SAINS Volume 15, Nomor 1/ Januari 2015 : 25 – 30

ESTIMASI PARAMETER MODEL COX PROPORTIONAL HAZARD MULTIRESPON DENGAN METODE MAXIMUM PARTIAL LIKELIHOOD ESTIMATION (MPLE)

Irfan Wahyudi
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahaun Alam, Universitas Cenderawasih Jayapura

ABSTRACT

Multirespon Cox proportional hazard models have ratio property, that is the ratio of hazard functions for two individuals with covariate vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are constant (time independent). In this study we talk about estimation of prameters on multirespon Cox model by using Maximum Partial Likelihood Estimation (MPLE) method. To determine the appropriate estimators that maximize the ln-partial likelihood function, after a score vector and a Hessian matrix are found, numerical iteration methods are applied. In this case, we use a Newton Raphson method. This numerical method is used since the solutions of the equation system of the score vector after setting it equal to zero vector are not closed form. Considering the studies about multirespon Cox model are limited, including the parameter estimation methods, but the methods are urgently needed by some fields of study related such as economics, engineering and medical sciences. For this reasons, the goal of this study is designed to develop parameter estimation methods from unirespon to multirespon cases.

Keywords : Multivariate Cox model, Maximum Partial Likelihood Estimation, Parameter Estimator, Newton Raphson Method.

PENDAHULUAN

Model Cox *proportional hazard* atau dikenal dengan analisis *survival* adalah metode statistika yang digunakan untuk menganalisis data dimana variabel responnya adalah waktu kejadian yang dianggap penting (Klembaum, 1996). Analisis waktu *survival* ini banyak diaplikasikan pada berbagai bidang ilmu seperti pada bidang ilmu kedokteran, ilmu ekonomi, dan ilmu rekayasa.

Pada bidang ilmu kedokteran, waktu *survival* bisa berupa waktu sakit, waktu sembuh, waktu terjadinya infeksi, dan sebagainya. Pada bidang ilmu ekonomi, waktu *survival* seperti kapan waktu yang tepat untuk mengganti produk dengan inovasi baru atau kapan inovasi baru dapat direalisasikan dalam proses siklus suatu produk. Bidang rekayasa juga turut berkontribusi pada perkembangan analisis *survival*, dimana pokok perhatian utamanya terletak pada pemodelan masa pakai berbagai mesin atau komponen elektronika (Lawless, 2003)

Metode-metode pada analisis *survival* telah dikembangkan oleh para peneliti pada bidang ilmu yang berbeda-beda, akibatnya muncul nama dan istilah yang berbeda-beda pula. Namun pada bidang-bidang tertentu yang sering dijumpai antara lain *survival analysis* (bidang kedokteran), *event history analysis* (bidang sosial), *failure time analysis* atau *lifetime analysis* (bidang rekayasa), *duration analysis* atau *transition analysis* (bidang ekonomi). Perbedaan nama ini tidak mengubah prinsip dan kaidah-kaidah keilmuannya, melainkan berbeda dalam hal istilah dan pendekatannya saja.

Pada data *survival,* model regresi linier biasa kurang tepat untuk digunakan, hal ini disebabkan data waktu *survival* jarang yang mengikuti distribusi normal (Klembaum,1996). Model Cox *proportional hazard* saat ini banyak

**Alamat Korespondensi*:
Jurusan Fisika, Kampus UNCEN-WAENA,
Jl. Kamp. Wolker, Jayapura Papua. 99358
e-mail: irfan_wahyudi06@yahoo.co.id

digunakan untuk analisis data *survival*. Model ini sangat populer untuk analisis data *survival* karena bentuknya sederhana dan tidak berdasarkan pada asumsi-asumsi distribusi.

Pada beberapa tahun terakhir ini teori yang berbasis model telah menjadi semakin kokoh, berbagai perkembangan memperlihatkan beberapa perluasan baru seperti model Cox terstratifikasi (Klembaum,1996) atau model Cox dengan variabel prediktor yang bergantung waktu (Collet, 1994). Metode alternatif lain untuk analisis data *survival* adalah model *accelerated failure time* (Collet, 1994).

Pada bidang statistika terdapat kajian khusus tentang analisis *survival* yaitu kajian tentang regresi Cox *propotional hazard*. Namun untuk kasus *multirespon*, hal ini masih tergolong sangat jarang ditemui, terutama pada bagian inferensi statistiknya seperti estimasi parameter dan statistik uji untuk pengujian hipotesis.

