

Analisis Penggunaan Metode *Kernel Density Estimation* pada *Loss Distribution Approach* untuk Risiko Operasional

Erwan Setiawan¹⁾, Hendri Murfi²⁾, Yudi Satria³⁾

1) Program Studi Pendidikan Matematika, FKIP Universitas Suryakancana
Jl. Dr. Muwardi, Kompleks Pasir Gede Raya Cianjur 43216

2), 3) Departemen Matematika, FMIPA Universitas Indonesia, Kampus UI Depok 16424
Email: 1) erwan85setiawan@gmail.com, 2) hmurfi@gmail.com, 3) y-satria@ui.ac.id

ABSTRAK

Loss Distribution Approach (LDA) merupakan metode yang populer untuk mengestimasi kebutuhan modal bagi risiko operasional pada bidang perbankan. Dalam LDA, bank harus mengestimasi *loss severity distribution* (*lsd*) dan *loss frequency distribution* (*lfd*) berdasarkan data internal bank. Permasalahan dari LDA saat ini adalah estimasi *lsd*-nya masih berbasis pada model distribusi tertentu, padahal banyak kasus dimana data tidak dapat diestimasi dengan sangat baik oleh model distribusi tertentu yang sudah ada. Oleh karena itu, akan dijelaskan solusi dari permasalahan tersebut dengan cara mengestimasi *lsd*-nya berbasis pada data. Metode yang digunakan adalah Kernel Density Estimation (KDE). Hasil dari penelitian adalah kebutuhan modal yang dihasilkan oleh LDA yang menggunakan KDE lebih kecil 1,6% – 3,2% dibandingkan dengan LDA yang menggunakan model distribusi tertentu.

Kata kunci: *risiko operasional, loss distribution approach, metode monte carlo, kernel density estimation*

ABSTRACT

Loss Distribution Approach (LDA) is a popular method to estimate a capital charge of operational risk in banking. In LDA, Bank must estimate loss severity distribution (*lsd*) and loss frequency distribution (*lfd*) based on the internal data. The problem of the current LDA is to estimate *lsd* still refers to a model on particular distribution whereas there are many cases which can not described a data well through a distribution model that has been there. Therefore, it would be describe the solution of the problem with the way the estimation of *lsd* based on the data. The method used is Kernel Density Estimation (KDE). The result is the capital charge produced by LDA using KDE is smaller 1,6 % - 3,2 % than LDA using a certain distribution model.

Keywords: *operational risk, loss distribution approach, monte carlo method, kernel density estimation.*

1. Pendahuluan

Risiko dalam konteks perbankan merupakan suatu kejadian potensial yang berdampak negatif terhadap pendapatan dan permodalan. Bank Indonesia dalam PBI 5/8/2003 membagi risiko perbankan ke dalam 8 jenis risiko, yaitu: risiko kredit, risiko pasar, risiko operasional, risiko likuiditas, risiko hukum, risiko reputasi, risiko strategik, dan risiko kepatuhan. Risiko yang penting untuk diperhatikan karena sifatnya yang melekat pada setiap aktifitas fungsional bank adalah risiko operasional. Dalam manajemen risiko operasional, bank dipersyaratkan untuk memperhitungkan kerugian yang diperkirakan (expected loss) dan kerugian yang tidak diperkirakan (un-expected loss) dalam kebutuhan modal bagi risiko operasional. Kebutuhan modal bagi risiko operasional dikenal sebagai Economic Capital (EC).

Komite Basel dalam aturan Basel II memberikan tiga pendekatan dalam perhitungan EC, yaitu: *Basic Indicator Approach* (BIA), *Standardized Approach* (SA), dan *Advanced Measurement Approach* (AMA). Untuk pendekatan BIA dan SA formulasinya sudah ditetapkan dalam aturan Basel II, sedangkan dalam pendekatan AMA pihak bank diberikan keleluasaan untuk mengembangkan metode perhitungan EC secara internal yang berbasis data internal bank, dengan mendapatkan persetujuan dari regulator lokal [1]. Metode AMA yang banyak digunakan adalah *Loss Distribution Approach* (LDA). Dalam LDA, bank harus mengestimasi *loss severity distribution* (*lsd*) dari besarnya kerugian pada satu kejadian dalam suatu periode dan *loss frequency distribution* (*lfd*) dari banyaknya kejadian rugi dalam suatu periode berdasarkan data internal bank, dan membentuk *aggregate loss distribution* dari gabungan kedua distribusi tersebut. Nilai EC dengan metode LDA didapat dari *Value at Risk* (VaR) pada *aggregate loss distribution* dengan tingkat kepercayaan 99,9% [3].

