

Fungsi Likelihood Pada Data Tersensor Interval Univariat (Likelihood Function For Univariat Interval Censored Data)

Dini Tresnawanti, Mohamad Fatekurohman, Alfian Futuhul Hadi
Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember (UNEJ)
Jln. Kalimantan 37, Jember 68121
E-mail: tresnawantidini@gmail.com

Abstrak

Analisis survival adalah metode statistika yang digunakan dalam mempelajari ketahanan hidup yang berhubungan dengan waktu, mulai waktu awal yang sudah ditentukan dalam penelitian sampai waktu akhir penelitian, namun ada beberapa kendala untuk mengestimasi fungsi tersebut yakni adanya data tersensor. Untuk mengestimasi fungsi dengan masalah demikian digunakan metode nonparametrik maksimum likelihood estimator dengan data tersensor interval univariat yakni data pasien kanker payudara di Rumah Sakit Baladhika Husada (DKT) berupa data interval $l_i = (L_i, R_i)$ dengan i adalah banyaknya pasien kanker Payudara serta. Pada metode NPMLE sesuai dengan usulan Turnbull perlu dicari terlebih dahulu bagaimana bentuk fungsi likelihood. Dalam mencari fungsi likelihood dengan data univariat, dilakukan pendekatan representasi petrie untuk menghasilkan matriks Clique sebagai matriks indikator (α_{ij}) . Hasil dari penelitian ini berupa fungsi non linear dengan derajat paling besar yaitu berderajat 5.

Kata Kunci: survival, nonparametrik, likelihood, matriks Clique, Turnbull.

Abstract

Survival analysis is a statistical method which in the context of time-related survival, is initiated in the study until the end time of the study, but there are some constraints to estimate the function of the lack of censored data and relatively small amount of data. To estimate the function with such issues used nonparametric maximum likelihood estimator method by which the data univariate breast cancer patients in Baladhika Husada hospital (DKT), form of the data $l_i = (L_i, R_i)$ over i is the number of patients with breast cancer. In NPMLE methods need to be sought in advance how the shape of the likelihood function. In search of the likelihood function with the data univariate, Petrie made representations approach to generate clique matrix as a matrix of indicators (α_{ij}) . The result of this research is non linear function with the highest degree that is degree of 5.

Keywords: survival, nonparametric, likelihood, clique matrix, Turnbull .

PENDAHULUAN

Analisis *survival* merupakan analisis statistika untuk mengolah data *survival* (ketahanan hidup). Hasil pengolahan data dalam analisis *survival* dapat berupa nilai numerik ataupun grafik. Hasil pengolahan tersebut dapat dicari dengan fungsi distribusi, salah satunya yakni dengan mengestimasi fungsi survival [2]. Pada analisis *survival* harus terlebih dahulu didefinisikan kejadian yang diinginkan. Kejadian yang diinginkan pada penelitian ini adalah sembuhnya pasien kanker payudara hanya dengan mengandalkan keoterapi saja. Data pengamatan berupa interval yang dibagi sebanyak 31 hari tiap satu interval waktu. Dengan demikian estimasi fungsi distribusi yang akan dilakukan pada penelitian ini yakni estimasi fungsi survival pasien kanker payudara.

Menurut data *Pathology Based Cancer Registry* yang dilakukan oleh ikatan patologi anatomi Indonesia yang bekerja sama dengan yayasan kanker Indonesia, kanker payudara menduduki peringkat kedua dari semua jenis kanker yang sering diderita. Karenanya peneliti tertarik untuk melakukan penelitian tentang kanker payudara.

Ada beberapa macam pengobatan yang dilakukan oleh pasien kanker payudara, salah satunya adalah

kemoterapi. Kemoterapi adalah pemberian obat untuk membunuh sel kanker. Tidak seperti radiasi atau operasi yang bersifat lokal, kemoterapi merupakan terapi sistemik, yang berarti obat menyebar ke seluruh tubuh dan dapat mencapai sel kanker yang telah menyebar jauh atau metastase ke tempat lain [8].

