Majalah Ilmiah Matematika dan Statistika Volume 16 Nomor 1 2016, 25 – 30
https://jurnal.unej.ac.id/index.php/MIMS/index ISSN 1411-6669

MODEL KISI BERSYARAT BEBAS UNTUK MENDUGA PELUANG SEL PADA TABEL KONTINGENSI TAK LENGKAP

(Model of Free Grading to Guess Cell Probability in Incomplete Contingency Table)

Yuliani Setia Dewi

Jurusan Matematika FMIPA Universitas Jember
Jl. Kalimantan 37, Jember 68121, Indonesia
Email: yulidewi.fmipa@unej.ac.id

Abstract: The problem of estimating cell probabilities from incomplete tables, that is when either the row or column variable missing for some of the subjects is very common. This article describes Lattice Conditional Independence Model to estimate cell probabilities in incomplete contingency table.

Keywords: incomplete contingency tables, Lattice Conditional Independence Model, likelihood function, MLE
MSC 2020: 62H17

1. Pendahuluan

Data tak lengkap pada tabel kontingensi sering terjadi dalam praktek. Pada tabel kontingensi dengan data tak lengkap, data diklasifikasikan ke dalam suatu tabel kontingensi multi arah dengan nilai dari semua variabel dicatat untuk sekumpulan contoh tertentu, sedangkan contoh sisanya mempunyai data hilang untuk satu atau lebih variabel. Ringkasan data dapat dinyatakan sebagai tabel yang mempunyai kategori penuh dengan beberapa tabel tambahan dengan kategori yang lebih rendah.

Dalam praktek, pendekatan umum untuk data hilang ini adalah :

1. Algoritma EM [4], [6] yaitu merupakan suatu algoritma iteratif. Metode ini memerlukan beberapa modifikasi dari sofware statistik yang ada untuk mendapatkan MLE.
2. Metode Complete Deletion (CD), yaitu dengan membuang observasi yang mengandung data hilang. Metode ini kurang efisien, khususnya apabila banyak observasi yang perlu dibuang. Di samping itu metode ini kemungkinan menghasilkan penduga parameter yang berbias [8].

Lipsitz, Parzen dan Molenberghs [5] memperkenalkan Model Linear Tergeneralisasi Poisson yang diaplikasikan untuk analisis tabel kontingensi tak lengkap R x C, metode ini tidak memerlukan tambahan pemrograman. Andersson dan Perlman [2] memperkenalkan Model Kisi Bersyarat Bebas untuk sebaran normal ganda, yang dapat diaplikasikan untuk analisis data hilang dengan pola tidak monoton, dan Model Kisi Bersyarat Bebas untuk tabel kontingensi tak lengkap [7]. Dalam tulisan ini akan dibahas mengenai Model Kisi Bersyarat Bebas untuk menduga peluang sel pada tabel kontingensi tak lengkap.

2. Hasil dan Pembahasan

2.1. Penarikan Contoh Poisson dan Multinomial

Misalkan kita mengobservasi jumlah $\{n_i, i=1,...,N\}$ dari N sel pada suatu tabel kontingensi, dalam hal ini bisa observasi untuk N level dari variabel kategorik tunggal atau untuk N = IJ sel pada tabel kontingensi dua arah. Jumlah yang diobservasi ini merupakan peubah acak. Masing-masing n_i merupakan bilangan integer tidak negatif dengan nilai harapan $m_i = E(n_i)$, nilai ini dinamakan sebagai frekuensi harapan.

Percobaan yang menghasilkan nilai-nilai bagi suatu peubah acak n_i, yaitu banyaknya hasil percobaan yang terjadi selama suatu selang waktu tertentu atau di suatu daerah tertentu , sering disebut percobaan Poisson. Bilangan n_i yang menyatakan banyaknya hasil percobaan dalam suatu percobaan Poisson disebut peubah acak Poisson dan sebaran peluangnya disebut sebaran Poisson [9]. Sebaran Poisson hanya tergantung pada parameter tunggal, yaitu nilai tengah (m_i). Fungsi kepekatan peluangnya adalah :

$$p(n_i; m_i) = \frac{\exp(-m_i) m_i^{n_i}}{n_i!} \qquad \text{untuk } n_i = 0,1,2,... \qquad (1)$$

Ragam $(n_i) = E(n_i) = m_i$.