Permasalahan dari metode LDA saat ini terletak dalam mengestimasi *lsd*. Pada saat ini, estimasi *lsd* masih berbasis pada model distribusi tertentu. Padahal sering kali data sebenarnya tidak dapat diestimasi dengan sangat baik oleh model distribusi tertentu yang sudah ada. Hal ini tentunya akan berpengaruh pada akurasi dari estimasi nilai EC. Oleh karena itu, dalam tulisan ini akan dijelaskan solusi dari permasalahan tersebut, yaitu dengan mengestimasi *lsd* berbasis pada data. Metode yang digunakan adalah *Kernel Density Estimation* (KDE). KDE merupakan suatu pendekatan statistika untuk mengestimasi fungsi distribusi probabilitas dari suatu variabel acak jika diasumsikan bentuk atau model distribusi dari variabel acak tersebut tidak diketahui.

2. Metode Penelitian

Jenis metode penelitian yang digunakan adalah eksperimen. Tahapan penelitian yang dilakukan adalah studi literatur yang berkaitan dengan topik penelitian seperti, risiko pada perbankan, konsep dasar dan penerapan LDA pada risiko operasional, konsep dasar KDE, metode Monte Carlo, Value at Risk, dan bahasa pemrograman Phyton. Setelah itu, peneliti membangun data (*data toys*) yang karakteristiknya menyerupai data risiko operasional untuk digunakan dalam simulasi. Simulasi dilakukan menggunakan bahasa pemrograman Python, dimana algoritma inti diunduh dari <http://matplotlib.org> [4], <http://scikit-learn.org> [5], dan <http://docs.scipy.org> [7].

1.1 Loss Distribution Approach (LDA)

Dalam LDA, bank mengestimasi fungsi distribusi probabilitas (pdf) dari dampak kejadian tunggal dan dari frekuensi kejadian untuk satu tahun kedepan dengan menggunakan data internal bank [3]. Secara matematis, besarnya kerugian tahunan diberikan oleh persamaan berikut:

$$Z = \sum_{i=1}^N X^{(i)} \quad (1)$$

dengan

Z = variabel acak dari besarnya kerugian per tahun

N = variabel acak dari banyaknya kejadian kerugian selama 1 tahun

$X^{(i)}$ = variabel acak dari besarnya kerugian pada kejadian ke- i

$X^{(i)}$ untuk $i = 1, 2, \dots, N$ bersifat *identically independent distribution* (iid) dan N saling bebas terhadap $X^{(i)}$. Dalam kasus ini, distribusi probabilitas dari variabel acak $X^{(i)}$ disebut *loss severity distribution* (*lsd*), dan distribusi probabilitas dari variabel acak N disebut *loss frequency distribution* (*lfd*). Karena Z dibentuk dari gabungan $X^{(i)}$ dan N maka Z akan membentuk suatu distribusi majemuk. Secara umum, tidak terdapat ekspresi secara analitik untuk menentukan distribusi majemuk sehingga perlu dilakukan suatu pendekatan secara numerik untuk mendapatkan aproksimasi dari distribusi majemuk. Beberapa metode numerik yang terkenal adalah metode Monte Carlo, Fast Fourier Transform dan Panjer Recursion.

Dalam penelitian ini, metode numerik yang digunakan adalah metode monte carlo karena lebih mudah untuk diimplementasikan dan dapat memberikan hasil yang baik. Algoritma dalam metode monte carlo adalah sebagai berikut:

- Untuk $k = 1$
 - Melakukan pengambilan bilangan secara acak dari *lfd*, misal bilangan yang terambil N .
 - Melakukan pengambilan bilangan secara acak sebanyak N dari *lsd*, misal $X^{(1)}, X^{(2)}, \dots, X^{(N)}$ dengan $X^{(1)}, X^{(2)}, \dots, X^{(N)}$ saling bebas.
 - Hitung $Z^{(k)}$ menggunakan persamaan (1)
- Ulangi langkah a) untuk $k = 2$ sampai dengan $k = K$

Pada akhirnya akan diperoleh $Z^{(1)}, \dots, Z^{(K)}$ yang merupakan sampel dari distribusi majemuk. Untuk mendapatkan estimasi nilai EC maka $Z^{(1)}, \dots, Z^{(K)}$ diurutkan sebagai berikut $\check{Z}^{(1)} \leq \dots \leq \check{Z}^{(K)}$, kemudian ditentukan persentil ke 99,9 % [6].

