Jurnal MIPA 40 (2) (2017): 141-146

Jurnal MIPA

http://journal.unnes.ac.id/nju/index.php/JM

Estimasi Distribusi Mixing dalam Model Mixture Poisson

N Dwidayati ✉

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia

Info Artikel

Sejarah Artikel:
Diterima 11 Juli 2017
Disetujui 23 September 2017
Dipublikasikan 1 Oktober 2017

Keywords:
mixing distribution. Poisson mixture model, Laplace inversion, IMSE, convergence

Abstrak

Model *mixture* telah dikenalkan sejak tahun 1986 (Farewell, 1986) dengan menganalisis penderita kanker payudara. Model ini diaplikasikan dalam ruang kerja yang berbeda sesuai dengan fleksibilitas pandangan kompleksnya situasi, walaupun generalisasi teori belum sepenuhnya dikembangkan. Studi klinik yang dilakukan difokuskan pada estimasi proporsi pasien yang sembuh dan distribusi *failure time* pasien yang tidak sembuh. Pada penelitian ini dikonstruksi estimator konsisten dari distribusi *mixing* dalam model *mixture* Poisson melalui inversi Laplace, baik untuk data tak tersensor maunpun tersensor. Berdasar estimasi tersebut ditentukan IMSE dan laju kekonvergenan yang berkorespondensi dengan inverse estimator.

Abstract

The mixture model has been introduced since 1986 (Farewell, 1986) by analyzing breast cancer patients. This model was applied in different workspaces according to the flexibility of the complex view of the situation, although the generalization of the theory has not been fully developed. Clinical studies undertaken focused on estimating the proportion of patients who recovered and the distribution of patient failure times that did not heal. In this research constructed a consistent estimator of the mixing distribution in the Poisson mixture model through Laplace inversion, both for uncensored and uncensored data. Based on these estimates, the IMSE and the convergence rate correspond to the inverse estimator.

✉ Alamat korespondensi:
E-mail: nurkaromah.mat@mail.unnes.ac.id

ISSN 0215-9945

PENDAHULUAN

Sejarah mengenai problem estimasi distribusi *mixing* dalam model *mixture* cukup panjang. Tercatat Robbins tahun 1956 (Klaassen and Mnatsakanov, 2003) yang pertama menggagas problematika tersebut. Gagasannya dimulai dengan mengkonstruksi barisan variabel random independen $X_1, X_2, \ldots$ yang dibangkitkan dari barisan densitas $f(.,\lambda_1), f(.,\lambda_1), \ldots$

Misalkan parameter tak diketahui, tak terobservasi, $\lambda_1, \lambda_2, \ldots$ i.i.d (*independent and identically distributed*) dengan distribusi G tak diketahui, maka $X_1, X_2, \ldots$ juga i.i.d dan densitas marginal $p(.)$ dari X_i adalah *mixture* dari $f(.,.)$ atas $G(.)$. Dengan kata lain

$$p(x) = f(x\,|\,G) = \int f(x,\lambda)\,dG(\lambda)$$

Distribusi G disebut distribusi *mixing*. Jadi dipunyai observasi yang tidak lengkap, dimana $\lambda_1, \lambda_2, \ldots$ tidak terobservasi dan akan diestimasi distribusi *mixing* yang tidak diketahui, hanya diberikan observasi $X_1, X_2, \ldots X_1, X_2, \ldots$ dari $f(.\,|\,G)$.

Ada beberapa tipe pendekatan yang telah dikenalkan untuk mengestimasi distribusi *mixing* . Pada metode nonparametrik dapat dilihat pada (1) Leonard (1970), dan Meeden (1993) melalui studi estimator tipe Bayes, (2) Simar (1976), Laird (1978), Redner (1981), Jewel (1982), dan Pfanzagl (1988) memperkenalkan MLE (*Maximum Likelihood Estimator*) untuk model *mixture,* (3) Lindsay (1989) yang mengkonstruksi estimator konsisten untuk *G* melalui estimasi dengan metode momen, serta (4) Loh & Zhang (1996) dan Fan (1991) memperkenalkan tipe kernel dari estimator berdasar transformasi Fourier. Masalah penelitian difokuskan pada rekonstruksi estimator konsisten dari distribusi *mixing* dalam model *mixture* Poisson melalui inversi Laplace.

METODE

Penelitian ini didesain dengan menggunakan pendekatan deduktif-analitis. Hasil penelitian pakar yang disajikan dalam jurnal internasional bereputasi, *proceeding* seminar, maupun *text book* dikaji secara analitik. Demikian pula dengan berbagai bahan yang dapat diakses dari internet maupun korespondensi secara langsung. Fokus kajian pada definisi, asumsi dan teorema-teorema yang telah ada.