Pasien kanker payudara menjalani sesi kemoterapi rata-rata selama 2-3 jam dan satu siklus kemoterapi berlangsung sekitar 3 minggu yaitu dari hari dimana pasien melakukan kemoterapi diperlukan sekitar 3 minggu lagi untuk sesi kemoterapi berikutnya. Siklus kemoterapi pasien rata-rata 4-6 kali. Namun mungkin saja tidak semua pasien sembuh hanya dengan mengandalkan kemoterapi, sehingga bisa jadi ada pasien yang sembuh karena adanya tindakan lain yaitu operasi pengangkatan kanker. Kejadian ini disebut sebagai kejadian data tersensor interval.

Pada kasus data tersensor interval dapat digunakan metode parametrik dan nonparametrik. Pada metode estimasi fungsi survival, metode parametrik membutuhkan beberapa asumsi-asumsi yang harus dipenuhi, sedangkan metode nonparametrik tidak diperlukan asumsi-asumsi tertentu.. Menurut [4] untuk mengestimasi distribusi F_0 yang terbentuk dari waktu survival, dengan masing-masing individu memiliki kemungkinan sederet waktu pemeriksaan

dari proses kejadian yang berbeda, dapat dilakukan metode nonparametrik maksimum *likelihood* estimator.

Analisis survival

Menurut [5] analisis *survival* adalah metode statistika yang digunakan dalam mempelajari ketahanan hidup yang berhubungan dengan waktu, mulai waktu awal yang sudah ditentukan dalam penelitian sampai waktu akhir penelitian. Kejadian yang terjadi dapat berupa kejadian meninggal, sakit, respon terhadap suatu percobaan yang dilakukan dalam penelitian yaitu sesuai dengan kejadian yang diteliti. Sedangkan menurut [2] analisis survival adalah metode yang digunakan untuk mendeskripsikan analisis data dalam bentuk waktu dari waktu awal pengamatan hingga kejadian terjadi pada sebagian kejadian atau pada titik akhir (end point).

Dalam penelitian kesehatan, waktu awal penelitian sering dikorespondensikan pada perekrutan individu dalam suatu studi dan titik akhir berupa meninggalnya pasien. Titik akhir sebagai kejadian yang dikehendaki tidak harus berupa kematian, namun juga dapat berupa hal lain seperti sembuhnya pasien dari penyakit, berkurangnya gejala penyakit atau kambuhnya pasien dari kondisi tertentu.

Kanker Payudara

Kanker payudara adalah gangguan dalam pertumbuhan sel normal mammae dimana sel abnormal timbul dari sel-sel normal, berkembang biak dan menginfiltrasi jaringan limfe dan pembuluh darah [1]. Menurut [9] Kanker payudara adalah suatu penyakit dimana terjadi pertumbuhan berlebihan atau perkembangan tidak terkontrol dari sel-sel (jaringan) payudara. Kanker payudara (carcinoma mammae) adalah sebuah tumor ganas yang tumbuh dalam jaringan payudara. Tumor ini dapat tumbuh dalam kelenjar jaringan susu maupun pada jaringan ikat payudara. Kanker ini memang tidak tumbuh dengan cepat tapi sangat berbahaya.

Kemoterapi

Menurut [8], kemoterapi adalah penggunaan preparat antineoplastik sebagai upaya untuk membunuh sel-sel tumor dengan mengganggu fungsi dan reproduksi seluler. Menurut [3], kemoterapi merupakan terapi modalitas kanker yang paling sering digunakan pada kanker stadium lanjut lokal, maupun metastatis dan sering menjadi satu-satunya pilihan metode terapi yang efektif. Kemoterapi dapat diberikan sebagai terapi utama, adjuvant (tambahan), dan neoadjuvant, yaitu kemoterapi adjuvant yang diberikan pada saat pra-operasi atau pra-radiasi.

Data Tersensor

Aspek penting dalam analisis *survival* yakni kejadian dan penyensoran. Sensor terjadi apabila ada informasi waktu ketahanan individu yang menjadi subjek penelitian, meskipun sesungguhnya kita tidak mengetahui lama waktu ketahanan secara pasti [5].