1.2 Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) merupakan suatu pendekatan statistika untuk mengestimasi fungsi distribusi probabilitas. Persamaan fungsi distribusi probabilitas dari KDE [2] sebagai berikut:

$$p_{KDE}(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h^D} K\left(\frac{\mathbf{x}-\mathbf{x}_n}{h}\right) \quad (2)$$

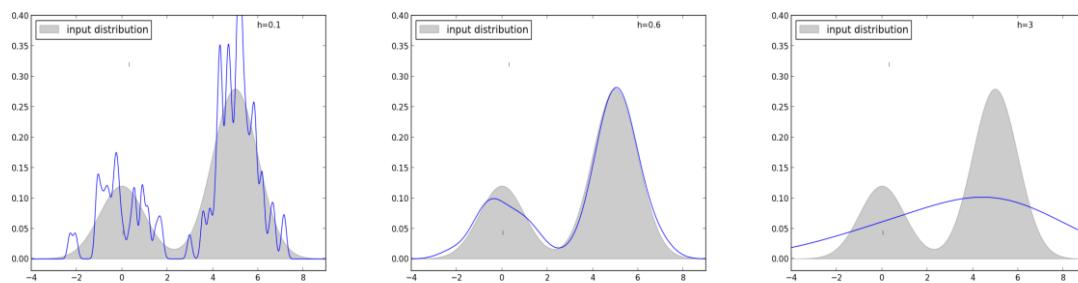
dengan

N = banyaknya sampel pengamatan
 h = bandwidth (lebar pita)
 K = fungsi kernel
 D = dimensi variabel pengamatan
 x = nilai data

Distribusi probabilitas $p_{KDE}(x)$ selalu berkaitan dengan fungsi kernel. Suatu fungsi $K(\cdot)$ dikatakan fungsi kernel jika K merupakan fungsi kontinu, bernilai riil, simetris, terbatas, dan $\int_{-\infty}^{\infty} K(u)du = 1$ [8]. Terdapat beberapa bentuk fungsi kernel diantaranya: uniform, segitiga, epanechnikov, dan gaussian. Dalam penelitian ini digunakan fungsi kernel gaussian karena dapat menghasilkan kurva yang lebih mulus. Bentuk fungsi kernel gaussian sebagai berikut:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), \text{ untuk } |u| < \infty \quad (3)$$

Parameter dalam distribusi probabilitas $p_{KDE}(x)$ adalah lebar pita h . Pemilihan nilai h yang terlalu kecil mengakibatkan grafik dari distribusi probabilitas $p_{KDE}(x)$ terlalu berduri, sehingga akan sulit untuk diinterpretasikan. Sedangkan nilai h yang terlalu besar mengakibatkan grafik dari distribusi probabilitas $p_{KDE}(x)$ terlalu mulus, sehingga menutupi struktur data yang sebenarnya. Untuk lebih jelas dapat dilihat dalam Gambar 1 berikut:

