



# **PEMODELAN AREA KECIL UNTUK MENDUGA ANGKA KEMATIAN BAYI MELALUI PENDEKATAN MODEL REGRESI POISSON BAYES BERHIRARKI DUA-LEVEL**

**NUSAR HAJARISMAN**



**SEKOLAH PASCASARJANA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2013**

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## PERNYATAAN MENGENAI DISERTASI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA\*

Dengan ini saya menyatakan bahwa disertasi berjudul *Pemodelan Area Kecil untuk Menduga Angka Kematian Bayi melalui Pendekatan Model Regresi Poisson Bayes Berhirarki Dua-Level* adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir disertasi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2013

*Nusar Hajarisman*  
NRP: G161080041

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## SUMMARY

**NUSAR HAJARISMAN.** Two-Level Hierarchical Bayesian Poisson Models for Small Area Estimation of Infant Mortality Rates. Supervised by **KHAIRIL ANWAR NOTODIPUTRO, I GUSTI PUTU PURNABA, dan KUSMAN SADIK.**

Hierarchical Bayes (HB) approach has been recently proposed for small area estimation problems. However, in this area, there is still a limited use of hierarchical Bayesian models though they have many advantages: (i) their specification is straightforward and allows to take into account the different sources of variation and (ii) inferences are clear-cut and computationally feasible in most cases by using standard Markov Chain Monte Carlo (MCMC) techniques. Within this approach, when the variable of interest is a count or a proportion, alternative model specifications can be considered. In the HB approach, a prior distribution on the model parameters is specified and the posterior distribution of the parameters of interest is obtained. Inferences are based on the posterior distribution, in particular, a parameter of interest is estimated by its posterior mean and its precision is measured by its posterior variance.

Even though the HB is straightforward, and HB inferences are clear-cut and 'exact', but the HB requires the specification of a subjective prior  $f(\theta)$  on the model parameter  $\theta$ . Priors on  $\theta$  maybe informative or 'diffuse'. Informative priors are based on substantial prior information, such previous studies judged relevant to the current data set  $\mathbf{y}$ . However, informative priors are seldom available in real HB application, particularly those related to public policy. This dissertation analyses and discusses the structure of alternative models when the variable of interest is a count, both on a theoretical side as well as by a simulation study. Bayesian specifications derived from classical models for SAE, e.g. the Fay-Herriot model (Fay and Herriot, 1979), and then consider a generalized linear Poisson model. The combination of these two models has formed a model called the hierarchical Bayes Poisson regression model. Development of HB models that extend the Fay-Herriot models and generalized linear model by using two different priors distribution on the model parameter of interest. First, we used gamma distribution as a conjugate priors for Poisson likelihood, and second we used inverse gamma distribution which is non conjugate prior for Poisson likelihood.

The proposed models are implemented using the Gibbs sampling method for fully Bayesian inference. We apply the proposed models to the analysis of infant mortality rate for sub-district level in Bojonegoro district, East Java Province. Based on the application result, we found that hierarchical Bayes Poisson regression model with inverse-gamma prior distribution gave the better prediction of infant mortality rate than model with gamma prior distribution. Our conclusion were made based on various criterion on diagnostic models, such as convergence of Markov chain diagnostic tests, a Bayesian measure of fit or adequacy, residual analysis, and other summary statistics (bias and mean square error).

**Keywords:** small area estimation, Fay-Herriot model, generalized linear models, Poisson distribution, Markov chain Monte Carlo, Gibbs sampling.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## RINGKASAN

**NUSAR HAJARISMAN.** Pemodelan Area Kecil untuk Menduga Angka Kematian Bayi melalui Pendekatan Model Regresi Poisson Bayes Berhirarki Dua-Level. Dibimbing oleh KHAIRIL ANWAR NOTODIPUTRO, I GUSTI PUTU PURNABA, dan KUSMAN SADIK.

Pembahasan mengenai pendekatan model Bayes berhirarki yang diterapkan pada pendugaan area kecil untuk data yang berbentuk cacahan relatif masih sedikit. Penggunaan pendekatan hirarki Bayes yang saat ini masih terbatas, padahal pendekatan ini mempunyai beberapa keuntungan, diantaranya yaitu: (1) spesifikasi modelnya adalah langsung dan dapat digunakan untuk memodelkan berbagai sumber keragaman yang berbeda, (2) masalah inferensinya relatif lebih jelas dan komputasinya juga relatif lebih mudah dengan menggunakan teknik rantai Markov Monte Carlo. Dalam pendekatan Bayes berhirarki, dinyatakan terlebih dahulu sebaran prior pada parameter model, kemudian diperoleh sebaran posterior dari parameter yang diamati. Secara khusus, inferensi yang didasarkan pada sebaran posterior untuk parameter tersebut diduga melalui rata-rata posterior, dan presisinya diukur melalui ragam posterior.

Walaupun pendekatan Bayes berhirarki ini mempunyai beberapa kelebihan, namun pendekatan ini memerlukan spesifikasi mengenai subjektif prior,  $f(\theta)$ , pada parameter model  $\theta$ . Prior pada parameter  $\theta$  bisa jadi merupakan prior yang bersifat informatif. Prior yang informatif yang berdasarkan pada informasi prior menurut hasil kajian penelitian sebelumnya yang relevan dengan gugus data yang sedang dikaji. Namun demikian, dalam prakteknya prior informatif ini jarang tersedia, terutama apabila diterapkan pada masalah kebijakan publik. Dalam disertasi ini menganalisis dan membahas tentang struktur model alternatif pada saat peubah respons yang diamati berbentuk data cacahan, baik secara teoritis maupun melalui studi simulasi. Spesifikasi Bayes diturunkan melalui model Fay-Harriot (Fay dan Harriot, 1979) dan model Poisson linear terampat. Kombinasi dari kedua model tersebut menghasilkan suatu model yang disebut dengan model regresi Poisson Bayes berhirarki. Pengembangan model Bayes berhirarki yang memadukan konsep model Fay-Harriot dan model linear menggunakan dua buah sebaran prior yang berbeda, yaitu sebaran prior gamma yang merupakan prior yang bersifat *conjugate* bagi likelihood Poisson, serta invers-gamma yang merupakan prior yang bersifat *non-conjugate* bagi likelihood Poisson.

Model yang diusulkan dalam disertasi ini diimplementasikan dengan menggunakan percontohan Gibbs. Model ini kemudian diaplikasikan untuk menganalisis angka kematian bayi level kecamatan di Kabupaten Bojonegoro, Provinsi Jawa Timur. Berdasarkan hasil aplikasi ini ditemukan bahwa model regresi Poisson Bayes berhirarki dengan sebaran prior invers-gamma memberikan prediksi yang lebih baik dibandingkan dengan model dengan sebaran prior gamma. Kesimpulan ini dibuat berdasarkan pada berbagai kriteria pada diagnostik model, seperti uji diagnostik konvergensi rantai Markov, ukuran kecocokan model Bayes, analisis residu, serta beberapa ringkasan statistik lainnya (bias dan kuadrat tengah galat).

**Kata Kunci:** pendugaan area kecil, model Fay-Herriot, model linear terampat, sebaran Poisson, rantai Markov Monte Carlo, Bayes berhirarki, angka kematian bayi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

© Hak Cipta milik IPB, Tahun 2013

Hak Cipta dilindungi Undang-Undang

*Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah; dan pengutipan tersebut tidak merugikan kepentingan IPB.*

*Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.*



# **PEMODELAN AREA KECIL UNTUK MENDUGA ANGKA KEMATIAN BAYI MELALUI PENDEKATAN MODEL REGRESI POISSON BAYES BERHIRARKI DUA-LEVEL**

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

**NUSAR HAJARISMAN**

**Disertasi**  
**sebagai salah satu syarat untuk memperoleh gelar Doktor**  
**pada Program Studi Statistika**

**SEKOLAH PASCASARJANA**  
**INSTITUT PERTANIAN BOGOR**  
**BOGOR**  
**2013**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pengujian pada Ujian Tertutup: 1. Dr. Ir. Anang Kurnia, MS.  
2. Dr. Ir. Bagus Sartono, MS.

Pengujian pada Ujian Terbuka: 1. Prof. Dr. Sutawanir Darwis  
2. Dr. Ir. Hari Wijayanto, MS.



Judul Disertasi : Pemodelan Area Kecil untuk Menduga Angka Kematian Bayi melalui Pendekatan Model Regresi Poisson Bayes Berhirarki Dua-Level  
Nama : Nusar Hajarisman  
NRP : G161080041

Hak Cipta Diliindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Disetujui oleh

Komisi Pembimbing

**Prof. Dr. Ir. Khairil Anwar Notodiputro, MS**  
Ketua

**Dr. Ir. I Gusti Putu Purnaba, DEA**  
Anggota

**Dr. Ir. Kusman Sadik, MS.**  
Anggota

Diketahui oleh

Koordinator Mayor Statistika

Dekan Sekolah Pascasarjana

**Dr. Ir. Aji Hamim Wigena, MSc.**

**Dr. Ir. Dahrul Syah, Msc.Agr**

Tanggal Ujian Terbuka: 28 Juni 2013

Tanggal Lulus: 30 Juli 2013

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.





#### Hak Cipta Diliindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## PRAKATA

*Bismillahirrahmanirrahim,*

Puji syukur penulis panjatkan kehadirat Allah *subhanahu wa ta'ala* atas kesehatan dan kekuatan yang diberikan kepada penulis selama proses penulisan disertasi ini. Topik utama dari disertasi ini adalah pengembangan model area kecil untuk data yang berbentuk cacahan yang diaplikasikan untuk menduga angka kematian bayi melalui pendekatan model Poisson Bayes berhirarki dua-level. Dalam penelitian ini telah ditunjukkan implementasi dari konsep pemodelan linear terampat pada masalah pendugaan area kecil. Model yang dikembangkan dalam penelitian ini adalah model regresi Poisson berhirarki dua-level dengan cara memadukan terminologi yang ada dalam model linear terampat dengan konsep metode Bayes berhirarki dua-level sedemikian rupa sehingga dapat diimplementasikan untuk menangani masalah pendugaan area kecil yang diwakili oleh model Fay-Herriot.

Pada kesempatan ini penulis ucapkan terima kasih yang sebesar-besarnya kepada komisi pembimbing, yaitu: Bapak Prof. Dr. Ir. Khairil Anwar Notodiputro, MS. (Ketua), Bapak Dr. Ir. I Gusti Putu Purnaba, DEA. (Anggota), dan Bapak Dr. Ir. Kusman Sadik, MS. (Anggota) yang telah banyak memberikan masukan dan arahan yang sangat berarti bagi penulis dalam menyelesaikan penyusunan disertasi ini. Penulis juga haturkan terima kasih atas saran dan bantuan yang sangat berharga dari Bapak Dr. Ir. Aji Hamim Wigena, MSc., selaku Koordinator Mayor Statistika, dan Prof. Dr. Ir. Marimin, MSc. selaku Sekretaris Program Doktor Program Pascasarjana, Institut Pertanian Bogor.

Penulis juga ucapkan banyak terima kasih kepada Dekan Sekolah Pascasarjana Institut Pertanian Bogor yang telah memberikan kesempatan kepada penulis untuk menimba ilmu di sekolah terbaik di Indonesia. Selain itu, penulis ucapkan terima kasih kepada dosen-dosen di Departemen Statistika IPB, rekan-rekan sejawat di Program Studi Statistika Unisba, serta rekan-rekan PS Statistika Pascasarjana IPB atas masukan dan diskusinya selama ini. Tak lupa penulis ucapkan banyak terima kasih kepada Bapak Heriawan di PS Statistika yang telah begitu banyak membantu penulis selama ini.

Penulis juga haturkan ucapan terima kasih kepada Mamah, Mah Enni, ibu mertua, serta adik-adikku dan keponakanku semua atas segala doa dan dukungannya. Akhirnya, kepada isteriku tercinta Dewi Rosiana dan anakku tersayang Syahmina Arkani Raihana, penulis ucapkan terima kasih yang sebesar-besarnya yang telah banyak memberikan doa, dorongan, semangat, kesabaran, serta kasih sayangnya yang penulis butuhkan selama penyusunan disertasi ini.

Semoga Allah SWT senantiasa memberikan rahmat dan hidayahnya kepada kita semua. Amin.

Bogor, Juni 2013

**Nusar Hajarisman**



## DAFTAR NOTASI DAN ISTILAH

AIC	:	<i>Akaike Information Criterion</i>
AKB	:	Angka Kematian Bayi
BIC	:	<i>Bayesian Information Criterion</i>
CAR	:	<i>Conditional AutoRegressive</i>
Dbar	:	<i>Posterior mean of deviance</i> (Rata-rata devians posterior)
DIC	:	<i>Deviance Information Criterion</i>
<i>Direct Estimation</i>	:	Pendugaan Langsung (Pendugaan parameter yang dilakukan hanya berdasarkan data survei pada masing-masing area)
Dmean	:	<i>Deviance evaluated at posterior mean</i> (Devians yang dihitung pada rata-rata posterior)
DPP	:	<i>Dirichlet Process Prior</i>
EB	:	<i>Empirical Bayes</i>
EBLU	:	<i>Empirical Best Linear Unbiased Prediction</i> (Prediksi Takbias Terbaik Linear Empirik)
$G(a, b)$	:	Sebaran gamma dengan parameter $a$ dan $b$
GAMM	:	<i>Generalized Additive Mixed Models</i> (Model Campuran Aditif Terampat)
GCV	:	<i>Generalized Variance Function</i> (Fungsi Ragam Terampat)
GLM	:	<i>Generalized Linear Model</i> (Model Linear Terampat)
GLMM	:	<i>Generalized Linear Mixed Models</i> (Model Campuran Linear Terampat)
GMLE	:	<i>Generalized Maximum Likelihood Estimate</i> (Penduga Kemungkinan Maksimum Terampat)
GREG	:	<i>Generalized Regression</i>
HB	:	<i>Hierarchical Bayes</i>
HPD	:	<i>Highest Posterior Density</i> (Fungsi Kepekatan Posterior Tertinggi)
$I\Gamma(a, b)$	:	Sebaran invers gamma dengan parameter $a$ dan $b$
IMR	:	<i>Infant Mortality Rate</i> (Angka Kematian Bayi)
<i>Indirect Estimation</i>	:	Pendugaan Tidak Langsung (Pendugaan parameter yang dilakukan dengan melibatkan informasi tambahan baik dari dalam area yang menjadi perhatian, area lain maupun survei lain)
IPM	:	Indeks Pembangunan Manusia

Hak Cipta Dilindungi Undang-Undang

Hak Cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



MAPE	:	<i>Mean Absolute Predictive Error</i> (Rata-rata Jumlah Absolut Galat Prediktif)
MCMC	:	<i>Markov Chain Monte Carlo</i>
MCSE	:	<i>Monte Carlo Standard Error</i> (Galat Baku Monte Carlo)
M-H	:	Algoritma Metropolis-Hastings
MSE	:	<i>Mean Square Error</i> (kuadrat tengah galat)
MSPE	:	<i>Mean Squared Predictive Error</i> (Rata-rata Jumlah Kuadrat Galat Prediktif)
Multiple-Run	:	Proses simulasi MCMC berganda
PB	:	<i>Parametric Bootstrap</i>
pD	:	<i>Effective number of parameters</i> (banyaknya parameter yang efektif)
PSRF	:	<i>Potential Scale Reduction Factor</i> (Faktor reduksi skala potensial)
REML	:	<i>Restricted Maximum Likelihood</i>
SAE	:	<i>Small Area Estimation</i> (pendugaan area kecil)
SD	:	<i>Standard Deviation</i> (Simpangan Baku)
SDKI2007	:	Survei Demografi dan Kesehatan Indonesia Tahun 2007
Single-Run	:	Proses simulasi MCMC tunggal
SP2010	:	Sensus Penduduk Tahun 2010
SUSENAS2010	:	Survei Sosial Ekonomi Nasional Tahun 2010

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



# DAFTAR ISI

<b>PRAKATA</b> .....	<b>X</b>
<b>DAFTAR NOTASI DAN ISTILAH</b> .....	<b>XI</b>
<b>DAFTAR ISI</b> .....	<b>XIII</b>
<b>DAFTAR TABEL</b> .....	<b>XIV</b>
<b>DAFTAR GAMBAR</b> .....	<b>XV</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>XV</b>
<b>1 PENDAHULUAN</b> .....	<b>1</b>
LATAR BELAKANG MASALAH .....	1
TUJUAN DAN MANFAAT PENELITIAN .....	6
KERANGKA KERJA PENELITIAN .....	6
KEBARUAN.....	8
KERANGKA PENULISAN DISERTASI.....	8
<b>2 TINJAUAN PUSTAKA</b> .....	<b>9</b>
PENDAHULUAN .....	9
PERKEMBANGAN METODE BAYES DALAM PENDUGAAN AREA KECIL .....	10
MODEL FAY-HERRIOT.....	13
INFERENSI BAYES BERBASIS MODEL .....	14
<i>Sebaran Prior</i> .....	14
<i>Inferensi Bayes</i> .....	16
<i>Model Bayes Berhirarki</i> .....	17
MODEL LINEAR TERAMPAT.....	20
METODE RANTAI MARKOV MONTE CARLO .....	21
<i>Permasalahan dalam MCMC</i> .....	23
<i>Besaran Posterior</i> .....	25
<b>3 MODEL REGRESI POISSON BAYES BERHIRARKI</b> .....	<b>26</b>
PENDAHULUAN .....	26
MODEL REGRESI POISSON BAYES BERHIRARKI DUA-LEVEL.....	28
<i>Level 1: Model Individu</i> .....	29
<i>Level 2: Model Terstruktur</i> .....	29
MASALAH KOMPUTASI.....	33
EVALUASI KEKONVERGENAN RANTAI MARKOV .....	37
<i>Metode Grafik</i> .....	38
<i>Metode Rasio Ragam</i> .....	40
<i>Metode Spectral</i> .....	41
UKURAN KECOCOKAN MODEL BAYES BERHIRARKI.....	42
<i>Deviance Information Criterion</i> .....	44
<i>Posterior Predictive Loss</i> .....	44
ANALISIS SISAAN PADA MODEL BAYES BERHIRARKI .....	45
<i>Sisaan Umum</i> .....	46
<i>Sisaan Bayes</i> .....	46

