Bidang Ilmu: MIPA

Executive Summary

PENELITIAN DISERTASI DOKTOR (APPD)

Pendekatan Teori Graf pada Data Tersensor Bivariat

Peneliti:
Mohamad Fatekurohman, M.Si.
NIDN. 0006066908

UNIVERSITAS JEMBER
Desember, 2013

Executive Summary

Penyusun Riset Hibah Doktor

Peneliti : Mohamad Fatekurohman
Sumber Dana : Dipa Universitas Jember
Kontak Email : m_fatkur@yahoo.com
Diseminasi : Telah di seminarkan pada Seminar Nasional MIPA dan Hasil Penelitian (21-22 September 2013 di FMIPA UGM)

Abstract

Data yang tersensor, univariat atau bivariat dan tersensor kiri, kanan atau interval dapat direpresentasikan dengan graf interseksi. Pada kasus bivariat, struktur *maximal clique* berkaitan dengan *Non Parametric Likelihood Estimate* (NPMLE) dari *Cummulative Distribution Function* (CDF) data tersebut. CDF NPMLE menempatkan massa pada representasi riil dari *clique* maksimal dan tidak pada tempat lain. Ada dua langkah untuk perhitungan NPMLE pada kasus data tersensor interval bivariat, pertama menentukan daerah interseksi persegi panjang dan kedua mencari *maximum likelihood.* Untuk pencarian interseksi persegi panjang dan *maximal clique* sebanyak n, dengan $n \geq 100$ dibutuhkan suatu algoritma.
Kata Kunci
Non Parametric Maximum Likelihood Estimator, Data tersensor, maximal clique, Cummulative Distribution Function.

Pendahuluan

1.1 Latar Belakang

Teori graf merupakan salah satu bagian penting dalam matematika. Pemakaian teori graf dapat diterapkan dalam berbagai ilmu pengetahuan seperti ekonomi, teknik, fisika, kimia, kedokteran dan riset operasi. Makalah pertama tentang teori graf ditulis oleh seorang ahli matematika dari Swiss, Euler, pada tahun 1736 (Chartrand dan Oellermann, 1993). Dalam kaitannya dengan data tersensor, Gentlemen dan Vandal (2002) menyatakan bahwa data tersensor dapat direpresentasikan sebagai graf interseksi., sedangkan Bogaerts dan Lesaffre (2004) berpendapat perhitungan *Non Parametric Maximum Likelihood Estimator* pada kasus data tersensor interval bivariat terdiri dari dua bagian pertama menentukan daerah irisan persegi panjang dan kedua mencari *maximum likelihood* Selanjutnya, menurut Liu dan Vandal (2011) batas-batas bawah dan batas-batas atas diberikan untuk massa probabilitas yang ditentukan oleh suatu *maximal clique* untuk mengestimasi *Cummulative Distribution Function* (fungsi distribusi komulatif, CDF) NPML *Estimator* untuk data tersensor interval multivariat (termasuk univariat), dengan asumsi bahwa mekanisme tersensor bertujuan untuk inferensi *likelihood* . Dari beberapa penelitian di atas, sampai saat ini belum ada kajian mendalam tentang data tersensor kiri untuk kasus bivariat dan akan ditunjukkan bagaimana struktur *maximal clique* berkaitan dengan peran NPML *Estimator* dari CDF. CDF dari NPML *Estimator* menentukan massa pada representasi riil dari *maximal clique* kemudian menerapkannya pada kasus Demam Berdarah Dengue.

1.2 Tujuan

Tujuan umum dari penelitian adalah menentukan ketunggalan NPML *Estimator* sedangkan tujuan khusus yang hendak dicapai dalam penelitian adalah:

a. Membuat suatu algoritma untuk data tersensor bivariat.