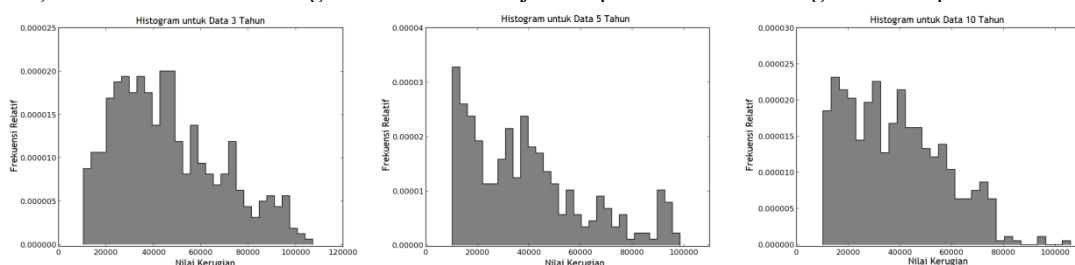


Gambar 1. Grafik distribusi probabilitas $p_{KDE}(x)$ dengan beberapa nilai h

Penelitian mengenai pemilihan nilai h yang optimal telah banyak dilakukan, namun sampai saat ini belum ada metode yang paling bagus untuk dapat digunakan dalam setiap situasi. Dalam banyak situasi, pemilihan nilai h secara subjektif dengan melihat pada grafik distribusi probabilitas yang dihasilkan dalam suatu rentang nilai h cukup baik untuk digunakan. Pertama, dimulai dengan memilih nilai h yang besar, kemudian turun ke nilai h yang lebih kecil sampai ditemukan grafik distribusi probabilitas yang cukup baik untuk mewakili data secara visual [8]. Cara ini hanya relevan pada kasus univariat.

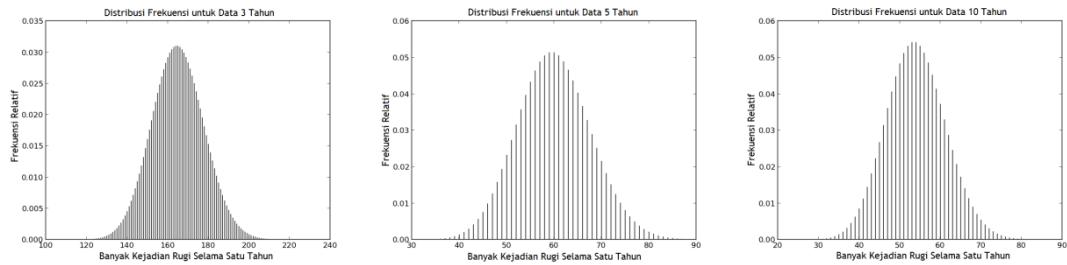
3. Hasil dan Pembahasan

Hasil dari penelitian ini didapat melalui simulasi menggunakan bahasa pemrograman python. Dalam penelitian ini, data yang digunakan adalah data *toys* yang karakteristiknya menyerupai data risiko operasional pada perbankan. Data yang digunakan sebanyak 3 data, yaitu data 3 tahun, data 5 tahun, dan data 10 tahun dengan satuan dalam juta rupiah. Berikut histogram data penelitian:



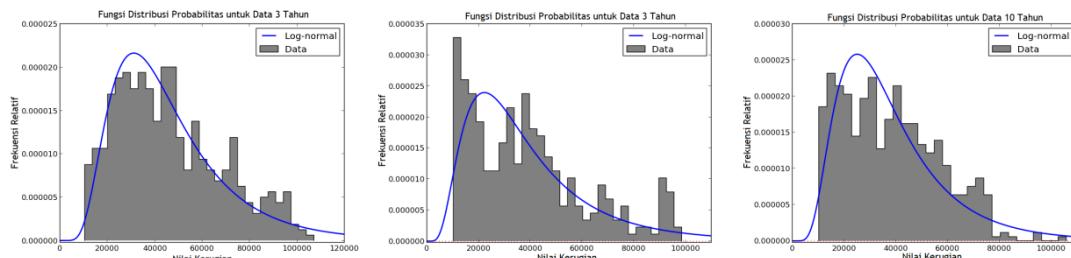
Gambar 2. Histogram data penelitian

Dari data tersebut kemudian dilakukan estimasi untuk *loss frequency distribution (lfd)*. Secara umum, banyaknya kejadian kerugian mengikuti distribusi poisson dengan parameternya adalah mean data. Diperoleh *lfd* sebagai berikut:



