METODE EXCESS-OF-LOSS MENGGUNAKAN BAYESIAN UNTUK MEMODELKAN TINGKAT KEPARAHAN KLAIM

Felivia Kusnadi[1], Benny Yong[2], Ferry Jaya Permana[3]
[1,2,3] Universitas Katolik Parahyangan, Jl. Ciumbuleuit 94, Bandung
[1]felivia@unpar.ac.id

Abstrak

Tingkat keparahan klaim merupakan variabel acak, terlebih pada data asuransi umum. Masalah utama dalam memodelkan data ini ialah kesenjangan yang besar antara nominal klaim. Metode penelitian yang digunakan bersifat kuantitatif dengan mencocokkan beberapa distribusi ekor panjang terhadap data simulasi besar keparahan klaim. Untuk menentukan distribusi yang cocok untuk digunakan, penulis melakukan pengujian menggunakan *Quantile-Quantile plots* beserta Akaike *Information Criterion* (AIC) untuk setiap model. Hasil dari penelitian ini menunjukkan bahwa gabungan distribusi *Lognormal*, *Inverse* Gaussian, dan Gamma merupakan model yang baik bagi data besar keparahan klaim. Dengan menggunakan model ini, penulis menghitung peluang *posterior* untuk setiap model menggunakan Teorema Bayes serta besar premi murni untuk *excess layer*.

Kata Kunci: Reasuransi *excess-of-loss*; Bayes; *posterior; likelihood*

Abstract

Claim severity experience was random, especially to general insurance dataset. The main problem in modelling these data was the huge discrepancies between the amounts. This research quantitatively fitted long-tailed distributions to the simulated claim severity data. To determine which distributions were best fit on data, we graphically used Quantile-Quantile plots and calculated the Akaike Information Criterion (AIC) for each model. Based on the result, we concluded that the combined distributions of Lognormal, Inverse Gaussian, and Gamma were the best model for fitting the claim severity data. By utilizing this model, we then calculated the posterior probabilities for each model using Bayesian Theory and determined the pure premium for excess layer.

Keywords: Excess-of-loss Reinsurance; Bayes; posterior; likelihood

Pendahuluan

Menurut Badan Pusat Statistik, jumlah unit kendaraan bermotor di Indonesia mencapai 146.858.759 pada tahun 2018. Sekitar 93% di antaranya merupakan kendaraan pribadi, yakni mobil penumpang dan sepeda motor. Menurut Dataku, angka kecelakaan lalu lintas di Indonesia mencapai 5.061 kejadian dengan kerugian materi sekitar Rp 407 miliar. Penetrasi asuransi di Indonesia bertumbuh sekitar 2,6% setiap tahunnya, yakni dari tahun 2014 hingga 2018, menurut laporan Asosiasi Asuransi Umum Indonesia tahun 2019. Pada laporan yang sama, pertambahan premi asuransi kendaraan bermotor selama tahun 2018 adalah sebesar 7,8%, sedangkan pertambahan premi asuransi secara keseluruhan adalah sebesar 9,8%.

Seiring bertambahnya jumlah pemegang polis di Indonesia, perusahaan asuransi umum menghadapi ketidakpastian dalam menghadapi jumlah klaim yang sangat banyak atau satu klaim yang tingkat keparahannya sangat besar karena diakibatkan oleh bencana alam pada suatu daerah. Perusahaan dapat mengalami kebangkrutan apabila tidak bisa mengantisipasi kemungkinan tersebut. Salah satu cara mengalihkan risiko ialah melalui reasuransi yang merupakan mekanisme pengalihan risiko dari perusahaan asuransi umum kepada reasuransi untuk perlindungan potensi kerugian besar yang tidak terduga.

Terdapat dua masalah inti dalam dunia reasuransi. Pertama, data tingkat keparahan klaim yang ada tidak cukup untuk menggambarkan harga dari *excess layer*, yakni perlindungan ekstra di atas nilai yang tertanggung dalam perjanjian. Kedua, data yang dimiliki tidak berkembang karena sebagian besar klaim dapat memakan waktu setidaknya beberapa tahun untuk diselesaikan. Hal tersebut diakibatkan bencana alam atau masalah hukum (Meyers, 2005). Tujuan penelitian ini adalah menyelesaikan dua permasalahan tersebut. (Tzougas et al., 2014) memodelkan tingkat keparahan klaim menggunakan distribusi ekor panjang untuk perhitungan peluang *posterior*. (Barnett, 2020) memakai Teorema Bayes untuk menghitung premi murni dari reasuransi *excess-of-loss* karena pembentukan peluang *prior* yang menyesuaikan dengan struktur data dan distribusi peluang *posterior* dikonstruksikan dengan menggunakan teorema tersebut untuk menghitung harga premi murni.