Mengingat kurangnya pembahasan tentang model Cox *multirespon* dan belum tersedianya bentuk estimator parameter, sedangkan perkembangan bidang-bidang tertentu yang terkait dengan *survival multirespon* dan membutuhkan alat-alat statistik yang relevan dirasa cukup pesat, untuk itu penelitian ini dibuat untuk memenuhi kebutuhan tersebut.

METODE PENELITIAN

Metode dan tahapan-tahapan yang dilakukan untuk mendapatkan estimator parameter model Cox *multirespon* dengan metode *Maximum Partial Likelihood* adalah sebagai berikut:

a. Menetapkan bentuk marginal model Cox *multirespon*, yaitu
   $\hbar_k(t,\mathbf{Z}_{ik}) = \hbar_0(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik})$, untuk baseline hazard yang identik, atau
   $\tilde{\hbar}_k(t,\mathbf{Z}_{ik}) = \hbar_{0k}(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik})$, untuk baseline hazard yang berbeda-beda.

b. Membentuk fungsi likelihood parsial model Cox multivariat, yaitu

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n}\sum_{l=1}^{K}Y_{jl}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(X_{ik}))}\right\}^{\Delta_{ik}}$$

   untuk baseline hazard yang identik, atau

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n}Y_{jl}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(X_{ik}))}\right\}^{\Delta_{ik}}$$

   untuk baseline hazard yang berbeda-beda Wong (1986).

c. Membentuk ln-likelihood parsial model Cox *multirespon* , yaitu

$$\ln PL(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\ln\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n}\sum_{l=1}^{K}Y_{jl}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(X_{ik}))}\right\}^{\Delta_{ik}}$$

   untuk baseline hazard yang identik, atau

$$\ln PL(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\ln\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n}Y_{jk}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(X_{ik}))}\right\}^{\Delta_{ik}}$$

   untuk baseline hazard yang berbeda-beda.

d. Memaksimumkan fungsi $\ln PL(\boldsymbol{\beta})$ dengan cara menurunkan $\ln PL(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ untuk mendapatkan vektor skor, dan selanjutnya mengatur vektor skor tersebut agar sama dengan nol, yaitu

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \ln PL(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

e. Menentukan matriks Hessian, yaitu matriks turunan parsial pertama dari vektor skor $\mathbf{U}(\boldsymbol{\beta})$ atau turunan kedua dari $\ln PL(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$, yaitu

$$\mathbf{H}(\boldsymbol{\beta}) = \frac{\partial^2 \ln PL(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}$$

f. Melakukan iterasi numerik, di sini digunakan metode Newton Raphson

$$\boldsymbol{\beta}^{(q)} = \boldsymbol{\beta}^{(q-1)} - \mathbf{H}^{-1}(\boldsymbol{\beta}^{(q-1)})\mathbf{U}(\boldsymbol{\beta}^{(q-1)}),\ q = 1,2,\ldots$$

g. Memperoleh estimator $\hat{\boldsymbol{\beta}}$ dengan menggunakan pemrograman MATLAB.

HASIL DAN PEMBAHASAN

Pada bagian ini dibahas cara mendapatkan estimator parameter pada model Cox *multirespon* dengan menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE), dan pengkajian sifat-sifatnya meliputi sifat kekonsistenan MPLE

$\hat{\boldsymbol{\beta}}$, *asymptotic normality* dari $\hat{\boldsymbol{\beta}}$, dan kekonsistenan estimator $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$.

A. Estimator Parameter Model Cox *multirespon*

Misalkan terdapat *n* unit (pasien/*cluster*) dan setiap unit berpotensi untuk mengalami K tipe *failure*. Misalkan T_{ik} adalah waktu terjadi tipe *failure* ke-*k* pada unit ke-*i*, dan misalkan C_{ik} adalah waktu sensoring untuk unit ke-*i* pada tipe *failure* ke-*k*. Didefinisikan $X_{ik} = \min(T_{ik}, C_{ik})$ dan fungsi indikator sensoring $\Delta_{ik} = I(T_{ik} \le C_{ik})$. Demikian juga dengan $\mathbf{Z}_{ik} = (Z_{1ik}, \dots, Z_{pik})^{\mathrm{T}}$ menyatakan vektor prediktor untuk unit ke-*i*, tipe *failure* ke-*k*. Vektor waktu *failure* $\mathbf{T}_i = (T_{i1}, \dots, T_{iK})$ dan vektor waktu sensoring $\mathbf{C}_i = (C_{i1}, \dots, C_{iK})$ diasumsikan saling bebas bersyarat terhadap vektor prediktor $\mathbf{Z}_i = (Z^{\mathrm{T}}{}_{i1}, \dots, Z^{\mathrm{T}}{}_{iK})$ $(i = 1, \dots, n)$. Untuk selanjutnya diasumsikan bahwa $(\mathbf{X}_i, \mathbf{C}_i, \mathbf{Z}_i)$ adalah elemen-elemen random yang saling bebas dan berdistribusi secara identik. Jika T_{ik} atau Z_{ik} hilang, maka ditetapkan $C_{ik} = 0$, yang berarti $X_{ik} = 0$ dan $\Delta_{ik} = 0$. Pada umumnya, kasus-kasus demikian tidak memberi kontribusi terhadap perhitungan statistik.

