ESTIMASI FUNGSI HAZARD DENGAN MENGGUNAKAN WAVELET

Oleh:
Fitriani Agustina
Bambang Avip Priatna Martadiputra
Jurusan Pendidikan Matematika FPMUPA UPI

ABSTRAK

Pada akhir-akhir ini dikalangan ahli-ahli statistik sedang ramai dibicarakan suatu *teknik baru untuk inferensi nonparametrik* yang dikenal dengan sebutan *teori wavelet*. Penggunaan wavelet dalam model regresi nonparametrik telah dilakukan oleh Antoniadis (1994) dengan mengunakan kernel wavelet $E_m(.,.)$. Selanjutnya dengan menunjukkan bahwa estimator wavelet analog dengan estimator deret ortogonal dan estimator kernel yang sudah dikenal serta dengan menggunakan asumsi yang sama untuk h seperti yang digunakan untuk r, Antoniadis (1994) mendefinisikan estimator wavelet untuk fungsi hazard sebagai

$$\hat{h}(t) = \int_0^1 E_m(t,s)\,d\hat{H}(s)$$

dengan $E_m(t,s) = \sum_{k\in Z} \phi_{m,k}(t)\phi_{m,k}(s)$

adalah kernel wavelet yang didefinisikan oleh Mayer (1990).
Hasil analisis lebih lanjut menunjukkan bahwa estimator wavelet untuk fungsi hazard bersifat takbias asimtotik, konsisten, dan $\sqrt{n2^{-m}}\left(\hat{h}_d(t) - \hat{h}(t)\right)$ normal asimtotik dengan rata-rata nol dan varians $(t)w_0^2(t)/\tau(t)$.

Kata kunci: estimator deret orthogonal**,** estimator kernel, estimator wavelet, dan fungsi hazard

---

1. Pendahuluan

Misalkan T adalah variabel random tunggal nonnegatif kontinu pada interval $[0,\infty)$ yang menunjukkan tahan hidup individu dalam suatu popuasi. Jika f(t) dan F(t) masing-masing adalah fungsi densitas dan fungsi distribusi dari T, maka *fungsi hazard* untuk T didefinisikan sebagai

$$h(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t S(t)} = \frac{F'(t)}{S(t)} = \frac{f(t)}{S(t)} \qquad \text{... (1.1)}$$

dengan S(t) = 1- F(t) adalah fungsi survival, dan F(t) < 1.

Untuk memperoleh data tahan hidup yang terbaik dari suatu populasi, dilakukan pengujian terhadap n benda pada kondisi (tegangan, stress) operasinya yang normal sampai semua benda mati (*sampel lengkap*) atau pengujian dihentikan sebelum semua benda mati (*sampel disensor*). Sensor dilakukan untuk memperpendek waktu atau memperkecil biaya pengujian, terutama bagi benda-benda yang cukup handal. Tahan hidup benda yang mati diamati dan digunakan untuk estimasi, misalnya estimasi fungsi hazard h.

Dalam estimasi h, beberapa teori nonparametrik berkonsentrasi pada fungsi hazard kumulatif (dalam Ramlau-Hansen (1983), h.455). Sebuah estimator alternatif untuk H yang lebih baik dalam tampilan untuk sampel berukuran kecil daripada estimator Kaplan-Meier yang diperoleh dengan menggunakan metode product limit estimasi, disebut estimator Nelson-Aalen untuk fungsi hazard kumulatif H yang diperoleh dengan menggunakan model intensitas multtiplikatif proses menghitung.

Selanjutnya dengan mengasumsikan bahwa model intestias multiplikatif Aalen berlaku untuk $t \in [0,1]$ serta dengan menggunakan fungsi-fungsi kernel untuk memperhalus estimator Nelson-Aalen untuk fungsi hazard kumulatif, Ramlau-Hansen (1983) mempelajari estimator kernel untuk fungsi hazard.