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



<b>4 STUDI SIMULASI .....</b>	<b>47</b>
PENDAHULUAN .....	47
SPESIFIKASI SIMULASI.....	48
<i>Model dengan Prior Gamma</i> .....	48
<i>Model dengan Prior Invers Gamma</i> .....	49
PEMBANGKITAN DATA .....	49
HASIL-HASIL SIMULASI .....	51
<i>Diagnostik Kekonvergenan Rantai Markov dalam Proses Tunggal</i> .....	52
<i>Diagnostik Kecocokan Model dan Analisis Sisaan</i> .....	53
<i>Sifat-sifat Penduga Parameter Model</i> .....	55
PEMBAHASAN .....	57
<b>5 APLIKASI MODEL POISSON BAYES BERHIRARKI UNTUK MENDUGA ANGKA KEMATIAN BAYI LEVEL KECAMATAN .....</b>	<b>60</b>
PENDAHULUAN .....	60
ANGKA KEMATIAN BAYI.....	61
<i>Konsep Angka Kematian Bayi</i> .....	61
<i>Faktor-faktor yang Mempengaruhi Kematian Bayi</i> .....	62
EVALUASI DATA DASAR .....	64
<i>Pengujian Kecocokan Sebaran Poisson</i> .....	66
<i>Pemilihan Peubah Tambahan</i> .....	67
HASIL-HASIL PENDUGAAN ANGKA MORTALITAS .....	70
<i>Diagnostik Kekonvergenan Rantai Markov</i> .....	70
<i>Diagnostik Kecocokan Model dan Analisis Sisaan</i> .....	73
<i>Hasil Prediksi Jumlah Kematian Bayi dan Pendugaan Angka Kematian Bayi</i> .....	75
<b>6 KESIMPULAN DAN SARAN .....</b>	<b>81</b>
KESIMPULAN .....	81
SARAN.....	82
<b>DAFTAR PUSTAKA.....</b>	<b>83</b>
<b>LAMPIRAN .....</b>	<b>87</b>
<b>RIWAYAT HIDUP .....</b>	<b>99</b>

## DAFTAR TABEL

4.1 Diagnostik kekonvergenan rantai Markov melalui statistik uji Geweke .....	52
4.2 Kriteria informasi devians untuk percobaan simulasi .....	53
4.3 Hasil analisis sisaan .....	54
4.4 Ringkasan statistik untuk sebaran posterior berdasarkan sebaran prior gamma dan invers gamma .....	55
4.5 Galat baku Monte Carlo dan Simpangan baku posterior .....	56
4.6 Ringkasan statistik posterior dari sebaran prior gamma .....	56
4.7 Ringkasan statistik posterior dari sebaran prior invers gamma .....	57
5.1 Definisi peubah penelitian .....	64

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



5.2	Data kesehatan tingkat kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur	65
5.3	Perhitungan statistik chi-kuadrat untuk uji kecocokan sebaran Poisson	66
5.4	Hasil pemilihan peubah tambahan melalui kriteria pemilihan himpunan bagian model regresi terbaik	68
5.5	Diagnostik kekonvergenan rantai Markov melalui statistik uji Geweke untuk data level kecamatan di Kabupaten Bojonegoro	73
5.6	Kriteria informasi devians untuk data level kecamatan di Kabupaten Bojonegoro	74
5.7	Hasil analisis sisaan untuk data level kecamatan di Kabupaten Bojonegoro	75
5.8	Galat baku Monte Carlo (GBMC) dan Simpangan baku posterior (SBP) untuk data level kecamatan Kabupaten Bojonegoro	75
5.9	Hasil prediksi kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur	76
5.10	Hasil pendugaan angka kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur	78

## DAFTAR GAMBAR

1.1	Kerangka kerja penelitian pengembangan model Poisson Bayes berhirarki dua-level	7
3.1	Trace plot untuk mengevaluasi kekonvergenan rantai Markov	39
4.1	Trace plot untuk studi simulasi yang berasal dari sebaran prior Gamma	50
4.2	Trace plot untuk studi simulasi yang berasal dari sebaran prior Invers Gamma	51
5.1	Trace plot untuk model Poisson Bayes berhirarki level kecamatan di Kabupaten Bojonegoro yang berasal dari sebaran prior Gamma	71
5.2	Trace plot untuk model Poisson Bayes berhirarki level kecamatan di Kabupaten Bojonegoro yang berasal dari sebaran prior invers gamma	72
L1.	Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatif sebaran gamma	88
L2.	Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatif sebaran invers gamma	90

## DAFTAR LAMPIRAN

1.	Sifat-sifat Sebaran Gamma dan Invers Gamma	87
2.	Program Makro SAS untuk Pemodelan Poisson Bayes Berhirarki	90
3.	Program Makro SAS untuk Perhitungan Statistik Gelman-Rubin	94

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



#### Hak Cipta Diliindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.





# 1 PENDAHULUAN

## Latar Belakang Masalah

Masalah pendugaan area kecil, atau *small area estimation* (SAE) dalam dua dekade terakhir ini mulai banyak diperhatikan oleh banyak kalangan peneliti. Hal ini mencerminkan perlunya untuk menghasilkan suatu penduga untuk area dengan otoritas lokal atau domain kecil. Kesulitan dalam menghasilkan suatu penduga seperti itu, terutama untuk area-area kecil, adalah ukuran contoh yang dicapai melalui rancangan survei tertentu untuk keperluan nasional adalah terlalu kecil sebagai penduga langsung (*direct estimator*) dengan presisi yang dapat diterima. Oleh karena itu perlu suatu penduga taklangsung (*indirect estimator*) yang ‘meminjam’ kekuatan informasi dari area sekitarnya yang dihubungkan melalui suatu pendekatan model.

Penduga taklangsung yang biasa, yang berdasarkan pada *implicit linking model*, termasuk penduga sintetik dan penduga komposit. Kedua penduga tersebut pada umumnya merupakan *design based* dan ragam rancangannya (yaitu ragam yang berhubungan dengan sebaran peluang yang disebabkan oleh rancangan penarikan contoh) akan bernilai relatif kecil dibandingkan dengan ragam rancangan dari penduga langsung. Akan tetapi, penduga taklangsung pada umumnya akan bias rancangan, dimana bias rancangan ini tidak akan mengecil sebagaimana meningkatnya ukuran contoh. Jika *implicit linking model* ini dianggap benar, maka bias rancangan akan kecil, sehingga akan menghasilkan kuadrat tengah galat (KTG) rancangan yang lebih kecil dibandingkan dengan KTG dari penduga langsung. Reduksi dalam KTG digunakan sebagai alasan utama untuk menggunakan penduga taklangsung.

Oleh karena kesulitan untuk memperoleh penduga langsung ini, seringkali dapat menggunakan penduga taklangsung dengan cara meminjam informasi dari area yang berhubungan melalui model penghubung (*linking*) eksplisit (atau implisit), dengan menggunakan data sensus dan data administratif yang berhubungan dengan area kecil. Dalam beberapa tahun terakhir ini penerapan dari penduga taklangsung yang didasarkan pada model penghubung eksplisit dalam masalah pendugaan area kecil semakin pesat karena mempunyai beberapa keuntungan dibandingkan dengan penduga taklangsung tradisional yang berdasarkan pada model implisit.

Menurut Rao (2003) beberapa keuntungan dari penggunaan penduga taklangsung itu diantaranya adalah (1) metode *model-based* eksplisit memungkinkan untuk melibatkan keragaman lokal melalui struktur galat kompleks di dalam model yang menghubungkan area kecil; (2) model dapat divalidasi melalui data contoh; (3) metode ini dapat menangani kasus yang kompleks seperti data *cross-sectional* dan data deret waktu, data biner, data cacahan, data spasial berkorelasi, serta data peubah ganda; serta (4) ukuran keragaman dari area tertentu yang berhubungan dengan suatu penduga parameter dapat diperoleh, tidak seperti ukuran keseluruhan yang biasa digunakan dalam penduga taklangsung biasa.

Model dasar untuk level area dan level unit sudah banyak dipelajari dalam berbagai literatur untuk menurunkan prediksi takbias linear terbaik empirik (*empirical best linear unbiased prediction*, EBLUP), Bayes empirik (*empirical Bayes*, EB), serta Bayes berhirarki (*hierarchical Bayes*, HB) untuk menduga area kecil dari total (atau rata-rata), serta ukuran keragamannya. Metode EBLUP ini dapat diterapkan untuk model linear campuran (*linear mixed models*) yang menangani masalah model dasar

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

Institut Pertanian Bogor



untuk level area dan level unit. Sebaliknya, metode EB dan HB merupakan metode yang lebih umum untuk diterapkan karena dapat diterapkan pada model campuran linear terampat (*generalized linear mixed models*, GLMM) yang digunakan untuk menangani data kategorik, seperti data biner dan data cacahan. Selanjutnya besaran KTG ini biasanya digunakan sebagai ukuran dari peubah di bawah pendekatan EBLUP dan EB, sedangkan pendekatan HB menggunakan ragam posterior sebagai ukuran keragamannya dengan mengasumsikan suatu sebaran prior tertentu pada parameter model.

Pendekatan model Bayes telah mulai banyak digunakan untuk menyelesaikan masalah pendugaan area kecil. Secara khusus, menurut Ghosh, et al. (1998) pendekatan hirarki Bayes dan empirik Bayes, sudah mulai diterapkan untuk memodelkan komponen sistematis dari area lokal. Teori dan aplikasi dari metode HB dan EB yang digunakan untuk pendugaan area kecil telah banyak dibahas oleh Ghosh dan Maiti (1999, 2008), Ghosh et al. (2008), Datta, et al. (1999, 2000), Torabi, et al. (2008), dan masih banyak lagi. Ghosh dan Rao (1994) telah memberikan ulasan yang cukup lengkap mengenai penerapan metode HB dan EB pada masalah pendugaan area kecil.

Pengembangan kedua pendekatan ini pada umumnya masih difokuskan pada data kontinu, padahal data survei seringkali juga berbentuk data diskrit atau kategorik, dimana analisis HB atau EB menjadi kurang tepat untuk diterapkan. Saat ini, telah banyak dikembangkan analisis Bayesian untuk data survei biner. Dempster dan Tomberlin (1980), Farrel et al. (1997), serta MacGibbon dan Tomberlin (1989) telah memperoleh penduga proporsi area kecil melalui teknik EB. Sedangkan Malec et al. (1997) membahas tentang inferensi dalam pendugaan area kecil untuk peubah biner melalui inferensi Bayes prediktif. Stroud (1991) juga mengembangkan metodologi HB untuk data biner, sedangkan Nandram dan Sedransk (1993) mengusulkan untuk menggunakan inferensi Bayesian prediktif untuk data biner pada contoh klaster dua-tahap. Stroud (1994) memberikan perlakuan yang lengkap dari data survei yang biner pada penarikan contoh acak sederhana, stratifikasi, klaster, serta penarikan contoh dua-tahap.

Pembahasan mengenai pendekatan model Bayes yang diterapkan pada pendugaan area kecil untuk data yang berbentuk cacahan relatif masih sedikit. Pada umumnya pendekatan model Bayes pada data cacahan ini diterapkan untuk menduga angka mortalitas. Misalnya, model alternatif untuk menduga angka kematian menurut peubah umur tertentu pada salah satu kategori penyakit kanker di Amerika Serikat juga telah dilakukan oleh Nandram et al. (1999). Mereka menggunakan metode Bayes yang diterapkan pada empat jenis model yang berbeda. Masing-masing model itu mengasumsikan bahwa banyaknya yang mati pada area tertentu, dan kelompok umur tertentu adalah bersebaran Poisson. Hasilnya menunjukkan bahwa model yang mereka usulkan mampu menangkap efek area kecil dan efek regional dengan baik, dan juga mampu mendeteksi korelasi spasial sisaannya, sehingga memudahkan proses pendugaan parameternya. Nandram et al. (2000) juga melakukan pemodelan yang hampir sama dengan apa yang telah dilakukan sebelumnya, akan tetapi di sini modelnya diterapkan untuk menduga angka kematian menurut peubah umur tertentu pada salah satu kategori penyakit *chronic obstructive pulmonary* di Amerika Serikat. Trevisani dan Torelli (2007) mengembangkan pendugaan area kecil untuk data cacahan melalui pendekatan HB. Penggunaan pendekatan hirarki Bayes yang saat ini masih terbatas, padahal pendekatan ini mempunyai beberapa keuntungan, diantaranya yaitu: (1) spesifikasi modelnya adalah langsung dan dapat digunakan untuk memodelkan berbagai sumber

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

variasi yang berbeda, (2) masalah inferensinya relatif lebih jelas dan komputasinya juga relatif lebih mudah dengan menggunakan teknik MCMC. Namun demikian hal yang harus diperhatikan dalam pendekatan ini adalah pemilihan prior yang tepat pada parameter model.

Walaupun implementasi model Bayes yang begitu kompleks memerlukan teknik komputasi yang rumit dari algoritma simulasi Rantai Markov Monte Carlo (*Markov Chain Monte Carlo*, MCMC), akan tetapi metode ini mempunyai beberapa keuntungan ketika diterapkan pada masalah pendugaan area kecil. Beberapa keuntungan itu diantaranya adalah kerangka kerja dari model Bayes dapat menangani sejumlah peubah sasaran yang berbeda (misalnya peubah kontinu, biner, ataupun cacahan), struktur efek acak yang berbeda (misalnya saling bebas atau berkorelasi secara spasial), area tanpa informasi survei secara langsung, atau bahwa model dapat digunakan untuk memuluskan penduga ragam contoh survei. Ketidakpastian mengenai seluruh parameter model secara otomatis dapat ditangkap melalui sebaran posterior dari penduga area kecil, dan melalui sebaran prediktif dari penduga area kecil yang tidak disertakan dalam contoh survei.

Menurut Rao (2003b) permasalahan yang masih relevan untuk dikaji sehubungan dengan masalah pendugaan area kecil ini, khusus melalui pendekatan EBLUP, EB, maupun HB meliputi enam hal yang seringkali menjadi isu utama dalam pendugaan area kecil ini, dimana keenam permasalahan itu adalah sebagai berikut

1. Masalah rancangan penarikan contoh. Masalah rancangan penarikan contoh merupakan hal penting yang harus diperhatikan dalam pendugaan area kecil. Sebuah resolusi yang tepat dari masalah rancangan dapat menyebabkan peningkatan dalam reliabilitas dari penduga langsung ataupun penduga taklangsung, baik untuk domain terencana dan tidak terencana (daerah).
2. Masalah pemilihan model dan validasi. Perkembangan metodologi dan aplikasi dari pendugaan are kecil, khususnya pada pendugaan berbasis model cukup pesat, namun demikian tetap harus memperhatikan asumsi-asumsi yang harus dipenuhi oleh model tersebut. Informasi tambahan yang baik yang dihubungkan pada peubah yang diamati memegang peranan penting dalam pemilihan model penghubung yang tepat. Oleh karena itu perhatian yang lebih khusus harus diberikan kepada kompilasi peubah tambahan yang menjadi prediktor yang baik bagi peubah yang sedang dipelajari. Dalam hal ini diagnostik model perlu dilakukan untuk mencari model yang cocok terhadap data. Model diagnostik semacam itu termasuk analisis sisaan untuk memeriksa pelanggaran asumsi dari model, pemilihan peubah tambahan, serta pemeriksaan data pencilan.
3. Masalah komputasi. Pendekatan hirarki Bayes menjadi semakin populer dalam beberapa tahun terakhir ini, karena pendekatan ini mampu menangani model yang begitu kompleks dengan menggunakan metode MCMC. Namun demikian hal yang harus diperhatikan dalam pendekatan ini adalah pemilihan prior yang tepat pada parameter model.
4. Model level area dan level unit. Model level area mempunyai cakupan yang lebih luas dibandingkan dengan model level unit, karena informasi tambahan level area biasanya lebih banyak tersedia dibandingkan dengan data tambahan pada level unit. Demikian juga, bobot rancangan akan dimasukkan melalui pemodelan penduga langsung rancangan terboboti, serta penduga EB atau HB yang dihasilkan adalah rancangan yang konsisten. Akan tetapi dalam hal ini asumsi mengenai ragam penarikan contoh diketahui menjadi sulit dipenuhi. Demikian juga penduga ragam

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

penarikan contoh yang dimuluskan yang diperoleh melalui pencocokan model fungsi ragam terampat (*generalized variance function*, GVF) juga dapat menyebabkan kesulitan dalam menduga KTG.

5. Masalah pendugaan parameter. Masalah pendugaan parameter dalam penelitian akan difokuskan pada pendugaan berbasis model dari total atau rata-rata, akan tetapi penduga seperti itu mungkin tidak tepat apabila tujuan utamanya adalah untuk menghasilkan suatu penduga parameter yang sebarannya dalam beberapa hal tidak jauh berbeda dengan sebaran dari parameter area tertentu. Dengan demikian perlu dicarikan suatu metode pendugaan yang tepat untuk mengatasi masalah tersebut. Beberapa metode pendugaan yang biasa digunakan misalnya adalah metode kemungkinan maksimum atau kemungkinan maksimum terbatas (*restricted maximum likelihood*, REML).
6. Masalah galat nonpenarikan contoh. Pada umumnya pada penelitian tentang pendugaan area kecil mengasumsikan bahwa tidak ada galat pengukuran dalam respons dan/atau pada kovariat dari nonrespons. Akan tetapi galat nonpenarikan contoh akan memberikan efek yang berarti pada pendugaan area kecil, sehingga perlu membangun suatu rancangan yang tepat dan dengan metode pendugaan yang tepat yang dapat memperhitungkan galat nonpenarikan contoh ini.

Pembahasan lebih lanjut dalam disertasi ini dibatasi untuk menangani dua masalah, yaitu masalah pemilihan model dan validasi, serta masalah komputasi yang menggunakan metode MCMC.