Estimator konsisten direkonstruksi dari distribusi *mixing* dalam model *mixture* Poisson melalui inversi Laplace, baik untuk data tak tersensor maunpun tersensor. Berdasar estimasi tersebut ditentukan IMSE dan laju kekonvergenan yang berkorespondensi dengan inverse estimator.

HASIL DAN PEMBAHASAN

1. Konstruksi Transformasi Invers

Persamaan Fredholm : $KG = p$

dengan:

$$(KG)(x) = \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} dG(\lambda).\ x = 0,1,\ldots \quad \ldots(1)$$

menyatakan distribusi *mixture* Poisson. Akan dikonstruksi estimator dari distribusi *mixing* *G* yang tak diketahui berdasar inversi Laplace (Feller, 1971) yang dapat ditulis sebagai:

$$K_\alpha^{-1} p)(z) = \sum_{k \le \alpha z} \frac{\alpha^k}{k!} \sum_{j=k}^{\infty} (-\alpha)^{j-k} \frac{1}{(j-k)!} (Qp)(j) \quad \ldots(2)$$

Transformasi Q diberikan oleh:

$$(Qp)(j) = \sum_{x=j}^{\infty} x^j p(x);\ j = 0,1,2,\ldots \quad \ldots(3)$$

dengan: $x^{(j)} = x!/(x-j)!$

Lemma 1

Untuk α menuju K_α^{-1} dari (2) menyatakan inversi *K* dari (1),

$$K_\alpha^{-1} KG \xrightarrow{w} G \quad \text{untuk } \alpha \to \infty$$

Bukti:

Berdasar Fubini:

$$(QKG)(j) = \int_0^\infty \lambda^j \, dG(\lambda)$$

Berdasar ketaksamaan Chebyshev untuk variabel random Poisson diperoleh:

$$\left|\sum_{k\le \alpha z} e^{-\alpha\lambda}\frac{(\alpha\lambda)^k}{k!} - I_{|\lambda\le z|}\right| \le \left(\frac{\lambda}{\alpha(z-\lambda)^2}\right)\wedge I$$

Akibatnya pada sebarang titik kontinu z dari fungsi distribusi G dipunyai:

$$\left(K_\alpha^{-1}KG\right)(z) = \int_0^\infty \sum_{k\le\alpha z}\frac{(\alpha\lambda)^k}{k!}\sum_{j=k}^{\infty}(-\alpha\lambda)^{j-k}\frac{1}{(j-k)!}dG(\lambda)$$

$$= \int_0^\infty \sum_{k\le \alpha z} e^{-\alpha\lambda}\frac{(\alpha\lambda)^k}{k!}dG(\lambda) \to G(z)$$

untuk $\alpha \to \infty$ (terbukti)

Transformasi K_α^{-1} mungkin juga bisa ditulis sebagai berikut.

$$K_\alpha^{-1}p)(z) = \sum_{k\le\alpha z}\frac{\alpha^k}{k!}\sum_{x=k}^{\infty}(1-\alpha)^{x-k}\frac{x!}{(x-k)!}p(x)$$

...(4)

Estimasi distribusi *mixture* $KG = p$ dari observasi dan aplikasi (4) serta lemma 1 akan diperoleh estimator konsisten dari distribusi *mixture G.* Hal ini diverifikasi untuk kasus variabel random tak tersensor i.i.d serta di bawah penyensoran random akan diuraikan di bawah ini.

2. Estimasi Distribusi *Mixing* Dalam Model *Mixture* Poisson dengan Data Tak-tersensor

Misalkan $X, X_1, X_2, \ldots, X_n$ variabel random i.id dengan distribusi (1) maka:

$$p(x) = P(X = x) = \int_0^\infty e^{-\lambda}\frac{\lambda^x}{x!}dG(\lambda) = (KG)(x), \quad x = 0,1,\ldots$$

...(5)

Penggantian distribusi marginal $p(x) = P(X = x)$ dalam (4) dengan versi empirik yang berkorespondensi

$$\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{|X_i = x|}$$

dan ambil $\alpha = \alpha_n$ bergantung pada n, maka diperoleh estimator dari G sebagai berikut.

$$\hat{G}_n(z) = \sum_{k\le\alpha z}\frac{\alpha_n^k}{k!}\sum_{x=k}^{\infty}(1-\alpha_n)^{x-k}\frac{x!}{(x-k)!}\hat{p}_n(x),\ z \ge 0$$

...(6)

serupa dengan lemma 1 yang akan konsisten jika dipilih α_n yang cenderung infinit. Estimator $\hat{G}_n(z)$ juga dapat dinyatakan sebagai berikut.