Pada akhir-akhir ini dikalangan ahli-ahli statistik sedang ramai dibicarakan suatu *teknik baru untuk inferensi nonparametrik* yang dikenal dengan sebutan *teori wavelet*. Penggunaan wavelet dalam model regresi nonparametrik telah dilakukan oleh Antoniadis (1994) dengan mengunakan kernel wavelet $E_m(.,.)$ yang didefinisikan oleh Mayer (1990). Selanjutnya Antoniadis menunjukkan bahwa estimator wavelet analog dengan estimator deret ortogonal dan estimator kernel yang sudah dikenal.

Permasalahan yang akan dikaji dalam tulisan ini adalah : (1) *bagaimana cara mengestimasi fungsi hazard dengan menggunakan wavelet*? Dengan kata lain, berdasarkan estimator kernel untuk fungsi hazard versi Ramlau-Hansen, akan dicari estimator wavelet untuk fungsi hazard h dengan menggunakan kernel wavelet, serta (2) *bagaimana sifat-sifat dari estimator yang diperoleh*?

2. Landasan Teori

2.1 Fungsi Survival, Fungsi Hazard, dan Fungsi Hazard Kumulatif

Misalkan T adalah variabel random tunggal nonnegatif kontinu pada interval $[0,\infty)$ yang menunjukkan tahan hidup individu dalam suatu popuasi. Jika f(t) dan F(t) masing-masing adalah fungsi densitas dan fungsi distribusi dari T maka selain fungsi hazard, dalam distribusi uji hidup didefinisikan beberapa fungsi-fungsi berikut.

Fungsi survival atau fungsi reliabilitas,

$$S(t) = \Pr(T > t) = \int_t^\infty f(x)dx = 1 - F(t) \qquad \text{... (2.1.1)}$$

Fungsi hazard kumulatif

$$H(t) = \int_0^t h(x)dx = -\log S(t) \qquad \text{... (2.1.2)}$$

2.2 Tipe-Tipe Penyensoran

Sensor dilakukan untuk memperpendek waktu atau memperkecil biaya pengujian, terutama bagi benda-benda yang cukup handal. Misalkan n benda diuji pada kondisi (tegangan, stress) operasinya yang normal, dalam *sensor tipe I* ; pengujian akan dihentikan jika telah dicapai waktu tertentu (waktu pensensoran). Dalam *sensor tipe II*, pengujian akan dihentikan setelah kegagalan atau kematian benda ke-r ($r<n$) diperoleh.

2.3 Sistem Ortonormal dari Fungsi-Fungsi

$L^2(R) = \left\{ f : R \to C; \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \right\}$ yang dilengkapi dengan hasilkali dalam dan norma yang didefinisikan sebagai

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx \quad \text{dan} \quad \|f\|_2 = \left\{ \int_{-\infty}^{\infty} |f(x)|^2 dx \right\}^{\frac{1}{2}}$$

adalah *ruang Hilbert*.

Dua fungsi f_1 dan $f_2 \in L^2(R)$ disebut *ortogonal* jika $<f_1,f_2> = 0$. Barisan fungsi $\{f_j\}$ disebut *ortonormal* jika $<f_i,f_j> = 0$ untuk setiap $i, j \in N$ dengan $i \neq j$ dan $\|f_j\|_2 = 1$ untuk setiap $j \in N$. Selanjutnya

$\{f_j\}$ disebut *sistem ortonormal lengkap (CONS)* jika $\{f_j\}$ ortonormal dan fungsi yang ortogonal pada setiap f_j hanyalah fungsi nol. Serta $\{f_j\}$ disebut *basis ortonormal* untuk $L^2(R)$ jika $\{f_j\}$ sistem ortonormal lengkap (CONS).

2.4 Teori Wavelet

Ogden (1997) mendefinisikan sebuah *wavelet* sebagai suatu *fungsi gelombang (wavy function) yang disederhanakan dan dikonstruksi dengan hati-hati sehingga memiliki sifat-sifat matematika yang dapat dipercaya*. Ide dari pengkontruksian wavelet adalah bagaimana memilih suatu fungsi tunggal ψ (yang disebut wavelet) kemudian dibentuk keluarga fungsi $\{\psi_{j,k}\}$ yang diperoleh dengan cara mendelasi dan mentranslasi ψ sehingga akan didapat basis di $L^2(R)$. Dengan kata lain $\{\psi_{j,k}\}$ membangun $L^2(R)$ dan setiap $f \in L^2(R)$ dapat dinyatakan sebagai kombinasi linear dari $\psi_{j,k}$.