Model yang lazim digunakan untuk merumuskan distribusi marginal pada setiap tipe *failure* adalah model *proportional hazard*. Model tersebut tergantung pada apakah *baseline hazard*-nya identik atau tidak identik diantara K tipe *failure*. Fungsi *hazard* dari unit ke-*i* pada tipe *failure* ke-*k* dengan *baseline hazard* identik atau berbeda-beda, secara berturut-turut, ditulis sebagai berikut

$$\hbar_k(t, Z_{ik}) = \hbar_0(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}) \qquad (1)$$

atau

$$\tilde{\hbar}_k(t, Z_{ik}) = \hbar_{0k}(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}) \qquad (2)$$

dimana $\hbar_0(t)$ dan $\hbar_{0k}(t)$ $(k = 1, \dots, K)$ keduanya adalah fungsi *baseline hazard*, dan $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^{\mathrm{T}}$ adalah vektor parameter (koefisien regresi) yang berukuran $p \times 1$. Pada kasus-kasus ketika fungsi *baseline hazard*-nya berbeda-beda diantara K tipe *failure*, digunakan $\hbar_{0k}(t)$ $(k = 1, \dots, K)$, jika tidak demikian, maka cukup diasumsikan fungsi *baseline hazard*-nya sama. Pada kedua model (1) dan (2), digunakan vektor $\boldsymbol{\beta}$ yang sama diantara model-model *marginal*-nya. Hal ini tidak akan mengurangi keumuman karena asumsi tersebut selalu dapat dipenuhi oleh pemakaian variabel prediktor yang tepat. Perlu diketahui bahwa Wei, et al. (1989) menggunakan model (2) dengan vektor parameter regresi tipe spesifik sedangkan Lee, et al. (1992) mempelajari model (1).

Jika observasi yang diamati diasumsikan independen maka fungsi *partial likelihood* untuk $\boldsymbol{\beta}$ dari model (1) dan model (2), secara berturut turut, adalah

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n}\sum_{l=1}^{K} Y_{jl}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(X_{ik}))}\right\}^{\Delta_{ik}} \qquad (3)$$

dan

$$PL_{*}(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\frac{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}))}{\sum_{j=1}^{n} Y_{jk}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(X_{ik}))}\right\}^{\Delta_{ik}} \qquad (4)$$

dengan indikator *survival* $Y_{ik}(t) = I(X_{ik} \ge t)$, yang berarti $Y_{ik}(t) = 1$ jika unit ke-*i* pada tipe *failure* ke-*k* *survive* dan beresiko dibawah pengamatan pada titik waktu t, dan bernilai nol untuk alternatif yang lainnya. Fungsi logaritma natural dari *partial likelihood* $PL(\boldsymbol{\beta})$ pada persamaan (3) dan (4), secara berturut-turut, adalah

$$\ln PL(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}) - \ln\left[\sum_{j=1}^{n}\sum_{l=1}^{K} Y_{jl}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(X_{ik}))\right]\right\} \qquad (5)$$

dan

$$\ln PL^{*}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(X_{ik}) - \ln\left[\sum_{j=1}^{n} Y_{jk}(X_{ik})\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(X_{ik}))\right]\right\} \qquad (6)$$