Pengembangan model pendugaan area kecil untuk data kategorik, seperti data biner atau data cacahan, dapat dilakukan dengan cara menambahkan faktor acak dalam model campuran ke model linear terampat (*generalized linear model*, GLM). Pada umumnya model campuran ini mengasumsikan bahwa faktor acak tersebut diasumsikan mengikuti sebaran normal. Akan tetapi, dalam prakteknya sering juga ditemukan bahwa asumsi kenormalan ini tidak terpenuhi sehingga harus dicarikan model pendugaan area kecil yang lebih fleksibel yang mampu mengatasi masalah ketidaknormalan dari faktor acak tersebut. Selain itu, Rao (2003a) dan juga Trevisani dan Torelli (2007) menyatakan bahwa pengembangan model HB untuk data kategorik dilakukan dengan mengadaptasi berbagai literatur yang membahas tentang pemetaan penyakit (*disease mapping*), dimana salah satunya misalnya banyak digunakan untuk menghitung angka kematian (mortalitas). Hal yang sama juga dilakukan oleh Christiansen dan Morris (1997), serta Nandram (2000) yang melakukan kajian tentang model Bayes berhirarki untuk data diskrit, termasuk data cacahan yang bersebaran Poisson.

Pendekatan model Bayes berhirarki dalam pendugaan area kecil yang didasarkan pada metode rantai Markov Monte Carlo (MCMC) mempunyai beberapa keterbatasan, diantaranya adalah pemilihan sebaran prior yang tepat, proses komputasi dalam bentuk proses tunggal (*single-run*) atau proses ganda (*multiple-run*), serta *burn-in length*. Dari ketiga masalah tersebut yang menjadi permasalahan utama adalah pemilihan sebaran prior yang 'tepat', karena apabila sebaran prior ini diketahui maka inferensi dapat dengan mudah dilakukan dengan cara meminimumkan galat posterior, menghitung daerah kepekatan sebaran posterior yang lebih tinggi dimensinya, atau mengintegrasikan parameter untuk menemukan sebaran prediktifnya. Sebaran prior merupakan kunci pada inferensi analisis Bayes, sehingga penetapan prior ini menjadi tahapan yang paling penting dalam menggambarkan inferensi ini. Dalam beberapa kasus pemilihan sebaran prior ini merupakan hal yang sangat sulit. Akan tetapi dalam prakteknya seringkali sudah tersedia informasi prior yang mempunyai presisi yang cukup baik sehingga akan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

membawa pada penetapan eksak dari suatu sebaran prior tertentu. Dalam hal ini, di dalam statistika tersedia banyak bentuk sebaran peluang yang sesuai dengan informasi tersebut.

Dalam pemodelan Bayes berhirarki suatu hal penting yang harus diperhatikan adalah pendugaan matriks koragam yang nantinya digunakan untuk keperluan inferensi. Hal ini menjadi penting karena galat baku dari efek tetap dan efek acak bergantung pada pendugaan matriks koragam dari sebaran efek acaknya. Berdasarkan hasil studi simulasi yang dilakukan oleh Daniels dan Kass (1999) menunjukkan bahwa penggunaan prior yang bersifat *non-conjugate* cukup efektif dalam memperkecil galat baku, sedangkan penggunaan prior yang *conjugate* bahkan kadang-kadang memberikan hasil dengan performa yang kurang baik. Selain itu, Natarajan dan Kass (2000) juga memberikan hasil yang serupa bahwa penggunaan prior yang *conjugate* dapat memberikan sifat-sifat sebaran posterior yang tidak baik. Sementara itu, Hobert dan Casella (1996) menyatakan bahwa penggunaan prior yang keliru pada pemodelan Bayes berhirarki dapat memberikan efek yang kurang baik terhadap galat baku penduga parameter terutama apabila dilakukan melalui teknik penarikan contoh Gibb.

Pada dasarnya berbagai model Bayes berhirarki alternatif ketika pendugaan area kecil yang berisi data cacahan sudah banyak dikembangkan. Spesifikasi Bayes diturunkan dari model klasik untuk pendugaan area kecil, seperti model Fay-Herriot (Fay dan Herriot, 1979), atau model yang lebih tepat dengan mempertimbangkan model linear terampat Poisson. You dan Rao (2002) membahas tentang model Normal-logNormal di dalam kelas yang disebut dengan *model unmatched*. Bahkan suatu penelitian yang relatif baru sebagaimana yang dilakukan oleh Trevisani dan Torelli (2007) mengusulkan penerapan model Normal-Poisson-logNormal dan model Gamma-Poisson-logNormal dalam menangani masalah pendugaan area kecil untuk data yang berbentuk cacahan.

Berdasarkan fakta di atas maka dalam penelitian ini akan mengkaji pengembangan model dengan menggunakan suatu sebaran prior yang bersifat *non-conjugate*. Gelman (2006) memperkenalkan suatu sebaran prior yang bersifat *conditionally conjugate*. Suatu keluarga sebaran prior  $p(\theta)$  merupakan *conditionally conjugate* untuk parameter  $\theta$  apabila sebaran posterior bersyarat,  $p(\theta|y)$  juga berada dalam kelas tersebut. Untuk keperluan komputasi, *conditionally conjugate* mempunyai makna bahwa apabila memungkinkan untuk mengambil  $\theta$  yang berasal dari kelas sebaran prior tertentu, maka hal ini juga memungkinkan untuk membentuk percontohan Gibbs bagi parameter  $\theta$  dalam sebaran posterior. Salah satu sebaran yang merupakan keluarga dari *conditionally conjugate* adalah sebaran invers gamma.

Dalam model Poisson Bayes berhirarki pada umumnya sebaran prior yang digunakan adalah gamma yang merupakan suatu sebaran *conjugate* bagi sebaran Poisson. Namun demikian, dalam penelitian ini akan dikembangkan suatu model Poisson Bayes berhirarki dua-level dengan menggunakan sebaran prior invers gamma. Pemilihan sebaran invers gamma dilakukan dengan pertimbangan bahwa sebaran dari invers gamma tidak jauh berbeda dengan sebaran gamma.

Masih tentang masalah pemilihan model dan validasi, perkembangan metodologi dan aplikasi dari pendugaan area kecil, khususnya pada pendugaan *model-based* cukup pesat, namun demikian tetap harus memperhatikan asumsi-asumsi yang harus dipenuhi oleh model tersebut. Informasi tambahan yang baik yang dihubungkan pada peubah yang diamati memegang peranan penting dalam pemilihan model penghubung yang tepat. Oleh karena itu perhatian yang lebih khusus harus diberikan kepada kompilasi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



peubah tambahan yang menjadi prediktor yang baik bagi peubah yang sedang dipelajari. Dalam hal ini diagnostik model perlu dilakukan untuk mencari model yang cocok terhadap data. Model diagnostik semacam itu termasuk analisis sisaan untuk memeriksa pelanggaran asumsi dari model, terutama misalnya yang berkaitan dengan masalah pemeriksaan data pencilan.

Selanjutnya, dalam pemodelan Bayes berhirarki, setelah dicapai kekonvergenan algoritma MCMC tidak senantiasa menghasilkan model yang baik. Terdapat banyak hal yang berkenaan dengan uji kecocokan model yang perlu dipertimbangkan. Dalam penelitian ini dikaji suatu model yang memadukan antara konsep Bayes, khususnya Bayes berhirarki, dengan konsep model linear terampat yang diterapkan pada masalah pendugaan area kecil. Oleh karena model Bayes berhirarki ini memerlukan teknik integrasi berdimensi tinggi, maka masalah komputasinya perlu dilakukan melalui metode MCMC. Dengan demikian, proses diagnostik yang perlu dilakukan sehubungan dengan pemodelan ini menyangkut dua hal, yaitu diagnostik konvergensi rantai Markov, serta diagnostik untuk mengevaluasi kecocokan model sebagaimana yang biasa dilakukan dalam model linear terampat. Diagnostik kecocokan model ini dilakukan melalui berbagai ukuran statistik (seperti devians) dan analisis sisaan.

### Tujuan dan Manfaat Penelitian

Pada dasarnya penelitian ini bertujuan untuk memberikan suatu alternatif kepada para pemerhati dan pengguna statistik yang tertarik pada analisis pendugaan area kecil yang menghasilkan suatu penduga untuk area dengan otoritas lokal atau domain kecil, khususnya untuk data survei yang berbentuk cacahan yang nantinya dapat diaplikasikan untuk menduga angka mortalitas. Ada tiga tujuan utama yang ingin dicapai dari hasil penelitian ini.

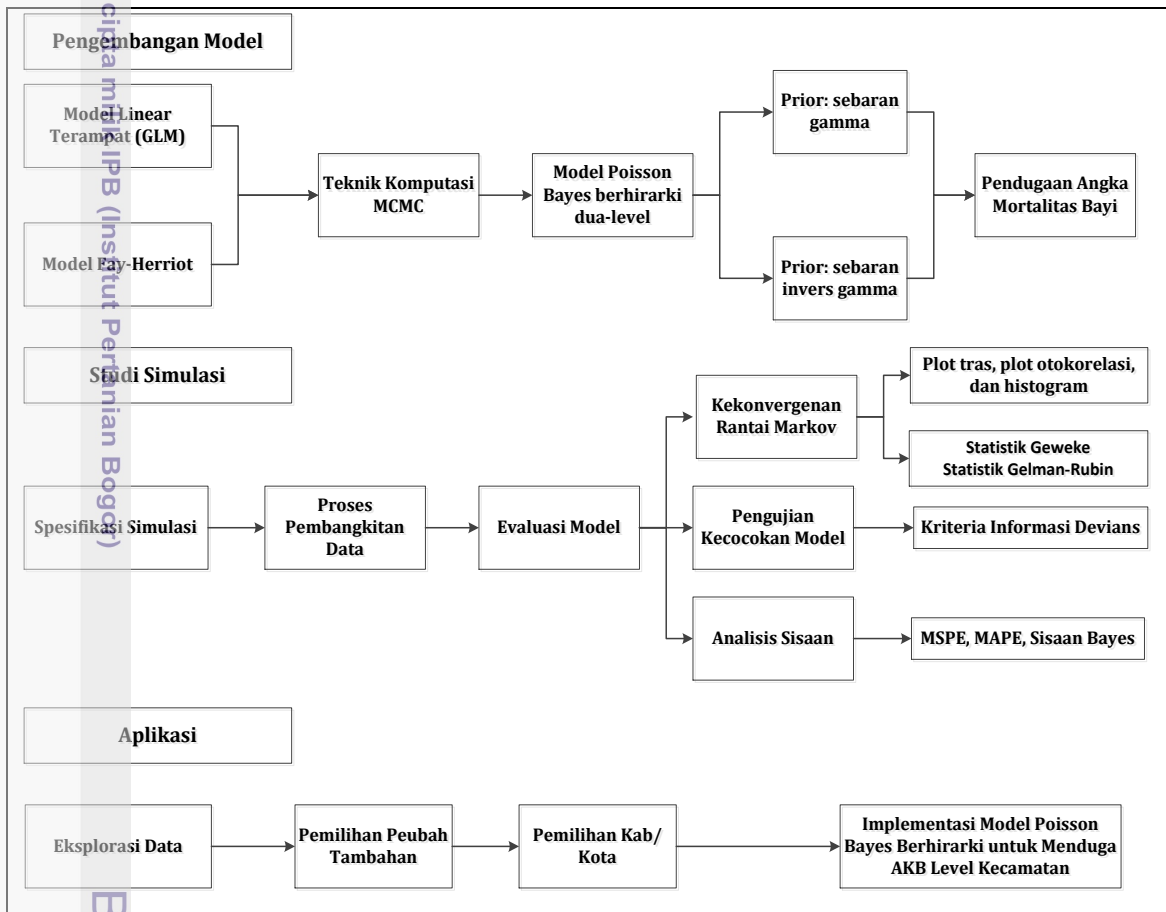
1. Mengembangkan model Bayes berhirarki untuk data cacahan yang bersebaran Poisson untuk menyelesaikan masalah pendugaan area kecil dengan cara memadukan konsep pemodelan linear terampat dengan model Fay-Herriot dengan menggunakan suatu sebaran prior yang bersifat *non-conjugate*.
2. Mengevaluasi kelayakan model yang dilakukan baik melalui plot maupun melalui berbagai ukuran statistik yang relevan melalui studi simulasi.
3. Mengaplikasikan model Poisson Bayes berhirarki dua-level untuk menduga angka kematian bayi pada level kecamatan.

Solusi statistik yang diperoleh akan sangat bermanfaat bagi para pengguna dan pemerhati statistik yang tertarik pada analisis pendugaan area kecil yang menghasilkan suatu penduga untuk area dengan otoritas lokal atau domain kecil, khususnya untuk data yang berbentuk cacahan. Manfaat yang diperoleh di sini tentunya tidak hanya dalam konteks yang sedang dibahas sekarang, akan tetapi secara umum dapat dipakai oleh peneliti di bidang lain, seperti: biologi, kedokteran, pertanian, peternakan, pendidikan, ekonomi, dan sosial.

### Kerangka Kerja Penelitian

Secara umum kerangka kerja penelitian yang ditulis dalam disertasi ini dibagi ke dalam tiga bagian utama. Pertama, pengembangan model yang dimulai dengan

mengasumsikan peubah acak yang diamati adalah bersebaran Poisson yang merupakan anggota dari keluarga eksponensial, sebagaimana yang muncul dalam konsep model linear terampat, bersyarat pada suatu parameter tertentu. Tujuan utama dari pengembangan model ini adalah membuat inferensi pada parameter tersebut yang juga dianggap sebagai peubah acak, baik secara analitik maupun secara numerik. Kemudian parameter tersebut dimodelkan dengan menggunakan model Fay-Herriot sebagai model dasar dalam konsep pendugaan area kecil dengan mengambil dua sebaran prior yang berbeda, yaitu sebaran gamma dan invers gamma. Selanjutnya, perpaduan dari kedua model tersebut akan distandarkan sedemikian rupa sehingga mewakili suatu model dalam kerangka kerja metode Bayes yang pada akhirnya akan terbentuk model Poisson Bayes berhirarki untuk menyelesaikan masalah dalam pendugaan area kecil.



Gambar 1.1 Kerangka kerja penelitian pengembangan model Poisson Bayes berhirarki dua-level.

Kedua, melakukan studi simulasi terhadap model yang sudah diperoleh sebelumnya. Tujuan utama dari studi simulasi ini adalah untuk mengevaluasi kelayakan model yang dilakukan melalui plot maupun melalui berbagai ukuran statistik yang relevan. Evaluasi yang dilakukan meliputi ketidakbiasan parameter beserta besaran galat bakunya, ukuran kecocokan model, serta analisis sisaan. Selain itu, oleh karena prosedur pengembangan model dilakukan melalui teknik komputasi MCMC, maka evaluasi model dilakukan untuk melihat kekonvergenan rantai Markovnya.

Ketiga, mengaplikasikan model Poisson Bayes berhirarki untuk menduga angka kematian bayi. Dalam penelitian ini pendugaan angka mortalitas dibatasi untuk

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Bogor Agricultural University



menduga angka kematian bayi (*infant mortality rate*, IMR) pada level kecamatan untuk beberapa kabupaten/kota di Provinsi Jawa Timur. Adapun sumber data yang digunakan dalam penelitian ini adalah Sensus Penduduk Tahun 2010 (SP2010), Survei Sosial Ekonomi Nasional Tahun 2010 (SUSENAS2010), serta Survei Demografi dan Kesehatan Indonesia Tahun 2007 (SDKI2007).

Secara ringkas seluruh rangkaian kerangka kerja penelitian ini disarikan dalam bagan sebagaimana yang ditunjukkan pada Gambar 1.1.

## Kebaruan

Implementasi model Bayes berhirarki dengan peubah respons yang bersebaran Poisson pada umumnya menggunakan prior yang bersebaran gamma. Dalam hal ini sebaran gamma merupakan sebaran yang bersifat *conjugate* bagi sebaran Poisson. Pemilihan sebaran yang bersifat *conjugate* ini akan sangat memudahkan untuk menemukan sebaran posterior secara analitik. Dalam penelitian ini dikembangkan model Poisson Bayes berhirarki dua-level dengan menggunakan sebaran prior yang bersifat *non-conjugate*, dimana prior yang dipilih adalah sebaran invers gamma. Walaupun secara analitik penggunaan sebaran prior invers gamma tidak berhasil menemukan sebaran posteriornya, akan tetapi secara numerik hasil dari pemodelan ini menunjukkan performa model yang tidak jauh berbeda pada saat menggunakan sebaran prior yang bersifat *conjugate*.

## Kerangka Penulisan Disertasi

Secara keseluruhan disertasi ini dirancang menjadi enam bab. Materi utama yang disajikan pada Bab 2 adalah peta perjalanan (*roadmap*) penelitian mengenai pendugaan area kecil, khususnya metode yang berhubungan dengan pemodelan data diskrit. Pada Bab 2 juga diperkenalkan konsep-konsep dasar mengenai model linear terampat dan model Fay-Herriot sebagai dasar untuk pengembangan model Poisson Bayes berhirarki dua-level. pendugaan area kecil, termasuk didalamnya berbagai metode untuk pengembangan pendugaan area kecil untuk data kategorik, seperti data biner dan data cacahan. Selain itu dibahas juga model dasar dari metode hirarki Bayes yang difokuskan pada pemodelan untuk level area, baik ketika ragam diketahui maupun yang tidak diketahui. Terakhir, akan dibahas pula suatu metode komputasi yang digunakan dalam penelitian ini, yaitu Rantai Markov Monte Carlo.

