STATISTIKA

FORUM TEORI DAN APLIKASI STATISTIKA

Volume 3, Edisi Khusus, Mei 2003 **ISSN : 1411 - 5891**

PROCEEDINGS

CONFERENCE ON STATISTICAL AND MATHEMATICAL SCIENCES OF ISLAMIC SOCIETY IN SOUTH EAST ASIA REGION

Bandung, April 25 -26, 2003

Diterbitkan oleh

Jurusan Statistika

Fakultas Matematika & Ilmu Pengetahuan Alam

Universitas Islam Bandung

Volume 3, Edisi Khusus, Mei 2003 ISSN : 1411 – 5891

STATISTIKA

FORUM TEORI DAN APLIKASI STATISTIKA

Terbit Minimal 1 kali Setahun

PENGANTAR REDAKSI

Bismillahirrahmanirrahim
Assalamu'alaikum Wr. Wb.

Dengan mengucapkan Alhamdulillah, segala puji dan syukur kami panjatkan ke hadirat Allah SWT bahwa Jurnal Statistika dapat terbit dihadapan pembaca.

Pada penerbitan kali ini merupakan edisi khusus. Semua artikel yang dimuat telah dipresentasikan pada Konfrensi Statistika dan Matematika Masyarakat Islam Asia Tenggara (Conference on Statistical and Mathematical Sciences of Islamic Society in South East Asia Region) yang berlangsung pada Tanggal 25 – 26 April 2003 yang diselenggarakan oleh Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung.

Artikel-artikel yang dimuat merupakan sebagian dari artikel yang dipresentasikan dalam kegiatan Konfrensi yang berkaitan dengan bidang Statistika.

Keseluruhan artikel yang disajikan pada terbitan keempat ini diharapkan dapat menambah wawasan pemikiran dan pengetahuan di bidang kajian statistika dan menambah wawasan penerapan statistika di bidang ilmu lainnya bagi para pembaca.

Ucapan terima kasih kami sampaikan kepada seluruh peserta Konfrensi yang telah berpartisipasi dan juga kepada Panitia penyelenggara yang telah memberikan kepercayaan untuk memuat artikel-artikel peserta dalam jurnal ini.

Wassalamu'alaikum Wr. Wb.
Redaksi

Alamat Redaksi
Jurusan Statistika FMIPA UNISBA
Jl. Tamansari No. 1 Bandung 40116
e-mail : fmipa_unisba@yahoo.com
Telp. (022)4203368 Pes.136 fax. (022)440678

Volume 3, Edisi Khusus, Mei 2003

ISSN : 1411 – 5891

STATISTIKA

FORUM TEORI DAN APLIKASI STATISTIKA

Daftar Isi

PEMODELAN COMPETING RISK DENGAN DISTRIBUSI EKSPONENSIAL YANG SALING BEBAS

ABDUL KUDUS

Jurusan Statistika Fakultas MIPA UNISBA
Jl. Tamansari No. 1 Telp. 4203368 pes. 135 Fax 4263895 Bandung 40116
akudus69@yahoo.com

NOOR AKMA IBRAHIM, ISA DAUD, MOHD. RIZAM ABU BAKAR

Institute for Mathematical Research and Department of Mathematics,
Universiti Putra Malaysia
e-mail : nakma@fsas.upm.edu.my, isa@fsas.upm.edu.my, mrizam@fsas.upm.edu.my

Abstrak : Dalam makalah ini dibahas pemodelan competing risk untuk dua penyebab kegagalan. Dua variabel competing risk tersebut diasumsikan berdistribusi eksponensial saling bebas satu sama lain. Selain pemodelan dengan distribusi eksponensial di sini juga dilakukan pemodelan dengan distribusi eskponensial campuran. Pada bagian akhir ditunjukkan metode penaksiran kemungkinan maksimum untuk mendapatkan taksiran parameternya.