Model penarikan contoh untuk jumlah $\{n_i\}$ mengasumsikan bahwa mereka merupakan peubah acak Poisson yang bebas. Fungsi peluang bersama untuk $\{n_i\}$ merupakan perkalian peluang dari persamaan (1) untuk N sel.

Ciri-ciri dari penarikan contoh Poisson adalah bahwa ukuran contoh total $n = \Sigma n_j$ adalah acak. Apabila kita memulai model Poisson tetapi dengan kondisi ukuran contoh total n, maka $\{n_i\}$ tidak menyebar menurut sebaran Poisson, karena masing-masing n_i tidak bisa lebih dari n . Apabila n diketahui, maka $\{n_i\}$ tidak bebas karena nilai yang satu akan mempengaruhi jumlah yang lain. Apabila $\Sigma n_j = n$, peluang bersyarat dari sekumpulan $\{n_i\}$ memenuhi kondisi berikut :

$$\mathrm{p}(n_i \text{observasi pada sel - } i, i = 1,..., N \mid \sum n_j = n)$$

$$= \frac{\mathrm{p}(n_i \text{observasi pada sel - } i, i = 1,..., N)}{\mathrm{p}(\sum n_j = n)} \qquad (2)$$

$$= \frac{\prod_i [\exp(-m_i) m_i / n_i!]}{\exp\left(-\sum m_j\right)\left(\sum m_j\right)^n / n!} = \left(\frac{n!}{\prod_i n_i!}\right) \prod_i \pi_i^{n_i}$$

dengan $\{\pi_i = m_i / (\sum m_j)\}$. Ini merupakan sebaran multinomial (n, $\{\pi_i\}$), yang dicirikan oleh ukuran contoh n dan peluang sel $\{\pi_i\}$. Sebaran Binomial (n, π_i) dengan indeks n dan peluang sukses π_1 merupakan kasus khusus dari Multinomial dengan N = 2 sel [1].

Majalah Ilmiah Matematika dan Statistika Volume 16 Nomor 1 2016, 25 – 30
https://jurnal.unej.ac.id/index.php/MIMS/index ISSN 1411-6669

2.2. Model Kisi Bersyarat Bebas

2.2.1. Pola Data Lengkap dan Tidak Lengkap

Misalkan (X_1, X_2, X_3) merupakan peubah kategori. Untuk contoh yang berukuan n, masing-masing observasi diklasifikasikan berdasarkan tiga kategori tersebut. Klasifikasi ini menghasilkan tabel kontingensi tiga arah jika datanya lengkap (jika nilai dari semua kategori dicatat untuk masing-masing observasi) (gambar 1(a)). Akan tetapi sering terjadi di dalam praktek, beberapa observasi hanya dapat diklasifikasikan secara parsial karena nilai dari satu atau lebih variabel hilang. Jadi pola data yang diobservasi mungkin seperti pada gambar 1(b) dan (c). Pola data pada gambar 1(a), 1(b) dan 1(c) dapat dinyatakan sebagai $\int_a = \{123\}$, $\int_b = \{1, 12, 123\}$ dan $\int_c = \{12, 13, 123\}$. Pola data $\int_a$, $\int_b$ dan $\int_c$ dinamakan pola data yang diobservasi [7].

Suatu pola data dinamakan monoton jika p kategori dapat dilabelkan kembali sehingga jika peubah ke-*i* hilang pada vektor x_j, maka peubah ke-i+1,…,p juga hilang pada x_j. Untuk selainnya pola data dinamakan tidak monoton [2]. Pola data tersebut dapat digambarkan sebagai berikut (untuk 3 kategori):

1 1 1 1 1 1	1 1 1 1 1 1	1 1 1 1 1 1
2 2 2 2 2 2	2 2 2 2	2 2 2 2
3 3 3 3 3 3	3 3	3 3 3 3
(a)	(b)	(c)
Lengkap	Monoton	Tidak monoton

Gambar 1.