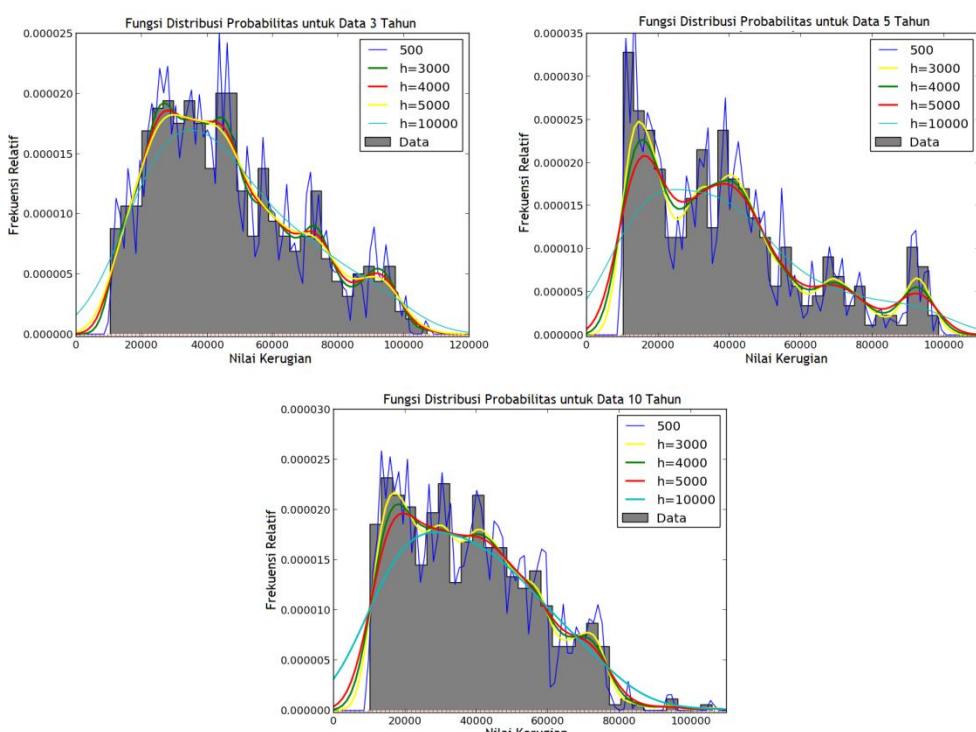
Gambar 3. *Loss frequency distribution (lfd)*

Setelah itu, dilakukan estimasi untuk *loss severity distribution (lsd)*. Dalam estimasi ini, dibagi menjadi dua cara, yang pertama estimasi menggunakan model distribusi probabilitas yang sudah ada. Dalam kasus ini, dilihat dari histogram data penelitian dalam Gambar 2, dipilih model distribusi log-normal. Hasilnya dapat dilihat pada Gambar 4.



Gambar 4. Grafik distribusi probabilitas model log-normal terhadap data penelitian

Cara kedua estimasi menggunakan metode KDE. Dalam hal ini, akan dilakukan pengujian terhadap beberapa nilai h diantaranya 500, 3000, 4000, 5000, dan 10000. Hasilnya sebagai berikut:

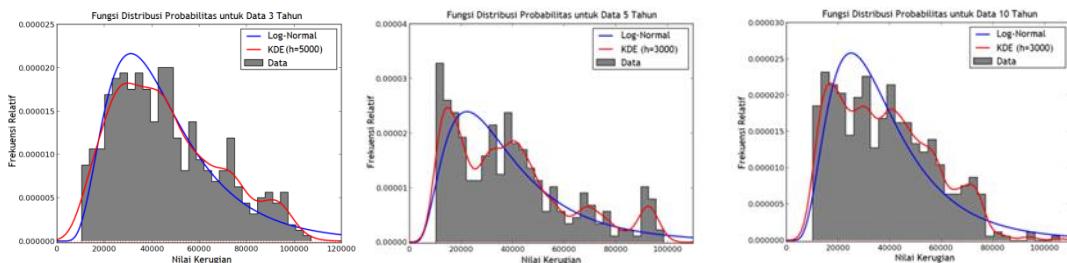


Gambar 5. Grafik distribusi probabilitas $p_{KDE}(x)$ dengan beberapa nilai h terhadap data.

Dari Gambar 5 dapat diambil beberapa kesimpulan yaitu untuk data 3 tahun dipilih nilai h optimal adalah $h=5000$ karena cukup baik menggambarkan data, untuk data 5 tahun dipilih nilai h optimal

adalah $h=3000$ karena tidak terlalu besar mengestimasi daerah yang lebih kecil dari nilai minimum data, dan untuk data 10 tahun dipilih nilai h optimal adalah $h=3000$ karena bisa menggambarkan data dengan baik termasuk tidak over estimasi untuk data-data yang bernilai besar.