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} B_n(z, X_i) \text{ dengan}$$

$$B_n(z,x) = \sum_{k=0}^{[\alpha_n z]\wedge x}\binom{x}{k}\alpha_n{}^k(1-\alpha_n)^{x-k} \quad ...(7)$$

Dalam hal ini [y] menyatakan bagian integer dari y.

Akan ditentukan IMSE (*Integrated Mean Square Error*)

$$\int_0^\infty E(\hat{G}_n(z) - G(z))^2\, dG(z)$$

Tulis

$$E(\hat{G}_n(z) - G(z))^2 = \operatorname{var}\hat{G}_n(z) + [E\hat{G}_n(z) - G(z)]^2$$

...(8)

dan $\operatorname{var}\hat{G}_n(z) = \dfrac{1}{n}\operatorname{var} B_n(z, X)$

...(9)

Pilih variabel random Λ dengan fungsi distribusi G, dalam hal ini distribusi bersyarat dari Λ adalah Poisson (Λ), sehingga:

$$P(X = x \mid \Lambda = \lambda) = e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0,1,\ldots$$

Momen kedua bersyarat dapat diestimasi sebagai berikut.

$$E\left(B_n^2(z, X) \mid \Lambda = \lambda\right) \le \sum_{x=0}^{\infty}\left(\sum_{k=0}^{x}\binom{x}{k}\alpha_n{}^k \mid I - \alpha_n\mid^{x-k}\right)^2 e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty}\left(\alpha_n + \mid I - \alpha_n\mid\right)^{2x} e^{-\lambda}\frac{\lambda^x}{x!} = \exp\left(\left\{(\alpha_n + \mid I - \alpha_n\mid)^2 - I\right\}\lambda\right)$$

...(10)

Karena $\alpha_n \ge 1$ untuk n besar dipunyai:

$$\operatorname{var} B_n(z,X) \le \int_0^\infty \exp\left(4\alpha_n(\alpha_n-1)\lambda\right)dG(\lambda)$$

...(11)

Selanjutnya dipunyai

$$E\left(B_n(z,X)\,|\,\Lambda=\lambda\right) \le \sum_{x=0}^{\infty}\sum_{k=0}^{[\alpha_n z]\wedge x}\binom{x}{k}\alpha_n{}^k(1-\alpha_n)^{x-k}e^{-\lambda}\frac{\lambda^x}{x!}$$

$$=\sum_{k=0}^{[\alpha_n z]}e^{-\lambda}\frac{(\alpha_n\lambda)^k}{k!}\sum_{x=k}^{\infty}\frac{1}{(x-k)!}((1-\alpha_n)\lambda)^{x-k}$$

...(12)

$$=\sum_{k=0}^{[\alpha_n z]}e^{-\lambda}\frac{(\alpha_n\lambda)^k}{k!}=P\left(U\alpha_{n\lambda}\le\alpha_n z\right)$$

dengan U_n berdistribusi Poisson (μ). Akibatnya, bias dari $\hat{G}_n$ sama dengan:

$$E[\hat{G}_n(z)-G(z)]=\int_0^\infty (P[U_{\alpha_n\lambda}\le\alpha_n z]-I_{|\lambda\le z|})dG(\lambda)$$

...(13)

Lemma 2

U_n berdistribusi Poisson (μ) dengan $\mu>0$, maka dipunyai

$$|\,P(U_\mu\le y)-I_{|\mu\le y|}\,|\le \exp(y-\mu+y\ln\frac{\mu}{y}),\ y\ge 0$$

...(14)

Lemma 3

Fungsi $\lambda\to\Psi_z(\lambda)=z-\lambda+z\ln(\lambda/z)$ *strictly concave* pada $(0,\infty)$ dengan maksimum sama dengan 0 tercapai pada $\lambda=z$

Proposisi

Ambil $\hat{G}_n$ seperti pada (6) dengan $\alpha_n\ge 1$ maka:

$$E(\hat{G}_n(z)-G(z))^2\le\int_0^\infty\left(\frac{1}{n}\exp(4\alpha_n(\alpha_n-1)\lambda+\exp(2\alpha_n\Psi_z(\lambda))\right)dG(\lambda)$$

...(15)

