p-ISSN: 2598-2400
e-ISSN: 2502-101X
Volume 4 Nomor 1 Tahun 2019

PENERAPAN MODEL PENYEMBUHAN DENGAN REGRESI COX HAZARD PROPORSIONAL PADA PENYAKIT KANKER KOLOREKTAL

Wirna Arifitriana[1], Danardono[2]
[1] Pendidikan Matematika, FKIP Universitas Graha Nusantara
[2] Departemen Matematika, FMIPA Universitas Gadjah Mada
email: wirnaariv3ana@gmail.com

Abstract
Survival analysis is a statistical technique used to analyze the data, aims to determine the variables that affect the outcome of a beginning to end the incident. One model of survival is a cure model is useful for estimating the proportion of patients who recover and the probability of survival of patients who did not recover until the deadline given. Analysis on Cox regression cure model Hazard Proportional with Maximum Likelihood Estimates and Algorithm Expectation Maximization (EM).

Keywords: Cox Proportional Hazard Cure Model, MLE, EM algorithm, likelihood ratio test, Wald test.

Abstrak

Analisis survival merupakan tehnik statistika yang digunakan untuk menganalisa data, bertujuan untuk mengetahui hasil dari variabel yang mempengaruhi suatu awal sampai akhir kejadian . Salah satu model survival yaitu model penyembuhan (*cure model)* yang berguna untuk mengestimasi proporsi pasien yang sembuh dan probabilitas survival pasien yang tak sembuh sampai pada batas waktu yang diberikan. Analisis pada model penyembuhan dengan regresi *Cox Hazard Proportional* dengan estimasi *Likelihood Maximum* dan Algoritma *Expectation Maximization (EM).*

Kata kunci: *Cox Proportional Hazard, Cure Model* , MLE, algoritma EM, uji rasio *likelihood*, uji Wald.

PENDAHULUAN

Salah satu analisis regresi yang terkenal untuk menganalisa data survival adalah regresi *Cox*. Regresi *Cox* termasuk ke dalam metode semiparametrik yang mana fungsi baseline hazard mengikuti model nonparametrik sedangkan variabel - variabel independennya mengikuti model parametrik. Tujuan dari metode regresi *Cox* adalah untuk mengetahui hubungan antara waktu survival dengan variabel-variabel yang diduga mempengaruhi waktu survival (Cox, 1972).

Analisis survival merupakan teknik statistik yang digunakan untuk menganalisis data yang bertujuan untuk mengetahui hasil dari variabel yang mempengaruhi suatu awal kejadian sampai akhir kejadian. Dalam analisis survival, ada istilah *failure* (meskipun peristiwa sebenarnya mungkin saja sukses) yaitu suatu kejadian dimana tercatatnya kejadian yang diinginkan.

Regresi *Cox* memiliki variabel *dependen* yaitu waktu survival dan variabel independen yaitu variabel yang diduga mempengaruhi waktu survival. *Cure models* merupakan model survival yang dikembangkan untuk estimasi proporsi pasien yang sembuh (*cure*) dalam studi klinik. Model ini selain digunakan untuk mengestimasi proporsi pasien yang sembuh juga digunakan untuk mengestimasi probabilitas survival pasien yang tak sembuh sampai pada batas waktu yang diberikan.

Model *mixture* dikatakan model (*cure*) *mixture* parametrik jika menggunakan distribusi probabilitas standar seperti distribusi eksponensial, Weibull, *Gompertz*, dan *generalized gamma*. Diskusi tentang model *mixture* parametrik beberapa dapat di lihat pada Boag (1949), Jones dkk (1981), Farewell (1982, 1986).

Estimator Proporsional *Hazard Cox Cure* Model

Misalkan Y adalah indikator dari individu mencapai suatu event $(Y = 1)$ atau tidak pernah $(Y = 0)$ dengan $p = \Pr(Y = 1)$. Misalkan T menunjukkan waktu untuk suatu kejadian event, didefinisikan hanya ketika Y = 1, dengan kepadatan $f(t|Y = 1)$ dan fungsi *survival* $S(t|Y = 1)$. Untuk individu tersensor, Y tidak diamati. Fungsi Margin *survival* dari T adalah $S(t) = (1 - p) + pS(t|Y = 1)$ untuk $t < \infty$. Perhatikan bahwa $S(t) \to 1 - p$ sebagai $t \to \infty$. Kita asumsikan *independent*, noninformatif, model tersensor acak dan tersensor statistik *independent* dari Y.