Ada tiga syarat yang harus dipenuhi dalam menentukan waktu gagal yaitu:

1. Ada waktu permulaan
2. Satuan pengukuran yang jelas
3. Definisi gagal harus jelas

Ada tiga alasan adanya data tersensor yakni :

1. Loss to follow up, terjadi ketika objek meinggal, pindah, atau menolak berpartisipasi
2. Drop out, terjadi apabila perlakuan dihentikan dengan alasan tertentu
3. Termination of study, terjadi apabila penelitian berakhir sementara objek yang diamati belum mencapai kejadian yang diinginkan [5].

Data Tersensor Interval

Data tersensor interval terjadi apabila informasi yang dibutuhkan telah dapat diketahui pada kejadian peristiwa di dalam selang pengamatan [6].

Data Tersensor Kanan

Menurut [6] data tersensor kanan terjadi apabila individu yang diamati masih hidup (apabila kejadian tidak tersensor adalah meninggal) pada saat waktu telah ditentukan. Ketika waktu penelitian telah berakhir, namun kejadian yang diinginkan dapat terjadi kapanpun setelah penelitian selesai.

Data Tersensor Kiri

Data tersensor kiri yakni kejadian ketika semua informasi yang ingin diketahui dari seorang individu telah dapat diperoleh pada awal studi [6]

Data Tersensor Random

Menurut [6] data tersensor random merupakan data tersensor yang terjadi apabila individu yang diamati mengalami kejadian karena sebab lain, bukan disebabkan dari tujuan utama penelitian. Dalam bidang kesehatan, misalkan seseorang individu yang menderita suatu penyakit meninggal setelah mengalami kecelakaan dan sebagainya, sehingga kematiannya bukan disebabkan oleh penyakit yang dideritanya.

Turnbull Estimator

Metode maksimum likelihood estimator sangat populer digunakan untuk menghasilkan suatu estimator. Ide dasar dari metode tersebut adalah mencari nilai parameter yang memberi kemungkinan (*likelihood*) yang paling besar untuk mendapatkan data yang teramati sebagai estimator. Tahun 1976 Richard Turnbull memformulasikan sebuah persamaan untuk membentuk fungsi *likelihood*. Sebelum membentuk fungsi *likelihood*, kelas ekuivalensi harus didefinisikan terlebih dahulu. Kelas ekuivalensi didefinisikan oleh L_i yang segera diikuti oleh R_i . Untuk menemukan kelas ekuivalensi perlu untuk mempertimbangkan semua interval (L_i, R_i) , untuk $i = 1, 2, \dots, n$.

Perlu untuk mencari matriks (α_{ij}) untuk membentuk fungsi *likelihood*. Matriks (α_{ij}) adalah matriks $n \times m$ dari variabel indikator. Persamaan dibawah ini merupakan prosedur yang digunakan Turnbull untuk membentuk fungsi *likelihood*.

$$L = \prod_{i=1}^n \left(\frac{\sum_{j=1}^m \alpha_{ij} s_j}{\sum_{j=1}^m s_j} \right)$$

[10]

Representasi Petrie

$M \subset X$ disebut *antichains* dari X jika $x \sim y$ untuk setiap $x, y \in M$. Matriks Petrie (kadang disebut matriks *clique*) dari kumpulan data interval adalah matriks indikator yang menghubungkan elemen X ke maksimal *antichains* dari $MA(X)$. Matriks Petrie dari X diberikan oleh :

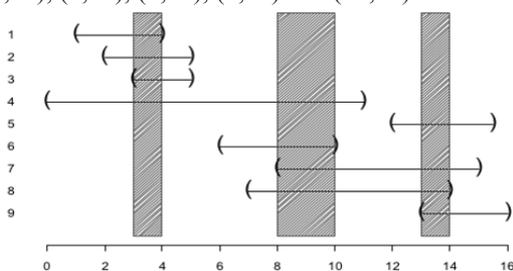
$$A = [\alpha_{ij}] \in \{0,1\}^{m \times n} \text{ dengan } \alpha_{ij} = 1 [x_j \in M_i] \tag{11}$$

$\alpha_{ij} = 1 [x_j \in M_i]$ didefinisikan sebagai fungsi berikut :