Vektor skor yang bersesuaian dengan persamaan (5) dan (6), berturut-turut adalah

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left\{\mathbf{Z}_{ik}(X_{ik}) - \frac{\overline{\mathrm{S}}^{(1)}(\boldsymbol{\beta}, X_{ik})}{\overline{\mathrm{S}}^{(0)}(\boldsymbol{\beta}, X_{ik})}\right\} \qquad (7)$$

dan

$$\tilde{\mathbf{U}}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\left\{\mathbf{Z}_{ik}(X_{ik}) - \frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, X_{ik})}{\mathrm{S}_k^{(0)}(\boldsymbol{\beta}, X_{ik})}\right\} \qquad (8)$$

dimana $\mathrm{S}_k^{(0)}(\boldsymbol{\beta},t) = n^{-1}\sum_{j=1}^{n} Y_{jk}(t)\, e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(t)}$, $\mathbf{S}_k^{(1)}(\boldsymbol{\beta},t) = n^{-1}\sum_{j=1}^{n} Y_{jk}(t)\, e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(t)}\mathbf{Z}_{jk}(t)$ $(k=1,\ldots,K)$ dan $\overline{\mathbf{S}}^{(r)}(\boldsymbol{\beta},t) = \sum_{k=1}^{K}\mathbf{S}_k^{(r)}(\boldsymbol{\beta},t)$ $(r=0,1)$. Pada kedua kasus di atas, diperoleh estimator tunggal $\hat{\boldsymbol{\beta}}$ dengan cara menyelesaikan Persamaan (7) atau (8) yaitu $\left\{\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0} \text{ atau } \tilde{\mathbf{U}}(\boldsymbol{\beta}) = \mathbf{0}\right\}$. Jika penyelesaian tersebut tidak menghasilkan solusi yang eksplisit (bukan bentuk tertutup), maka estimator $\hat{\boldsymbol{\beta}}$ diperoleh dengan pendekatan secara numerik. Matriks Hessian atau matriks turunan parsial kedua dari $\ln PL(\boldsymbol{\beta})$ atau matriks turunan pertama dari vektor skor $\mathbf{U}(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ adalah

$$\mathbf{H}(\boldsymbol{\beta}) = \frac{\partial^2 \ln PL(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} = -\sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\mathbf{V}_k(\boldsymbol{\beta},t)$$

dimana

$$\mathbf{V}_k(\boldsymbol{\beta},t) = \frac{\mathbf{S}_k^{(2)}(\boldsymbol{\beta}, X_{ik})}{\mathrm{S}_k^{(0)}(\boldsymbol{\beta}, X_{ik})} - \left(\frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, X_{ik})}{\mathrm{S}_k^{(0)}(\boldsymbol{\beta}, X_{ik})}\right)^{\otimes 2}$$

adalah matriks kovariansi empiris dengan $\mathbf{S}_k^{(2)}(\boldsymbol{\beta},t) = n^{-1}\sum_{j=1}^{n} Y_{jk}(t)\, e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jk}(t)}\mathbf{Z}_{jk}(t)^{\otimes 2}$.

Sebagai penjelasan, bentuk $\mathbf{a}^{\otimes 2}$ di sini adalah notasi *outer-product* yang berarti matriks $\mathbf{aa}^{\mathrm{T}}$, ini berlaku untuk sebarang vektor kolom $\mathbf{a}$. Selanjutnya MPLE $\hat{\boldsymbol{\beta}}$ diperoleh melalui iterasi numerik, di sini digunakan metode Newton Raphson, yaitu dirumuskan sebagai berikut

$$\boldsymbol{\beta}^{(q)} = \boldsymbol{\beta}^{(q-1)} - \mathbf{H}^{-1}(\boldsymbol{\beta}^{(q-1)})\mathbf{U}(\boldsymbol{\beta}^{(q-1)}),\ q = 1,2,\ldots$$

Proses iterasi dimulai dengan penentuan nilai awal $\boldsymbol{\beta}^{(0)}$ dan iterasi dihentikan pada langkah ke-*m* jika $\left\|\boldsymbol{\beta}^{(m)} - \boldsymbol{\beta}^{(m-1)}\right\| \le \varepsilon$ dimana ε adalah bilangan yang sangat kecil. Maka MPLE $\hat{\boldsymbol{\beta}}$ dari $\boldsymbol{\beta}$ adalah $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m-1)}$.

B. Sifat kekonsistenan MPLE $\hat{\boldsymbol{\beta}}$

Pada bagian ini dibahas sifat kekonsistenan **MPLE** $\hat{\boldsymbol{\beta}}$ dari parameter $\boldsymbol{\beta}$. Secara ringkas hal ini dijelaskan pada proposisi berikut ini.