Artikel ini akan membahas tentang perhitungan premi tipe perjanjian reasuransi *excess-of-loss*. Besarnya premi harus dapat bersaing dengan kompetitor lainnya, namun harus juga memperhitungkan solvabilitas dari premi yang diperoleh dengan probabilitas terjadinya klaim di masa mendatang. Pada reasuransi *excess-of-loss*, perusahaan asuransi menetapkan nilai risiko maksimal yang ditanggung sendiri. Apabila terjadi kerugian di atas nilai tersebut, maka perusahaan mengalihkan kelebihan risiko pada reasuransi sampai dengan batas tertentu, sesuai dengan perjanjian. Reasuransi *excess of loss* biasanya diatur dalam beberapa lapisan (*layers*) demi proteksi reasuransi yang lebih besar dan sekaligus memperkecil preminya.

Metode Penelitian

Pada bagian ini, akan dijelaskan langkah-langkah yang dilakukan dalam penelitian ini.

1. Mencocokkan beberapa distribusi ekor panjang terhadap data simulasi besar keparahan klaim.
2. Mencari tiga distribusi terbaik dengan nilai AIC terkecil.
3. Menghitung nilai *log-likelihood* dari ketiga distribusi tersebut terhadap data simulasi besar keparahan klaim.
4. Mencari peluang *posterior* berdasarkan nilai *log-likelihood* yang telah diurutkan.
5. Menghitung rata-rata besar premi berdasarkan peluang *posterior* yang telah dicari sebelumnya.

Berikut ini merupakan konsep-konsep dasar yang perlu diketahui untuk memahami langkah-langkah pengerjaan dalam penelitian ini.

Teorema dan Model Bayes

Untuk mengestimasi harga dari *excess layer* akan digunakan Teorema Bayes untuk menghitung probabilitas *posterior* untuk serangkaian distribusi keparahan klaim yang akan dipilih. (Pitt & Guillén, 2012) melihat estimasi kepadatan pada bagian ekor distribusi berdasarkan Metode Bayesian dalam data klaim yang univariat, yakni pada data asuransi yang tera-filiasi dengan kendaraan bermotor.

Teorema Bayes digunakan untuk menghitung suatu peluang bersyarat (*conditional probability*) antar dua kejadian, yakni (Hoff, 2009):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ ; } P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Berlaku hubungan:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(A|B)$ proporsional terhadap $P(B|A) \cdot P(A)$. Hubungan ini dinotasikan sebagai berikut.

$$P(A|B) \propto P(B|A) \cdot P(A)$$

Terdapat 3 sifat dasar Bayesian, yaitu distribusi *prior*, distribusi *posterior*, dan fungsi *likelihood*. Distribusi *prior* adalah informasi parameter yang diperoleh dari suatu asumsi. Selain itu, fungsi *likelihood* merupakan informasi parameter dari data. Distribusi *posterior* adalah informasi parameter yang bergantung pada data yang telah diamati dan juga *prior* yang telah diasumsikan sebelumnya. A_i diasumsikan sebagai suatu kejadian yang telah terjadi, sehingga

$P(A_i)$ ialah peluang *prior*. *Likelihood* dari kejadian B ialah peluang bersyarat $P(B|A_i)$ dan peluang *posterior* dinotasikan $P(A_i|B)$. Hubungan antara peluang *prior*, fungsi *likelihood*, dan peluang *posterior* pada Bayesian dapat dinyatakan dalam persamaan berikut.

$$P[(A_i|B)|B] \propto P[B|(A_i|B)] \cdot P(A_i|B)$$

Metode Penaksiran Parameter dengan *Maximum Likelihood*

Salah satu metode untuk menaksir parameter adalah *Maximum Likelihood*. Misal X_i merupakan sampel ke-i dari suatu populasi dengan pdf $f(x;\theta)$ dan θ adalah parameter yang tidak diketahui, fungsi *likelihood* yang merupakan pdf gabungan dari sampel dapat didefinisikan (Hogg, 2013).