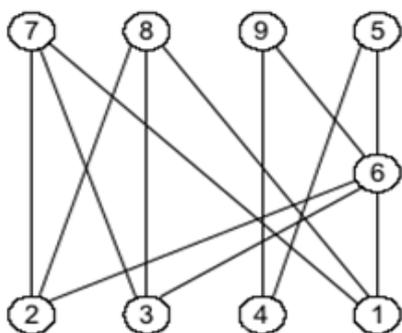
$$\alpha_{ij} = \begin{cases} 1 & \text{jika } x_j \in M_i \\ 0 & \text{jika } x_j \notin M_i \end{cases} \tag{12}$$

Contoh

Diberikan data interval: (1,4), (2,5), (3,5), (0,11), (12,16), (6,10), (8,15), (7,14) dan (13,16)



Gambar 1. Maksimal antichains dari data



Gambar 2. Hasse Diagram

Matriks Petrie atau Matriks *Clique* :

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

METODE PENELITIAN

Penelitian ini menggunakan metode dengan menganalisis data pasien kanker payudara di Rumah Sakit Baladhika Husada (DKT). Data pasien kanker payudara ditentukan data tersensornya. Selanjutnya ditentukan fungsi likelihoodnya dengan langkah-langkahnya yakni dibuat grafik maksimal antichains dari data sehingga akan diperoleh matriks Petrie atau matriks *Clique* sebagai nilai α_{ij}

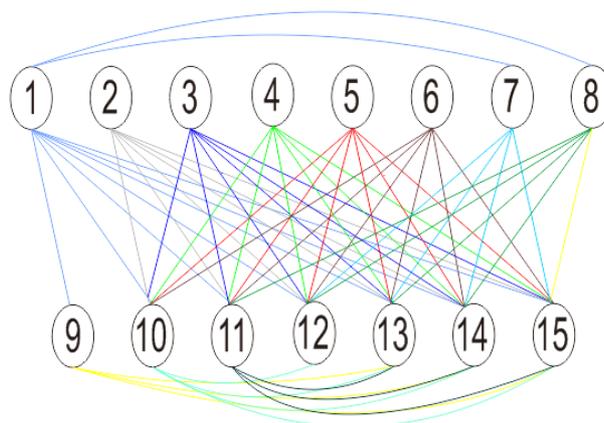
HASIL DAN PEMBAHASAN

Berikut data pasien kanker payudara dengan bentuk data interval. Setiap satu interval dibagi sebanyak 31 hari

Tabel Data L_i, R_i

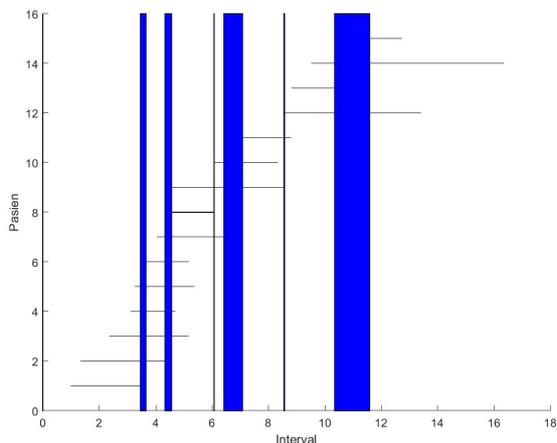
No	L_i	R_i
1	1	3.67
2	1.35	4.58
3	2.38	5.16
4	3.12	4.7
5	3.29	5.38
6	3.45	5.19
7	4.06	7.09
8	4.32	6.09
9	4.32	8.58
10	6.06	8.35
11	6.41	8.8
12	8.54	13.41
13	8.83	11.61
14	9.54	16.35
15	10.35	12.74

Dari data interval pasien kanker payudara, dibentuk sebuah diagram yaitu Hasse diagram. Hasse diagram adalah jenis diagram matematis yang digunakan untuk merepresentasikan himpunan berurutan sebagian yang terbatas. Hasse diagram pada penelitian ini merupakan pasangan pasien ke pasien lain yang dihubungkan dengan garis dimana setiap garis melambangkan kedmisto.fmipa@gmail.comua pasien tersebut tidak saling beririsan.



