PENDUGAAN STATISTIK AREA KECIL DENGAN METODE *EMPIRICAL CONSTRAINED BAYES* [1]

Kismiantini
Jurusan Pendidikan Matematika
FMIPA Universitas Negeri Yogyakarta

Abstrak

Metode *empirical Bayes* (EB) merupakan metode yang lebih aplikatif pada pendugaan area kecil. Metode ini mampu menangani data kontinu, biner maupun cacahan serta mampu menurunkan galat baku dibandingkan penduga langsung. Penduga *empirical Bayes* diperoleh dengan menduga parameter model melalui fungsi kepekatan peluang marjinal lalu disubstitusikan dalam penduga Bayes. Namun penduga Bayes ini akan mengalami underdispersi pada kuadrat galat dengan model dua tahap. Untuk mengatasi permasalahan ini, dapat dengan memasukkan kendala (*constraint*) pada *posterior expected squared error loss*, sehingga penduga yang diperoleh disebut penduga *constrained Bayes*. Selanjutnya dengan mensubstitusikan nilai dugaan parameternya diperoleh penduga *empirical constrained Bayes* (ECB). Kebaikan ketiga penduga statistik area kecil yaitu penduga langsung, *empirical Bayes* dan *empirical constrained Bayes* dengan melihat besarnya kuadrat tengah galat.

Kata-kata kunci : pendugaan area kecil, *empirical Bayes*, *empirical constrained Bayes*

PENDAHULUAN

Pendugaan area kecil (*small area estimation*) adalah suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil (Rao 2003). Pendugaan sederhana area kecil yang didasarkan pada penerapan model desain penarikan contoh (*design-based*) dengan ukuran contoh dari subpopulasi disebut sebagai pendugaan langsung (*direct estimation*). Bila ukuran contohnya kecil maka statistik yang dihasilkan akan memiliki ragam yang besar bahkan pendugaan tidak dapat dilakukan.

Berbagai metode pendugaan area kecil (*small area estimation*) telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based area estimation*) sebagai alternatif dari pendugaan langsung. Metode tersebut adalah *empirical best linear unbiased prediction* (EBLUP), *empirical Bayes* (EB), dan *hierarchical Bayes* (HB). Metode EB dan HB merupakan metode yang lebih umum yang mampu menangani data kontinu, biner maupun cacahan.

[1] Makalah ini disampaikan pada Seminar Nasional Matematika 2007 yang diselenggarakan oleh Jurusan Pendidikan Matematika FPMIPA UPI pada tanggal 8 Desember 2007

Pada model area kecil, pendugaan dengan metode *empirical Bayes* dimulai dengan mengasumsikan model dua tahap. Selanjutnya dengan memaksimumkan fungsi marjinal akan diperoleh nilai dugaan parameternya, yang kemudian disubstitusikan dalam penduga Bayes. Menurut Rao (2003) penduga Bayes ini akan menunjukkan underdispersi dengan model dua tahap tersebut, yang dapat dilihat pada kuadrat galatnya. Untuk mengatasi permasalahan ini dapat dilakukan dengan memasukkan suatu kendala (*constraint*) pada kuadrat galatnya (Ghosh, 1992). Dalam makalah ini akan dibahas pendugaan statistik area kecil dengan menggunakan penduga *empirical constrained Bayes* (ECB) berdasarkan asumsi normal.

METODE *EMPIRICAL BAYES*

Metode *empirical Bayes* (EB) merupakan metode yang lebih umum untuk menangani model dengan data kontinu, biner maupun cacahan. Berdasarkan asumsi normal, model area kecil (*basic area level*) dapat diekspresikan sebagai model dua tahap (Rao, 1999) sebagai berikut :

$$\text{(i) } \hat{\theta}_i = \theta_i + e_i \text{ dengan } e_i \overset{iid}{\sim} N(0, \psi_i),\ i = 1,2,\cdots,m \tag{1}$$

$$\text{(ii) } \theta_i \overset{iid}{\sim} N(\mu_i, \sigma_v^2) \text{ sebagai prior, } \mu_i = x_i^T \beta \tag{2}$$

Berdasarkan teorema Bayes maka diperoleh sebaran posterior yaitu :

$$\theta_i \Big| \hat{\theta}_i, \beta, \sigma_v^2 \overset{ind}{\sim} N\left(\hat{\theta}_i^B, g_{1i}(\sigma_v^2) = \gamma_i \psi_i\right) \text{ dengan } \gamma_i = \sigma_v^2 / (\sigma_v^2 + \psi_i) \tag{3}$$