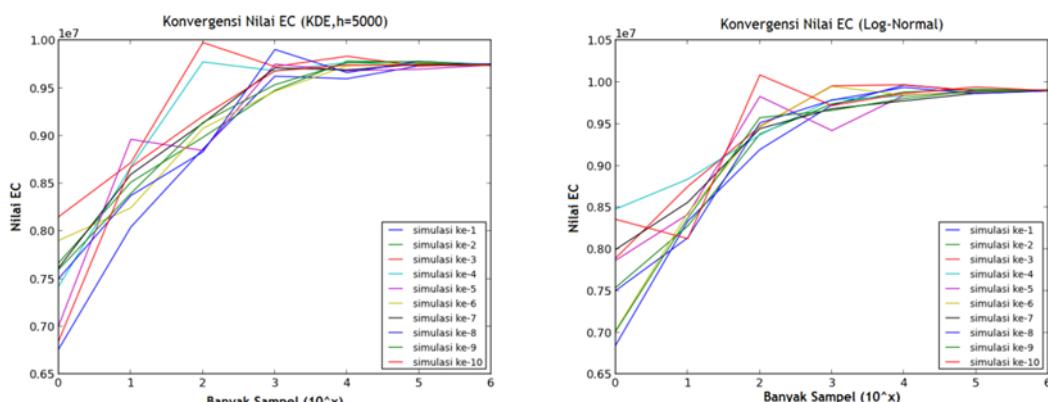
Berdasarkan hasil estimasi lsd dengan menggunakan dua cara tersebut, akan dilihat perbandingan hasil estimasi secara visual. Berikut perbandingannya:



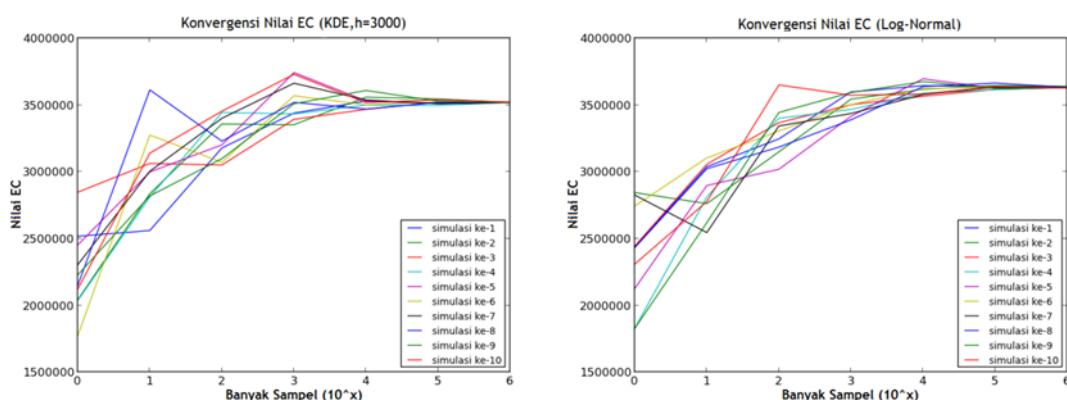
Gambar 6. Perbandingan grafik estimasi lsd antara KDE dan log-normal.

Dari Gambar 6 terlihat bahwa ketiga data yang digunakan dalam penelitian ini, estimasi lsd dengan menggunakan KDE lebih baik dalam menggambarkan data penelitian karena bisa mengestimasi daerah-daerah lokal tempat berkumpulnya data (daerah yang berupa tonjolan-tonjolan) dibandingkan dengan model distribusi log-normal.