Gambar 3. Hasse diagram

Maksimal *Antichains* pasien kanker payudara dilihat dari grafik representasi maksimal *antichains* data pasien serta Hasse diagram pada gambar 3



Gambar 4. Grafik Representasi Maksimal Antichains

$$\begin{aligned}
 M_1 &= \{x_1, x_2, x_3, x_4, x_5, x_6\} \\
 M_2 &= \{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} \\
 M_3 &= \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\} \\
 M_4 &= \{x_7, x_9, x_{10}, x_{11}\} \\
 M_5 &= \{x_9, x_{11}, x_{12}\} \\
 M_6 &= \{x_{12}, x_{13}, x_{14}, x_{15}\}
 \end{aligned}$$

Dari maksimal antichains yang telah didapat sebelumnya, dibentuk sebuah matriks indikator yaitu matriks Petrie atau matriks Clique.

$$\begin{bmatrix}
 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 1 & 1 & 1 & 0 & 0 \\
 0 & 1 & 1 & 0 & 0 & 0 \\
 0 & 1 & 1 & 1 & 1 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 1
 \end{bmatrix}$$

Setelah matriks indikator ditemukan, selanjutnya dibentuk fungsi likelihood sesuai dengan persamaan yang diusulkan oleh Turnbull

$$L = \prod_{i=1}^{15} \left(\frac{\sum_{j=1}^6 \alpha_{ij} s_j}{\sum_{j=1}^6 s_j} \right)$$

$$\begin{aligned}
 L &= [s_1 \times (s_1 + s_2)^5 \times (s_2 + s_3 + s_4) \times \\
 &\quad (s_2 + s_3) \times (s_2 + s_3 + s_4 + s_5) \times \\
 &\quad (s_3 + s_4) \times (s_4 + s_5) \times (s_5 + s_6) \times (s_6)^3] / \\
 &\quad [s_1 + s_2 + s_3 + s_4 + s_5 + s_6]
 \end{aligned}$$

KESIMPULAN

Fungsi likelihood yang diperoleh berupa fungsi non linear. Derajat yang paling tinggi yaitu berderajat 5. matriks variabel indikator diperlukan untuk membentuk fungsi likelihood.

SARAN

Agar mendapatkan fungsi likelihood pada data tersensor seperti pada penelitian ini, terlebih dahulu dipertimbangkan data yang akan digunakan harus berisikan.

DAFTAR PUSTAKA

- [1] Carpenito, L.J. 2000. *Buku Diagnosa Keperawatan*. Editor Monica Ester. Jakarta: EGC.
- [2] Collett, D. 2003. *(Text in Statistical) Science Modelling Survival Data in Medical Research*. 2Nd ed. London: Chapman dan Hall/CRC.
- [3] Desen, W. 2008. *Buku Ajar Onkologi Klinis*, edisi 2. Jakarta : Balai Penerbit FKUI.
- [4] Gentleman, R., dan G. J. Geyer. 1994. *Maximum likelihood for interval censored data: consistency and computation*. Biometrika.
- [5] Kleinbaum, D.G. dan M. Klein. 2012. *Survival Analysis a Self-Learning Text (Third Edition)*. New York: Springer.
- [6] Klein J. P dan M. L Moeschberger. 2003. *Survival Analysis: Technique for Censored and Truncated Data*. New York: Springer.
- [7] Rasjidi, I. 2007. *Kemoterapi Kanker Ginekologi Dalam Praktik Sehari-hari*. Jakarta: Sagung Seto.
- [8] Smeltzer, S.C. dan Bare, B.G, 2002, *Buku Ajar Keperawatan Medikal Bedah Brunner dan Suddarth* (Ed.8, Vol. 1,2), Alih bahasa oleh Agung Waluyo. EGC, Jakarta.
- [9] Suryaningsih, K.E. 2009. *Kupas tuntas kanker payudara*. Yogyakarta: Paradigm Indonesia.
- [10] Turnbull, B. W. 1976, The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data, *Journal of the Royal Statistical Society, Series B*.
- [11] Vandall, A. C. dan R. Gentleman. 1998. *Order theory and nonparametric maximum likelihood for interval censored data*. New Zealand: departement of Statistic, University of Aucland.
- [12] Vandall, A. C. dan R. Gentleman. 1998. *Weak order partitioning of interval orders with applications to interval censored data*. New Zealand: departement of Statistic, University of Aucland.