Penduga Bayes bagi θ_i adalah nilai harapan dari sebaran posterior sebagai berikut :

$$E\left(\theta_i \Big| \hat{\theta}_i, \beta, \sigma_v^2\right) = \hat{\theta}_i^B = \gamma_i \hat{\theta}_i + (1-\gamma_i)\mu_i\ ,\ \gamma_i = \sigma_v^2 / (\sigma_v^2 + \psi_i),\ \mu_i = x_i^T \beta \tag{4}$$

Penduga Bayes $\hat{\theta}_i^B$ tergantung pada parameter β dan σ_v^2, yang dapat diperoleh dari sebaran marjinal : $\hat{\theta}_i \overset{ind}{\sim} N(\mu_i, \sigma_v^2 + \psi_i)$. Penduga *empirical Bayes* (EB) bagi θ_i diperoleh dengan mensubstitusikan β dengan $\hat{\beta}$ dan σ_v^2 dengan $\hat{\sigma}_v^2$, yaitu :

$$\hat{\theta}_i^{EB} = \hat{\theta}_i^B(\hat{\beta}, \sigma_v^2) = \hat{\gamma}_i \hat{\theta}_i + (1-\hat{\gamma}_i)\hat{\mu}_i\ ,\ \hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \hat{\psi}_i) \text{ dan } \hat{\mu}_i = x_i^T \hat{\beta} \tag{5}$$

Pada makalah ini akan diasumsikan bahwa ragam sampling sama. Menurut Morris (1983) bila ragam sampling sama yaitu $\psi_i = \psi$ maka penduga tak bias bagi $1-\gamma_i = 1-\gamma$ adalah

$1-\gamma^{*}=\psi(m-p-2)/S$ dengan $S=\sum_i\left(\hat{\theta}_i-x_i^T\hat{\beta}_{LS}\right)^2$, $\hat{\beta}_{LS}$ merupakan penduga kuadrat terkecil bagi β. Sehingga penduga EB bagi θ_i adalah

$$\hat{\theta}_i^{EB}=\gamma^{*}\hat{\theta}_i+\left(1-\gamma^{*}\right)x_i^T\hat{\beta}_{LS} \tag{6}$$

Pendugaan kuadrat tengah galat (MSE) bagi penduga EB ($\hat{\theta}^{EB}$) dengan menggunakan metode jackknife yang dikemukakan oleh Jiang, Lahiri dan Wan (2002) yaitu

$$\begin{aligned} MSE\left(\hat{\theta}_i^{EB}\right) &= E\left(\hat{\theta}_i^{EB}-\hat{\theta}_i^{B}\right)^2+E\left(\hat{\theta}_i^{B}-\theta_i\right)^2 \\ &= E\left(\hat{\theta}_i^{EB}-\hat{\theta}_i^{B}\right)^2+g_{1i}\left(\sigma_v^2\right) \\ &=: M_{2i}+M_{1i} \end{aligned} \tag{7}$$

METODE *EMPIRICAL CONSTRAINED BAYES*

Metode *constrained Bayes* merupakan metode yang digunakan untuk mengatasi permasalahan underdispersi pada penduga *Bayes*. Misalkan model dua tahap berikut $\hat{\theta}_i\middle|\theta_i \overset{ind}{\sim} f\left(\hat{\theta}_i\middle|\theta_i,\lambda_1\right)$ dan $\theta_i \overset{iid}{\sim} f\left(\theta_i\middle|\lambda_2\right)$, maka dapat diperoleh penduga Bayes bagi θ_i yaitu $\hat{\theta}_i^{B}=E\left(\theta_i\middle|\hat{\theta}_i,\lambda\right)$. Persamaan (8) menunjukkan bahwa penduga Bayes mengalami underdispersi pada kuadrat tengah galatnya.

$$\begin{aligned} E\left[\frac{1}{m-1}\sum_i\left(\theta_i-\theta_.\right)^2\right] &= \frac{1}{m-1}\sum_i V\left(\theta_i-\theta_.\middle|\hat{\theta}\right)^2+\frac{1}{m-1}\sum_i\left(\hat{\theta}_i^{B}-\hat{\theta}_.^{B}\right)^2 \\ &> \frac{1}{m-1}\sum_i\left(\hat{\theta}_i^{B}-\hat{\theta}_.^{B}\right)^2 \end{aligned} \tag{8}$$

dengan $\hat{\theta}=\left(\hat{\theta}_1,\cdots,\hat{\theta}_m\right)^T$, $\hat{\theta}_.^{B}=\sum_i\hat{\theta}_i^{B}/m$.