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta)$$

Akaike *Information Criterion* (AIC)

Untuk mengetahui apakah model yang dibentuk cocok untuk menggambarkan data, AIC digunakan oleh (Achieng, 2010; Guiahi, 2000) sebagai metode evaluasi tersebut. AIC didefinisikan sebagai:

$$AIC = -2\ln L + 2k$$

dengan $\ln L$ merupakan nilai *log-likelihood* dan k menyatakan banyaknya parameter model.

Quantile-Quantile* (QQ) *plot

QQ*plot* merupakan metode grafis untuk mengevaluasi kecocokan suatu distribusi dalam memodelkan data (Achieng, 2010). Artikel ini menggunakan QQ*plot* untuk melihat seberapa baik suatu distribusi ekor panjang dalam memodelkan besaran klaim.

Distribusi Ekor Panjang (*Long-tailed distribution*)

Asuransi kendaraan bermotor umumnya memiliki tipe distribusi ekor panjang karena waktu penyelesaian klaim yang lama dan besarnya tidak menentu tapi lebih dapat diprediksi dibandingkan klaim asuransi bencana alam. (Fiete, 2004) membuat pemodelan besar klaim menggunakan distribusi ekor panjang dan metode *maximum likelihood* untuk memodelkan distribusi besar klaim. Berikut ini beberapa distribusi ekor panjang (Hogg, 2013).

Tabel 1 Distribusi ekor panjang

No.	Distribusi Ekor Panjang	Parameter	Fungsi Kepadatan Peluang
1.	*Lognormal* $X \sim LN(\mu,\sigma^2)$	$\mu \in (0,+\infty)$: rata-rata $\sigma > 0$: standar deviasi	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$
2.	Gamma	$k > 0$: parameter bentuk	$f(x) = \frac{1}{\Gamma(k)\cdot\theta^k} x^{k-1} \cdot e^{-\frac{x}{\theta}}$

	$X{\sim}Gamma(k,\theta)$	$\theta > 0$: parameter skala	
3.	Weibull $X{\sim}Wbl(\lambda,k)$	$\lambda \epsilon (0,+\infty)$: parameter skala $k \epsilon (0,+\infty)$: parameter bentuk	$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, x \geq 0 \\ 0, x < 0 \end{cases}$
4.	Burr $X{\sim}Burr(c,k)$	$c > 0$: parameter bentuk $k > 0$: parameter skala	$f(x) = c \cdot k \cdot \frac{x^{c-1}}{(1+x^c)^{k+1}}$
5.	*Inverse* Gaussian $X{\sim}IG(\mu,\lambda)$	$\mu > 0$: rata-rata $\lambda > 0$: parameter bentuk	$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$
6.	Loglogistic $X{\sim}LL(\alpha,\beta)$	$\alpha > 0$: parameter skala $\beta > 0$: parameter bentuk	$f(x) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2}$

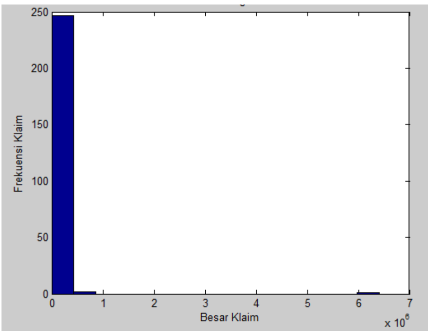
Premi Murni per Lapisan (*Layer Pure Premium*)

Berikut ini merupakan rumusan perhitungan premi murni pada lapisan 5 juta hingga 10 juta (Klugman, 2012; Meyers, 2005):

$$\mathrm{PM}_{5-10} = \int_{5.000.000}^{10.000.000} S(x)\, dx = \int_{5.000.000}^{10.000.000} [1 - F(x)]\, dx \quad (1)$$