Kata Kunci : Data Masa Hidup, Competing Risk, Eksponensial Campuran

1. PENDAHULUAN

Dalam pendekatan competing risk kita memodelkan data sebagai barisan pasangan iid (Z_i,D_i), i = 1, 2, ... Setiap Zi adalah minimum dari dua atau lebih variabel competing risk. Dalam banyak kasus, kita dapat menyederhanakan masalah menjadi analisis dua competing risk, yang digambarkan oleh dua variabel acak X dan Y sedemikian sehingga Z = min(X,Y). Biasanya X merupakan waktu kegagalan minimum diantara sekian banyak penyebab kegagalan, sedangkan Y adalah waktu penyensoran yakni waktu kegagalan oleh sebab lain. Disamping pengamatan variabel Z, juga diamati variabel indikator D=I(X<Y) yang menjelaskan penyebab berhentinya waktu pengamatan. Untuk kemudahan kita asumsikan bahwa P(X=Y)=0.

Dari pengamatan (Z,D) kita hanya dapat mengidentifikasi fungsi subsurvival dari X dan Y,

$$S_X^*(t) = P(X > t, X < Y) = P(Z > t, D = 1) \tag{1}$$

$$S_Y^*(t) = P(Y > t, Y < X) = P(Z > t, D = 0) \tag{2}$$

tetapi fungsi survival $S_X(t)$ dan $S_Y(t)$ tidak dapat diidentifikasi. Kita lihat bahwa $S_X^*(t)$ tergantung dari Y, walaupun hal ini tidak tampak dalam notasi, juga bahwa $S_X^*(0) = P(X < Y) = P(D = 1)$ dan $S_Y^*(0) = P(Y < X) = P(D = 0)$, sehingga $S_X^*(0) + S_Y^*(0) = 1$.

Fungsi subsurvival bersyarat (*conditional subsurvival*) didefinisikan sebagai fungsi survival dengan syarat bahwa satu penyebab kegagalan telah terjadi. Dengan mengasumsikan kontinuitas dari $S_X^*(t)$ dan $S_Y^*(t)$ di nol, fungsi subsurvival bersyarat adalah

$$CS_X^*(t) = P(X > t \mid X < Y) = P(Z > t \mid D = 1) = \frac{S_X^*(t)}{S_X^*(0)} \tag{3}$$

$$CS_Y^*(t) = P(Y > t \mid Y < X) = P(Z > t \mid D = 0) = \frac{S_Y^*(t)}{S_Y^*(0)} \tag{4}$$

Fungsi lain yang berkaitan dengan fungsi subsurvival adalah peluang penyensoran di atas t,

$$\Phi(t) = P(Y < X \mid Z > t) = P(D = 0 \mid Z > t) = \frac{S_Y^*(t)}{S_X^*(t) + S_Y^*(t)} \tag{5}$$

Fungsi ini tampaknya berguna untuk patokan pemilihan model competing risk yang cocok dengan data. Kita lihat bahwa $\Phi(0) = P(D = 0) = S_Y^*(0)$.

Misal kita mempunyai data (Z1, D1), ..., (Zn, Dn). Fungsi subsurvival empiris dan fungsi subsurvival bersyarat adalah :

$$\hat{S}_X^*(t) = \frac{\#\,\text{pasangan } (Z_i, D_i) \text{ dengan } Z_i > t, D_i = 1}{n} \tag{6}$$

$$\hat{S}_Y^*(t) = \frac{\#\,\text{pasangan } (Z_i, D_i) \text{ dengan } Z_i > t, D_i = 0}{n} \tag{7}$$

$$\mathcal{C}S_X^*(t) = \frac{\#\,\text{pasangan } (Z_i, D_i) \text{ dengan } Z_i > t, D_i = 1}{\sum_{i=1}^n D_i} \tag{8}$$

$$\mathcal{C}S_Y^*(t) = \frac{\#\,\text{pasangan } (Z_i, D_i) \text{ dengan } Z_i > t, D_i = 0}{n - \sum_{i=1}^n D_i} \tag{9}$$

Sebagaimana sudah dijelaskan, tanpa tambahan asumsi tentang distribusi bersama X dan Y, maka tidak mungkin untuk mengidentifikasi fungsi survival marjinal $S_X(t)$ dan $S_Y(t)$. Akan tetapi, dengan tambahan asumsi, kita dapat membatasi pada sekelompok model dimana fungsi survivalnya dapat diidentifikasi.