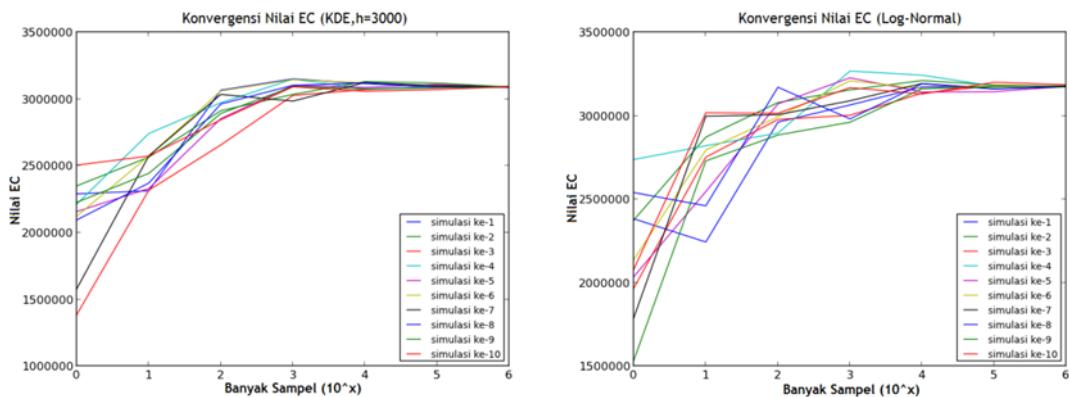
Berikutnya adalah tahap perhitungan EC dari distribusi majemuk (distribusi gabungan lfd dan lsd) yang dilakukan secara numerik menggunakan metode monte carlo. Tujuan dari simulasi ini untuk mengetahui seberapa besar perbedaan nilai EC yang dihasilkan oleh LDA yang menggunakan KDE dengan LDA yang menggunakan log-normal. Jumlah sampel yang digunakan adalah $1, 10, 10^2, 10^3, 10^4, 10^5$, dan 10^6 . Simulasi dilakukan sebanyak 10 kali untuk masing-masing jumlah sampel, hasilnya sebagai berikut:



Gambar 7. Perbandingan nilai EC antara KDE dan model log-normal pada data 3 tahun



Gambar 8. Perbandingan nilai EC antara KDE dan model log-normal pada data 5 tahun



Gambar 9. Perbandingan nilai EC antara KDE dan model log-normal pada data 10 tahun

Dari Gambar 7,8, dan 9 dapat dilihat untuk setiap simulasi akan mulai konvergen ke suatu nilai ketika jumlah sampelnya sebanyak 10^5 , hal ini sesuai dengan karakteristik dari metode monte carlo yang membutuhkan sampel lebih besar 10^5 agar konvergen ke suatu nilai. Untuk nilai EC yang diperoleh lebih rinci diberikan oleh tabel berikut:

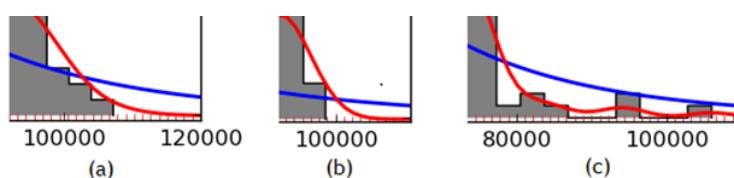
Tabel 1. Nilai EC yang dihasilkan dari KDE dan log-normal pada data 3 tahun, 5 tahun, dan 10 tahun

Metode	Nilai EC	Banyak Sampel (10^x)	Standar Deviasi	Metode	Nilai EC	Banyak Sampel (10^x)	Standar Deviasi
KDE	9,747,878.30	5	26,312	KDE	3,521,863.00	5	14,715
	9,743,212.00	6	6,553		3,520,330.50	6	2,217
Log-Normal	9,893,008.39	5	22,765	Log-Normal	3,634,601.30	5	14,731
	9,901,079.80	6	4,819		3,632,659.70	6	3,483

Metode	Nilai EC	Banyak Sampel (10^x)	Standar Deviasi
KDE	3,094,578	5	13,752
	3,091,953	6	3,152
Log-Normal	3,174,357	5	15,369
	3,178,200	6	3,337

Dari Tabel 1 dapat dilihat bahwa untuk data 3 tahun nilai EC yang dihasilkan dengan menggunakan KDE lebih kecil 1,6%, untuk data 5 tahun nilai EC yang dihasilkan dengan menggunakan KDE lebih kecil 3,2%, dan untuk data 10 tahun nilai EC yang dihasilkan dengan menggunakan KDE juga lebih kecil 2,8%.