2.2.2 Model Kisi Bersyarat Bebas untuk Tabel Kontingensi I Arah

Misalkan tabel kontingensi I arah dinyatakan sebagai $\zeta := x(J_i \mid i \in I)$, dengan I merupakan indeks sekumpulan kategori yang terbatas dan J_i adalah sekumpulan level pada kategori ke-i. Keluarga sebaran peluang positif pada ζ dinotasikan dengan P(I).

$$P(I) := \{P \equiv (p(x) \mid x \in \zeta) p(x) > 0, \sum p(x) = 1\}$$

P(I) adalah model penuh untuk tabel kontingensi ζ, yakni tidak ada batasan untuk P(I), kecuali peluang sel-selnya adalah positif dan berjumlah 1. Untuk beberapa sel $x \equiv (x_i \mid i \in I) \in \zeta$ dan beberapa subset $K \subseteq I$, misalkan $x_k := (x_i \mid i \in K)$ menotasikan koordinat proyeksi dari x pada tabel marjinal K-arah $\zeta_k := x(J_i \mid i \in K)$. Untuk beberapa $P \in P(I)$ dan beberapa $K \subseteq I$, sebaran peluang marjinal $P_K \equiv (p(x_K) \mid x_K \in \zeta_K) \in P(K)$ dinyatakan dengan :

$$p(x_K) := \sum_{x_{I\setminus K} \in \zeta_{I\setminus K}} p(x), x_K \in \zeta_K, \quad (3)$$

dengan $P(K) = \{P_K \equiv (p(x_K) | x_K \in \zeta_K) p(x_K) > 0, \sum p(x_K) = 1\}$ untuk beberapa $K \subset I$.

Anggaplah bahwa X ~ P ∈ P(I). Misalkan $\mathcal{K}$ merupakan kisi dari I. Untuk masing-masing K ∈ J($\mathcal{K}$), sebaran peluang bersyarat dari $X_{[K]}$ dengan $X_{\langle K\rangle} = x_{\langle K\rangle}$ dinyatakan sebagai berikut :

$$P_{[K]}^{x_{\langle K\rangle}} \equiv (p(x_{[K]} \mid x_{\langle K\rangle}) \mid x_{[K]} \in \zeta_{[K]}) \in P([K]),$$

dengan

$$p(x_{[K]} \mid x_{\langle K\rangle}) = \frac{p(x_K)}{p(x_{\langle K\rangle})} = \frac{p(x_K)}{\sum_{x_{[K]}\in\zeta_{[K]}} p(x_K)}$$

Untuk P ∈ P(I), keluarga dari peluang bersyarat ($P_{[K]}^{x_{\langle K\rangle}} \mid x_{\langle K\rangle} \in \zeta_{\langle K\rangle}$, K∈J($\mathcal{K}$)) dinamakan keluarga $\mathcal{K}$-parameter dari P.

Misal P($\mathcal{K}$) menyatakan sub-famili dari P(I) yang diperoleh dengan menetapkan kendala kisi dengan syarat bebas $X_L \perp X_M | X_{L\cap M}$, ∀L,M∈ $\mathcal{K}$, pada model penuh P(I) . Maka P($\mathcal{K}$) dinamakan Model Kisi Bersyarat Bebas untuk tabel kontingensi ζ I arah.
Catatan : P(I) = P({Ø,I}).
Proposisi 1. : Untuk P ∈ P(I), tiga kondisi berikut ini ekuivalen :
1. P ≡ (p(x)|x∈ζ)∈P($\mathcal{K}$);
2. Untuk masing-masing x ∈ ζ,
$$p(x) = \prod_{K\in J(\mathcal{K})} p(x_{[K]} \mid x_{\langle K\rangle}); \qquad (4)$$
3. Untuk masing-masing L ∈ $\mathcal{K}$ dan $x_L \in \zeta_L$,
$$p(x_L) = \Pi(p(x_{[K]} \mid x_{\langle K\rangle}) \mid K \in J(\mathcal{K}), K \subseteq L).$$