Selanjutnya untuk mengatasi permasalahan underdispersi tersebut adalah dengan meminimumkan *posterior expected squared error loss* $E\left[\sum_i\left(\theta_i-t_i\right)^2\middle|\hat{\theta}\right]$ terhadap kendala (*constraint*) berikut :

$$t_.=\hat{\theta}_.^{B} \tag{9}$$

$$\frac{1}{m-1}\sum_i\left(t_i-t_.\right)^2=E\left[\frac{1}{m-1}\sum_i\left(\theta_i-\theta_.\right)^2\middle|\hat{\theta}\right] \tag{10}$$

dengan $t_{.} = \sum_i t_i / m$. Dengan perkalian Lagrange, dapat diperoleh penduga *constrained Bayes* (CB) sebagai solusi masalah minimisasi berikut (Rao, 2003) :

$$t_{i,opt} = \hat{\theta}_i^{CB} = \hat{\theta}_i^{B} + a(\hat{\theta},\lambda)\left(\hat{\theta}_i^{B} - \hat{\theta}_{.}^{B}\right) \tag{11}$$

dengan

$$a(\hat{\theta},\lambda) = \left[1 + \frac{(1/m)\sum_i V\left(\theta_i \middle| \hat{\theta}_i, \lambda\right)}{\{1/(m-1)\}\sum_i \left(\hat{\theta}_i^{B} - \hat{\theta}_{.}^{B}\right)^2}\right]^{1/2} \tag{12}$$

Dari persamaan (11) dapat diketahui bahwa $\sum_i \left(\hat{\theta}_i^{CB} - \hat{\theta}_{.}^{CB}\right)^2 > \sum_i \left(\hat{\theta}_i^{B} - \hat{\theta}_{.}^{B}\right)^2$ karena $a(\hat{\theta},\lambda) > 1$ dan $\hat{\theta}_{.}^{CB} = \hat{\theta}_{.}^{B}$ (Rao, 2003).

Pada makalah ini akan dibahas untuk model berdasarkan asumsi normal dengan ragam sampling sama. Misalkan model dua tahap yaitu $\hat{\theta}_i = \theta_i + e_i$ dengan $e_i \overset{iid}{\sim} N(0,\psi)$ dan saling bebas dengan $\theta_i \overset{iid}{\sim} N(\mu_i, \sigma_v^2)$ sebagai prior. Sehingga sebaran posteriornya adalah $\theta_i \middle| \hat{\theta}_i, \beta, \sigma_v^2 \overset{ind}{\sim} N\left(\hat{\theta}_i^{B}, g_{1i}(\sigma_v^2) = \gamma\psi\right)$. Penduga Bayes diberikan oleh $\hat{\theta}_i^{B} = \gamma\hat{\theta}_i + (1-\gamma)\mu_i$ dengan $\gamma = \sigma_v^2 / (\sigma_v^2 + \psi)$, $\mu_i = x_i^T \beta$. Menurut Rao (2003), penduga *constrained Bayes* pada model dua tahap ini adalah :

$$\hat{\theta}_i^{CB} = \left[\gamma\hat{\theta}_{.} + (1-\gamma)\mu_i\right] + \left[1 + \frac{\psi/\gamma}{\{1/(m-1)\}\sum_i \left(\hat{\theta}_i - \hat{\theta}_{.}\right)^2}\right]^{1/2} \gamma\left(\hat{\theta}_i - \hat{\theta}_{.}\right) \tag{13}$$

Diketahui bahwa $\hat{\theta}_i \overset{iid}{\sim} N(\mu_i, \sigma_v^2 + \psi)$, bila $m \to \infty$ maka $\hat{\theta}_{.}$ konvergen peluang terhadap μ_i dan $\frac{1}{m-1}\sum_i \left(\hat{\theta}_i - \hat{\theta}_{.}\right)^2$ konvergen peluang terhadap $\psi + \sigma_v^2 = \psi/(1-\gamma_i)$, sehingga penduga *constrained Bayes* bagi θ_i dapat dinyatakan sebagai berikut :

$$\hat{\theta}_i^{CB} \approx \gamma^{1/2}\hat{\theta}_i + \left(1-\gamma^{1/2}\right)\mu_i \tag{14}$$

Selanjutnya penduga *empirical constrained Bayes* bagi θ_i diperoleh dengan mensubstitusikan μ_i dengan $\hat{\mu}_i$ dan γ dengan $\hat{\gamma}$ pada persamaan (14), yaitu :

$$\hat{\theta}_i^{ECB} = \hat{\gamma}^{1/2}\hat{\theta}_i + \left(1-\hat{\gamma}^{1/2}\right)\hat{\mu}_i \tag{15}$$

dengan $\hat{\gamma} = \hat{\sigma}_v^2 / \left(\hat{\sigma}_v^2 + \hat{\psi}\right)$ dan $\hat{\mu}_i = x_i^T \hat{\beta}$. Kuadrat tengah galat bagi $\hat{\theta}^{ECB}$ diperoleh dengan menggunakan metode jackknife pula.