Teknik Pengumpulan dan Analisis Data

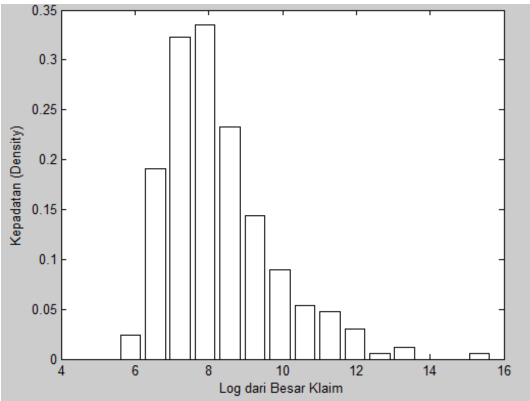
Data yang akan digunakan berasal dari laman COTOR (*Casualty Actuarial Society Committee on Theory of Risk*) *Round* 2. Data tersebut terdiri dari 250 besaran klaim yang disimulasikan dari suatu distribusi *heavy-tailed*, yang tidak diketahui jenisnya dan bersifat acak. Data ini disimulasikan oleh seorang aktuaris senior, Stuart Klugman. Terdapat asumsi bahwa klaim ini telah selesai dan nominalnya tidak akan bertambah lagi, serta tidak ada klaim yang bernilai nol. Dapat dilihat dalam Gambar 1 bahwa rata-rata besaran klaim kurang dari 1.000.000. Hanya terdapat 1 klaim yang bernilai sekitar 6.000.000. Data ini memiliki pola yang serupa dengan asuransi kendaraan bermotor pada umumnya yang memiliki besar klaim yang bernilai relatif sama dan satu klaim dengan tingkat keparahan yang besar. Terdapat kesenjangan yang besar antara klaim yang termasuk pencilan dengan klaim-klaim lainnya.



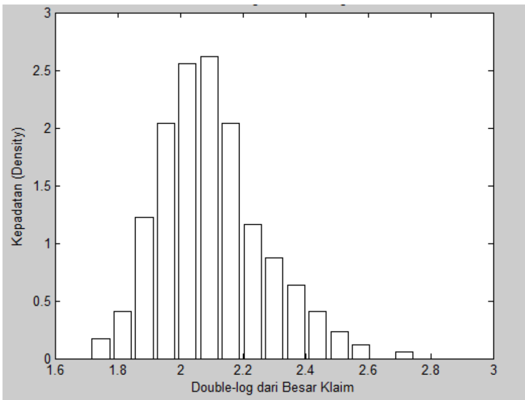
Gambar 1 Grafik histogram data COTOR

Hasil dan Pembahasan

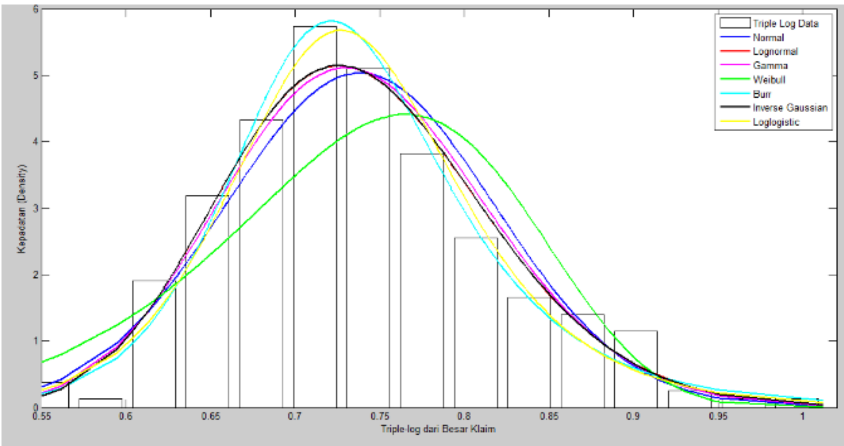
Dari data yang diperoleh, terdapat pencilan klaim sebesar 6,4 juta. Untuk memperoleh visualisasi grafik yang lebih terstruktur, (Meyers, 2005) menyarankan mentransformasikan data dengan menggunakan *log*. Hasil visualisasi gambar setelah ditransformasikan *log* sebanyak satu kali dan dua kali dapat dilihat pada Gambar 2 dan Gambar 3. Pada kedua gambar tersebut, dapat dilihat bahwa hasil transformasi merapatkan data. Transformasi *log* dilakukan satu kali lagi untuk mereduksi kecondongan (*skewness*) dan hasil transformasi besaran klaimnya masih positif, lalu beberapa distribusi ekor panjang dicocokkan pada data sehingga menghasilkan Gambar 4. Distribusi normal juga dicocokkan pada data sebagai pembanding karena kemiripan distribusi transformasi data tersebut secara visual.