Jika kita membatasi fungsi survival $S_X(t)$ sebagai berikut

$$P(X \le t, X \le Y) \le P(X \le t) \le P(\min(X, Y) \le t), \tag{10}$$

berarti bahwa

$$S_X^*(t) + S_Y^*(0) \ge S_X(t) \ge S_X^*(t) + S_Y^*(t) \tag{11}$$

Dapat dilihat bahwa kuantitas sisi kiri dan sisi kanan selalu dapat ditaksir dari data. Laju kegagalan $r_X(t)$ bagi X adalah

$$r_X(t) = -\frac{\frac{dS_X(t)}{dt}}{S_X(t)} = -\frac{d\{\ln[S_X(t)]\}}{dt} \tag{12}$$

Dengan demikian rata-rata waktu laju kegagalan sebagai fungsi dari t adalah

$$\frac{1}{t}\int_0^t r_X(u)\,du = -\frac{1}{t}\ln[S_X(t)] \tag{13}$$

sehingga batas bagi fungsi survival adalah

$$-\frac{\ln[S_X^*(t) + S_Y^*(0)]}{t} \le \frac{\int_0^t r_X(u)\,du}{t} \le -\frac{\ln[S_X^*(t) + S_Y^*(t)]}{t} \tag{14}$$

Hal yang sama juga berlaku bagi laju kegagalan Y. Batas-batas ini dapat memberikan indikasi berharga tentang kemonotonan laju kegagalan X dan Y.

2. EKSPONENSIAL YANG SALING BEBAS

Dengan asumsi bahwa X dan Y saling bebas, maka kita dapat menentukan fungsi survival X dan Y secara tunggal (*unique*) dari fungsi distribusi bersama (Z,D). Dalam hal ini fungsi survival X dan Y dapat diidentifikasi dari data tersensor (Z,D). Dengan demikian, model yang saling bebas selalu konsisten dengan data, walaupun model eksponensial yang saling bebas ini bukan merupakan bentuk umum. Kita dapat menurunkan kriteria yang sangat khusus bagi kebebasan dan ke-eksponensial-an yang dinyatakan dalam fungsi subsurvival (Cooke, 1996).

Teorema 1: Jika X dan Y adalah variabel masa hidup yang saling bebas, maka dua dari keadaan-keadaan berikut berakibat terhadap yang lainnya :

$$S_X(t) = \exp(-\lambda t) \tag{15}$$

$$S_Y(t) = \exp(-\gamma t) \tag{16}$$

$$S_X^*(t) = \frac{\lambda}{\lambda+\gamma}\exp\left[-(\lambda+\gamma)t\right] \tag{17}$$

$$S_Y^*(t) = \frac{\gamma}{\lambda+\gamma}\exp\left[-(\lambda+\gamma)t\right] \tag{18}$$

Dengan demikian, jika X dan Y variabel masa hidup eksponensial yang saling bebas dengan laju kegagalan λ dan γ, maka fungsi subsurvival bersyarat dari X dan Y adalah sama yaitu distribusi eksponensial dengan laju kegagalan $\lambda + \gamma$ dan peluang tersensor di atas waktu t adalah konstan. Jadi

$$CS_X^*(t) = CS_Y^*(t) = \exp\left[-(\lambda+\gamma)t\right] \tag{19}$$

$$\Phi(t) = \frac{\gamma}{\lambda+\gamma} \tag{20}$$

3. MODEL EKSPONENSIAL CAMPURAN

Misalkan bahwa $S_X(t)$ adalah campuran dari dua distribusi eksponensial dengan parameter λ_1, λ_2 dan koefisien pencampuran p, dan distribusi survival tersensor $S_Y(t)$ adalah eksponensial dengan parameter λ_y:

$$S_X(t) = p\,\exp(-\lambda_1 t) + (1-p)\exp(-\lambda_2 t) \tag{21}$$

$$S_Y(t) = \exp(-\lambda_y t) \tag{22}$$

Sifat-sifat dari model competing risk-nya dijelaskan dalam teorema-teorema berikut :