Berlaku untuk $\Psi_z(\lambda)=z-\lambda+z\ln(\lambda/z),\ \lambda,z>0$

Teorema 1

Jika $P_G(\Lambda\le D)=1$ dan *G* mempunyai densitas *g* yang terbatas dengan *C*, maka IMSE dari $\hat{G}_n$ memenuhi

$$\int_0^\infty E(\hat{G}_n(z)-G(z))^2\,dG(z)\le 4C\sqrt{D}\,(\ln n)^{-1/2}(1+o(1))$$

....(16)

Untuk $n\to\infty$ jika dipilih α_n akan memenuhi

$$S\alpha_n^2 D=2\ln n+\ln(8C^2D)-\ln\ln(8C^2Dn^2)$$

....(17)

3. Estimasi Distribusi *Mixing* Pada Model *Mixture* Poisson Di Bawah Penyensoran Random

Misalkan $X, X_1, X_2, \ldots, X_n$ variabel random i.id dengan distribusi

$$p(x)=P(X=x)$$

sebagaimana pada (1) dan $Y, Y_1, Y_2, \ldots Y_n$ variabel random i.i.d dengan distribusi seperti suatu fungsi distribusi H. Diasumsikan bahwa *X* dan *Y* independen dan satu observasi $Z_i=\min\{X_i,Y_i\}$ dan $\Delta_i=I_{|x_i\le y_i|}$. Akan diestimasi fungsi distribusi *mixing* tak diketahui *G* dalam model penyensoran random.

Untuk kasus pertama, diambil fungsi distribusi penyensoran *H* yang diketahui.

$$p_1(x0=P(Z_i=x,\Delta_i=1)=P(X_i=x,X_i\le Y_i)$$

$$=P(X_i=x)(1-H(x-))=p(x)(1-H(x-)),\ x=0,1,\ldots$$

...(18)

Digunakan observasi Z_i dan Δ_i untuk mengestimasi $p(x)$ dengan ekspresi empirik sebagai berikut.

$$\hat{p}_n(x) = \frac{\hat{p}_{1n}(x)}{1 - H(x-)}$$ dengan

$$\hat{p}_{1n}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{|Z_i = x, \Delta_i = 1|}$$

...(19)

Selanjutnya akan dikonstruksi estimator dari fungsi distribusi *mixing* G yang tak diketahui dengan transformasi invers dari (4) dengan $\hat{p}_n$, diperoleh:

$$\hat{G}_n(z) = \sum_{k \le \alpha z} \frac{\alpha_n^k}{k!} \sum_{x=k}^{\infty} (1-\alpha_n)^{x-k} \frac{x!}{(x-k)!} \hat{p}_n(x),\ z \ge 0$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} \sum_{k \le \alpha z} \frac{\alpha_n^k}{k!} \sum_{x=k}^{\infty} (1-\alpha_n)^{x-k} \frac{x!}{(x-k)!} I_{|Z_i = x} |$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} \sum_{k=0}^{|\alpha_n z| \Lambda Z_i} \binom{Z_i}{k} \alpha_n{}^k (1-\alpha_n)^{Z_i - k}$$

...(20)

Jadi diperoleh estimator dalam bentuk:

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)} B_n(z, Z_i)$$

...(21)

dengan $B_n(z,x)$ sebagaimana didefinisikan pada (7).

Jika variabel random penyensoran Y_i mempunyai dukungan terbatas, maka untuk $\alpha_n z$ besar, estimator direduksi ke: $\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{1 - H(Z_i -)}$, dimana berdasar hukum bilangan besar akan konvergen ke 1.

Diasumsikan bahwa terdapat konstanta finit $\beta > 1$ dengan

$$P[Y > x] = 1 - H(x-) \ge \beta^{-x},\ x > 0$$

...(22)

Berdasar argumen yang sama dalam teorema 1 dapat diperoleh konsistensi $\hat{G}_n$ dari (20) dan (21) pada laju yang sama seperti dalam kasus tak tersensor.