Kajian dari pengembangan model HB untuk pendugaan area kecil untuk data cacahan akan dibahas pada Bab 3. Pembahasan dalam bab ini dimulai dengan berbagai permasalahan yang biasa ditemui dalam pemodelan Bayes berhirarki dan berbagai besaran posterior. Fokus utama yang dibahas dalam bab ini adalah pengembangan model Poisson Bayes berhirarki dengan dua buah sebaran prior yang berbeda, yaitu menggunakan prior gamma dan invers gamma, baik masalah inferensi maupun masalah komputasinya. Pada bab ini juga dibahas tentang masalah diagnostik model yang telah dikembangkan pada bab sebelumnya. Pembahasan pada bab ini lebih ditekankan pada masalah diagnostik model yang menyangkut kekonvergenan rantai Markov yang diperoleh melalui metode MCMC, serta masalah diagnostik untuk mengevaluasi kecocokan model Bayes berhirarki melalui berbagai ukuran statistik, seperti: *deviance*



*information criterion* (DIC) dan *Posterior Predictive Loss*. Selain itu akan dibahas pula berbagai ukuran sisaan yang tepat, khususnya untuk model Bayes berhirarki untuk data cacahan yang bersebaran Poisson. Hasil-hasil dan pembahasan untuk studi simulasi terhadap model yang dikembangkan pada penelitian ini diberikan pada Bab 4. Bab 5 dibahas tentang hasil-hasil dan pembahasan tentang kajian pendugaan area kecil untuk data cacahan yang akan diterapkan untuk menduga angka kematian bayi level Kecamatan di Provinsi Jawa Timur. Akhirnya, berbagai temuan dari hasil penelitian ini yang diberikan dalam bentuk kesimpulan dan saran disajikan pada Bab 6.

## 2 TINJAUAN PUSTAKA

### Pendahuluan

Pendugaan area kecil (*small area estimation*, SAE) merupakan teknik statistika yang tujuannya untuk memperoleh penduga pada area (*domain*) kecil, dimana penduga survei langsung tidak dapat diandalkan, bahkan kadang-kadang tidak dapat dihitung yang disebabkan oleh terbatasnya ukuran contoh yang tersedia. Pembahasan mengenai berbagai metode statistika untuk memperoleh pendugaan area kecil sudah banyak dilakukan dalam Rao (2003). Pembahasan yang paling utama adalah memperhatikan penggunaan model area kecil yang eksplisit melalui kekuatan peminjaman (*borrow strength*) dari area yang berdekatan menurut ruang atau waktu atau melalui informasi tambahan yang diperkirakan berkorelasi dengan peubah yang diamati.

Penduga taklangsung berdasarkan pada area kecil ini disebut juga penduga *model-based*. Secara umum, model area kecil ini biasanya diklasifikasikan ke dalam dua jenis, yaitu model untuk level area dan model untuk level unit. Model untuk level area akan menghubungkan penduga langsung area kecil ke kovariat dari area tertentu, dan biasanya pemodelan ini dilakukan ketika data untuk level unit tidak tersedia. Sedangkan pada model untuk level unit akan menghubungkan nilai-nilai unit dari peubah yang sedang diamati pada kovariat dari unit tertentu. Asumsi penting dari model level unit adalah bahwa nilai contoh membolehkan model populasi yang diasumsikan, artinya tidak adanya bias seleksi contoh. Sementara itu, untuk model level area diasumsikan tidak adanya penarikan contoh informatif dari area dalam situasi dimana hanya ada beberapa area yang dipilih sebagai contoh, artinya nilai dari area contoh (sebagai penduga langsung) mengasumsikan model populasi.

Berbagai model dasar dalam pendugaan area kecil ini pada umumnya digunakan ketika peubah respons yang diamati adalah berbentuk kontinu. Padahal seringkali dalam menganalisis hubungan antara beberapa peubah, terdapat sejumlah fenomena dimana peubah responnya berbentuk data cacahan. Dalam mengamati suatu fenomena dimana peubah responnya berbentuk cacahan, maka fenomena seperti ini menyangkut banyaknya suatu kejadian yang biasanya diasumsikan mengikuti sebaran Poisson. Selanjutnya, inferensi dari penduga *model-based* ini merujuk pada suatu sebaran tertentu yang harus terpenuhi dari model yang diasumsikan, misalnya mengikuti sebaran Poisson tadi. Oleh karena itu, pemilihan model dan validasi model akan memegang peranan penting dalam pendugaan model-based ini. Apabila model yang diasumsikan tidak memberikan kecocokan yang baik terhadap data, maka penduga *model-based* akan

menjadi bias yang pada akhirnya nanti akan membawa pada kegagalan dalam membuat inferensi yang baik.

Salah satu model yang biasa digunakan dalam pendugaan area kecil untuk data cacahan adalah model campuran loglinear. Namun demikian, saat ini telah mulai banyak dikembangkan beberapa model alternatif dalam pendugaan area kecil untuk data cacahan, dimana salah satunya adalah dengan menggunakan pendekatan Bayesian hirarki (*hierarchical Bayes*, HB). Walaupun penerapan dari metode Bayesian hirarki pada masalah pendugaan area kecil sudah mulai banyak berkembang, akan tetapi masih sedikit apabila dibandingkan dengan pendekatan klasik utama. Sebagai contoh, misalnya tidak banyak pembahasan mengenai spesifikasi model nonlinear mana yang digunakan pada saat penduga area kecil diperlukan untuk peubah non-kontinu (misalnya untuk peubah kategorik atau cacahan).

Pada bab ini akan dibahas mengenai beberapa model yang dijadikan dasar sebagai pengembangan model area kecil yang diterapkan untuk data yang berbentuk cacahan. Pembahasan dimulai dengan perkembangan metode pendugaan area kecil, khususnya yang berkaitan dengan model SAE untuk data cacahan melalui pendekatan Bayes. Pembahasan mengenai model SAE dimulai dari model Fay-Herriot sebagai model dasar dalam masalah pendugaan area kecil. Kemudian, model dasar dalam pengembangan pendugaan area kecil untuk data cacahan adalah model hirarki Bayes, dimana pembahasan difokuskan pada masalah bentuk sebaran prior dan inferensinya. Selanjutnya, dibahas tentang konsep dasar dari model linear terampat serta kaitannya dengan masalah pendugaan area kecil.

### Perkembangan Metode Bayes dalam Pendugaan Area Kecil

Sebagaimana yang telah disebutkan sebelumnya bahwa walaupun berbagai metode statistika untuk memperoleh penduga area kecil sudah banyak dibahas dalam Rao (2003), namun menurut Torelli dan Trevisani (2007) sekarang ini pembahasan mengenai spesifikasi model nonlinear yang tepat pada saat penduga area kecil diperlukan untuk peubah kategorik relatif masih sedikit. Penggunaan pendekatan hirarki Bayes yang saat ini masih terbatas, padahal pendekatan ini mempunyai beberapa keuntungan, diantaranya yaitu: spesifikasi modelnya adalah langsung dan dapat digunakan untuk memodelkan berbagai sumber variasi yang berbeda, masalah inferensinya relatif lebih jelas dan komputasinya juga relatif lebih mudah dengan menggunakan teknik MCMC.

Spesifikasi model Bayes yang diturunkan melalui model klasik SAE, seperti model Fay-Harriot atau lebih tepatnya lagi menggunakan model Poisson linear terampat. Yu dan Rao (2002) memperkenalkan model Normal-logNormal di dalam kelas yang disebut dengan *unmatched model*. Sementara itu, Trevisani dan Torelli (2007) mengusulkan model Normal-Poisson-logNormal dengan alasan bahwa model ini lebih tepat digunakan karena secara eksplisit memperhitungkan sifat-sifat dari peubah yang diamati. Di bawah kondisi yang sesuai berbagai model yang dispesifikasikan bisa memberikan beberapa manfaat. Akan tetapi ketepatan penggunaan model ini sangat bergantung pada berbagai kondisi, seperti ukuran area, ketersediaan peubah penjelas yang baik di tingkat area, akurasi dari penduga ragam penarikan contoh. Pengembangan dari model HB ini menjadi dimungkinkan karena ketersediaan perangkat lunak yang menyediakan fasilitas untuk melakukan simulasi MCMC, sehingga proses pendugaan

model menjadi relatif lebih mudah. Permasalahan yang masih relevan untuk dibahas adalah untuk melakukan inferensi terutama yang berkaitan dengan pemilihan spesifikasi model yang ‘terbaik’.

Trevisani dan Torelli (2007) menyatakan bahwa pengembangan model HB untuk data kategorik dilakukan dengan mengadaptasi berbagai literatur yang membahas tentang pemetaan penyakit (*disease mapping*). Terdapat beberapa artikel tentang penerapan model hirarki Bayes untuk data kategorik ini pada bidang pemetaan penyakit, kesehatan masyarakat, masalah pengangguran, atau masalah mortalitas. Misalnya, Malec, et al. (1997) menggunakan model hirarki untuk melakukan survei interview kesehatan nasional di Amerika Serikat, dimana data yang digunakan adalah data biner. Malec et al. (1997) menggunakan metode Bayes hirarki dibandingkan metode Bayes empiris karena mampu memberikan penduga titik dan ukuran keragaman yang baik, serta mampu meningkatkan presisi dari penduganya. Sementara itu Datta, et al. (1999) menerapkan pendekatan Bayes hirarki untuk menduga tingkat pengangguran di Amerika Serikat. Mereka mengusulkan metode Bayes hirarki dengan menggunakan generalisasi deret waktu dari model cross-sectional yang banyak digunakan dalam pendugaan area kecil.

Model alternatif untuk menduga angka kematian menurut peubah umur tertentu pada salah satu kategori penyakit kanker di Amerika Serikat juga telah dilakukan oleh Nandram, et al. (1999). Mereka menggunakan metode Bayes yang diterapkan pada empat jenis model yang berbeda. Masing-masing model itu mengasumsikan bahwa banyaknya yang mati pada area tertentu, dan kelompok umur tertentu adalah bersebaran Poisson. Hasilnya menunjukkan bahwa model yang mereka usulkan mampu menangkap efek area kecil dan efek regional dengan baik, dan juga mampu mendeteksi korelasi spasial sisanya, sehingga memudahkan proses pendugaan parameternya. Nandram, et al. (2000) juga melakukan pemodelan yang hampir sama dengan apa yang telah dilakukan sebelumnya, akan tetapi di sini modelnya diterapkan untuk menduga angka kematian menurut peubah umur tertentu pada salah satu kategori penyakit *chronic obstructive pulmonary* di Amerika Serikat. Sementara itu, Daniels dan Gatsonis (1999) menerapkan model linear terampat melalui pendekatan Bayes berhirarki pada masalah kesehatan masyarakat di Amerika Serikat.

Diketahui bahwa walaupun penduga sintetik mempunyai ragam yang kecil dibandingkan dengan penduga langsung, tetapi penduga itu cenderung bias karena penduga itu tidak menggunakan informasi pada karakteristik yang diamati yang secara langsung diperoleh dari survei contoh. Kompromi antara penduga langsung dan pendugaan sintetik adalah metode komposit yang menggunakan data contoh survei dalam rangka menghubungkan data sensus dan data administratif yang berbeda. Kemudian Jiang dan Lahiri (2001) mengusulkan suatu metode pendugaan area kecil komposit yang efisien yang dapat diterapkan untuk data biner ketika rancangan penarikan contohnya diabaikan. Metode ini menggunakan metodologi Bayes hirarki, namun secara komputasi agak rumit untuk diterapkan.

Kemudian, Nandram dan Choi (2002) mempertimbangkan tiga buah model hirarki Bayes untuk data nonrespons biner, yang terklusterkan dalam sejumlah domain atau area. Sebagaimana dalam pendugaan area kecil, satu hal yang penting adalah bahwa setiap model ‘meminjam kekuatan’ menurut area melalui pendekatan pemilihan pada nonrespons. Model pertama mengasumsikan bahwa mekanisme nonrespons diabaikan, dan model yang kedua bahwa ini tidak dapat diabaikan. Mereka memberikan alasan bahwa perluasan model diskrit menjadi kurang tepat untuk memodelkan ketidakpastian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

mengenai data yang terabaikan ini. Oleh karena itu mereka mengusulkan model ketiga melalui perluasan model kontinu pada odds rasio untuk setiap area. Pada saat odds rasio sama dengan satu, maka model dapat diabaikan, dan jika tidak maka model tidak dapat diabaikan. Namun dari hasil analisis menunjukkan bahwa inferensinya tidak begitu sensitif terhadap asumsi sebaran yang penting.

Telah banyak artikel yang membahas kelebihan dari model hirarki Bayes dalam pemodelan pemetaan penyakit dibandingkan dengan model Bayes empiris. Alasan utama mengapa Bayes empiris kurang baik adalah bahwa model tersebut tidak memperhitungkan keragaman dalam pendugaan pada *hyperparameter*. Namun demikian, MacNab, et al. (2004) telah mengusulkan suatu metodologi Bayes empiris bootstrap untuk membuat inferensi tentang resiko relatif yang nantinya menghasilkan selang kepercayaan Bayes empiris yang akurat. Model tersebut kemudian diterapkan pada pendugaan angka kematian bayi di suatu provinsi di Kanada.

Saat ini, penerapan pendugaan area kecil untuk data kategorik, terutama untuk data biner, sudah mulai banyak dilakukan. Sebagai contoh, misalnya Munnich, et al. (2009) menggunakan metodologi pendugaan area kecil untuk menduga hasil sensus penduduk di Jerman dengan menggunakan model binomial campuran. Sementara itu Bizier, et al. (2009) menerapkan pendugaan area kecil untuk memperoleh penduga bagi banyaknya orang dengan angka ketidakmampuan dan kemampuan pada daerah kesehatan tertentu. Mereka menggunakan log linear unmatched model pada pendekatan Bayes hirarki bersamaan dengan metode penarikan contoh Gibb.

Berbagai penelitian mengenai penerapan model SAE untuk data kategorik melalui pendekatan model Bayes berhirarki di Indonesia relatif sudah mulai banyak dikembangkan. Kismiantini et al. (2006) menggunakan model SAE untuk menduga resiko demam berdarah dengan menggunakan model Poisson-gamma tetapi melalui pendekatan EB. Kurnia dan Notodiputro (2007, 2008) menduga suatu parameter melalui pendekatan *generalized additive mixed models* (GAMM). Selain itu, Kurnia dan Notodiputro juga melakukan kajian tentang penerapan metode Bayes pada pendugaan area kecil, serta Kurnia dan Notodiputro (2006) menerapkan pendekatan EB dan EBLUP dalam masalah pendugaan kuadrat tengah galat pada data BPS. Sementara itu, Sadik dan Notodiputro (2007) mengkaji tentang model *state space* pada *generalized linear mixed model* untuk menyelesaikan masalah pendugaan are kecil. Selain itu, Sadik dan Notodiputro (2009) menerapkan pendekatan Bayes untuk data deret waktu dan *cross-sectional*.

Perkembangan terakhir dari pengembangan model SAE adalah melalui pendekatan Bayes adalah sebagaimana yang dilakukan oleh Rumiati, et al. (2012) yang memfokuskan pada pengembangan model SAE untuk peubah respon binomial dan multinomial berbasis pada penarikan contoh berpeluang tidak sama. Penelitian yang dilakukan oleh Rumiati et al. (2012) dilakukan dengan tujuan untuk mengembangkan model SAE untuk menduga Indeks Pendidikan yang merupakan salah satu komponen dalam Indeks Pembangunan Manusia (IPM). Indeks Pendidikan diukur angka melek huruf dan rata-rata lama sekolah di suatu wilayah. Angka melek huruf diukur dengan proporsi penduduk berusia 10 tahun ke atas yang bisa baca tulis, sedangkan rata-rata lama sekolah diukur dari proporsi penduduk berusia 10 tahun ke atas di tiap level pendidikan tertentu. Proporsi penduduk berusia 10 tahun ke atas yang bisa baca tulis adalah parameter dari sebuah sebaran binomial, sedangkan proporsi penduduk berusia 10 tahun ke atas di tiap level pendidikan tertentu merupakan parameter dari sebaran multinomial. Sementara itu, Hajarisman et al. (2012) mengimplementasikan model

Poisson Bayes berhirarki dua-level untuk menduga angka kematian Bayi level kecamatan di Provinsi Jawa Barat, namun dalam penelitian tersebut hanya difokuskan pada suatu sebaran prior yang bersifat *conjugate*. Dalam penelitian yang ditulis dalam disertasi ini adalah mengembangkan model Bayes berhirarki untuk data cacahan yang bersebaran Poisson untuk menduga angka kematian level kecamatan di Indonesia dengan memperhatikan dua buah sebaran prior yang berbeda, yaitu yang bersifat *conjugate* dan *non-conjugate*.

### Model Fay-Herriot

Perhatikan suatu populasi berukuran  $N$  yang dipartisi ke dalam  $m$  himpunan bagian yang masing-masing berukuran  $N_1, \dots, N_m$ , yang selanjutnya disebut sebagai area atau domain. Misalkan  $Y_{ij}$  merupakan suatu nilai dari peubah yang diamati pada unit ke- $j$  untuk area ke- $i$ . Tujuannya adalah untuk memperoleh rata-rata dari  $m$  area yang memenuhi persamaan berikut:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}, \text{ untuk } i = 1, 2, \dots, m \quad \dots (2.1)$$

Suatu contoh berukuran  $n$  diambil dari populasi dengan menggunakan rancangan penarikan contoh tertentu. Penduga langsung, sebut saja  $\hat{\theta}_i$ , bagi parameter  $\mu_i$  merupakan penduga yang hanya menggunakan data contoh dari area  $m$ . Melalui rancangan seperti ini maka menurut Fay dan Herriot (1979) penduga langsung untuk area-area dengan data contoh yang kecil akan mempunyai ragam yang terlalu besar. Untuk mengurangi ragam tersebut, dalam konteks pendugaan area kecil, Fay-Herriot (1979) mengasumsikan bahwa parameter  $\theta_i = g(\mu_i)$  untuk beberapa  $g(\cdot)$  yang dihubungkan pada data tambahan dalam area tertentu  $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})^T$  melalui model linear

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i, \text{ untuk } i = 1, 2, \dots, m \quad \dots (2.2)$$

dimana  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  adalah vektor dari koefisien regresi berukuran  $p \times 1$ , serta  $v_i$  adalah efek acak dari area tertentu yang saling bebas dan bersebaran identik dengan nilai harapan model  $E_m(v_i) = 0$  dan ragam model  $V_m(v_i) = \sigma_v^2 (\geq 0)$ .