Teorema 2: Misal X dan Y adalah variabel masa hidup yang saling bebas dengan distribusi seperti di atas, maka

$$S_X^*(t) = p\frac{\lambda_1}{\lambda_y+\lambda_1}\exp\left[-(\lambda_y+\lambda_1)t\right] + (1-p)\frac{\lambda_2}{\lambda_y+\lambda_2}\exp\left[-(\lambda_y+\lambda_2)t\right] \tag{23}$$

$$S_Y^*(t) = p\frac{\lambda_y}{\lambda_y+\lambda_1}\exp\left[-(\lambda_y+\lambda_1)t\right] + (1-p)\frac{\lambda_y}{\lambda_y+\lambda_2}\exp\left[-(\lambda_y+\lambda_2)t\right] \tag{24}$$

$$CS_X^*(t) = \frac{\exp\left[-(\lambda_y+\lambda_1)t\right] + \dfrac{1-p}{p}\dfrac{\lambda_2}{\lambda_1}\dfrac{\lambda_y+\lambda_1}{\lambda_y+\lambda_2}\exp\left[-(\lambda_y+\lambda_2)t\right]}{1+\dfrac{1-p}{p}\dfrac{\lambda_2}{\lambda_1}\dfrac{\lambda_y+\lambda_1}{\lambda_y+\lambda_2}} \tag{25}$$

$$CS_Y^*(t) = \frac{\exp\left[-(\lambda_y+\lambda_1)t\right] + \frac{1-p}{p}\frac{\lambda_y+\lambda_1}{\lambda_y+\lambda_2}\exp\left[-(\lambda_y+\lambda_2)t\right]}{1+\frac{1-p}{p}\frac{\lambda_y+\lambda_1}{\lambda_y+\lambda_2}} \tag{26}$$

$$CS_X^*(t) \le CS_Y^*(t) \tag{27}$$

Dan $\Phi(t)$ bernilai minimum pada titik pangkal (*origin*), dan menaik saat $\lambda_1 \neq \lambda_2$. Pembuktian lihat di lampiran.

4. PENUTUP

Model untuk analisis data survival dengan fenomena competing risk dalam makalah ini dibentuk dalam kasus dimana fungsi subsurvival bersyarat dari variabel yang tersensor mendominasi fungsi subsurvival bersyarat dari penyebab kegagalan yang lain.

DAFTAR PUSTAKA

Cooke, R. M., 1996. The design of reliability databases Part I and II, *Reliability Engineering and System Safety*, 51, 137 – 146 dan 209 – 223.

Lampiran

Bukti :

Dengan menggunakan kebebasan antara X dan Y maka ditunjukkan

$$S_X^*(t) = -\int_t^{\infty} S_Y(u)\,dS_X(u) \tag{1.a}$$

$$S_Y^*(t) = -\int_t^{\infty} S_X(u)\,dS_Y(u) \tag{2.a}$$

Empat rumus pertama dalam teorema 2 berdasarkan integral ini.

Untuk membuktikan ketidaksamaan antara fungsi subsurvival bersyarat kita dapat menuliskan fungsi subsurvival bersyarat dalam bentuk yang ringkas sebagai berikut

$$CS_X^*(t) = \frac{A+\frac{\lambda_2}{\lambda_1}BC}{1+\frac{\lambda_2}{\lambda_1}B}, \tag{3.a}$$

$$CS_Y^*(t) = \frac{A+BC}{1+B} \tag{4.a}$$

Dengan demikian,

$$\frac{S_X^*(t)}{S_X^*(0)} \le \frac{S_Y^*(t)}{S_Y^*(0)} \tag{5.a}$$

yang ekivalen dengan

$$A\left(1-\frac{\lambda_2}{\lambda_1}\right) \le C\left(1-\frac{\lambda_2}{\lambda_1}\right) \tag{6.a}$$

atau

$$\exp(-\lambda_1 t)(\lambda_1-\lambda_2) \quad \le \quad \exp(-\lambda_2 t)(\lambda_1-\lambda_2) \tag{7.a}$$

Ini berlaku untuk semua λ_1, λ_2.