Proposisi 1: Misalkan $\hat{\boldsymbol{\beta}}$ adalah MPLE dan $\boldsymbol{\beta}_0$ adalah *true value* dari $\boldsymbol{\beta}$ pada model Cox *multirespon*. Jika $n \to \infty$ maka $\hat{\boldsymbol{\beta}} \xrightarrow{\mathrm{P}} \boldsymbol{\beta}_0$, yaitu MPLE $\hat{\boldsymbol{\beta}}$ konsisten.

Bukti: Bentuk $PL(\boldsymbol{\beta})$ pada persamaan (3) jika dinyatakan dalam notasi *counting process*, maka bentuk tersebut menjadi

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{u\ge 0}\left\{\frac{Y_{ik}(u)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(u))}{\sum_{j=1}^{n}\sum_{l=1}^{K}Y_{jl}(u)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(u))}\right\}^{dN_{ik}(u)}$$

dengan $dN_{ik}(u) = 1$ jika $N_{ik}(u) - N_{ik}(u-) = 1$ dan $dN_{ik}(u) = 0$ untuk alternatif yang lainnya. misalkan $\mathcal{L} \equiv \ln PL,$ maka $\mathcal{L}(\boldsymbol{\beta},t)$ dapat diuraikan menjadi

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta},t) &= \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(u)\, dN_{ik}(u) \\ &\quad - \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t \ln\left\{\sum_{j=1}^{n}\sum_{l=1}^{K}Y_{jl}(u)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(u)\right)\right\} dN_{ik}(u) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t \left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ik}(u) - \ln n\ \mathrm{S}_{j\square}^{(0)}(\boldsymbol{\beta},u)\right\} dN_{ik}(u)\end{aligned}$$

Definisikan

$$\tilde{G}(\boldsymbol{\beta},t) = \frac{1}{n}\left(\mathcal{L}(\boldsymbol{\beta},t) - \mathcal{L}(\boldsymbol{\beta}_0,t)\right) = \sum_{k=1}^{K}\tilde{G}_k(\boldsymbol{\beta},t)$$

dengan

$$\tilde{G}_k(\boldsymbol{\beta},t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t\left\{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}_{ik}(u) - \ln\frac{\mathrm{S}_{j\square}^{(0)}(\boldsymbol{\beta},u)}{\mathrm{S}_{j\square}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}dN_{ik}(u),$$

$$\mathrm{S}_{j\square}^{(0)}(\boldsymbol{\beta},u) = \sum_{l=1}^{K}\mathrm{S}_{jl}^{(0)}(\boldsymbol{\beta},u) \text{ dan } \mathrm{S}_{jl}^{(0)}(\boldsymbol{\beta},u) = \frac{1}{n}\sum_{j=1}^{n}Y_{jl}(u)\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{jl}(u)\right).$$

Definisikan pula $G(\boldsymbol{\beta},t) = \sum_{k=1}^{K}G_k(\boldsymbol{\beta},t)$ dengan

$$G_k(\boldsymbol{\beta},t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t\left\{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{Z}_{ik}(u) - \ln\frac{\mathrm{s}_{j\square}^{(0)}(\boldsymbol{\beta},u)}{\mathrm{s}_{j\square}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}dN_{ik}(u),$$

$$\mathrm{s}_{j\square}^{(0)}(\boldsymbol{\beta},u) = \sum_{l=1}^{K}\mathrm{s}_{jl}^{(0)}(\boldsymbol{\beta},u) \text{ dan } \mathrm{s}_{jl}^{(0)}(\boldsymbol{\beta},u) = \mathrm{E}\left(\mathrm{S}_{jl}^{(0)}(\boldsymbol{\beta},u)\right).$$

Untuk $\mathbf{S}_{jl}^{(d)}(\boldsymbol{\beta},t)$, $d = 0,1,2,$ seperti yang telah dirumuskan pada bagian 3.1 di atas, terdapat persekitaran $\mathbf{B}$ dari $\boldsymbol{\beta}_0$ dan, mencakup $\mathbf{s}_{jl}^{(d)}(\boldsymbol{\beta},t) = \mathrm{E}\left(\mathbf{S}_{jl}^{(d)}(\boldsymbol{\beta},t)\right)$ secara berturut-turut, fungsi scalar, fungsi vector, dan fungsi matriks