Gambar 2 Grafik histogram *log* data COTOR

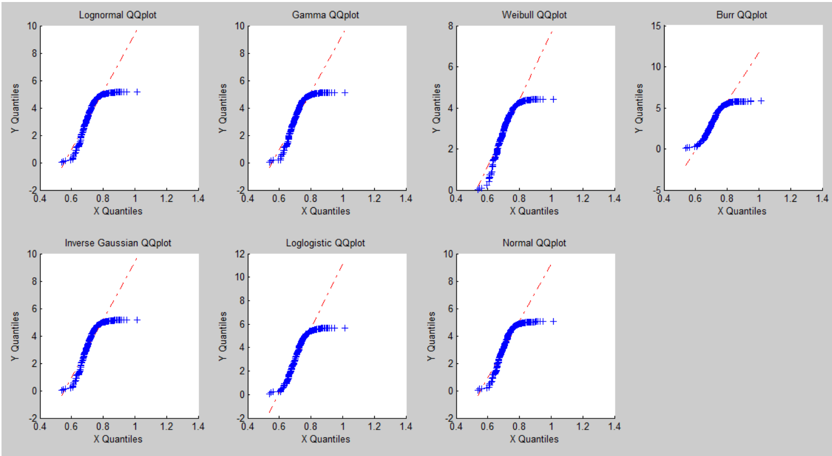


Gambar 3 Grafik histogram *double-log* data COTOR



Gambar 4 Grafik pencocokan distribusi terhadap *triple-log* data COTOR

Grafik memperlihatkan pencocokan yang baik dari distribusi terhadap data. Untuk melihat kecocokan grafis secara lebih mendalam, perhatikan grafik QQ*plot* berikut ini.



Gambar 5 *QQplot* distribusi terhadap *triple-log* COTOR

Plot yang dapat mencocokkan ekor kiri pada data adalah distribusi *Lognormal*, Gamma, *Inverse* Gaussian, dan Normal. Untuk mengetahui distribusi terbaik untuk digunakan, AIC dari setiap distribusi dihitung sehingga diperoleh hasil sebagai berikut. Untuk pencocokan ekor kanan pada data, semua distribusi melakukan performansi tersebut dengan cukup baik dikarenakan terdapat data pencilan.

Tabel 2 Nilai AIC berbagai distribusi

No.	**Distribusi**	**AIC**
1.	*Lognormal*	−562,99
2.	Gamma	−561,24
3.	Weibull	−513,40
4.	Burr	−560,33
5.	*Inverse* Gaussian	−562,99
6.	*Loglogistic*	−559,34
7.	Gaussian	−554,92

Distribusi *Lognormal*, *Inverse* Gaussian, dan Gamma memberikan hasil AIC terkecil. Oleh karena itu, ketiga distribusi ini merupakan kandidat pencocokan distribusi pada data.

Setelah melakukan pencocokan distribusi, Metodologi Bayesian yang penulis bangun berdasarkan (Achieng, 2010; Meyers, 2005) Semua perhitungan dilakukan dengan menggunakan program Matlab dan Excel.

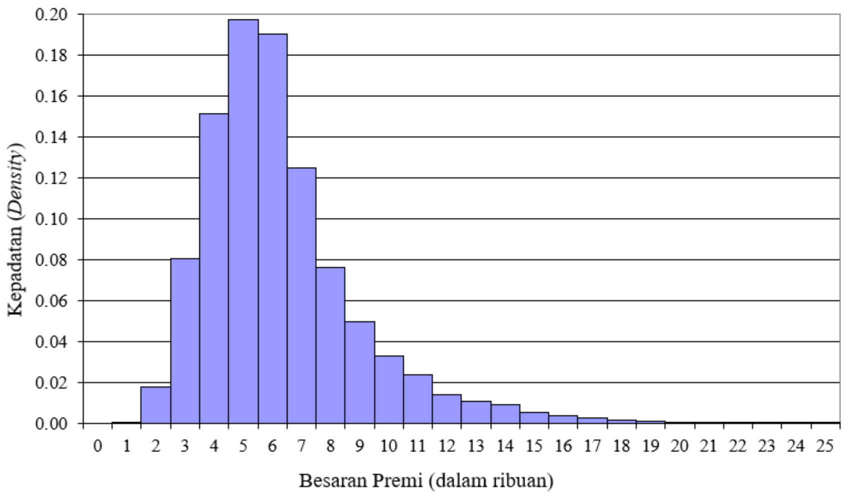
1. Penaksiran parameter setiap distribusi ekor panjang yang direkomendasikan dengan selang kepercayaan 0,1% untuk masing-masing parameter.
2. Untuk setiap distribusi, masing-masing selang kepercayaan dibagi menjadi 50 bagian yang sama besar sehingga menghasilkan sebanyak 51×51 kombinasi parameter model.
3. Ketiga distribusi tersebut masing-masing menghasilkan 2.601 model, sehingga terdapat 7.803 model secara keseluruhan. Kemudian semua model tersebut dihitung nilai *log-likelihood* yang menghasilkan nilai AIC dengan menggunakan fungsi *erf* dalam Matlab terhadap *likelihood* pada penjumlahan fungsi pdf distribusi.
4. Hitung besar premi lapisan 5 juta hingga 10 juta dengan menggunakan persamaan (1). Semua hasil perhitungan selanjutnya dilakukan di Excel. Seluruh model diurutkan dari nilai premi murni terkecil hingga terbesar.
5. Asumsikan peluang *prior* untuk semua model bernilai sama. Untuk mencari peluang *posterior*, hitung selisih nilai *log-likelihood* dari setiap model dengan nilai *log-likelihood* terbesar kemudian transformasikan hasilnya dengan eksponen. Normalisasikan hasilnya agar penjumlahan semua peluang *posterior* adalah 1.
6. Hitung rata-rata besar premi menggunakan penjumlahan dari besar premi murni yang dikalikan dengan peluang *posterior* yang telah dinormalisasikan.