Farewell (1982) menggunakan model regresi logistik untuk kejadian $p(x) = \Pr(Y = 1; x) = exp(x'b)/(1 + exp(x'b))$, dimana kovarian dari x vektor termasuk *intercept* dan model *survival* parametrik untuk $S(t|Y = 1)$. Kuk dan Chen (1992) menggeneralisasi ini dengan menggunakan model *Cox Proportional Hazard* dengan fungsi *hazard* $\lambda(t|Y = 1; z) = \lambda_0(t|Y = 1)\exp(x'\beta)$, dimana z adalah vektor kovarian intercept yang lain dan $\lambda_0(t|Y = 1)$ adalah fungsi dasar *hazard* bersyarat. Melalui b dan β, model tersebut mampu memisahkan efek kovariat pada kejadian dan latensi dan, dalam arti bahwa, menyediakan kelas yang fleksibel dari model ketika terdapat keyakinan apriori dalam kelompok yang tidak rentan. Untuk model Kuk and Chen, fungsi *hazard* kumulatif bersyarat adalah $\Lambda(t|Y = 1; z) = \Lambda_0(t|Y = 1)\exp(z'\beta)$, dimana $\Lambda_0(t|Y = 1; z) = \int_0^t \lambda_0(u|Y = 1)du$. Fungsi *survival* bersyarat adalah $S(t|Y = 1; z) = S_0(t|Y = 1)^{\exp(z'\beta)}$ dimana $S_0(t|Y = 1)$ adalah fungsi dasar *survival* bersyarat.

Model standar Proportional *Hazard* adalah sebuah kasus khusus dari model Cure Proportional *Hazard* dimana $p(x) = 1$ untuk setiap x, model Cure Proportional *Hazard* adalah kasus khusus dari model kelemahan perkalian, dimana bahaya untuk individual, kondisional pada Y, dapat dituliskan $\lambda(t|Y; z) = Y\lambda(t|Y = 1; z)$. Sebagai sebuah variabel kelemahan, Y tidak sepenuhnya teramati sejak individu dianggap sebagai $Y = 1$ jika suatu kejadian teramati.

Maksimum likelihood estimation

Data yang diobservasi menunjukkan nilai i oleh $(t_i, \delta_i, z_i), i = 1, \dots, n$, dimana $t_i$adalah kejadian yang diamati atau waktu tersensor, $\delta_i = 1$ jika $t_i$tidak tersensor dan $\delta_i = 0$ untuk yang lain, dan z_i adalah suatu vektor kovarian. Untuk faktor keamanan, kita misalkan $x_i = (1, z_i')$, meskipun kovariat pada x_i dan z_i tidak harus identik. Nilai k menunjukkan nilai waktu kejadian yang berbeda dimana $t_{(i)} < \cdots < t_{(k)}$. Hal ini menunjukkan bahwa, jika $\delta_i = 1, y_i = 1$ dan, jika $\delta_i = 0$, y_i tidak teramati, dimana y_i adalah nilai yang diambil dari variabel acak Y_i. Kontribusi kemungkinan dari i adalah $p_i f(t_i|Y = 1; z_i)$ untuk $\delta_i = 1$ dan $(1 -$

$p_i) + p_i S(t_i|Y=1; z_i)$ untuk $\delta_i = 0$ dimana $p_i = pr(Y=1; x_i)$. Untuk Proportional *Hazard* Cure Model, kemungkinan teramati secara keseluruhan adalah :

$$L(b,\beta,\Lambda_0) = \prod_{i=1}^{n} \left\{p_i \Lambda_0(t_i|Y=1) e^{z_i'\beta} e^{-\Lambda_0(t_i|Y=1)\exp(z_i'\beta)}\right\}^{\delta_i}$$

$$\times \left\{(1-p_i) + p_i e^{-\Lambda_0(t_i|Y=1)\exp(z_i'\beta)}\right\}^{1-\delta_i}$$

Algoritma Expectation Maximization

Algoritma *EM* merupakan metode optimisasi iteratif untuk mendapatkan estimasi *likelihood* maksimum yang berguna dalam permasalahan data hilang atau tidak lengkap. Data keseluruhan ditunjukkan oleh $(t_i, \delta_i, z_i, y_i), i = 1, \ldots, n$ termasuk data yang diamati dan $y_i's$ yang tidak diamati. Keseluruhan data dari full *hazard* adalah :

$$L_C(b,\beta,\Lambda_0; y) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \prod_{i=1}^{n} \left\{\lambda_0(t_i|Y=1) e^{z_i'\beta}\right\}^{\delta_i y_i} \times e^{-y_i \Lambda_0(t_i|Y=1)\exp(z_i'\beta)}$$
$$= L_1(b; y) L_2(\beta, \Lambda_0; y)$$

dimana y adalah vektor dari nilai y_i. Faktor *hazard* merupakan logistik dan komponen Proportional *Hazard*. Kita menggunakan notasi L untuk *hazard* dan l untuk log-likehood.