$s_{jl}^{(0)}(\boldsymbol{\beta},t)$, $\mathbf{s}_{jl}^{(1)}(\boldsymbol{\beta},t)$ dan $\mathbf{s}_{jl}^{(2)}(\boldsymbol{\beta},t)$ terdefinisi pada $\mathbf{B}\times[0,\tau]$ sedemikian sehingga, untuk $d=0,1,2,\ \sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathbf{B}}\left\|\mathbf{S}_{jl}^{(d)}(\boldsymbol{\beta},t)-\mathbf{s}_{jl}^{(d)}(\boldsymbol{\beta},t)\right\|\xrightarrow{\mathrm{P}}0$ jika $n\to\infty$, $j=1,\ldots,n$; $l=1,\ldots,K$. (9)

Selanjutnya, akan ditunjukkan bahwa $\tilde{G}(\boldsymbol{\beta},\tau)$ ekuivalen dengan $G(\boldsymbol{\beta},\tau)$ secara asimtotik, dalam arti bahwa $\tilde{G}(\boldsymbol{\beta},\tau)-G(\boldsymbol{\beta},\tau)\xrightarrow{\mathrm{P}}\mathbf{0}$ jika $n\to\infty$.

Dari definisi $\tilde{G}_k(\boldsymbol{\beta},\tau)$ dan $G_k(\boldsymbol{\beta},\tau)$,

$$\tilde{G}_k(\boldsymbol{\beta},\tau)-G_k(\boldsymbol{\beta},\tau)=-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\left\{\ln\frac{S_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{S_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}-\ln\frac{s_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{s_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}dN_{ik}(u)$$

$$=-\int_0^{\tau}\left\{\ln\frac{S_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{S_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}-\ln\frac{s_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{s_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}d\left(\frac{N_{\Box k}(u)}{n}\right)$$

Karena $\mathbf{s}_{jl}^{(d)}(\boldsymbol{\beta},t)$ adalah fungsi-fungsi dari $\boldsymbol{\beta}\in\mathbf{B}$ terbatas pada $\mathbf{B}\times[0,\tau]$, kontinu di $\boldsymbol{\beta}_0$ dan berlaku persamaan (9), maka untuk setiap $u\in[0,\tau]$ dan $\boldsymbol{\beta}\in\mathbf{B}$,

$$\ln\frac{S_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{S_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}-\ln\frac{s_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{s_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}$$

$$=\ln\frac{S_{j\Box}^{(0)}(\boldsymbol{\beta},u)}{s_{j\Box}^{(0)}(\boldsymbol{\beta},u)}-\ln\frac{S_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}{s_{j\Box}^{(0)}(\boldsymbol{\beta}_0,u)}\xrightarrow{\mathrm{P}}0\quad\text{jika n}\to\infty$$

sehingga diperoleh bahwa

$$\mathrm{P}\left\{\frac{N_{\Box k}(\tau)}{n}\geq\eta\right\}\leq\frac{\delta}{\eta}+\mathrm{P}\left\{\int_0^{\tau}\frac{1}{n}\sum_{i=1}^{n}Y_{ik}(u)\ \hbar_{ik}(u)\ du\geq\delta\right\}$$

$$=\frac{\delta}{\eta}+\mathrm{P}\left\{\int_0^{\tau}\hbar_{0k}(u)\ S_k^{(0)}(\boldsymbol{\beta}_0,u)\ du\geq\delta\right\}$$

Pilih $\delta>\int_0^{\tau}\hbar_{0k}(u)\ s_k^{(0)}(\boldsymbol{\beta}_0,u)\ du,$ maka

$$\mathrm{P}\left\{\int_0^{\tau}\hbar_{0k}(u)\ S_k^{(0)}(\boldsymbol{\beta}_0,u)\ du\geq\delta\right\}\xrightarrow{\mathrm{P}}0\quad\text{jika}$$

$n\to\infty$, sehingga $\delta/\eta\to 0$ jika $\eta\to\infty$.

Karena

$$\lim_{\eta\to\infty}\lim_{n\to\infty}\mathrm{P}\left\{\frac{N_{\Box k}(\tau)}{n}\geq\eta\right\}=0$$

Jadi, $\tilde{G}_k(\boldsymbol{\beta},\tau)-G_k(\boldsymbol{\beta},\tau)\xrightarrow{\mathrm{P}}\mathbf{0}$. Akibatnya $\tilde{G}(\boldsymbol{\beta},\tau)$ ekuivalen dengan $G(\boldsymbol{\beta},\tau)$ secara asimtotik. Karena itu, untuk data besar ($n\to\infty$), MPLE $\hat{\boldsymbol{\beta}}$ di $\tilde{G}(\boldsymbol{\beta},\tau)$ konvergen dalam probabilitas ke $\boldsymbol{\beta}_0$ (*true value* yang memaksimumkan $\mathcal{L}(\boldsymbol{\beta},t)$ di $G(\boldsymbol{\beta},\tau)$). Hal ini menunjukkan bahwa jika $n\to\infty$, maka MPLE $\hat{\boldsymbol{\beta}}\xrightarrow{\mathrm{P}}\boldsymbol{\beta}_0$.