Hasil perhitungan premi diperoleh sebagai berikut:

Tabel 3: Statistik premi murni

Rata-rata besar premi murni	5.815,119
Standar deviasi premi murni	2.802,655

Selain itu, berikut ini merupakan grafik histogram dari prediksi premi murni. Secara grafis, terlihat bahwa rata-rata besar premi terletak di antara 5.000 dan 6.000. Hal ini sesuai dengan hasil perhitungan yang penulis peroleh. Distribusi ekor panjang memodelkan tingkat keparahan klaim dengan baik, terutama jika terdapat data pencilan yang nilainya sangat besar dan berbeda jauh dari sebaran data pada umumnya.



Gambar 6 Histogram distribusi prediktif untuk premi murni per lapisan

Simpulan dan Saran

Hasil perhitungan premi berdasarkan distribusi ekor panjang (*Lognormal*, *Inverse* Gaussian, dan Gamma) menghasilkan prediksi yang lebih akurat dibandingkan dengan menggunakan distribusi yang dipilih oleh (Meyers, 2005) yakni gabungan antara distribusi ekor panjang (*Lognormal* dan Gamma) serta distribusi ekor pendek (Gaussian). Hal ini terlihat dari QQ*plot* dan AIC yang diperoleh. Data yang digunakan diperoleh dari simulasi dari sebuah distribusi yang tidak diketahui jenisnya yang bentuknya mirip dengan data asuransi kendaraan bermotor pada umumnya. Penulis menyarankan untuk menerapkan metode ini pada data industri asuransi kendaraan bermotor di Indonesia untuk memodelkan tingkat keparahan klaim dan juga terhadap reasuransi untuk menghitung premi murni pada *excess layer*. Pemilihan distribusi dapat berbeda dengan yang telah dipilih penulis karena struktur data yang berbeda-beda. Distribusi yang digunakan dapat berupa distribusi ekor panjang lainnya agar hasil yang diperoleh lebih akurat.

Referensi

Achieng, O. M. (2010). Actuarial Modeling For Insurance Claim Severity in Motor Comprehensive Policy Using Industrial Statistical Distributions. *International Congress of Actuaries*, 1–29.

Barnett, J. (2020). A Bayesian Approach to Excess of Loss Pure Premium Rating. *Variance*, *13*(4), 54–79. http://www.variancejournal.org/issues/13-01/54.pdf

Fiete, S. (2004). COTOR Challenge Round 3. *Casualty Actuarial Society Forum*. https://www.casact.org/research/cotor/Fiete.doc

Guiahi, F. (2000). Fitting to Loss Distributions with Emphasis on Rating Variables. *Casualty Actuarial Society Forum*.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer.

Hogg, R. (2013). *Introduction to Mathematical Statistics* (7th ed.). Pearson.

Klugman, S. A. et al. (2012). *Loss Models From Data To Decisions* (4th ed.). John Wiley & Sons.

Meyers, G. (2005). On Predictive Modeling for Claim Severity. *Casualty Actuarial Society Forum*, 215–254. http://www.casact.org/pubs/forum/05spforum/05spf215.pdf

Pitt, D., & Guillén, M. (2012). An Introduction to Parametric and Non-Parametric Models for Bivariate Positive Insurance Claim Severity Distributions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1815127

Tzougas, G., Vrontos, S., & Frangos, N. (2014). Optimal bonus-malus systems using finite mixture models. *ASTIN Bulletin*, *44*(2), 417–444. https://doi.org/10.1017/asb.2013.31