Teorema 2

Jika terdapat konstanta finit C dan D sehingga distribusi *mixing* G mempunyai densitas yang terbatas pada C dan termuat dalam $[0,D]$ dan distribusi penyensoran H memenuhi asumsi di atas maka IMSE dari $\hat{G}_n$ adalah order $(\log n)^{-1/2}$

$$\int_0^{\infty} E(\hat{G}_n(z) - G(z))^2\, dG(z) \le 4C\sqrt{\beta D}\,(\log n)^{-1/2}(1 + o(1)$$

...(23)

Untuk $n \to \infty$ jika dipilih α_n akan memenuhi

$$S\alpha_n^2 \beta D = 2\log n + \log(8\beta C^2 D) - \log\log(8\beta C^2 D n^2)$$

Selanjutnya akan dikonstruksi untuk kasus dengan fungsi distribusi H. Dari (19) $p(.)$ dapat diestimasi dengan:

$$\hat{p}_n(x) = \frac{\hat{p}_{1n}(x)}{S_n(x-)}$$

...(24)

dimana S_n estimator Kaplan_Meier dari fungsi survival $S(x) = 1 - H(X)$

Didefinisikan:

$$\hat{S}_n(x) = \begin{cases} 1 & ,0 \le x \le Z_{(1)} \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1} \right)^{i - \Lambda_i}, & Z_{(k-1)} < x \le Z_{(k)}, k = 2,\ldots,n \\ 0 & ,Z_{(n)} < x \end{cases}$$

...(25)

Dalam hal ini $Z_{(i)}$ dan Λ_i menyatakan Z_i terurut yang berkorespondensi dengan Λ_i.

Berdasar (25) dengan menggunakan formula transformasi invers diperoleh estimator G

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{S_n(Z_i -)} B_n(z, Z_i)$$

...(26)

Perhatikan bahwa penyebut pada (26) mungkin bernilai nol maka perlu ditambahkan konstanta positip (sebagai kesepakatan) ε_n sehingga diperoleh:

$$\hat{G}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{S_n(Z_i -) + \varepsilon_n} B_n(z, Z_i) \quad ...(27)$$

Teorema 3

Jika asumsi pada teorema 2 di atas dipenuhi dan diasumsikan bahwa $E_n Y$ finit serta terdapat konstanta positip finit γ maka

$$\liminf_{t\to\infty} e^{\gamma t}(1 - G(t)) > 1)$$

Jika dipilih

$$\alpha_n = \sqrt{\frac{4\log\log n - 5\log\log\log n}{8\beta^3 D}},\ \varepsilon_n = \frac{\log\log n}{\log n},\ n \geq 3$$

maka IMSE dari $\hat{G}_n$ dari (27) memenuhi:

$$\limsup_{n\to\infty} \sqrt{\log\log n}\int_0^\infty E\left(\hat{G}_n(z) - G(z)\right)^2 dG(z) < \infty$$

SIMPULAN

Berdasar uraian di atas maka telah dapat dikonstruksi estimator konsisten dari distribusi *mixing* dalam model *mixture* Poisson melalui inversi Laplace, baik untuk data tak tersensor maunpun tersensor. Berdasar estimasi tersebut ditentukan telah ditentukan IMSE dan laju kekonvergenan yang berkorespondensi dengan inverse estimator.

DAFTAR PUSTAKA

Fan, J. 1991. On The Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Ann.Statist* 19: 1257-1272

Farewell, V.T. 1986. Mixture models in survival analysis. Are they worth the risk ? *The Canadian Journal of Statistics* 14: 257-262.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications,*Vol.II. New York: Wiley

Jewel, NP. 1982. Mixtures of Exponential Distribution. *Ann Statist.* 10:479-484

Klaassen, C.A.J & Mnatsakanov, R.M. 2003. Estimating the Structural Distribution Function of Cell Probabilities. *Austrian J.Statist* 32: 85-98

Laird, NM. 1978. Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *J.A.S.A.* 73: 805-811

Leonard, T. 1970. Some Data-analytic Modifications to Bayes-Stein Estimation. *Ann.Inst.Statist. Math* Part A.36: 11-21

Lindsay, BG. 1989. Moment Matrics: Applications in Mixture. *Ann. Statist* 17: 722-740

Loh, WL & Zhang, CH. 1996. Global Properties of Kernel Estimators for Mixing Densities in Discrete Exponential Family Models. *Statis. Sinica.* 6: 561-578

Meeden, G. 1993. Noninformative Nonparametric Bayesian Estimation on Quantiles. *Statistics & Probability Letters* 16: 103-109

Pfanzagl, J. 1988. Consistency of Maximum Likelihood for Certain Nonparametric Families, in Particular: Mixture. *J. Stat. Plan. Inf.* 19: 137-158

Redner, R. 1981. Not on the Consistency of the Maximum Likelihood Estimate for Non Identifiable Distribution. *Ann. Stat.* 9: 225-228

Simar, L. 1976. Maximum Likelihood Estimation of a Compound Poisson Process. *Ann. Statist* 4: 1200-1209