Dalam beberapa aplikasi, tidak seluruh area dipilih ke dalam contoh. Misalkan terdapat  $M$  area dalam populasi dan hanya terdapat  $m$  area yang dipilih dalam contoh. Diasumsikan suatu model dalam bentuk seperti yang dinyatakan dalam (2.2) untuk populasi, yaitu  $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$ , untuk  $i = 1, 2, \dots, M$ . Lebih jauh diasumsikan bahwa contoh area memenuhi model populasi, artinya tidak terjadi bias dalam contoh yang terpilih sehingga model dalam Persamaan (2.2) terpenuhi untuk contoh area.

Untuk membuat inferensi mengenai rata-rata area area kecil  $Y_{ij}$  di bawah model (2.2), diasumsikan bahwa telah tersedia penduga langsung  $\hat{\mu}_i$ . Menurut Rao (2003) sebagaimana metode James-Stein diasumsikan bahwa

$$\hat{\theta}_i = g(\hat{\mu}_i) = \theta_i + e_i, \text{ untuk } i = 1, 2, \dots, m \quad \dots (2.3)$$

dimana  $e_i$  adalah galat penarikan contoh yang saling bebas dengan nilai harapan dan ragamnya masing-masing adalah  $E_m(e_i|\theta_i) = 0$  dan  $V_m(e_i|\theta_i) = \psi_i$ .

Pada umumnya diasumsikan pula bahwa ragam penarikan contoh,  $\psi_i$  adalah diketahui. Dalam beberapa aplikasi tertentu asumsi bahwa ragam penarikan contoh

diketahui sulit dipenuhi. Sebagai contoh, misalnya penduga langsung  $\hat{\theta}_i$  biasa jadi bias rancangan (*design-biased*) untuk  $\theta_i$  apabila fungsi  $g(\cdot)$  merupakan fungsi yang non-linear dan contoh area  $n_i$  adalah kecil. Galat penarikan contoh  $e_i$  tidak selalu harus saling bebas apabila area kecil diambil menurut rancangan penarikan contoh. Asumsi mengenai ragam  $\psi_i$  diketahui dapat dilonggarkan dengan cara menduga  $\psi_i$  dari data contoh level unit, kemudian lakukan pemulusan terhadap penduga ragam  $\hat{\psi}_i$  untuk memperoleh penduga  $\psi_i$  yang lebih stabil. Asumsi mengenai kenormalan bagi penduga  $\hat{\theta}_i$  juga sering harus dipenuhi, akan tetapi asumsi ini tidak terlalu ketat sebagaimana asumsi kenormalan dari efek acak. Hal ini bisa terjadi karena adanya dalil limit pusat terhadap  $\hat{\theta}_i$ .

Model deterministik pada  $\theta_i$  diperoleh dengan cara menetapkan  $\sigma_v^2 = 0$ , sehingga diperoleh  $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ . Model tersebut akan membawa pada penduga sintetik yang sama sekali tidak memperhitungkan keragaman lokal dibandingkan dengan keragaman yang dicerminkan dalam peubah tambahan  $\mathbf{x}_i$ . Dengan mengkombinasikan model dalam Persamaan (2.2) dan (2.3) diperoleh

$$\hat{\theta}_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i, \text{ untuk } i = 1, 2, \dots, m \quad \dots (2.4)$$

Perlu dicatat bahwa model dalam (2.4) melibatkan dua buah galat, yaitu galat rancangan terinduksi (*design-induced*)  $e_i$  dan galat model  $v_i$ , dimana dalam hal ini diasumsikan bahwa  $e_i$  dan  $v_i$  adalah saling bebas.

Asumsi bahwa  $E_p(e_i|\theta_i) = 0$  dalam model penarikan contoh yang diberikan dalam Persamaan (2.3) bisa jadi tidak valid apabila ukuran contoh  $n_i$  dalam area ke- $i$  dan parameter  $\theta_i$  merupakan fungsi yang non-linear dari total  $Y_i$  bahkan jika penduga langsung  $\hat{\mu}_i$  merupakan penduga yang takbias rancangan. Model penarikan contoh yang lebih realistis adalah

$$\hat{Y}_i = Y_i + e_i^*, \text{ untuk } i = 1, 2, \dots, m \quad \dots (2.5)$$

dengan  $E_p(e_i^*|\theta_i) = 0$ , artinya bahwa  $\hat{Y}_i$  adalah takbias rancangan bagi total  $Y_i$ . Dalam hal ini model penarikan contoh dan model penghubung tidak dipasangkan. Akibatnya, model yang diberikan dalam (2.5) tidak dapat dikombinasikan dengan model dalam (2.2) untuk menghasilkan model campuran linear sebagaimana yang diberikan dalam Persamaan (2.4). Oleh karena itu pemodelan berdasarkan teori model campuran linear yang menggunakan pendekatan EBLUP tidak dapat diterapkan di sini. Menurut Rao (2003) untuk menangani masalah penarikan contoh yang tidak dipasangkan (*unmatched penarikan contoh*) dan model penghubung ini dapat diselesaikan melalui pendekatan Bayes berhirarki.

## Inferensi Bayes Berbasis Model

### Sebaran Prior

Statistik Bayes berbeda dengan teori statistik klasik karena seluruh parameter yang tidak diketahui dipandang sebagai suatu peubah acak. Berdasarkan alasan tersebut, sebaran prior harus didefinisikan pada awal pemodelan. Sebaran prior ini menyatakan informasi yang tersedia bagi peneliti sebelum sembarang data dilibatkan dalam analisis statistika. Dalam hal ini yang menjadi perhatian utama adalah perhitungan sebaran

posterior  $f(\theta|\mathbf{y})$  dari parameter  $\theta$  bersyarat pada data pengamatan  $\mathbf{y}$ . Menurut teorema Bayes, sebaran posterior dapat dituliskan sebagai

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})} \propto f(\mathbf{y}|\theta)f(\theta) \quad \dots (2.6)$$

Sebaran posterior mewakili informasi prior maupun data pengamatan, yang masing-masing dinyatakan oleh sebaran prior  $f(\theta)$  dan fungsi likelihoodnya.

Spesifikasi sebaran prior dalam inferensi Bayes adalah sangat penting karena sebaran prior ini akan mempengaruhi inferensi mengenai sebaran posteriornya. Pada umumnya spesifikasi sebaran prior ditekankan pada parameter rata-rata dan ragamnya. Rata-rata prior memberikan suatu penduga titik prior untuk parameter yang diamati, sedangkan ragam prior menyatakan ketidakpastian mengenai penduga tersebut. Apabila secara a priori peneliti mempunyai keyakinan yang kuat bahwa penduga tersebut adalah akurat, maka ragam seharusnya mempunyai nilai yang kecil. Sementara itu dengan mengabaikan atau ketidakpastian yang begitu besar berkenaan dengan rata-rata prior dapat dilihat dari ragam yang besar. Apabila informasi mengenai prior ini tersedia, maka hal ini dapat diringkas dengan tepat menurut sebaran priornya.

Namun demikian, pada umumnya informasi mengenai priornya tidak tersedia. Dalam kasus seperti ini perlu dispesifikasikan suatu sebaran prior tertentu yang tidak akan berpengaruh terhadap sebaran posterior. Sebaran semacam itu seringkali disebut sebagai sebaran prior *noninformative*. Banyak peneliti menggunakan sebaran prior yang *noninformative* ini karena sebaran tersebut dianggap lebih objektif. Namun demikian, prior yang bersifat *noninformative* ini tidak sepenuhnya dapat menggambarkan total kekeliruan atau galat dari parameter yang sedang diamati. Dalam beberapa kasus, prior yang *noninformative* dapat membawa pada sebaran posterior yang bersifat *improper*, artinya fungsi kepekatan peluang yang tidak dapat diintergralkan, yang pada akhirnya tidak dapat membuat inferensi berdasarkan pada sebaran posterior yang bersifat *improper*.

Sifat sebaran prior lainnya adalah sebaran prior yang bersifat *improper*. Istilah *improper* di sini maksudnya adalah bahwasanya tidak terintegrasi pada satu. Dengan kata lain suatu sebaran prior  $f(\theta)$  disebut *improper* apabila  $\int f(\theta)d\theta = \infty$ . Prior yang bersifat *improper* sering juga digunakan dalam inferensi Bayes, karena prior semacam ini biasanya dapat menghasilkan prior yang bersifat *noninformative* dan sebaran posterior yang bersifat *proper*. Namun dalam beberapa kasus, sebaran prior yang *improper* ini dapat membawa pada sebaran posterior yang *improper*. Untuk menentukan apakah suatu sebaran posterior adalah *proper*, dapat dilakukan dengan cara konstanta yang dinormalkan melalui  $\int f(\mathbf{y}|\theta)f(\theta)d\theta$  adalah *finite* untuk semua  $\mathbf{y}$ . Apabila suatu sebaran prior yang *improper* menghasilkan sebaran posterior yang *improper*, maka inferensi Bayes berdasarkan sebaran posterior yang *improper* ini menjadi tidak valid.

Selanjutnya, suatu prior dikatakan sebagai suatu prior yang bersifat *conjugate* untuk keluarga dari sebaran apabila sebaran prior dan posteriornya berasal keluarga yang sama. Artinya bahwa bentuk dari sebaran posterior mempunyai bentuk sebaran yang sama sebagai sebaran prior. Lebih jauh, dapat dikatakan bahwa sebaran prior merupakan anggota dari keluarga sebaran  $D$  dengan parameter  $\theta$  dikatakan *conjugate* pada sebaran  $f(\mathbf{y}|\theta)$  jika menghasilkan sebaran posterior  $f(\theta|\mathbf{y})$  juga merupakan anggota dari keluarga sebaran yang sama. Dengan demikian jika  $f(\theta) \sim D(\omega_0)$ , maka  $f(\theta|\mathbf{y}) \sim D(\omega)$ , dimana  $\omega_0$  dan  $\omega$  masing-masing menyatakan parameter prior dan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Institut Pertanian Bogor (IPB) Bogor Agricultural University

posterior dari sebaran  $D$ . Dalam beberapa kasus sederhana, parameter posterior dapat dinyatakan sebagai rata-rata terboboti dari parameter prior dan penduga kemungkinan maksimum. Sebaliknya, apabila suatu prior dikatakan sebagai suatu prior yang bersifat *nonconjugate* untuk keluarga dari sebaran tertentu apabila sebaran prior dan posteriornya bukan berasal keluarga yang sama.

### Inferensi Bayes

Setelah sebaran prior dispesifikasikan, kemudian dengan menggunakan teknik tertentu melalui teorema Bayes akan diperoleh sebaran posterior dari parameter model atau untuk memperoleh fungsi densitas prediktif dari pengamatan yang akan datang. Untuk melihat perilaku dari sebaran posterior ini pertama dapat dilihat secara grafik melalui fungsi densitas (atau sebaran kumulatifnya) sebagai dasar inferensi untuk semua posterior. Akan tetapi fungsi ini dalam beberapa hal sulit diinterpretasikan. Oleh karena itu perlu suatu pendekatan untuk membuat ringkasan dari sebaran posterior ini. Secara khusus, menurut Carlin dan Louis (2000) pengembangan inferensi Bayes analog dengan apa yang dilakukan dalam teknik *frequentist*, yaitu melakukan analisis yang meliputi masalah pendugaan titik dan pendugaan selang.

Pertama, untuk memperoleh penduga titik  $\hat{\theta}(\mathbf{y})$  bagi parameter  $\theta$ , maka perlu memilih ringkasan statistik bagi sebaran posterior  $f(\theta|\mathbf{y})$ , seperti rata-rata, median, atau modus. Menurut Gelman (2004) perlu dicatat bahwa ketika prior  $f(\theta)$  adalah berbentuk datar (*flat*), maka modus posterior akan sama dengan penduga kemungkinan maksimum bagi parameter  $\theta$ . Oleh karena itu modus posterior kadang-kadang dirujuk juga sebagai penduga kemungkinan maksimum terampat (*generalized maximum likelihood estimate, GML*) bagi parameter  $\theta$ .

Untuk densitas posterior yang simetris, maka rata-rata dan median tentu saja akan identik. Sedangkan untuk posterior yang tidak simetris, pemilihan ringkasan statistiknya menjadi kurang jelas, walaupun median biasanya lebih banyak dipilih, karena nilai median akan berada diantara modus (yang merupakan nilai maksimum dari densitasnya) dan rata-rata (yang seringkali memberikan bobot yang tinggi pada suatu nilai yang ekstrim).

Untuk memperoleh ukuran akurasi dari penduga titik  $\hat{\theta}(\mathbf{y})$  bagi parameter  $\theta$ , maka perlu menghitung ragam posterior bagi  $\hat{\theta}(\mathbf{y})$ ,  $E_{\theta|\mathbf{y}}(\theta - \hat{\theta}(\mathbf{y}))^2$ . Misalkan rata-rata posterior dinotasikan sebagai

$$E_{\theta|\mathbf{y}}(\theta) = \mu(\mathbf{y}) = \mu \quad \dots (2.7)$$

maka akan diperoleh

$$\begin{aligned} E_{\theta|\mathbf{y}}(\theta - \hat{\theta}(\mathbf{y}))^2 &= E_{\theta|\mathbf{y}}(\theta - \mu + \mu - \hat{\theta}(\mathbf{y}))^2 \\ &= E_{\theta|\mathbf{y}}\left[(\theta - \mu)^2 + 2(\theta - \mu)(\mu - \hat{\theta}(\mathbf{y})) + (\mu - \hat{\theta}(\mathbf{y}))^2\right] \\ &= V_{\theta|\mathbf{y}}(\theta) + 2[\mu - \hat{\theta}(\mathbf{y})][E_{\theta|\mathbf{y}}(\theta) - \mu] + (\mu - \hat{\theta}(\mathbf{y}))^2 \\ &= V_{\theta|\mathbf{y}}(\theta) + (\mu - \hat{\theta}(\mathbf{y}))^2 \end{aligned}$$

Dengan demikian dapat ditunjukkan bahwa rata-rata posterior  $\mu$  akan meminimumkan ragam posterior terhadap  $\hat{\theta}(\mathbf{y})$  untuk seluruh penduga titik  $\hat{\theta}(\mathbf{y})$ . Lebih jauh nilai minimum ini adalah ragam posterior bagi parameter  $\theta$ , yang dinyatakan sebagai

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



$$V_{\theta|y}(\theta) = E_{\theta|y}[\theta - E_{\theta|y}(\theta)]^2 \quad \dots (2.8)$$

Menurut Carlin dan Louis (2000) pernyataan bahwa rata-rata posterior dipandang sebagai penduga titik, hal ini juga berlaku karena secara historis hanya rata-rata dan ragam posterior yang ditampilkan sebagai hasil dari analisis Bayes.

Kedua, menurut Carlin dan Louis (2000) selang kepercayaan Bayes analog dengan selang kepercayaan *frequentist*, dimana dalam konteks Bayes selang kepercayaan ini dirujuk sebagai *credible set*. Jadi, suatu selang kepercayaan  $100 \times (1 - \alpha)\%$  (atau disebut juga sebagai *credible set*) bagi parameter  $\theta$  merupakan himpunan bagian  $C$  dari ruang parameter  $\Theta$  sedemikian rupa sehingga memenuhi persamaan berikut:

$$1 - \alpha \leq P(C|y) = \int_C f(\theta|y)d\theta \quad \dots (2.9)$$

Definisi mengenai selang kepercayaan Bayes ini dapat diinterpretasikan sebagai pernyataan peluang mengenai likelihood bagi parameter  $\theta$  yang jatuh pada wilayah  $C$ , yaitu: "peluang bahwa parameter  $\theta$  berada dalam  $C$  bersyarat pada data pengamatan  $y$  paling tidak sebesar  $(1 - \alpha)$ ". Interpretasi ini sedikit berbeda dengan pernyataan selang kepercayaan yang biasa, dimana selang kepercayaan frequentist ini dapat diinterpretasikan sebagai berikut: "jika  $C$  dapat dihitung kembali unruk sejumlah besar gugus data yang dikumpulkan dengan cara yang sama, maka sekitar  $(1 - \alpha) \times 100\%$  akan berisi nilai sebenarnya bagi parameter  $\theta$ . Lebih jauh, dalam konteks pemodelan Bayes berhirarki, selang kepercayaan Bayes disebut juga dengan selang *highest posterior density* (HPD).

### Model Bayes Berhirarki

Menurut Congdon (2010) inferensi melalui pendekatan Bayes difokuskan pada pemutakhiran informasi tentang parameter  $\theta$  yang tidak diketahui, yang berdasarkan pada data pengamatan  $y$ , bersamaan dengan informasi yang telah direvisi yang dinyatakan melalui kepekatan posterior,  $f(\theta|y)$ . Contoh pengamatan  $y$  yang dianalisis memberikan informasi baru mengenai parameter  $\theta$ , sedangkan kepekatan prior  $f(\theta)$  merupakan informasi yang terakumulasi mengenai  $\theta$  sebelum analisis terhadap  $y$  dilakukan. Pengujian hipotesis pada parameter didasarkan pada peluang posterior bersyarat pada data pengamatan.

Perkembangan metode Bayes dalam berbagai aplikasi semakin meningkat seiring dengan meningkatnya perkembangan algoritma MCMC untuk pendugaan model (Gilks et al., 1996), dengan cara mengambil parameter contoh secara berulang berdasarkan pada sebaran posterior dari model statistika yang telah dispesifikasikan sebelumnya. Pendugaan parameter berbasis penarikan contoh melalui MCMC memberikan kepekatan posterior lengkap dari parameter sehingga sembarang pola sebaran yang non-normal menjadi lebih jelas terlihat, serta pengujian hipotesis mengenai parameter atau penduga selang dapat ditentukan melalui contoh MCMC tanpa harus memenuhi asumsi kenormalan yang biasa ditemui dalam pendugaan *frequentist*.