Langkah E memperkirakan $l_C(b, \beta, \Lambda_0; y)$ sehubungan dengan distribusi $y_i's$ teramati, diberikan nilai-nilai parameter dan data O yang ter amati, dimana $O = \{y_i's\ teramati, (t_i, \delta_i, z_i, y_i), i = 1, \ldots, n\}$. Perhatikan bahwa, untuk kasus tersensor, yang $y_i's$ adalah linear dalam data lengkap log-likehood sehingga kita hanya perlu menghitung

$$\pi_i^{(m)} = E\left(Y_i|\theta^{(m)}, O\right)$$
$$= pr(Y_i = 1|T_i > t_i, \delta_i = 0, z_i; \theta^{(m)})$$
$$= pr(Y_i = 1; b) S_0(t_i|Y=1)^{\exp(z_i'\beta)} \div \Big[1 - pr(Y_i = 1; b) +$$
$$pr(Y_i = 1; b) S_0(t_i|Y=1)^{\exp(z_i'\beta)}\Big] \Big|_{\theta=\theta^{(m)}}$$

untuk kasus tersensor, dimana $\theta = (b, \beta, \Lambda_0)$, $\theta^{(m)}$ menunjukkan nilai parameter yang ada pada iterasi mth, dan $S_0(t_i|Y=1) = \exp\{-\Lambda_0(t_i|Y=1)\}$. Untuk i tidak tersensor, $E\left(Y_i|\theta^{(m)}, O\right) = y_i = 1$. Dengan demikian, langkah E menggantikan $y_i's$ dengan $w_i^{(m)}$, yang sama dengan satu jika i tidak tersensor dan sama dengan $\pi_i^{(m)}$ jika i tersensor. Perkiraan dari loglikehood oleh $\tilde{l}_C\left(b, \beta, \Lambda_0; w^{(m)}\right) = \tilde{l}_1\left(b; w^{(m)}\right) + \tilde{l}_2\left(\beta, \Lambda_0; w^{(m)}\right)$, dimana $w^{(m)} = \left\{w_i^{(m)} : i = 1, \ldots, n\right\}$.

Langkah M melibatkan algoritma maksimum dari $\tilde{l}_C$ berdasarkan b dan β dan fungsi Λ_0, diberikan $w^{(m)}$. Untuk menghadapi fungsi gangguan $\Lambda_0(t|Y=1)$ atau $S_0(t|Y=1)$, kita menambahkan tahap penyempurnaan tambahan pada langkah M menggunakan tehnik *hazard*. Dua metode dari *Cox* Proporsional *Hazard* dapat diperluas : estimator tipe Breslow untuk $\Lambda_0(t_i|Y=1)$ dan estimator batasan untuk $S_0(t_i|Y=1)$.

Aspek Komputasi

Taylor (1995) mengemukakan bahwa kendala $S_0\left(t_{(k)}|Y=1\right)=0$ adalah bentuk khusus dari Proportional Hazard mixture model dengan $\beta=0$. Kendala terjadi secara otomatis ketika bobot $w_l^{(m)}$untuk pengamatan tersensor setelah $t_{(k)}$ menjadi nol pada langkah E, pada dasarnya mengklasifikasikan bahwa itu tidak rentan. Solusi dengan kendala ini memiliki sifat statistik yang lebih baik dan konvergen lebih cepat dari MLE yang tidak terbatas.

Maximization algoritma dalam langkah M. Dalam langkah M, kita menggunakan prosedur Newton-Raphson (NR) untuk memaksimalkan $\tilde{l}_1\left(\beta;w^{(m)}\right)$ untuk mendapatkan $\hat{b}$. Simulasi NR pada (β,α) menggunakan $\tilde{l}_1\left(\beta,\alpha;w^{(m)}\right)$, bagaimanapun sensitif terhadap nilai awal dam akan mudah gagal untuk konvergen. Metode yang ditemukan untuk menjadi efisien adalah dua langkah dari NR yang disarankan oleh Prentice dan Gloeckler (1978) dalam model Proportional Hazard dalam memperbaharui dari β dan α yang diperoleh secara bergantian. Kita menggunakan parameterisasi $\lambda_i=-log\alpha_i$. Dari pengalaman bahwa algoritma di atas tidak didapatkan. Data yang diamati dari likelihood meningkat dari setiap iterasi EM., dan nilai awal berbeda memberikan mode yang sama. Hal ini dimungkinkan untuk estimasi b dan/atau β menjadi tak terbatas. Hal ini sangat jarang terjadi dan hanya jika ukuran sampel kecil dan ada jumlah yang sangat kecil dari setiap kejadian atau survivor.