Suatu fungsi $\psi \in L^2(R)$ disebut *wavelet ortogonal* jika keluarga fungsi $\{\psi_{j,k}\}$ yang didefinisikan sebagai $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x\text{-}k)$ merupakan basis ortonormal untuk $L^2(R)$. Keluarga ruang bagian tertutup $\{V_j\text{: } j \in Z\}$ dari $L^2(R)$ yang memenuhi sifat : inklusi, maksimalitas, densiti, skala, dan basis disebut *Analisis Multiresolusi (AMR)* pada $L^2(R)$.

Sedangkan ϕ yang memenuhi sifat basis disebut *fungsi skala* pada AMR. Akibatnya jika $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x\text{-}k)$, $j,k \in Z$ maka untuk setiap $j \in Z$, $\{\phi_{j,k}(x);\ k \in Z\}$ merupakan basis ortonormal untuk V_j. Jika $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x\text{-}k)$ dengan $j,k \in Z$ maka $\{\psi_{j,k} : j,k \in Z\}$ merupakan basis ortonormal untuk $L^2(R)$.

2.5 Kernel Wavelet $E_m(.\,,.)$

Misalkan $h \in L^2(R)$ dan V_j adalah subruang tertutup dari $L^2(R)$. Proyeksi dari h pada V_j dapat dituliskan sebagai

$$h \to E_j(h)(u) \sum_{k \in Z} c_{j,k}\phi_{j,k}(u) \quad ..(2.5.1)$$

dengan $c_{j,k} = \left\langle h, \phi_{j,k} \right\rangle = \int_R h(v)\phi_{j,k}(v)dv$

Misalkan diperkenalkan proyektor yang berhubungan dengan integral kernel, yaitu

$$h \to E_j(h)(u) = \int_R E_j(u,v)h(v)dv \quad ...(2.5.2)$$

yang juga menyatakan proyeksi dari h pada V_j .
Dari (2.5.1) dan (2.5.2) diperoleh

$$E_j(u,v) = \sum_{k \in Z} \phi_{j,k}(u)\phi_{j,k}(v) \quad ...(2.5.3)$$

yaitu sebuah fungsi yang didefinisikan oleh Mayer (1990) sebagai **kernel wavelet.**
Karena $\phi_{j,k}(u) = 2^{j/2}\ \phi(2^j u - k)$ dan $\phi_{j,k}(v) = 2^{j/2}\ \phi(2^j v - k)$ maka kernel wavelet (2.5.3) dapat dinyatakan dalam bentuk

$$E_j(u,v) = 2^j \sum_{k \in Z} \phi(2^j u - k)\phi(2^j v - k) \quad ...(2.5.4)$$

2.6 Proses Menghitung (Counting Process)

Proses menghitung adalah suatu pendekatan alternatif untuk mengembangkan prosedur inferensi data uji hidup tersensor. Dalam proses menghitung didefinisikan proses intensitas untuk proses menghitung; proses intensitas kumulatif; martingale proses menghitung; martingale umum; serta integral stokastik.

2.7 Model Intensitas Multiplikatif Aalen

Model intensitas multiplikatif proses menghitung diberikan oleh

$$\lambda(t) = h(t)Y(t), \text{ untuk } t \in [0,1] \qquad \text{... (2.7.1)}$$

dengan

h(.) adalah fungsi nonstokastik tak diketahui, disebut *fungsi intensitas (fungsi hazard)*

Y(.) adalah proses stokastik terobservasi.