Kaitannya pendekatan Bayes dengan masalah pendugaan area kecil, Rao (2003) telah membahas secara rinci tiga jenis pendekatan Bayes yang biasa diterapkan pada masalah pendugaan area kecil. Ketiga pendekatan itu adalah prediksi takbias linear terbaik empirik (*empirical best linear unbiased prediction*, EBLUP), Bayes empirik (*empirical Bayes*, EB), serta Bayes berhirarki (*hierarchical Bayes*, HB). Metode

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Bogor Agricultural University

EBLUP ini dapat diterapkan untuk model campuran linear (*linear mixed models*) yang menangani masalah model dasar untuk level area dan level unit. Sebaliknya, metode EB dan HB merupakan metode yang lebih umum untuk diterapkan karena dapat diterapkan pada GLMM yang digunakan untuk menangani data kategorik, seperti data biner dan data cacahan. Besaran KTG biasanya digunakan sebagai ukuran keragaman dari peubah di bawah pendekatan EBLUP dan EB, sedangkan pendekatan HB menggunakan ragam posterior sebagai ukuran keragamannya dengan mengasumsikan suatu sebaran prior tertentu pada parameter model.

Dalam penelitian ini pembahasan difokuskan pada pendekatan HB, dimana saat ini pemodelan Bayes secara berhirarki dan model efek acak melalui teknik MCMC telah banyak diterapkan dalam analisis data modern. Dalam pendekatan HB, ditentukan terlebih dahulu sebaran prior subjektif  $f(\theta)$  pada parameter model  $\theta$ , kemudian untuk gugus data  $y$  tertentu akan diperoleh sebaran posterior  $f(\mu|y)$  dari parameter area kecil (acak)  $\mu$  yang diamati. Model dua-tahap, katakan  $f(y|\mu, \theta_1)$  dan  $f(y|\mu, \theta_2)$  dikombinasikan dengan prior subjektif pada  $\theta = (\theta_1^T, \theta_2^T)$ , dengan menggunakan teorema Bayes, sehingga diperoleh posterior  $f(\mu|y)$ . Inferensi yang berdasarkan pada  $f(\mu|y)$ , secara khusus pada suatu parameter yang diamati, sebut saja  $\phi = h(\mu)$ , diduga dengan menggunakan rata-rata posteriornya sebagaimana yang ditunjukkan dalam Persamaan (2.7), serta ragam posterior yang dinyatakan dalam Persamaan (2.8) yang digunakan sebagai ukuran presisi penduga.

Pendekatan HB merupakan metode yang bersifat langsung, dan inferensi dari model HB ini relatif jelas dan bersifat eksak, akan tetapi memerlukan spesifikasi prior subjektif  $f(\theta)$  pada parameter model  $\theta$ . Prior pada parameter  $\theta$  dapat menjadi informatif atau sebaliknya bersifat ‘membaur’. Prior yang informatif dapat diperoleh dari informasi prior yang lengkap, misalnya berdasarkan hasil studi sebelumnya yang relevan dengan gugus data yang sedang dipelajari. Namun demikian, dalam prakteknya pada penerapan model HB prior yang informatif jarang tersedia, khususnya yang berhubungan dengan kebijakan publik. Prior yang noninformatif atau baur dirancang untuk mencerminkan kurangnya informasi mengenai  $\theta$ . Pemilihan prior baur bersifat tidak unik, dan beberapa prior baur yang keliru tentu saja akan menghasilkan posterior yang juga keliru. Oleh karena itu adalah sangat penting untuk memilih prior baur,  $f(\theta)$ , secara hati-hati sehingga akan menghasilkan posterior  $f(\mu|y)$ . Selain itu, diperlukan pemilihan prior baur sedemikian rupa sehingga mampu menghasilkan inferensi yang valid di bawah kerangka kerja yang umum. Dalam prakteknya, bias  $E(\hat{\phi}^{HB} - \phi)$  dari penduga HB dan bias relatif dari ragam posterior yang merupakan penduga dari  $MSE(\hat{\phi}^{HB})$  akan bernilai kecil. Selain itu, cakupan dari selang kepercayaan HB pada parameter  $\phi$  harus mendekati nilai nominalnya.

Berdasarkan teorema Bayes diperoleh

$$f(\mu, \theta|y) = \frac{f(y, \mu|\theta)f(\theta)}{f_1(y)} \quad \dots (2.10)$$

dimana  $f_1(y)$  merupakan fungsi densitas marginal dari  $y$  yang dinyatakan dalam bentuk

$$f_1(y) = \int \int f(y, \mu|\theta)f(\theta)d\mu d\theta \quad \dots (2.11)$$

Selanjutnya, fungsi kepekatan posterior  $f(\mu|y)$  diperoleh melalui Persamaan (2.10) sebagai

$$f(\mu|y) = \int f(\mu, \theta|y) d\theta = \int f(\mu|y, \theta)f(\theta|y) d\theta \quad \dots (2.12)$$

Bentuk yang diberikan dalam Persamaan (2.12) menunjukkan bahwa  $f(\mu|y)$  merupakan campuran dari fungsi densitas bersyarat  $f(\mu|y, \theta)$ , dan perlu dicatat bahwa  $f(\mu|y, \theta)$  juga digunakan dalam inferensi Bayes empirik (EB).

Berdasarkan Persamaan (2.10) dan (2.12) terlihat bahwa evaluasi dari  $f(\mu|y)$  dan besaran posterior lainnya seperti  $E[h(\mu)|y]$  melibatkan integrasi multi-dimensi. Namun demikian, seringkali dimungkinkan untuk membentuk integrasi secara analitik dengan memperhatikan beberapa komponen dari  $\mu$  dan  $\theta$ . Apabila masalah reduksi melibatkan integrasi satu atau dua dimensi maka integrasi numerik secara langsung dapat digunakan untuk menghitung besaran posterior yang diinginkan. Akan tetapi untuk masalah yang lebih kompleks diperlukan mengevaluasi integral berdimensi tinggi. Saat ini telah banyak dikembangkan metode Markov Chain Monte Carlo (MCMC) untuk mengatasi masalah kesulitan komputasinya. Metode ini membangkitkan contoh dari sebaran posterior, kemudian menggunakan contoh yang disimulasikan tersebut untuk memperoleh besaran posterior yang diinginkan. Beberapa algoritma yang dikembangkan untuk proses numerik dalam metode MCMC ini diantaranya adalah algoritma Gibbs penarikan contoh dan algoritma Metropolis-Hastings (M-H).

Model Bayes berhirarki salah satunya banyak diterapkan pada bidang kesehatan, terutama yang berkaitan dengan pemetaan penyakit, sebagaimana yang dibahas secara rinci oleh Lawson (2009). Salah satu contoh penerapan model Bayes berhirarki yang banyak dikembangkan dalam pemetaan penyakit adalah dimana likelihood data bersebaran Poisson dan parameter resikonya menggunakan sebaran prior gamma.

$$f(\theta|y) = L(y|\theta)f(\theta)$$

dimana  $f(\theta)$  adalah kepekatan dari sebaran gamma dengan parameter,  $a$  dan  $b$ , serta  $L(y|\theta) = \prod_{i=1}^m \{(e_i\theta)^{y_i} \exp(-e_i\theta)\}$  adalah fungsi likelihood yang hanya bergantung pada data. Model ini dapat dituliskan sebagai

$$y_i|\theta \sim \text{Poisson}(e_i\theta)$$

$$\theta \sim G(a, b)$$

Parameter  $a$  dan  $b$  diasumsikan mempunyai suatu nilai tertentu, akan tetapi pada umumnya para peneliti tidak punya keyakinan yang kuat dalam menentukan nilai parameter priornya. Oleh karena itu, sebagaimana parameter dalam model dipandang sebagai suatu proses stokastik. Dengan demikian parameter tersebut mempunyai sebaran peluang yang dikenal sebagai sebaran *hyperprior*, dan parameternya disebut juga sebagai *hyperparameter*.

Gagasan bahwa suatu nilai dari parameter dapat muncul dari sebaran merupakan konsep dasar dari metodologi Bayes dan secara alamiah akan membawa pada penggunaan model dimana parameternya muncul dalam hirarki. Dalam model Poisson-gamma terdapat dua level hirarki, yaitu  $\theta$  mempunyai sebaran  $G(a, b)$  pada level pertama dari hirarki, serta parameter  $a$  dan  $b$  masing-masing akan mempunyai sebaran *hyperprior* ( $h_a$ ) dan ( $h_b$ ) pada level kedua dari bentuk hirarkinya. Hal ini dapat dituliskan sebagai:

$$y_i|\theta \sim \text{Poisson}(e_i\theta)$$

$$\theta|a, b \sim G(a, b)$$

$$a|v \sim h_a(v)$$

$$b|\rho \sim h_b(\rho)$$

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Bogor Agricultural University

Lebih jauh Lawson (2009) menyebutkan bahwa sangatlah penting untuk menghentikan bentuk hirarki pada tempat yang tepat, dan jika ini tidak dilakukan maka tentu saja akan mempunyai bentuk hirarki yang tidak terbatas. Pada umumnya yang dijadikan titik potongnya adalah apakah keragaman dalam parameter tidak berpengaruh pada level dari model hirarki yang lebih bawah. Pada titik tersebut maka parameter diasumsikan sebagai tetap. Sebagai contoh dalam model Poisson-gamma apabila  $a$  dan  $b$  diasumsikan tetap, maka prior gamma juga akan tetap dan pemilihan parameter  $a$  dan  $b$  menjadi tidak terinformasikan, yang berarti bahwa data tidak memberikan informasi mengenai sebarannya secara keseluruhan. Namun demikian, dengan memandang adanya keragaman pada level yang lebih tinggi, yaitu pada *hyperprior*  $a$  dan  $b$ , maka nilai  $\rho$  dan  $\rho$  dapat ditetapkan tanpa mempengaruhi keragaman pada level yang bawahnya.

### Model Linear Terampat

Istilah model linear terampat (*generalized linear model*, GLM) merujuk pada kelas model yang lebih luas yang dipopulerkan oleh McCullagh dan Nelder (1983). Dalam model ini, peubah respons  $Y_i$  diasumsikan mengikuti sebaran dari keluarga eksponensial dengan rata-rata  $\mu_i$ , yang biasanya diasumsikan sebagai suatu fungsi (seringkali bentuknya nonlinear) dari  $\mathbf{x}_i^T \boldsymbol{\beta}$ . Beberapa penulis menyatakan bentuknya adalah nonlinear karena  $\mu_i$  seringkali merupakan fungsi non-linear dari kovariat, tetapi McCullagh dan Nelder (1983) mempertimbangkan fungsi tersebut sebagai bentuk yang linear, karena kovariat ini mempengaruhi sebaran dari  $Y_i$  hanya melalui kombinasi linear dari  $\mathbf{x}_i^T \boldsymbol{\beta}$ .

Model linear terampat didefinisikan dalam bentuk segugus peubah acak  $Y_1, \dots, Y_N$  yang saling bebas, dan masing-masing peubah acak tersebut mempunyai sebaran yang sama dari keluarga eksponensial dengan sifat-sifat sebagai berikut:

1. Sebaran dari masing-masing  $Y_i$  adalah berbentuk kanonik dan bergantung pada satu parameter tunggal  $\theta_i$ , yang fungsi peluangnya dapat dituliskan sebagai berikut:

$$f(y | \theta, \phi) = \exp[(y\theta - b(\theta))/a(\phi) + c(y, \phi)] \quad \dots (2.13)$$

2. Sebaran dari semua peubah acak  $Y_i$  mempunyai bentuk yang sama (misalnya semuanya bersebaran normal atau semuanya binomial).

Jika parameter  $\phi$  diketahui, maka hal ini merupakan model keluarga eksponensial dengan parameter kanonik  $\theta$ . Akan tetapi apabila  $\phi$  tidak diketahui, maka model ini merupakan keluarga eksponensial dua-parameter.

Untuk pembentukan model, parameter  $\theta_i$  bukan merupakan parameter yang diamati secara langsung. Untuk menspesifikasikan model linear terampat, perhatikan segugus parameter yang lebih kecil  $\beta_1, \dots, \beta_p$  ( $p < N$ ). Misalkan bahwa  $E(Y_i) = \mu_i$ , dimana  $\mu_i$  merupakan suatu fungsi bagi  $\theta_i$ . Dalam GLM dilakukan transformasi bagi  $\mu_i$  sedemikian rupa sehingga diperoleh

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

dimana  $g$  merupakan fungsi monoton dan mempunyai turunan yang disebut juga sebagai fungsi hubung (*link function*);  $\mathbf{x}_i$  merupakan vektor dari peubah penjelas yang berukuran  $p \times 1$  (kovariat dan peubah boneka (*dummy variables*) untuk taraf faktor); serta  $\boldsymbol{\beta}$  adalah vektor parameter berukuran  $p \times 1$ , dimana  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$ .

Model linear terampat mempunyai tiga buah komponen, yaitu:

1. Peubah respons  $Y_1, \dots, Y_N$  yang diasumsikan merupakan anggota dari sebaran keluarga eksponensial;
2. Segugus parameter  $\beta$  dan peubah penjelas  $\mathbf{X} = [x_1^T, x_2^T, \dots, x_N^T]$ ;
3. Fungsi hubung  $g$  yang bersifat monoton sedemikian rupa sehingga  $g(\mu_i) = \mathbf{x}_i^T \beta$ , dimana  $\mu_i = E(Y_i)$ .

Dalam kaitannya model linear terampat ini dengan model pendekatan Bayes, maka Ntzoufras (2009) menuliskan fungsi likelihood untuk contoh berukuran  $n$  sebagai berikut:

$$f(y|\theta, \phi) = \exp\left(\frac{n\bar{y}\theta - nb(\theta)}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi)\right) \quad \dots (2.14)$$

Dengan menggunakan sebaran prior seperti yang dinyatakan dalam persamaan berikut:

$$f(\theta|\theta_0, \tau_0, \phi) = \exp\left(\frac{\theta\theta_0 - \tau_0 b(\theta)}{a(\phi)}\right) \quad \dots (2.15)$$

dimana  $\theta_0$  dan  $\tau_0$  masing-masing menyatakan parameter lokasi dan parameter dispersi dari sebaran prior. Dengan demikian akan diperoleh sebaran posterior yang dinyatakan sebagai berikut:

$$f(\theta|y, \phi) = \exp\left(\frac{(n\bar{y} + \theta_0)\theta - (n + \tau_0)b(\theta)}{a(\phi)} + \sum_{i=1}^n c(y_i, \phi)\right) \quad \dots (2.16)$$

Sebaran posterior ini mempunyai bentuk yang sama sebagai sebaran prior yang diberikan dalam Persamaan (2.15) dengan parameter  $(n\bar{y} + \theta_0)$  dan  $(n + \tau_0)$ . Sebaran yang bersifat *conjugate* yang telah ditunjukkan di atas adalah bersyarat pada parameter dispersi  $\phi$  yang diasumsikan diketahui dan tetap (*fixed*).

Model pendugaan area kecil untuk data cacahan pada awalnya menggunakan model linear terampat campuran. Sebagaimana yang diusulkan oleh Ghosh, dkk (1998), dimana mereka mengusulkan model linear terampat dengan efek wilayah acak. Bersyarat pada  $\theta_i$ , statistik contoh  $y_i$  ( $i = 1, \dots, m$ ) diasumsikan bersebaran saling bebas dengan fungsi kepekatan peluang yang merupakan anggota dari keluarga eksponensial dengan parameter kanonik  $\theta_i$  sebagaimana yang dinyatakan dalam Persamaan (2.13). Berbagai sebaran yang merupakan anggota dari keluarga eksponensial yang dinyatakan dalam persamaan tersebut diantaranya adalah sebaran normal, binomial, dan Poisson.

Misalkan peubah respons,  $y_{ij}$ , merupakan data cacahan yang mengikuti sebaran Poisson dengan parameter  $\lambda_i$ , atau dapat dinyatakan bahwa  $y_i \sim \text{Poisson}(\lambda_i)$ . Akibatnya akan diperoleh bahwa  $\theta_i = \log(\lambda_i)$  dan  $\phi_i = 1$ . Pengembangan model SAE untuk data cacahan, yang peubah responnya mengikuti sebaran Poisson dapat dilakukan dengan cara memodelkan  $\theta_i$  sebagaimana model Fay-Herriot yang dinyatakan dalam Persamaan (2.4). Model tersebut melibatkan dua buah galat, yaitu galat rancangan terinduksi (*design-induced*)  $e_i$  dan galat model  $v_i$ , dimana dalam hal ini diasumsikan bahwa  $e_i$  dan  $v_i$  adalah saling bebas, serta diketahui bahwa  $v_i \sim N(0, \sigma_v^2)$  dan  $e_i \sim N(0, \sigma_e^2)$ .

### Metode Rantai Markov Monte Carlo

Metode Rantai Markov Monte Carlo (*Markov chain Monte Carlo*, MCMC) merupakan metode simulasi untuk penarikan contoh dari sebaran posterior dan perhitungan besaran yang diamati. Metode MCMC mengambil contoh secara berturut-

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

turut dari suatu sebaran sarasannya. Untuk setiap contoh yang terambil bergantung pada contoh yang terambil sebelumnya, sehingga membentuk suatu rantai Markov. Rantai Markov itu sendiri adalah sederet peubah acak, sebut saja  $\theta^{(1)}, \theta^{(2)}, \dots$ , yang mana peubah acak  $\theta^{(k)}$  bergantung pada peubah acak yang sebelumnya, yaitu  $\theta^{(k-1)}$ .

Monte Carlo, sebagaimana dalam integrasi Monte Carlo, utamanya digunakan untuk membuat suatu pendekatan pada suatu nilai harapan dengan menggunakan contoh rantai Markov. Dalam bentuk yang paling sederhana dapat dirumuskan sebagai:

$$\int_S g(\theta)f(\theta)d\theta \cong \frac{1}{n} \sum_{k=1}^n g(\theta^{(k)}) \quad \dots (2.17)$$

dimana  $g(\cdot)$  merupakan suatu fungsi yang sedang dianalisis dan  $\theta^{(k)}$  adalah contoh yang diambil dari  $f(\theta)$ .

