Statistical Planning and Inference*, 141(3), 1153-1160.

[8] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. Jurnal Riset Statistika, 1(1), 35-42.