Jika digunakan teori-teori integral stokastik, martingale dan proses menghitung maka berdasarkan model (2.7.1) akan diperoleh :

Estimator Nelson-Aalen untuk H(t), yaitu

$$\hat{H}(t) = \int_0^t \left(\frac{J(s)}{Y(s)} \right) dN(s) \qquad \text{... (2.7.2)}$$

Integral stokastik untuk proses J(u)/Y(u) yang dapat diramalkan berkenaan dengan suatu matringale, yaitu

$$W(t) = \int_0^t \left(\frac{J(s)}{Y(s)} \right) dM(s) \qquad \text{... (2.7.3)}$$

W(t) juga martingale.

Besaran random untuk data tersensor yang sama dengan H(t) dalam suatu range dimana data diperoleh dengan mengabaikan W(t), yaitu

$$H^*(t) = \int_0^t [J(s)h(s)]d(s) \qquad \text{... (2.7.4)}$$

Perlu dicatat bahwa H(t) adalah *estimator nonparametrik untuk besaran random* H*(t).

3. Estimasi Fungsi Hazard dengan Menggunakan Fungsi Kernel

Untuk memecahkan masalah pertama, dilakukan dengan mempelajari hasil pekerjaan Aalen (1978) yang menunjukkan bahwa estimasi h untuk data uji hidup tersensor dapat digambarkan dalam suatu konteks inferensi untuk suatu model intensitas multiplikatif proses menghitung (N) yang diberikan oleh

$$\lambda(t) = h(t)Y(t) \qquad \text{… (3.1)}$$

dengan Y(t) adalah proses stokastik terobservasi.

Berdasarkan model (3.1), Aalen memperoleh estimator fungsi hazard kumulatif H yang selanjutnya disebut estimator Nelson-Aalen. Estimator ini merupakan penyempurnaan untuk estimator fungsi hazard kumulatif yang ditemukan oleh Nelson dalam kontek reliabilitas yang lebih baik penampilannya untuk sampel berukuran kecil daripada estimator yang diperoleh dengan menggunakan metode Kaplan-Meier (product limit estimator).

Setelah mempelajari estimator Nelson-Aalen, kemudian dilanjutkan dengan mempelajari estimator kernel untuk fungsi hazard versi Ramlau-Hansen. Berdasarkan estimator Nelson-Aalen dan mentransformasikan waktu-waktu kegagalan atau kematian kedalam interval satuan [0,1], Ramlau-Hansen (1983), menemukan suatu metode estimasi fungsi hazard proses menghitung dengan menggunakan fungsi-fungsi kernel, yaitu

$$\hat{h}(t) = \frac{1}{b} \int_0^1 K\left(\frac{t-s}{b} \right) d\hat{H}(s) \qquad \text{… (3.2)}$$

dengan K(.) adalah fungsi tebatas dengan integral 1

b adalah sebuah parameter positif (bandwidth)

$\hat{H}(t) = \int_0^1 \left(\frac{J(s)}{Y(s)} \right) dN(s)$ adalah estimator Nelson-Aalen untuk H … (3.3)

4. Estimasi Fungsi Hazard dengan Menggunakan Wavelet

Sebelum memperoleh estimator wavelet untuk fungsi hazard, perlu dipelajari lebih dahulu kernel-wavelet $E_J(.\,,.)$ yang didefinsikan oleh Mayer (1990), yaitu

$$E_j(u,v)=\sum_{k\in Z}\phi_{j,k}(u)\phi_{j,k}(v) \qquad \dots (4.1)$$

dengan $\phi_{j,k}(x) = 2^{j/2}\,\phi(2^j x - k)$, $k\in Z$ adalah fungsi skala yang didilasi dan ditranslasi untuk keluarga wavelet tertentu.

Ternyata, kernel-wavelet tersebut dapat digunakan untuk mengestimasi fungsi regresi r yang dinyatakan dalam model regresi nonparametrik standar. Setelah mempelajari hubungan antara estimator deret ortogonal, estimator kernel, dan estimator wavelet untuk r, diketahui bahwa estimator wavelet analog dengan estimator deret ortogonal dan estimator kernel yang sudah dikenal. Hubungan antara ketiga estimator tersebut telah dipelajari oleh Antoniadis(1994) dan Ogden (1997).