Metode rantai Markov saat ini banyak digunakan dalam komputasi Bayes modern. Dalam model Bayes yang sederhana, biasanya dapat ditemukan bentuk analitik dari sebaran posterior dan menghitung ringkasan statistiknya secara langsung. Akan tetapi model Bayes yang lebih kompleks, fungsi kepekatan posterior akan terlalu sulit ditemukan secara analitik. Oleh karena itu melalui metode MCMC dimungkinkan untuk membangkitkan contoh dari sembarang fungsi kepekatan posterior,  $f(\theta|y)$ , kemudian menggunakan contoh tersebut untuk menghitung nilai harapan dari besaran posterior yang akan dikaji. Satu hal yang penting dalam penggunaan MCMC adalah jika algoritma simulasi diimplementasikan dengan benar, maka rantai Markov akan konvergen ke sebaran sasaran  $f(\theta|y)$ .

Dalam metode MCMC ini terdapat dua algoritma yang populer, yaitu algoritma Metropolis-Hasting dan penarikan contoh Gibbs. Pertama, algoritma Metropolis-Hasting dapat digunakan untuk memperoleh contoh acak dari sembarang sebaran sasaran untuk sembarang dimensi. Misalkan untuk memperoleh  $k$  buah contoh dari suatu sebaran univariat dengan fungsi kepekatan peluang  $f(\theta|y)$ . Misalkan pula  $\theta^{(k)}$  adalah contoh ke- $k$  yang diambil dari fungsi tersebut. Untuk menggunakan algoritma M-H perlu menetapkan nilai awal  $\theta^{(0)}$  dan fungsi densitas usulan yang simetris,  $q(\theta^{(k+1)}|\theta^{(k)})$ . Pada iterasi ke- $(k + 1)$ , algoritma M-H akan membangkitkan contoh  $q(\cdot|\cdot)$  berdasarkan pada contoh sebelumnya  $\theta^{(k)}$ , kemudian membuat keputusan apakah menerima atau menolak contoh baru tersebut. Apabila contoh yang baru ini diterima, maka algoritma akan melanjutkan kembali yang dimulai pada contoh yang baru tersebut. Sebaliknya, apabila contoh baru ditolak, maka algoritma akan dimulai pada titik saat ini, kemudian mengulangnya kembali. Dalam prakteknya, peneliti dapat memutuskan banyaknya contoh yang diperlukan dan memberhentikan proses penarikan contoh setelah mencapai pada iterasi yang ditetapkan. Algoritma M-H mendefinisikan suatu rantai Markov dari peubah acak dimana sebarannya akan konvergen pada suatu sebaran yang diinginkan. Dalam terminologi rantai Markov, sebaran semacam itu disebut sebagai sebaran yang stasioner, dan dalam statistik Bayes, ini merupakan sebaran posterior bagi parameter model. Sudah banyak literatur yang membahas tentang kerangka kerja dari algoritma M-H ini, salah satunya adalah Robert (1996).

Kedua, algoritma penarikan contoh Gibbs dipopulerkan oleh Geman dan Geman (1984) merupakan kasus khusus dari algoritma Metropolis-Hasting, dimana sebaran yang diusulkan adalah akan persis sama dengan sebaran bersyarat posterior. Kerangka kerja dari Gibbs penarikan contoh adalah mendekomposisi sebaran posterior bersama ke dalam sebaran bersyarat untuk setiap parameter dalam model, kemudian contoh diambil dari proses tersebut. Contoh akan menjadi efisien ketika parameter untuk setiap contoh

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

yang terambil tidak berkorelasi tinggi. Beberapa peneliti lebih menyukai menggunakan algoritma Gibbs penarikan contoh karena algoritma ini tidak memerlukan instrumen dari sebaran proposalnya sebagaimana yang dilakukan pada algoritma M-H. Akan tetapi dalam menurunkan sebaran bersyaratnya relatif tidak mudah. Artinya tidak selalu mungkin untuk menemukan cara yang efisien untuk mengambil contoh dari sebaran bersyarat ini. Pembahasan secara rinci mengenai Gibbs penarikan contoh dikemukakan oleh Casella dan George (1992).

**Permasalahan dalam MCMC**

Pendekatan model Bayes berhirarki dalam pendugaan area kecil yang didasarkan pada metode rantai Markov Monte Carlo (MCMC) mempunyai beberapa keterbatasan yang akan dibahas pada bagian ini. Beberapa keterbatasan dalam MCMC ini diantaranya adalah pemilihan sebaran prior yang tepat, proses komputasi dalam bentuk proses tunggal atau proses ganda, serta *burn-in length*.

**A. Pemilihan Sebaran Prior**

Permasalahan utama yang sering dihadapi dalam analisis Bayes adalah pemilihan sebaran prior, karena apabila sebaran prior ini diketahui maka inferensi dapat dengan mudah dilakukan dengan cara meminimumkan galat posterior, menghitung daerah kepekaan sebaran posterior yang lebih tinggi dimensinya, atau mengintegrasikan parameter untuk menemukan sebaran prediktifnya. Sebaran prior merupakan kunci pada inferensi analisis Bayes, sehingga penetapan prior ini menjadi tahapan yang paling penting dalam menggambarkan inferensi ini. Dalam beberapa kasus pemilihan sebaran prior ini merupakan hal yang sangat sulit. Akan tetapi dalam prakteknya seringkali sudah tersedia informasi prior yang mempunyai presisi yang cukup baik sehingga akan membawa pada penetapan eksak dari suatu sebaran prior tertentu. Dalam hal ini, di dalam statistika tersedia banyak bentuk sebaran peluang yang sesuai dengan informasi tersebut.

Sebaran prior,  $f(\theta)$ , yang bersifat *noninformative* mencerminkan kurangnya informasi mengenai parameter model  $\theta$ , yang biasa digunakan dalam pendekatan Bayes berhirarki pada pendugaan area kecil. Akan tetapi pada saat *noninformative* prior ini tidak tepat, yaitu dalam kondisi dimana  $\int f(\theta)d\theta = \infty$ , maka Gibbs sampler akan membawa pada hasil inferensi yang sepertinya masuk akal mengenai nonexistent posterior,  $f(\mu, \theta|y)$ . Sebagaimana yang diungkapkan oleh Natarajan dan McCulloch (1995) dan juga Hobert dan Casella (1996) bahwa hal tersebut dapat terjadi ketika sebaran posteriornya tidak tepat dan belum semua Gibbs bersyaratnya adalah tepat. Untuk menunjukkan masalah ini, perhatikan model galat tersarang sederhana tanpa kovariat berikut ini:

$$y_i = \mu + v_i + e_i \quad \dots (2.18)$$

dimana  $v_i \sim N(0, \sigma_v^2)$  dan saling bebas dengan  $e_i \sim N(0, \sigma_e^2)$ . Apabila dipilih sebaran prior yang tidak tepat dalam bentuk

$$f(\mu, \sigma_v^2, \sigma_e^2) = f(\mu)f(\sigma_v^2)f(\sigma_e^2) \quad \dots (2.19)$$

dengan  $f(\mu) \approx 1$ ,  $f(\sigma_v^2) \approx 1/\sigma_v^2$ , dan  $f(\sigma_e^2) \approx 1/\sigma_e^2$ , sehingga diperoleh suatu sebaran posterior bersama bagi  $\mu, \mathbf{v} = (v_1, \dots, v_m)$ ,  $\sigma_v^2$  dan  $\sigma_e^2$  yang bersifat improper. Sebaliknya, seluruh Gibbs bersyarat akan bersifat proper untuk pemilihan sebaran prior semacam itu.

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Untuk mengatasi masalah ini, suatu perangkat lunak statistika yang biasa digunakan untuk memodelkan data melalui pendekatan Bayes seperti BUGS, menggunakan sebaran yang tepat berdifusi dalam bentuk  $\mu \sim N(0, \sigma_0^2)$ ,  $\sigma_v^{-2} \sim G(a_0, b_0)$ , serta  $\sigma_e^{-2} \sim G(a_0, b_0)$  sebagai bentuk sebaran prior yang baku, dimana dalam hal ini  $\sigma_0^2$  dipilih sebesar mungkin (katakan saja sebesar 10,000), serta  $a_0$  dan  $b_0$  dipilih sekecil mungkin (katakan saja masing-masing bernilai 0.001) untuk mencerminkan bahwa kurangnya informasi mengenai prior pada  $\mu$ ,  $\sigma_v^2$ , dan  $\sigma_e^2$ . Perlu dicatat bahwa  $G(a, b)$  menyatakan sebaran gamma dengan parameter bentuk  $a$  dan parameter skala  $b$ , serta ragam  $G(a_0, b_0)$  adalah  $1/a_0$  akan bernilai besar pada saat  $a_0 \rightarrow 0$ . Oleh karena itu, sebaran posterior akan mendekati tidak tepat (atau hampir tidak tepat) untuk  $a_0$  yang bernilai sangat kecil, dimana masalah ini akan mempengaruhi pada masalah kekonvergenan dari contoh Gibbs. Sebagai alternatifnya, dapat dipilih sebaran prior dalam bentuk  $\sigma_v^2 \sim \text{uniform}(0, 1/a_0)$  dan  $\sigma_e^2 \sim \text{uniform}(0, 1/a_0)$  untuk mengatasi permasalahan tersebut dengan tetap menjaga diperolehnya suatu sebaran posterior yang tepat ketika  $a_0 \rightarrow 0$ .

Browne dan Draper (2006) telah membandingkan sifat dari besaran posterior di bawah model galat tersarang sederhana dan  $G(a_0, b_0)$  atau  $\text{uniform}(0, 1/a_0)$  sebagai prior pada parameter ragam. Secara khusus, untuk  $\sigma_v^2$  telah diperiksa bias dari rata-rata, median, dan modus posteriornya, serta cakupan interval Bayes dalam penarikan contoh berulang (*repeated sampling*). Seluruh Gibbs bersyarat mempunyai bentuk tertutup, sehingga penarikan contoh Gibbs dapat digunakan untuk membangkitkan contoh dari sebaran posteriornya. Hasil pemeriksaan pada bias ini menunjukkan bahwa median posterior lebih baik dibandingkan dengan rata-rata posterior (penduga Bayes berhirarki) untuk prior gamma, sedangkan modus posterior lebih baik dibandingkan dengan rata-rata posterior pada prior uniform. Sementara itu hasil pemeriksaan pada interval Bayes untuk prior uniform atau gamma tidak mampu mencapai cakupan nominal pada saat banyaknya area,  $m$ , dan/atau rasio ragam  $\tau = \sigma_v^2/\sigma_e^2$  adalah kecil. Sementara itu, Datta dan Sweeting (2005) mengembangkan prior yang bersifat *noninformative* untuk diterapkan pada model galat tersarang sederhana yang memfokuskan pada rasio ragam  $\tau$ . Prior semacam ini akan menghasilkan cakupan dari interval Bayes untuk  $\tau$  yang mendekati taraf nominal pada saat  $m \rightarrow \infty$ . Sedangkan Rao (2003) mempertimbangkan prior untuk model dasar level area yang menghasilkan ragam posterior dari rata-rata area kecil,  $\mu_i$ , mempunyai sifat takbias untuk kuadrat tengah galat dari penduga EB/HB pada saat  $m \rightarrow \infty$ .

## B. Proses Tunggal atau Proses Ganda

Sebagaimana yang telah dibahas pada bagian sebelumnya bahwa untuk mengevaluasi dari  $f(\mu|y)$  dan besaran posterior lainnya seperti  $E[h(\mu)|y]$  melibatkan integrasi multi-dimensi, dimana salah satu metode komputasi yang digunakan adalah melalui metode rantai markov Monte Carlo. Metode ini dilakukan dengan cara membentuk rantai Markov  $\{\eta^{(k)}, k = 0, 1, 2, \dots\}$  sedemikian rupa sehingga sebaran dari  $\eta^{(k)}$  konvergen pada sebaran stasioner yang unik yang sama dengan  $f(\eta|y)$ , yang dinotasikan dengan  $\pi(\eta)$ . Jadi, setelah dilakukan 'burn-in' yang relatif besar, katakan saja sebesar  $l$ , maka  $\eta^{(l+1)}, \dots, \eta^{(l+L)}$  dapat dianggap sebagai  $L$  buah contoh yang takbebas dari sebaran target  $f(\eta|y)$  dengan mengabaikan nilai awal  $\eta^{(0)}$ . Proses pembangkitan  $\eta^{(l+1)}, \dots, \eta^{(l+L)}$  ini dilakukan dalam proses tunggal (*single-run*). Proses

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



tunggal dapat memberikan penduga Monte Carlo bagi besaran posterior yang andal pada saat memilih  $D$  yang cukup besar. Akan tetapi hal ini masih menyisakan porsi yang signifikan dari ruang yang dibangkitkan oleh posterior  $f(\eta|y)$  yang secara keseluruhan belum tereksplorasi.

Untuk mengatasi masalah tersebut, Gelman dan Rubin (1992) menggunakan proses ganda (*multiple-run*) dengan menggunakan beberapa nilai awal yang berbeda sehingga menghasilkan contoh paralel. Nilai awal yang baik sangat diperlukan untuk membangkitkan contoh melalui proses ganda ini, dan menurut Rao (2003) penentuan nilai awal yang baik bukan merupakan pekerjaan yang mudah. Sebaliknya, dalam proses tunggal, penduga parameter model melalui metode REML ( $\lambda$ ) dan penduga EB dari parameter area kecil ( $\mu$ ) dapat digunakan sebagai penduga awal yang baik (Rao, 2003). Sementara itu, Gelman dan Rubin (1992) merekomendasikan nilai awal dari sebaran yang kelebihan dispersi (*overdispersion*) dibandingkan dengan sebaran target  $\pi(\eta|y)$ .

Sementara itu, proses ganda dapat sia-sia disebabkan oleh periode *burn-in* awal diabaikan dari setiap proses, walaupun hal ini bukan merupakan keterbatasan yang serius apabila menggunakan processor paralel untuk membangkitkan contoh paralel. Gelfand dan Smith (1990) menggunakan banyak proses jangka pendek, dimana setiap proses berisi  $(l + 1)$  buah nilai  $\eta$ , dan hanya menyimpan observasi terakhir pada setiap proses. Metode ini memberikan  $\eta^{(l+1)}(1), \dots, \eta^{(l+1)}(L)$ , dimana  $\eta^{(l+1)}(l)$  adalah nilai  $\eta$  terakhir dari proses ke- $l$  (untuk  $l = 1, \dots, L$ ). Di lain pihak, Gelman et al. (2004) menggunakan  $L$  berukuran kecil (biasanya  $L = 10$ ), dan kemudian dibangkitkan  $2l$  nilai untuk setiap proses dan menyimpan  $l$  nilai terakhir dari setiap proses. Tahapan ini akan menghasilkan segugus  $L$  nilai  $\eta$  yang saling bebas. Perlu dicatat bahwa dalam hal ini tidak perlu membangkitkan contoh bebas untuk memperoleh penduga Monte Carlo bagi besaran posterior karena proses ini membutuhkan teori rantai Markov yang tinggi. Namun demikian, contoh bebas ini dapat digunakan untuk perhitungan galat baku Monte Carlo.

### C. Burn-in Length

Panjang atau periode *burn-in*,  $d$ , sangat bergantung pada nilai awal  $\eta^{(0)}$  dan derajat konvergensi  $P^{(k)}(\cdot)$  pada sebaran stasioner  $\pi(\cdot)$ . Derajat konvergensi ini telah banyak dipelajari dalam berbagai literatur, akan tetapi tidak mudah menggunakan derajat konvergensi tersebut dalam menentukan  $d$ . Perlu diketahui pula bahwa derajat konvergensi untuk algoritma MCMC yang berbeda juga akan sangat bervariasi dan juga bergantung pada sebaran target  $\pi(\cdot)$ . Dalam prakteknya diagnostik terhadap derajat kekonvergenan, berdasarkan pada luaran MCMC  $\{\eta^{(k)}\}$ , sering digunakan untuk menentukan panjang *burn'in*. Salah satu yang paling banyak digunakan adalah ukuran diagnostik berdasarkan pada proses ganda dan juga analisis ragam, sebagaimana yang diusulkan oleh Gelman dan Rubin (1992).

### Besaran Posterior

Duaran yang dihasilkan melalui metode MCMC  $\{\eta^{(k)}, k = l + 1, \dots, l + L\}$  dari suatu proses tunggal dapat digunakan untuk menghitung besaran posterior yang sedang dikaji, yang misalkan bahwa  $\{\mu^{(k)}\}$  merupakan contoh dari posterior marginal  $f(\mu|y)$ . Secara khusus, rata-rata posterior (atau merupakan penduga Bayes berhirarki) bagi  $\phi = h(\mu)$  dapat diduga melalui rata-rata ergodic

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$\hat{\phi}^{HB} = \frac{1}{L} \sum_{k=l+1}^{l+L} \phi^{(k)} = \phi^{(\cdot)} \quad \dots (2.20)$$

dimana  $\phi^{(k)} = h(\mu^{(k)})$ . Dengan cara yang sama, ragam posterior bagi  $\phi$  diduga melalui

$$\hat{V}(\phi|y) = \frac{1}{L-1} \sum_{k=l+1}^{l+L} (\phi^{(k)} - \phi^{(\cdot)})^2 \quad \dots (2.21)$$

Berdasarkan teorema ergodic untuk rantai Markov, maka penduga  $\hat{\phi}^{HB}$  akan konvergen ke  $E(\phi|y)$  dan  $\hat{V}(\phi|y)$  konvergen ke  $V(\phi|y)$  pada saat  $D \rightarrow \infty$ . Namun demikian galat baku Monte Carlo  $\hat{\phi}^{HB}$  dan  $\hat{V}(\phi|y)$  tidak mudah ditentukan pada saat nilai-nilai  $\phi^{(l+1)}, \dots, \phi^{(l+L)}$  yang disimulasikan tidak saling bebas.