Standard Error Dan Inferensi

Diperoleh perkiraan dari variansi asymptotik dari $\left(\hat{b},\hat{\beta}\right)$ berdasarkan kebalikan dari informasi matriks $I(b,\beta,\lambda)$ lengkap teramati. Perhitungan didasarkan pada likelihood yang teramati penuh yang diparameterkan berdasarkan model Proportional Hazard diskrit

$$L(b,\beta,\lambda)=\prod_{i=0}^{k}\left[\prod_{l\in D_i}\left\{1-e^{-\lambda_i\exp(z_l'\beta)}\right\}\times\exp\left(-\sum_{j:t_{(j)}\le t_{(i-1)}}\lambda_j e^{z_l'\beta}\right)\right]\times\left[\prod_{l\in C_i}\left\{(1-p)+\ p_l\exp\left(\sum_{j:t_{(j)}\le t_{(i)}}\lambda_j e^{z_l'\beta}\right)\right\}\right]$$

STUDI KASUS

Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Karya Ilmiah Paripurna fakultas Kedokteran UGM pada periode Januari sampai dengan Desember 1999. Rahmad (2006) melakukan penelitian terhadap 78 pasien penderita kanker kolorektal dan didapatkan sebanyak 43 pasien masih bertahan hidup dan 35 pasien meninggal. Mayoritas penderita adalah laki-laki sebanyak 50 pasien dan 28 pasien perempuan.

Tabel 1. Data pasien penyakit kanker kolorektal

No	Jenis Kelamin	Hidup	Meninggal
1.	Pasien Laki-laki	30	20
2.	Pasien Perempuan	13	15
	Pasien total	**43**	**35**

Berikut ini adalah variabel – variabel yang digunakan yaitu :

a. Variabel Dependen :

b. Variabel Independen

1. Jenis kelamin pasien (1= Laki-laki, 0= Perempuan)
2. Umur Pasien
3. Stadium *Dukes*
   Stadium tumor menurut *Dukes* modifikasi Turnbull ditetapkan dari laporan operasi meliputi *Dukes* A, B, C dan D
4. Jenis Operasi (1=Elektif, 0=*Emergency*)
5. Terapi

ANALISIS DATA

Model *Cure*

Secara umum cure model dinyatakan sebagai berikut:

$$\pi(z) = \frac{\exp(bz)}{1 + \exp(bz)}$$

Tabel 2. Estimasi dengan nboot = 10, nboot = 20 dan nboot = 30

	SE (nboot = 10)	SE (nboot = 20)	SE (nboot = 30)
Cure probability model			
JK	0,0010820292	0,0009876189	3,293356
UMUR	0,0002082207	0,0007327752	2,077682
J.OPERASI	5,7661580736	6,6905967834	8,096843
T.LKP	0,9953950148	0,6301596925	1,641129
T.TLKP	7,3346243716	6,8305137339	8,138435
STADIUM2	4,1275725792	8,7698624813	10,616308
STADIUM3	2,8276805922	8,3719106121	10,628397
Failure time distribution model			
JK	0,5153363	0,5174826	0,7532487
UMUR	0,4135528	0,1867372	0,2849693
J.OPERASI	0,7392011	0,6681020	0,7008350
T.LKP	0,8234033	1,0749470	1,5002460
T.TLKP	1,4814685	1,1680694	1,6234455
STADIUM2	0,7252889	0,4686885	0,6414544
STADIUM3	0,8838928	0,7041828	1,2673194

Dari ketiga nboot pada tabel 2 dapat kita lihat bahwa nboot = 10 memiliki 4 nilai *standar error* yang kecil dibandingkan nboot yang lain.