Berdasarkan kenyataan bahwa estimator deret ortogonal wavelet ekuivalen dengan estimator kernel berdasarkan wavelet, serta dengan menggunakan asumsi-asumsi yang sama pada fungsi hazard h seperti yang digunakan pada fungsi regresi nonparametrik r, **Antoniadis (1994)** mendefinisikan estimator kernel untuk fungsi hazard versi Ramlau-Hansen (1983) dalam kernel-wavelet, yaitu sebagai berikut.

Definisi 4.1

Misalkan $E_m(.\,,.)$ adalah kernel-wavelet, yaitu proyeksi dari fungsi hazard h pada subruang V_m dari $L^2(R)$. Berdasarkan estimator kernel untuk fungsi hazard versi Ramlau-Hansen, estimator wavelet untuk fungsi hazard didefinisikan sebagai

$$\hat{h}(t)=\int_0^1 E_m(t,s)dH(s) \qquad \dots (4.2)$$

Jika diberikan suatu barisan proses menghitung dimensi satu apabila populasi yang dipelajari terdiri dari n individu $\{N_n(t)\}$ dengan setiap proses intensitasnya berbentuk $\lambda_n(t) = h(t)Y_n(t)$ maka barisan dari estimator wavelet yang berkorespondensi dinyatakan dengan $h_n(t)$, yaitu

$$\hat{h}_n(t)=\int_0^1 E_m(t,s)dH_n(s) \qquad \dots (4.3)$$

dengan

$$\hat{H}_n(t)=\int_0^1\left(\frac{J_n(s)}{Y_n(s)}\right)dN_n(s) \qquad \dots (4.4)$$

adalah estimator Nelson-Aalen untuk fungsi hazard kumulatif $H_n(t)$ dan

$$J_n(s) = I(Y_n(s) > 0) \text{ dan } Y_n(t)=\sum_{j=1}^{n} I(T_j\ge t, T_j\ge C_j) \qquad \dots (4.5)$$

5. Sifat-sifat Estimator Wavelet untuk Fungsi Hazard

Sifat-sifat dari estimator wavelet untuk fungsi hazard disajikan dalam bentuk teorema-teorema berikut.

Teorema 5.1

Misalkan didefinisikan $\delta_n=\operatorname*{Sup}_{0<s<1}E[1-J_n(s)]$. Jika $\hat{h}_n(t)=\int_0^1 E_m(t,s)d\hat{H}_n(s)$ adalah estimator wavelet untuk fungsi hazard h(t) maka bias dari h(t) dinyatakan sebagai

$$E\left[\hat{h}(t)-h(t)\right]=O(\eta_m)+2^{\frac{m}{2}}O\left(\delta_n^{\frac{1}{2}}\right) \qquad \dots (5.1)$$

dengan

$$O(\eta_m) = \begin{cases} (1/2^m)^{\upsilon - 1/2} & \text{, jika } 1/2 < \upsilon < 3/2 \\ \sqrt{m} / 2^m & \text{, jika } \upsilon = 3/2 \\ 1/2^m & \text{, jika } \upsilon > 3/2 \end{cases} \qquad \dots (5.2)$$

Dari teorema 5.1 diperoleh sifat pertama dari estimator wavelet untuk fungsi hazard, yaitu takbias asimtotis.

Teorema 5.2

Misalkan $h_d(t) = h(t^{(m)})$ adalah suatu aproksimasi untuk h berdasarkan pada nilai-nilai h pada titik-titik diadik berordo m.

Jika (1) $nE(J_n(t)/Y_n(t)) \to 1/\tau(t)$ secara seragam dalam persekitaran dari t dan

(2) h dan τ kontinu pada titik t,

maka MSE dari $h_d(t)$ dapat dinyatakan sebagai

$$E\left[\hat{h}_d(t) - h(t)\right]^2 = \frac{2^m}{n}\frac{h(t)}{\tau(t)} w_0^2 + 2^m o\left(n^{-1}\right) + O\left(\left[\eta_m\right]^2\right) + O\left(2^{-m\gamma}\right) \qquad \dots (5.3)$$

dengan $w_0^2 = \int_0^1 E_0^2(0,s)ds = \sum_{k \in Z} \phi^2(k)$ … (5.4)

Dari teorema 5.2 diperoleh sifat kedua dari estimator wavelet untuk fungsi hazard, yaitu konsisten.