Dari $\dfrac{S_X^*(t)}{S_X^*(0)} \le \dfrac{S_Y^*(t)}{S_Y^*(0)}$ berarti bahwa $\Phi(t)$ bernilai minimum pada titik pangkal, dan bahwa $\Phi(t)$ bersifat menaik dengan melihat

$$\Phi'(t) = C\frac{pS_1(t)\ \cdot\ (1-p)S_2(t)}{\left[pS_1(t)\ +\ (1-p)S_2(t)\right]^2} \tag{8.a}$$

dengan $C = \dfrac{\lambda_y(\lambda_1-\lambda_2)^2}{\left[(\lambda_y+\lambda_1)(\lambda_y+\lambda_2)\right]}$ dan $S_1(t)=\exp(-\lambda_1 t)$, $S_2(t)=\exp(-\lambda_2 t)$

Menarik untuk diketahui bahwa fungsi survival bagi Z menjadi

$$P(Z>t) = P(X>t, Y>t) = P(X>t)P(Y>t) = p\exp[-(\lambda_1+\lambda_y)t] + (1-p)\exp[-(\lambda_2+\lambda_y)t] \tag{9.a}$$

Ini juga merupakan campuran dari dua distribusi eksponensial dengan parameter $\lambda_1 + \lambda_y$ dan $\lambda_2 + \lambda_y$, dan koefisien pencampuran p.

Nilai ekspektasi dan varians dari Z masing-masing adalah

$$E(Z) = \int_0^\infty z f_z(z)dz = \int_0^\infty P(Z>t)dt = p\frac{1}{\lambda_y+\lambda_1} + (1-p)\frac{1}{\lambda_y+\lambda_2} \tag{10.a}$$

$$Var(Z) = \frac{1}{(\lambda_y+\lambda_2)^2} + 2p\frac{1}{\lambda_y+\lambda_1}\left(\frac{1}{\lambda_y+\lambda_1} - \frac{1}{\lambda_y+\lambda_2}\right) - p^2\left(\frac{1}{\lambda_y+\lambda_1} - \frac{1}{\lambda_y+\lambda_2}\right)^2 \tag{11.a}$$

Dengan menggunakan persamaan di atas, $S_Y^*(0)$ dan $\Phi'(t)$ kita dapat memperoleh taksiran dari parameter. Misalnya $\lambda_y = \dfrac{S_Y^*(0)}{E(Z)}$ dan nilai-nilai p, λ_1, λ_2 dapat diperoleh secara numerik. Dapat dilihat bahwa solusi yang sama bagi λ_y dapat diperoleh dengan metode kemungkinan maksimum.

Bagi pengamatan $(Z, D) = (t, 1)$ kontribusi kemungkinannya adalah

$$-\frac{dS_X^*(t)}{dt} = \exp(-\lambda_y t)\left[p\lambda_1\exp(-\lambda_1 t) + (1-p)\lambda_2\exp(-\lambda_2 t)\right] \tag{12.a}$$

sedangkan bagi pengamatan $(Z, D) = (t, 0)$ kontribusinya adalah

$$-\frac{dS_Y^*(t)}{dt} = \lambda_y\exp(-\lambda_y t)\left[p\exp(-\lambda_1 t) + (1-p)\exp(-\lambda_2 t)\right] \tag{13.a}$$

Dengan demikian dengan data $(Z_1, D_1), \ldots, (Z_n, D_n)$ kita mendapatkan fungsi kemungkinan

$$\lambda_y^{n-\sum D_i}\exp\left(-\lambda_y\sum Z_i\right)\prod_i\left[p\lambda_1\exp(-\lambda_1 t) + (1-p)\lambda_2\exp(-\lambda_2 t)\right]^{D_i}\left[p\exp(-\lambda_1 t) + (1-p)\exp(-\lambda_2 t)\right]^{1-D_i} \tag{14.a}$$

Pemaksimuman fungsi di atas terhadap parameter akan mendapatkan solusi $\hat{\lambda}_y = \dfrac{n-\sum D_i}{\sum Z_i}$. Untuk mendapatkan taksiran parameter yang lain diperlukan metode numerik.