Tabel 3. Estimasi parameter model Cure

Variabel	Estimasi	P-Value
Jenis Kelamin	-1,388927	0,000000e+00
Umur	1,719028	0,000000e+00
Jenis Operasi	10,781749	6,150711e-02
T. Lengkap	24,356449	0,000000e+00
T. Tidak Lengkap	5,428664	4,592138e-01
Stadium2	1,121343	7,858748e-01
Stadium3	-21,680278	1,754152e-14

Berdasarkan hasil estimasi yang di tunjukkan pada tabel 3 variabel jenis kelamin, umur, terapi lengkap dan stadium3 atau stadium D berpengaruh secara signifikan terhadap model *cure*. Hasil ini dapat diinterpretasikan sebagai berikut:

1. Seorang pasien berjenis kelamin laki-laki maka akan menurunkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebesar exp(-1,388927)=0,2493. Namun untuk seorang pasien berjenis kelamin perempuan maka akan menurunkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebanyak seorang pasien. Jadi pasien dengan jenis kelamin perempuan lebih beresiko terjangkit kanker kolorektal.
2. Setiap penambahan umur seorang pasien maka akan menaikkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebesar exp(1,719028)=5,579 dengan kata lain semakin bertambah umur seorang pasien maka akan menambah 6 pasien yang hidup terhadap penyakit kanker kolorektal.
3. Seorang pasien mengikuti terapi lengkap maka akan menambah banyak pasien yang hidup terhadap penyakit kanker kolorektal sebesar exp(24,356449)=3,7833$\times 10^{10}$. Namun untuk setiap seorang pasien yang tidak mengikuti terapi lengkap maka akan menaikkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebanyak seorang pasien. Jadi semakin banyak pasien yang mengikuti terapi lengkap maka akan menambah banyak pasien yang hidup terhadap penyakit kanker kolorektal.
4. Seorang pasien yang dinyatakan stadium 3 maka akan menurunkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebesar exp(-21,680278)=3,840$\times 10^{-10}$. Namun untuk setiap seorang pasien yang tidak dinyatakan stadium 3 maka akan menurunkan banyak pasien yang hidup terhadap penyakit kanker kolorektal sebanyak seorang pasien.

KESIMPULAN

Berdasarkan pembahasan dan studi kasus pada bab sebelumnya, maka dapat ditarik kesimpulan sebagai berikut:
Secara umum cure model dinyatakan sebagai berikut:

$$\pi(z) = \frac{\exp(bz)}{1+\exp(bz)}$$

Hasil estimasi parameter untuk cure model menghasilkan persamaan dugaan sebagai berikut:

$$\pi(x) = \frac{\begin{array}{c}\exp(-1{,}388927X_{JK} + 1{,}719028X_{umur} + 10{,}781749X_{J.Op} + 24{,}356449X_{T.Lkp} \\ +5{,}428664X_{T.TLkp} + 1{,}121343X_{St.2} - 21{,}680278X_{St.3})\end{array}}{1+\begin{array}{c}\exp(-1{,}388927X_{JK} + 1{,}719028X_{umur} + 10{,}781749X_{J.Op} + 24{,}356449X_{T.Lkp} \\ +5{,}428664X_{T.TLkp} + 1{,}121343X_{St.2} - 21{,}680278X_{St.3})\end{array}}$$

Variabel prediktor yang berpengaruh signifikan terhadap banyak pasien yang hidup pada penyakit kanker kolorektal pada model *Cure* yaitu jenis kelamin, Umur, terapi lengkap dan jenis stadium 3 sedangkan variabel prediktor lainnya yaitu jenis operasi, terapi tidak lengkap dan jenis stadium 2 tidak berpengaruh secara signifikan terhadap banyak pasien yang hidup terhadap penyakit kanker kolorektal.

REFERENSI

Boag, J.W. 1949. *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 11, No. 1, pp. 15-53

Cox, D.R., 1972, *Regression models and life tables (with Discussion)*, Journal of The Royal Statistical Society B 34, 187 – 220.

Danardono, 2012, *Analisis Data Survival*, Diktat Kuliah, UGM, Yogyakarta.

Farewell, V.T.1982. The use ofmixture models for analysis of survival data with long-term survivors. *Biometrics* 38, 1041 - 1046.

Farewell, V. T., 1986. The use of mixture models for the analysis: Are they worth the risk? *Canadian Journal of Statistics* 14, 257 – 262

Jones, D.R., Powles, R.L., Machin, D., and Sylvester, R.J. 1981. On Estimating the proportion of cured patients in clinical studies. *Biometrie-Praximetrie* 21, 1-11.

Kuk, A. Y. C. and Chen, C. H., 1992. *A mixture model combining logistic regression with proportional hazards regression. Biometrika 79,* 531-541.

Taylor, J.M.G. 1995. *Semi-parametric Estimation in Failure-Time Mixture Models.* Biometrics. Vol.51: 899-907