Jadi, dengan P($\mathcal{K}$) persamaan (4) menyatakan bahwa masing-masing peluang sel dapat difaktorkan sebagai perkalian dari $\mathcal{K}$ parameternya.
Proposisi 2. :Faktorisasi ruang parameter. Pemetaan

$$\pi : P(\mathcal{K}) \to \times(P([K])^{\zeta_{\langle K\rangle}} \mid K \in J(\mathcal{K})) := \Pi(\mathcal{K})$$

$$\mathrm{P} \to (\mathrm{P}_{[K]}^{x_{\langle K\rangle}} \mid x_{\langle K\rangle} \in \zeta_{\langle K\rangle}, K \in J(\mathcal{K})) := \pi(\mathrm{P})$$

adalah bijektif. (Perlman & Wu, 1999)

2.2.3. Faktorisasi dari Fungsi Kemungkinan dan MLE dalam Model Kisi Bersyarat Bebas P($\mathcal{K}$).

Misal $x^1, x^2, \ldots, x^n$ merupakan contoh yang bebas dan identik dari ζ, dengan masing-masing $x^j \equiv x_i^j \mid i \in I) \in \zeta$. Masing-masing observasi x^j dianggap diobservasi secara penuh, sehingga dapat diklasifikasikan dengan lengkap menurut I kategori. Contoh ini kemudian menghasilkan tabel kontingensi I arah ζ. Jumlah yang diobservasi dalam sel x pada tabel kontingensi I arah ζ dinotasikan dengan n(x). Jumlah yang diobservasi pada sel x_K pada tabel marjinal K-arah (ζ_K) kemudian dinyatakan dengan:

$$n(x_K) = \sum_{x_{I\setminus K} \in \zeta_{I\setminus K}} n(x). \qquad (5)$$

Fungsi kemungkinan bersama untuk model multinomial dapat dinyatakan sebagai berikut :

$$\lambda(P) := \lambda(P; x^1, \ldots, x^n) = \prod_{x \in \zeta} p(x)^{n(x)} \qquad (6)$$

Di bawah model penuh P(I), apabila semua n(x)>0, MLE dari p(x) ada dan dinyatakan dengan : $\hat{p}(x) = \dfrac{n(x)}{n}, x \in \zeta.$ (7)

Di bawah Model Kisi Bersyarat Bebas P ∈ P($\mathcal{K}$), dengan proposisi 1, fungsi kemungkinan dapat difaktorkan sebagai berikut :

$$\begin{aligned}
\lambda_{\mathcal{K}}(P) &:= \prod_{x \in \zeta} p(x)^{n(x)} \\
&= \prod_{x \in \zeta} \prod_{K \in J(\mathcal{K})} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x)} \\
&= \prod_{K \in J(\mathcal{K})} \prod_{x \in \zeta} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x)} \\
&= \prod_{K \in J(\mathcal{K})} \prod_{x_K \in \zeta_K} \prod_{x_{I\setminus K} \in \zeta_{I\setminus K}} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x)} \\
&= \prod_{K \in J(\mathcal{K})} \prod_{x_K \in \zeta_K} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x_K)} \\
&= \prod_{K \in J(\mathcal{K})} \prod_{x_{\langle K \rangle} \in \zeta_{\langle K \rangle}} \prod_{x_{[K]} \in \zeta_{[K]}} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x_K)} \\
&=: \prod_{K \in J(\mathcal{K})} \prod_{x_{\langle K \rangle} \in \zeta_{\langle K \rangle}} \lambda_{K, x_{\langle K \rangle}}(P_{[K]}^{x_{\langle K \rangle}}),
\end{aligned} \qquad (8)$$

dengan

$$\lambda_{K, x_{\langle K \rangle}}(P_{[K]}^{x_{\langle K \rangle}}) := \prod_{x_{[K]} \in \zeta_{[K]}} p(x_{[K]} \mid x_{\langle K \rangle})^{n(x_K)} \qquad (9)$$