PENERAPAN PADA DATA BPS

Peubah yang diamati dan menjadi perhatian dalam ilustrasi ini adalah rata-rata pengeluaran perkapita rumah tangga. Sumber data yang digunakan adalah SUSENAS 2003 dengan materi informasi berbasis rumah tangga, serta PODES 2003 sebagai sumber data peubah penyerta. Peubah penyertanya adalah peubah-peubah yang diasumsikan mempengaruhi dan atau menggambarkan pengeluaran rumah tangga pada suatu wilayah, meliputi: persentase rumah tangga prasejahtera dan sejahtera 1, persentase pengangguran, persentase rumah tangga pelanggan listrik PLN, dan persentase rumah tangga pelanggan telepon.

Analisis menggunakan SAS 9.1 meliputi: PROC TABULATE untuk memperoleh penduga langsung, PROC MIXED untuk mendapatkan penduga β, σ_v^2, ψ, dan PROC IML untuk mendapatkan penduga EB dan ECB.

Tabel 1. Pendugaan rata-rata pengeluaran per kapita (× Rp.100.000,-) berdasarkan *design-based*, *empirical Bayes* (EB) dan *empirical constrained Bayes* (ECB)

Kecamatan	**Ukuran Contoh**	**Design-Based**		**EB**		**ECB**	
		Theta_hat	**MSE**	**Theta_hat**	**MSE**	**Theta_hat**	**MSE**
Mantrijeron	32	3.707	0.326	3.841	0.421	3.826	0.237
Kraton	32	3.738	0.274	3.949	0.514	3.927	0.311
Mergangsan	64	4.023	0.312	4.056	0.359	4.052	0.188
Umbulharjo	128	4.456	0.206	4.325	0.872	4.339	0.598
Kotagede	32	3.608	0.422	3.828	0.547	3.804	0.338
Gondokusuman	112	5.607	0.339	5.516	0.395	5.526	0.219
Danurejan	32	3.184	0.550	3.564	0.878	3.523	0.601
Pakualaman	16	2.483	0.509	2.602	0.408	2.590	0.227
Gondomanan	16	3.243	0.286	3.219	0.353	3.221	0.185
Ngampilan	16	4.583	0.59	4.047	1.43	4.105	1.05

			2		5		7
Wirobrajan	48	4.212	0.332	3.723	1.270	3.776	0.924
Gedong tengen	32	2.596	0.198	2.645	0.389	2.640	0.212
Jetis	48	3.609	0.277	3.560	0.359	3.566	0.190
Tegalrejo	64	3.740	0.267	3.909	0.544	3.891	0.335

Kajian empirik pada Tabel 1 memperlihatkan bahwa pendugaan dengan metode *empirical constrained Bayes* memberikan hasil yang lebih baik dibanding metode *empirical Bayes* yang ditunjukkan oleh nilai kuadrat tengah galat (MSE) yang relatif lebih kecil. Pendugaan langsung berdasarkan *design-based* untuk kasus data Susenas di kota Yogyakarta relatif memberikan hasil yang baik, hal ini mengindikasikan bahwa ukuran contoh untuk area kecamatan di kota Yogyakarta cukup memadai untuk digunakan dalam pendugaan langsung. Namun pendugaan langsung ini belum memasukkan unsur pembobot padahal pembobot merupakan salah satu hal penting pada pendugaan berdasarkan *design-based*.

SIMPULAN

Pendugaan statistik area kecil dengan metode *empirical constrained Bayes* (ECB) memberikan hasil yang lebih baik dibanding metode *empirical Bayes* (EB). Pada pendugaan langsung berdasarkan metode *design-based* perlu dilakukan pengkajian tentang besarnya pembobot.

DAFTAR PUSTAKA

Ghosh, M. 1992. Constrained Bayes estimation with applications. *Journal of the American Statistical Association* 87: 533-540.

Jiang, J., Lahiri, P., & Wan, S.M. 2002. A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics* 30:1782-1810.

Morris, C.A. 1983. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* 78: 47-54.

Rao, J.N.K. 1999. Some recent advances in model-based small area estimation. *Survey Methodology* 25: 175-186.

Rao, J.N.K. 2003. *Small area estimation.* New York: John Wiley and Sons.