C. Sifat *asymptotic normality* dari $\hat{\boldsymbol{\beta}}$

Untuk data besar, estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ konvergen dalam distribusi ke $\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$, dimana $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)=\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{n}\sum_{l=1}^{K}\mathrm{E}\left\{\mathbf{D}_{1ik}(\boldsymbol{\beta}_0)\mathbf{D}_{1jl}^{\mathrm{T}}(\boldsymbol{\beta}_0)\right\}$.

Hal ini dijelaskan pada proposisi berikut ini.

Proposisi 2: Misalkan $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ adalah matriks kovariansi. jika $n\to\infty$, maka $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{\mathbf{D}}\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$.

Bukti: Ekspansi Taylor orde pertama untuk vektor skor $\mathbf{U}(\boldsymbol{\beta})$ dengan pusat *true value* $\boldsymbol{\beta}_0$ dari $\boldsymbol{\beta}$ dapat ditulis sebagai berikut

$$\mathbf{U}(\boldsymbol{\beta})=\mathbf{U}(\boldsymbol{\beta}_0)+\frac{\partial}{\partial\boldsymbol{\beta}}\mathbf{U}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}(\boldsymbol{\beta}-\boldsymbol{\beta}_0),\qquad(10)$$

dimana $\boldsymbol{\beta}^*$ terletak pada segmen garis antara $\hat{\boldsymbol{\beta}}$ dan $\boldsymbol{\beta}_0$. Karena $\hat{\boldsymbol{\beta}}$ adalah MPLE maka $\mathbf{U}(\hat{\boldsymbol{\beta}})=\mathbf{0}$, sehingga persamaan (10) menjadi

$$\frac{1}{\sqrt{n}}\mathbf{U}(\boldsymbol{\beta}_0)=\frac{1}{n}\left(-\frac{\partial}{\partial\boldsymbol{\beta}}\mathbf{U}(\boldsymbol{\beta})\right)\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\sqrt{n}\ (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0).$$

Dari teorema limit sentral multivariat Puri dan Sen (1971) diperoleh bahwa $\frac{1}{\sqrt{n}}\mathbf{U}(\boldsymbol{\beta}_0)\xrightarrow{\mathrm{P}}\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$. Dari proposisi 1 diketahui bahwa $\hat{\boldsymbol{\beta}}$ konsisten, ini berarti $\boldsymbol{\beta}^*\xrightarrow{\mathrm{P}}\boldsymbol{\beta}_0$ jika $n\to\infty$, sehingga $\frac{1}{n}\left(-\frac{\partial}{\partial\boldsymbol{\beta}}\mathbf{U}(\boldsymbol{\beta})\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\xrightarrow{\mathrm{P}}\mathbf{I}(\boldsymbol{\beta}_0)$ jika $n\to\infty$. Jadi $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{\mathrm{P}}\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$. Akibatnya $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{\mathbf{D}}\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$.

3.4. Sifat kekonsistenan estimator $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$

Penjelasan sifat kekonsistenan estimator $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$ secara ringkas dijelaskan pada proposisi berikut ini.

Proposisi 3: Misalkan $\hat{\boldsymbol{\beta}}$ adalah MPLE dan $\boldsymbol{\beta}_0$ adalah *true value* dari $\boldsymbol{\beta}$ pada model Cox *multirespon*. Jika $n \to \infty$ maka $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$.