Catatan bahwa untuk K dan $x_{\langle K \rangle}$ tetap, $\lambda_{K, x_{\langle K \rangle}}(P_K^{x_{\langle K \rangle}})$ mempunyai bentuk fungsi kemungkinan untuk model penuh P([K]) didasarkan pada ukuran contoh $n(x_{\langle K \rangle})$. Pembatasan-pembatasan yang ditentukan dalam Model Kisi Bersyarat Bebas P($\mathcal{K}$) pada $\mathcal{K}$-parameter $P(x_{[K]}|x_{\langle K \rangle})$ yang digunakan dalam memfaktorkan $\lambda_{K, x_{\langle K \rangle}}(.)$ adalah sederhana, yaitu :

$$p(x_{[K]} \mid x_{\langle K \rangle}) > 0, \qquad \sum_{x_{[K]} \in \zeta_{[K]}} p(x_{[K]} \mid x_{\langle K \rangle}) = 1, \qquad (10)$$

dan pembatasan (10) adalah bebas untuk masing-masing K ∈ J($\mathcal{K}$). Jadi di bawah Model Kisi Bersyarat Bebas P($\mathcal{K}$), fungsi kemungkinan $\lambda_{\mathcal{K}}$(P) adalah perkalian dari fungsi kemungkinan untuk model penuh P([K]) yang berbeda, K ∈ J($\mathcal{K}$), yang memiliki peluang-peluang sel $\mathcal{K}$-parameter $p(x_{[K]}|x_{\langle K \rangle})$. Oleh karena itu $\lambda_{\mathcal{K}}$(P) dapat dimaksimumkan dengan memaksimisasi masing-masing faktor $\lambda_{K, x_{\langle K \rangle}}(.)$ secara terpisah, sehingga dengan persamaan (7), MLE dari $\mathcal{K}$ parameter dinyatakan dengan :

$$\hat{p}(x_{[K]} \mid x_{\langle K \rangle}) = \frac{n(x_K)}{n(x_{\langle K \rangle})}, \qquad x_{[K]} \in \zeta_{[K]},$$

masing-masing $n(x_K)>0$. Sebagai catatan $n(x_{\emptyset}) = n$.

MLE dari parameter asal p(x) dinyatakan dengan :

$$\hat{p}(x) = \prod_{K \in J(\mathcal{K})} \frac{n(x_K)}{n(x_{\langle K \rangle})}, \qquad x \in \zeta.$$

dalam Model Kisi Bersyarat Bebas P($\mathcal{K}$). Nilai maksimum dari fungsi kemungkinan $\lambda_{\mathcal{K}}(P)$ di bawah P($\mathcal{K}$) dinyatakan dengan :

$$\lambda_{\mathcal{K}}(\hat{P}) = \prod_{x \in \zeta} \hat{p}(x)^{n(x)} = \prod_{x \in \zeta} \prod_{K \in J(\mathcal{K})} \left(\frac{n(x_K)}{n(x_{\langle K \rangle})} \right)^{n(x)}$$

Daftar Pustaka

[1] Agresti, A. 1990. *Categorical Data Analysis.* Wiley, New York.

[2] Andersson,S.A. & Perlman, M.D. 1991. *Lattice-ordered Conditional Independence Models for Missing Data.* Statist.Probab.Lett. 12:465-486.

[3] Andersson,S.A. & Perlman, M.D. 1993. *Lattice Models for Conditional Independence in a multivariate normal distribution.* Ann.Statist. 21:1318-1358.

[4] Dempster, A.P., Laird, N.M. & Rubin, D.B. 1977. Maximum Likelihood Estimation from Incomplete Data via The EM Algorithm. J. Roy.Statist. Soc. B, 39, 1-38.

[5] Lipsitz, S.R., Parzen, M. & Molenberghs, G. 1998. *Obtaining the Maximum Likelihood Estimates in Incomplete R x C Contingency Tables Using a Model LinearTerampat Poisson.* J. Comp.& Graph. Statist. 7, 356 – 376.

[6] Little, R.T & Rubin D.B. 1987. *Statistical Analysis with Missing Data*. Wiley, New York.

[7] Perlman, M.D. and Wu,L. 1999. *Kisi Bebas Bersyarat Models for Contingency Tables with Non-Monoton Missing Data Patterns.* J.Statist.Planning and Inference 79, 259-287.

[8] Rubin, D.B. 1974. *Characterizing the Estimation of Parameters in Incomplete Data Problems*. J. Amer.Statist.Assoc.69,467-474.

[9] Walpole, R.E. 1982. *Introduction to Statistics 3rd edition*. Prentice Hall.