Perbedaan nilai EC yang dihasilkan ini, disebabkan oleh estimasi yang berlebih dari log-normal untuk daerah yang jauh dari nilai maksimum data sehingga nilai EC yang menggunakan log-normal nilai lebih besar dibandingkan nilai EC yang menggunakan KDE, untuk lebih jelas mengetahui estimasi berlebih yang dilakukan oleh log-normal dapat dilihat pada gambar 10 berikut:



Gambar 10. Perbandingan tail grafik dari KDE dan log-normal (a) 3 tahun (b) 5 tahun (c) 10 tahun

Selain perbandingan nilai EC antara KDE dan log-normal, akan dilihat juga perbandingan nilai EC yang menggunakan KDE dengan beberapa nilai h . Berikut tabelnya

Tabel 2. Perbandingan nilai EC yang dihasilkan KDE dengan beberapa nilai h

Bandwidth (h)	Nilai EC	Standar Deviasi	Bandwidth (h)	Nilai EC	Standar Deviasi	Bandwidth (h)	Nilai EC	Standar Deviasi
500	9,731,174.67	7,296.34	500	3,513,631.70	8,029.72	500	3,086,040.10	6,177.87
2000	9,734,267.03		2000	3,514,677.00		1500	3,081,622.20	
3000	9,735,190.15		3000	3,521,863.00		3000	3,094,578.40	
4000	9,741,654.05		4000	3,518,947.10		4000	3,096,651.10	
5000	9,747,878.30		5000	3,522,754.50		5000	3,095,726.80	
6810	9,750,494.27		7802	3,536,074.70		5606	3,094,041.70	
10000	9,769,592.46		10000	3,552,739.60		10000	3,120,605.70	

Simulasi ini dilakukan untuk mengetahui seberapa besar pengaruh pemilihan nilai h terhadap nilai EC yang dihasilkan. Hasil simulasi pada ketiga data adalah pemilihan nilai h yang berbeda-beda dengan catatan tidak terlalu berduri atau terlalu mulus berpengaruh kecil terhadap perbedaan nilai EC hal ini dapat dilihat dari nilai standar deviasi dari nilai EC yang diberikan oleh beberapa nilai h .

4. Simpulan

Simpulan dari hasil penelitian adalah a) KDE sangat baik dalam menggambarkan struktur data yang bentuknya tidak dapat digambarkan oleh model distribusi yang sudah ada, b) nilai EC yang dihasilkan oleh LDA yang menggunakan KDE lebih kecil 1,6% – 3,2% dibandingkan nilai EC yang dihasilkan oleh LDA yang menggunakan model distribusi log-normal, c) konvergensi tail ke $p(x) = 0$ dari grafik *lsd* berpengaruh pada besar kecilnya nilai EC yang dihasilkan, d) perbedaan nilai lebar pita h dengan catatan tidak terlalu berduri atau terlalu mulus memberikan pengaruh yang kecil terhadap perbedaan nilai EC yang dihasilkan.

Daftar Pustaka

- [1]. Basel Committee on Banking Supervision. 2006. *International Convergence of Capital Measurement and Capital Standards: a revised framework*. Basel.
- [2]. Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer: New York.
- [3]. Frachot, A. Georges, P. dan Roncalli, T. 2001. *Loss Distribution Approach for Operational Risk. Working Paper, Groupe de Recherche Operationnelle*: France. Diunduh pada: 25 Mei 2013.
- [4]. Hunter, J. D. 2007. *Matplotlib: A 2D graphics environment*. *Computing In Science & Engineering*, vol 9(3), pages 90-95.
- [5]. Pedregosa *et al.* 2011. *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research* 12, pp. 2825-2830.
- [6]. Shevchenko, P.V. 2009. *Implementing Loss Distribution Approach for Operational Risk*. *Applied Stochastic Models in Business and Industry*, vol. 26(3), pages 277–307.
- [7]. Walt, S.v.d. Colbert, S.C. Varoquaux, G. 2011. *The NumPy Array: A Structure for Efficient Numerical Computation*. *Computing in Science & Engineering*, vol 13, pages 22-30.
- [8]. Zambom, A. Z. dan Dias, R. 2012. *A review of Kernel Density Estimation with Applications to Econometrics*. *arXiv:1212.2812v1 [stat.ME]*. diunduh pada: 26 Oktober 2013.