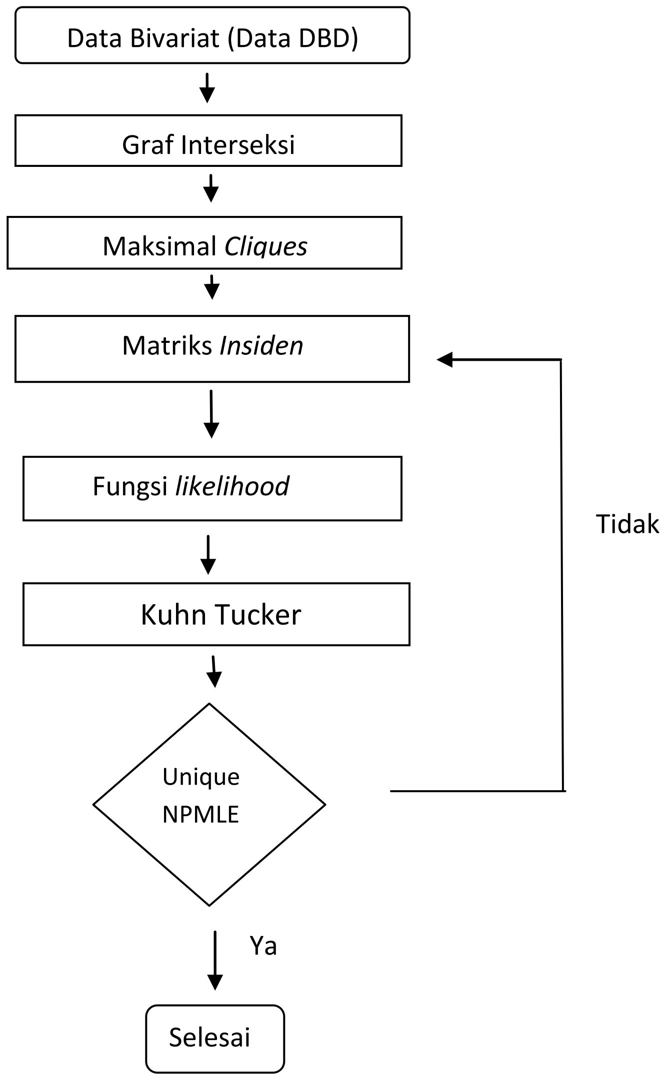
b. Menghitung kompleksitas waktu yang dibutuhkan oleh algoritma pada data tersensor bivariat.

Metode

Materi pokok untuk dasar penelitian ini adalah karya-karya ilmiah, hasil penelitian para pakar yang telah dimuat dalam buku, artikel, dan jurnal. Metode penelitian yang digunakan adalah metode teoritis. Secara teoritis, penelitian dilakukan dengan cara mempelajari karya-karya ilmiah yang telah dihimpun, yang berkaitan dengan *Nonparametric Maximum Likelihood Estimator* , data tersensor bivariat dan teori graf.. Untuk mencapai tujuan penelitian, diperlukan pengertian/konsep tentang estimasi parameter (maksimum *likelihood*), fungsi *concave*, metode sensor, fungsi tahan hidup, teori graf, graf interseksi, graf interval, *clique,* matriks insiden, metode Kuhn Tucker. Pada kajian terapan, dilakukan pengambilan data pasien demam berdarah. Untuk keperluan simulasi, estimasi, dan perhitungan akan mengunakan bantuan program komputer. Setelah penelitian dilakukan, langkah selanjutnya adalah publikasi hasil penelitian sebagai sarana berkomunikasi dan mendiskusikan hasil penelitian pada forum ilmiah yang diikuti matematikawan atau statistikawan.

Kerangka Pemecahan Permasalahan

Adapun tahapan dalam menyelesaikan penelitian yang akan dikerjakan sebagai berikut:



Hasil dan Pembahasan

Penelitian sejenis pernah dilakukan peneliti terdahulu, yaitu Lee (1983), kompleksitas waktu *worst case* $O(n^2)$, Gentleman & Vandall (2001) kompleksitas waktu *worst case* $O(n^5)$, Bogaerts dan Lesaffre (2004) kompleksitas waktu *worst case* $O(n^3)$. Adapun perbedaan penelitian ini dengan penelitian yang telah dilakukan oleh Lee (1983) yang ditulis oleh Maathuis(2003), meskipun mempunyai kompleksitas waktu *worst case* yang sama yaitu $O(n^2)$, adalah persegi panjang R_i dengan koordinat (x_{1i},x_{2i} ,y_{1i},y_{2i}), dengan (x_{1i}, y_{1i}) adalah pojok kiri atas dan (x_{2i}, y_{2i}) pojok kanan bawah. Untuk koordinat sumbu *x* yaitu (x_{1i}, x_{2i}), sedangkan koordinat sumbu *y* yaitu (y_{1i} ,y_{2i}), dan mempertimbangkan berbagai kemungkinan bentuk irisan dari persegi panjang. Diasumsikan semua koordinat *x* dan semua koordinat *y* berbeda tapi kemungkinan ada yang sama karena ada individu yang mempunyai waktu kejadian yang sama, sehingga dimungkinkan akan berimpit. Algoritma ini bekerja mengacu konsep kombinasi pada teori probabilitas, yaitu C(*n*;*r*) atau $\binom{n}{r}=\binom{n}{1}+\binom{n}{2}+\binom{n}{3}+\ldots+\binom{n}{r}$ sehingga berbagai kemungkinan irisan dari persegi panjang akan terjadi. Kompleksitas waktu yang diperoleh adalah $O(n^2)$.