Bukti: Untuk menunjukkan sifat kekonsistenan estimator $\hat{\boldsymbol{\Sigma}}$, diperlukan dua hal yang harus diperlihatkan, yaitu jika $n \to \infty$ maka:

1) $n^{-1}\sum_{i=1}^{n} \mathbf{B}_{ik}(\hat{\boldsymbol{\beta}})\mathbf{B}_{il}^{T}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} E\left\{\mathbf{D}_{1k}(\boldsymbol{\beta}_0)\mathbf{D}_{1l}^{T}(\boldsymbol{\beta}_0)\right\}$ dan
2) $\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} \mathbf{I}(\boldsymbol{\beta}_0)$

dimana

$$\mathbf{B}_{ik}(\boldsymbol{\beta}) = \int_0^{\tau} \left\{ \mathbf{Z}_{ik}(u) - \frac{\mathbf{S}_{j\square}^{(1)}(\boldsymbol{\beta},u)}{\mathbf{S}_{j\square}^{(0)}(\boldsymbol{\beta},u)} \right\} \left\{ dN_{ik}(u) - \frac{Y_{ik}(u)e^{\boldsymbol{\beta}^T \mathbf{Z}_{ik}(u)}}{n\mathbf{S}_{j\square}^{(0)}(\boldsymbol{\beta},u)} dN_{\square k}(u) \right\}$$

$$\mathbf{D}_{ik}(\boldsymbol{\beta}) = \int_0^{t} \left\{ \mathbf{Z}_{ik}(u) - \frac{\mathbf{s}_{j\square}^{(1)}(\boldsymbol{\beta},u)}{\mathbf{s}_{j\square}^{(0)}(\boldsymbol{\beta},u)} \right\} dM_{ik}(u)$$

dan

$$\mathbf{I}(\boldsymbol{\beta}) = \sum_{k=1}^{K} \int_0^{\tau} \mathbf{v}_k(\boldsymbol{\beta},t)\ s_k^{(0)}(\boldsymbol{\beta},t)\ h_{0k}(t)\ dt.$$

Untuk menunjukkan bagian pertama, cukup dibuktikan bahwa setiap komponen dari $n^{-1}\sum_{i=1}^{n} \mathbf{B}_{ik}(\hat{\boldsymbol{\beta}})\mathbf{B}_{il}^{T}(\hat{\boldsymbol{\beta}})$ konvergen dalam probabilitas ke setiap komponen yang sesuai dengan $E\left\{\mathbf{D}_{1k}(\boldsymbol{\beta}_0)\mathbf{D}_{1l}^{T}(\boldsymbol{\beta}_0)\right\}$. Bagian kedua dapat ditunjukkan dengan cara membuktikan $\left\|\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}) - \mathbf{I}(\boldsymbol{\beta}_0)\right\| \xrightarrow{P} \mathbf{0}$ jika $n \to \infty$. Dari kedua hal tersebut, diperoleh bahwa $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$ adalah estimator yang konsisten, dengan $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}) \xrightarrow{P} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$.

KESIMPULAN

Dari uraian singkat dengan penjelasan-penjelasan sederhana di atas dapat ditarik beberapa kesimpulan, yaitu:

a) Estimator-estimator parameter model Cox *multirespon* dapat dicari dengan cara yang cukup sederhana menggunakan metode MPLE. Nilai estimator-estimator tersebut dapat didekati dengan menggunakan metode iterasi Newton Raphson dengan algoritma yang cukup sederhana pula.
b) Estimator MPLE bersifat konsisten dan *asymptotic normality*.
c) Estimator matriks varian kovarian $\hat{\Sigma}(\hat{\boldsymbol{\beta}})$ bersifat konsisten.

SARAN

Saran bagi para peneliti yang bekerja pada bidang survival *multirespon*, khususnya pada model Cox *multirespon* untuk menggunakan metode iterasi selain Newton Raphson, hal ini untuk membandingkan metode mana yang lebih akurat dan lebih efisien dalam memperoleh nilai-nilai hampiran estimator dan dalam proses iterasi.

DAFTAR PUSTAKA

Collett, D. (1994). *Modeling Survival Data in Medical Research,* chapman and Hall, London.

Klembaum, D.G. (1996). *Survival Analysis: A Self learning text.* Springer, New York.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2nd edition, John Wiley & Sons. Inc., Hoboken, New Jersey.

Lee, E.W., Wei, L.J. dan Amato, D.A. (1992). *Cox-type regression analysis for large numbers of small groups of correlated failure time observations,* in Klein, J.P. and Goel, P.K. (eds), *Survival analysis*: *State of the art*, Kluwer Academic Publishers, Dordrecht, pp. 237-247.

Puri, M.L. dan Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*, New York: Chapman and Hall.

Wei, L.J., Lin, D.Y. dan Weissfeld, L. (1989). *Regression analysis of multivariate incomplete failure time data by modeling marginal distributions*, Journal of the America Statistical Association, 84, 1065-1073.

Wong, W.H. (1986). *Theory of partial likelihood. The Annals of Statistics,* Vol. 14. No. 1, 88-123.