Teorema 5.3

Jika (1) $nE(J_n(t)/Y_n(t)) \to 1/\tau(t)$ secara seragam dalam persekitaran dari t dan (2) h dan τ kontinu pada titik t, maka $\sqrt{n2^{-m}}\left[\hat{h}_d(t) - h(t)\right]$ adalah normal asimtotik dengan mean 0 dan varians $h(t)w_0^2/\tau(t)$ bila $n2^{-m} \to \infty$, $n2^{-m\upsilon} \to 0$, dan $\delta_n = o(1/n)$. ... (4.8)

Dari teorema 4.3 diperoleh sifat kedua dari estimator wavelet untuk fungsi hazard, yaitu normal asimtotik.

DAFTAR PUSTAKA

Aalen, O. O. (1978).“Nonparametric Inference for a Family of Counting Processes”. *The Annal of Statistics, 6, 701-726*

Antoniadis, A., Gregoire, G., Mc.Keague, W. (1994). “Wavelets Methods for Curve Estimation”. *Journal of the American Statistical Association, Dec, Vol.84, No.428, 1340-1353.*

Bain, L. J. (1992). *“Introduction to Probability and Mathematical Statistics”*. Belmont, California : Duxbury Press.

Bruce, A., Gao, H. Y. (1996). *”Applied Wavelet Analysis with S-Plus”*. New York: Springer - Verlag.

Chui, K. (1992). *“An Introduction to Wavelet”*. Boston : Academic Press, Inc.

Cox, D. R., Oakes, D. (1992). *“Analysis of Survival Data”*. New York : Chapman & Hall.

Eubank, R. L. (1988). *“Spline Smoothing and Nonparametric Regression”*. New York: Marcel Dekker.Inc.

Gupta, V. P. (1986). *“Lebesgue Measure and Integration”*. New Delhi : Wiley Eastern Limited.

Hardle, W. (1991). *“Smoothing Techniques with Implementation in S”*. New York: Springer-Verlag.

Issogi, E. (1990). “Nonparametric Estimation of a Regressi Function by Delta Sequences”. *Ann. Inst. Statist. Math., 42, 699-708.*

Kaiser, G. (1994). *"A Friendly Guide to Wavelets".* Boston : Birkhauser.
Klein, J. P., Moeschberger, M. L. (1997). *"Survival Analysis: Techniques for Censored and Truncated Data".* New York : Springer-Verlag.
Lawless, J. F. (1982). *"Statistical Models and Method for Life Time Data".* New York: John Willey & Sons.
Liptser, R. Sh., Shiryayev, A. N. (1980). "A Functional Central Limit Theorem for Semimartingaes". *Theorm. Probab. Appl. 25, 4, 667-688.*
Mallat, S. (1989). "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(R)$". *Transactions of the American Mathematical Society, 315, 69-87.*
MathSoft. (1993). *"User's Manual S-Plus, version 3.2".* Seattle : Stat Sci, a divition of MathSoft, Inc.
Mayer, Y. (1990). *"Ondelettes et Operateurs I: Ondelettes".* Paris : Hermann.
Ogden, T.R. (1997). *"Essensial Wavelets for Statistical Application and Data Analysis".* Boston : Birkhauser.
Ramlau-Hansen, H. (1983). "Smoothing Counting Processes by Mean of Kernel Function". *The Annals of Statistics, 11, 453-466.*
Schomburg, B. (1990). "On the Approximation of the Delta Distribution in Sobolev Spaces of Negative Order". *Applicable Analysis, 36, 89-93.*
Serfling, R.J. (1990). *"Approximation Theorems of Mathematical Statistics".* New York: John Willey & Sons.
Soejoeti, Z. (1996). "Suatu Studi Tentang Uji Hidup Dipercepat Tegangan Bertingkat : Perkembangan Mutakhir". *Medan: Makalah pada Konfrensi Matematika IX*.