Untuk menguji kebutuhan waktu dari algoritma dibutuhkan alat bantu komputer (*note book*) dengan spesifikasi Prosesor AMD E 450, Memory 4 Gb, dengan *software* Free Pascal. Hasil yang didapat data pasien demam berdarah sebanyak 36 adalah:

$p_1 = -4.88928903957242\ 10^{-18}$, $p_4 = -7.20202654616397\ 10^{-18}$,

$p_6 = 6.66827222402300\ 10^{-18}$, $p_7 = 0.500000000000009$, $p_8 = -1.06917601309375\ 10^{-18}$,

$p_{10} = 1.23639680098755\ 10^{-17}$, $p_{12} = 0.500000000000009$, $p_{15} = -8.64191452180740\ 10^{-19}$,

$p_{18} = -3.69073414672994\ 10^{-19}$, $p_{21} = 2.78958550436792\ 10^{-19}$, $p_{24} = -9.26179107305430\ 10^{-20}$,

$p_{27} = 1.13233747810143\ 10^{-19}$, $p_{28} = 8.29749200467129\ 10^{-20}$, $p_{29} = 1.31160712422215\ 10^{-20}$,

$p_{30} = -5.14057680360240\ 10^{-20}$, $p_{32} = 1.13696850422374\ 10^{-19}$, yang lain nilainya nol.

dan titik maksimumnya adalah 1.

Kesimpulan

Dari hasil dan pembahasan diperoleh kesimpulan:

1. Algoritma sejenis pernah dilakukan peneliti terdahulu, yaitu Lee (1983) yang ditulis oleh Maathuis(2003), kompleksitas waktu *worst case* $O(n^2)$, Gentleman & Vandall (2002) kompleksitas waktu *worst case* $O(n^5)$, Bogaerts dan Lesaffre (2004) kompleksitas waktu *worst case* $O(n^3)$. Adapun perbedaan penelitian ini dengan penelitian yang telah dilakukan

oleh Lee (1983) meskipun mempunyai kompleksitas waktu *worst case* yang sama yaitu $O(n^2)$, adalah persegi panjang R_i dengan koordinat (x_{1i},x_{2i} ,y_{1i},y_{2i}), dengan (x_{1i}, y_{1i}) adalah pojok kiri atas dan (x_{2i}, y_{2i}) pojok kanan bawah. Untuk koordinat sumbu x yaitu (x_{1i}, x_{2i}), sedangkan koordinat sumbu y yaitu (y_{1i} , y_{2i}), dan mempertimbangkan berbagai kemungkinan bentuk irisan dari persegi panjang.

2. Kompleksitas waktu yang dibutuhkan adalah $O(n^2)$.

Referensi

Bogaerts, K. and Lesaffre, E. (2004). A new, Fast Algorithm to find the Regions of Possible Support for Bivariate Interval-Censored Data. *J Comp Graph Statist* 13, 330–340

Chartrand, G. and Oellermann, O.R., 1993, *Applied and Algorithmic Graph Theory*, McGraw-Hill, Inc., New York.

Gentleman, R. and Vandal, A.C. 2002. Graph-Theoritical Aspects of Bivariate Censored Data. *Can. J Statist* 10. 557-571.

Liu, X and Vandal, A.C. 2011. Bounds for Self-consistent CDF Estimators for Univariate and Multivariate Censored Data**.** 7*th International Symposium on Imprecise Probability*: *Theories and Applications*, Innsbruck, Austria

Maathuis, M.H. 2003. *Nonparametric Maximum Likelihood Estimation for Bivariate Cencored Data.* Master’s thesis, Delft University of Technology, Delft