Apabila  $\eta^{(k)}$  merupakan vektor yang identik dan saling bebas, sebagaimana yang diusulkan melalui metode atau proses ganda, maka penduga bagi  $\phi$  akan mereduksi galat baku yang diperoleh apabila bentuk matematis untuk ekspektasi bersyarat bagi  $\phi$  pada data  $y$  dan parameter model  $\theta$  diketahui. Misalkan bahwa bentuk  $E(\phi|y, \theta)$  diketahui, maka penduga bagi  $E(\phi|y)$  dapat dirumuskan kembali sebagai berikut

$$\tilde{\phi}^{HB} = \frac{1}{L} \sum_{k=l+1}^{l+L} E(\phi|y, \theta^{(k)}) \quad \dots (2.22)$$

dimana  $\{\theta^{(k)}\}$  merupakan contoh identik dan saling bebas dari posterior marginal  $f(\theta|y)$ . Berdasarkan hasil tersebut juga dapat diperoleh ragam sebagai berikut:

$$V(\phi|y) = E[V(\phi|y, \theta)|y] + V[E(\phi|y, \theta)|y] \geq V[E(\phi|y, \theta)|y]$$

dan ragam Monte Carlo bagi  $\hat{\phi}^{HB}$  dan  $\tilde{\phi}^{HB}$  tersebut masing-masing diberikan oleh  $D^{-1}V(\phi|y)$  dan  $D^{-1}V[E(\phi|y, \theta)|y]$ . Perlu diketahui bahwa penduga  $\tilde{\phi}^{HB}$  yang diberikan dalam Persamaan (2.22) disebut juga penduga Rao-Blackwell karena prosesnya diperoleh didasarkan pada teorema Rao-Blackwell (Robert dan Casella, 2004). Penduga ragam posterior untuk  $\tilde{\phi}^{HB}$  diberikan oleh

$$\tilde{V}(\phi|y) = \frac{1}{L} \sum_{k=l+1}^{l+L} V(\phi|y, \theta^{(k)}) + \frac{1}{L-1} \sum_{k=l+1}^{l+L} [E(\phi|y, \theta^{(k)}) - \tilde{\phi}^{HB}]^2 \quad \dots (2.23)$$

Untuk kasus proses tunggal dengan contoh yang tidak saling bebas, maka hasil yang diberikan dalam Persamaan (2.23) tidak akan terpenuhi. Akan tetapi menurut Rao (2003) jika  $D$  contoh “dirampingkan” dengan cara memilih setiap  $\eta^{(k)}$  yang ke-5 atau yang ke-10, maka akan menghasilkan luaran MCMC tereduksi dengan suatu contoh yang identik dan saling bebas. Proses ini kadang-kadang dapat menghemat ruang simpan dan waktu komputasi yang diperlukan, khususnya ketika contoh  $\eta^{(k)}$  berkorelasi tinggi.

### 3 MODEL REGRESI POISSON BAYES BERHIRARKI

#### Pendahuluan

Penduga taklangsung berdasarkan pada area kecil disebut juga penduga berbasis model (*model-based*). Secara umum, model area kecil ini biasanya diklasifikasikan ke dalam dua jenis, yaitu model untuk level area dan model untuk level unit. Model untuk level area akan menghubungkan penduga langsung area kecil ke kovariat dari area tertentu, dan biasanya pemodelan ini dilakukan ketika data untuk level unit tidak

tersedia. Sedangkan pada model untuk level unit akan menghubungkan nilai-nilai unit dari peubah yang sedang diamati pada kovariat dari unit tertentu. Asumsi penting dari model level unit adalah bahwa nilai contoh membolehkan model populasi yang diasumsikan, artinya tidak adanya bias seleksi contoh. Sementara itu, untuk model level area diasumsikan tidak adanya penarikan contoh informatif dari area dalam situasi dimana hanya ada beberapa area yang dipilih sebagai contoh, artinya nilai dari area contoh (sebagai penduga langsung) mengasumsikan model populasi.

Berbagai model dasar dalam pendugaan area kecil ini pada umumnya digunakan ketika peubah respons yang diamati adalah berbentuk kontinu. Padahal seringkali dalam menganalisis hubungan antara beberapa peubah, terdapat sejumlah fenomena dimana peubah responnya berbentuk data cacahan. Dalam mengamati suatu fenomena dimana peubah responnya berbentuk cacahan, maka fenomena seperti ini menyangkut banyaknya suatu kejadian yang biasanya diasumsikan mengikuti sebaran Poisson. Selanjutnya, inferensi dari penduga berbasis model ini merujuk pada suatu sebaran tertentu yang harus terpenuhi dari model yang diasumsikan, misalnya mengikuti sebaran Poisson tadi. Oleh karena itu, pemilihan model dan validasi model akan memegang peranan penting dalam pendugaan berbasis model ini. Apabila model yang diasumsikan tidak memberikan kecocokan yang baik terhadap data, maka penduga berbasis model akan menjadi bias yang pada akhirnya nanti akan membawa pada kegagalan dalam membuat inferensi yang baik.

Pada saat ini metode Bayes telah banyak digunakan untuk menangani masalah pendugaan area kecil. Metode Bayes yang sudah mulai banyak dikembangkan dalam hal ini adalah metode Bayes empirik dan Bayes berhirarki, yang secara khusus cukup baik dalam menggambarkan hubungan sistematis dari area lokal melalui model. Namun demikian, perkembangan metode Bayes untuk masalah pendugaan area kecil saat ini masih difokuskan pada peubah yang kontinu. Padahal seringkali data yang diperoleh melalui survei ini berbentuk diskrit atau kategorik, sehingga metode Bayes empirik dan Bayes berhirarki yang dirancang untuk data kontinu menjadi tidak tepat lagi untuk diterapkan.

Pada saat data (atau dalam hal ini peubah responsnya) yang diamati berbentuk data diskrit atau kategorik, lebih khusus lagi berbentuk data cacahan yang bersebaran Poisson, maka model yang dapat diterapkan adalah melalui model linear terampat (*generalized linear model*, GLM). Penggunaan pendekatan metode Bayes pada model linear terampat ini pada dasarnya sudah banyak dilakukan. Akan tetapi bagaimana penggunaan metode Bayes, khususnya metode Bayes berhirarki, pada model linear terampat yang secara langsung dapat diterapkan pada penanganan masalah pendugaan area kecil. Masih sedikit literatur yang membahas tentang masalah ini. Oleh karena itu, dalam disertasi ini akan dikembangkan model linear terampat Bayes berhirarki untuk data cacahan yang berbentuk Poisson yang akan diterapkan pada masalah pendugaan area kecil.

Model Poisson merupakan sub-bagian dari model linear terampat yang banyak digunakan, selain model logistik atau binomial. Pada bagian ini akan dikembangkan suatu model linear terampat Bayes berhirarki yang bersebaran Poisson. Model Bayes berhirarki yang akan dikembangkan adalah model Bayes berhirarki dua-level, atau selanjutnya model ini disebut juga sebagai model regresi Poisson Bayes berhirarki dua level yang nantinya akan diterapkan pada masalah pendugaan area kecil. Pembahasan dalam bab ini akan dimulai dengan berbagai permasalahan yang sering ditemukan dalam pemodelan Bayes berhirarki, seperti pemilihan sebaran prior, penentuan proses

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

tunggal atau ganda, serta penetapan panjang ‘burn-in’. Isu lain yang sering muncul dalam metode Bayes adalah pembentukan berbagai besaran posterior yang diperlukan. Bahasan utama pada bab ini adalah mengembangkan model regresi Poisson Bayes berhirarki yang akan diterapkan pada pendugaan area kecil. Terminologi yang ada dalam model linear terampat akan dipadukan dengan konsep metode Bayes, khususnya Bayes berhirarki, sedemikian rupa sehingga dapat diimplementasikan untuk menangani masalah pendugaan area kecil untuk data survei.

Selanjutnya, isu penting yang akan dibahas dalam bab ini adalah masalah komputasi. Syarat cukup yang diberikan untuk sebaran posterior dari parameter yang diamati diharapkan akan bersifat proper di bawah model hirarki yang diusulkan ini. Prosedur Bayes yang diimplementasikan dalam penelitian ini dilakukan melalui teknik integrasi rantai Markov Monte Carlo, khususnya menggunakan teknik penarikan contoh Gibbs. Terakhir, untuk mengevaluasi performa dari model yang diusulkan akan dilakukan studi simulasi. Adapun performa model yang akan menjadi perhatian utama dari studi simulasi ini adalah ketidakbiasan dari penduga parameter dan galat bakunya.

### Model Regresi Poisson Bayes Berhirarki Dua-Level

Model regresi Poisson berhirarki telah banyak digunakan untuk menganalisis berbagai jenis data yang berbentuk cacahan. Kebanyakan analisis yang dilakukan untuk pemetaan penyakit (*disease mapping*) dimulai dengan proses penarikan contoh Poisson. Clayton dan Klador (1987) menggambarkan pendekatan Bayes empirik yang memperhatikan kemiripan spasial antar angka kematian penyakit tertentu. Sementara itu Ghosh et al. (2009) membandingkan metode Bayes empirik dan Bayes berhirarki, yang diaplikasikan pada masalah kesehatan untuk subpopulasi yang bersifat minoritas. Sementara itu, Breslow dan Clayton (1993) menggunakan model campuran linear terampat untuk mempelajari masalah pemetaan penyakit ini. Sedangkan, Waller, et al. (1997) mengusulkan model Bayes berhirarki *spatio-temporal* untuk memodelkan angka kematian regional menurut ruang dan waktu termasuk didalamnya interaksi antara ruang dan waktu itu sendiri.

Saat ini sudah mulai banyak penelitian mengenai penerapan model Bayes berhirarki untuk menangani masalah pendugaan area kecil ini. Torabi dan Rao (2008) mengembangkan masalah pendugaan area kecil dua-level melalui penduga *generalized regression* (GREG). Sementara itu, You dan Chapman (2006) membahas tentang pendugaan area kecil untuk level area. Mereka mengembangkan model untuk memperoleh ragam penarikan contoh melalui pendekatan model Bayes berhirarki dengan algoritma Gibbs penarikan contoh. Penggunaan model Bayes berhirarki juga dilakukan oleh Souza et al. (2009) dalam rangka memprediksi populasi area kecil, dimana model yang dikembangkan dilakukan melalui model pertumbuhan terstruktur secara spasial. Pengembangan model spasial yang dikombinasikan dengan model Bayes berhirarki untuk menangani masalah pendugaan area kecil juga dilakukan oleh You dan Zhou (2011). Hasil pengembangan model yang dilakukan oleh You dan Zhou (2011) ini diaplikasikan pada masalah data kesehatan.

Dalam penelitian ini akan diusulkan pengembangan model regresi Poisson berhirarki yang pertama kali diusulkan oleh Christiansen dan Morris (1997), dimana model ini pada awalnya tidak dirancang untuk digunakan dalam masalah survei penarikan contoh. Sekali lagi, model ini dikembangkan dengan cara memadukan

terminologi yang ada dalam model linear terampat dengan konsep metode Bayes, khususnya metode Bayes berhirarki, sedemikian rupa sehingga dapat diimplementasikan untuk menangani masalah pendugaan area kecil untuk data survei yang berbentuk data cacahan. Pengembangan model ini dimulai dengan mengasumsikan peubah acak yang diamati merupakan anggota dari keluarga eksponensial, sebagaimana yang muncul dalam konsep pemodelan linear terampat, bersyarat pada suatu parameter tertentu. Tujuan utama dari pengembangan model ini adalah membuat inferensi pada parameter tersebut yang juga dianggap sebagai peubah acak. Kemudian parameter tersebut dimodelkan dengan menggunakan model Fay-Herriot sebagai model dasar dalam konsep pendugaan area kecil. Selanjutnya, perpaduan dari kedua model tersebut akan standarkan sedemikian rupa sehingga mewakili suatu model dalam kerangka kerja metode Bayes yang pada akhirnya akan terbentuk model Poisson Bayes berhirarki untuk menyelesaikan masalah dalam pendugaan area kecil.

Berikut ini pembahasan mengenai pengembangan model regresi Poisson Bayes berhirarki dua-level. Misalkan  $y_i$  menyatakan banyaknya peristiwa ‘sukses’ atau dalam hal ini banyaknya kejadian yang mati pada area ke- $i$ ,  $n_i$  menyatakan populasi dalam area ke- $i$ , serta  $\theta_i$  menyatakan angka mortalitas pada area ke- $i$ , dimana  $\theta_i = y_i/n_i$  (untuk  $i = 1, 2, \dots, m$ ), serta  $m$  menunjukkan banyak area kecil yang diamati. Dalam hal ini  $y_i$  adalah peubah acak yang saling bebas dengan fungsi kepekatan peluang yang merupakan anggota dari keluarga eksponensial sebagaimana yang dirumuskan dalam Persamaan (2.14). Kemudian, fungsi kepekatan tersebut diparameterisasi terhadap parameter kanonik,  $\theta_i$ , dan parameter skala  $\phi_i$ , dimana  $\phi_i > 0$  dan diasumsikan diketahui.

Parameter kanonik  $\theta_i$  akan dimodelkan dengan menggunakan model Fay-Herriot seperti yang dirumuskan dalam Persamaan (2.4). Untuk merumuskan model regresi Poisson Bayes berhirarki dua-level, dilakukan melalui dua level, yaitu level 1 dari model deskriptif menyatakan sebaran dari vektor data yang diamati,  $\mathbf{y} = (y_1, \dots, y_m)^T$ , dengan syarat pada parameter individu  $\{\theta_i\}$ ; Pada level 2 terdapat dua sebaran prior yang dipertimbangkan, yaitu untuk menyatakan sebaran gamma untuk  $\{\theta_i\}$  dengan syarat pada *hyperparameter*  $\alpha = (\tau, \boldsymbol{\beta})$  dan level 2 untuk menyatakan sebaran invers-gamma untuk  $\{\theta_i\}$  dengan syarat pada *hyperparameter*  $\alpha = (\tau, \boldsymbol{\beta})$ .

### Level 1: Model Individu

Asumsi dasar yang harus dipenuhi adalah bahwa peubah respons  $y_i$  mengikuti sebaran Poisson untuk parameter  $\theta_i$  yang tetap (*fixed*), yaitu:

$$y_{ij} | \theta_i \sim \text{Poisson}(n_i \theta_i), \text{ untuk } i = 1, \dots, m \text{ dan } j = 1, 2, \dots, n_i \quad \dots (3.1)$$

Diketahui bahwa angka mortalitas pengamatan,  $\mathbf{z}$ , dimana  $z_i = y_i/n_i$ , dimana angka mortalitas pengamatan ini mempunyai nilai harapan  $E(z_i) = \theta_i$ .

### Level 2: Model Terstruktur

Parameter Poisson individu  $(\theta_1, \dots, \theta_m)$  mengikuti sebaran gamma yang bersifat *conjugate* untuk  $i = 1, \dots, m$  yang saling bebas dengan syarat pada vektor *hyperparameter* yang tidak diketahui  $\alpha = (\tau, \beta_0, \dots, \beta_{p-1})$ , dimana  $p$  menyatakan banyaknya koefisien regresi. Dengan demikian

$$\theta_i | \tau, \boldsymbol{\beta} \sim \text{Gamma}(a, b) \quad \dots (3.2)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

dimana  $a = \tau$  dan  $b = \tau/\mu_i$ . Fungsi hubung log (yang merupakan fungsi hubung alamiah untuk sebaran Poisson) diasumsikan untuk rata-rata terstruktur, sehingga  $\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$  untuk kovariat yang bersifat tetap (*fixed*)  $\mathbf{x}_i = (x_{i0}, \dots, x_{i,k-1})^T$  dan  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{k-1})^T$ . Dalam hal ini parameter  $\tau$  merupakan cacahan prior yang tidak teramati, dan tidak harus berupa bilangan bulat.

Untuk mendukung pernyataan bahwa sebaran gamma merupakan sebaran prior yang bersifat *conjugate* bagi parameter Poisson individu  $(\theta_1, \dots, \theta_m)$  diberikan suatu teorema sebagai berikut:

**Teorema 1.** Misalkan peubah acak  $y$  mengikuti sebaran Poisson dengan parameter  $\theta_i$  yang bersifat tetap (*fixed*),  $y_{ij}|\theta_i \sim \text{Poisson}(n_i\theta_i)$ , untuk  $i = 1, \dots, m$  dan  $j = 1, 2, \dots, n_i$ . Kemudian diasumsikan bahwa sebaran prior bagi parameter  $\theta_i$  adalah sebaran gamma dengan parameter  $a$  dan  $b$ , maka sebaran posterior bagi parameter  $\theta_i$  adalah Gamma dengan parameter  $(n_i\bar{y} + a)$  dan  $(n_i + b)$ ,  $\theta_i|\mathbf{y}_i \sim \text{Gamma}(n_i\bar{y} + a, n_i + b)$ . Sedangkan rata-rata dan ragam untuk sebaran posterior  $\theta_i$  masing-masing diberikan oleh

$$E(\theta_i|\mathbf{y}_i) = \frac{n_i\bar{y} + a}{n_i + b}$$

dan

$$\text{var}(\theta_i|\mathbf{y}_i) = \frac{n_i\bar{y} + a}{(n_i + b)^2}$$

**Bukti.**

Diketahui bahwa apabila  $y_{ij}|\theta_i \sim \text{Poisson}(n_i\theta_i)$ , untuk  $i = 1, \dots, m$  dan  $j = 1, 2, \dots, n_i$ , maka akan menghasilkan fungsi likelihood sebagai berikut:

$$f(\mathbf{y}_i|\theta_i) = \prod_{j=1}^{n_i} \frac{e^{-\theta_i} \theta_i^{y_{ij}}}{y_{ij}!} = \frac{e^{-n_i\theta_i} \theta_i^{\sum_{j=1}^{n_i} y_{ij}}}{\prod_{j=1}^{n_i} y_{ij}!}$$

Perhatikan sebaran prior gamma bagi parameter  $\theta_i$  dengan parameter  $a$  dan  $b$ , dimana fungsi kepekatannya diberikan oleh

$$f(\theta_i) = \frac{b^a}{\Gamma(a)} \theta_i^{a-1} e^{-b\theta_i}$$

Dengan menggunakan aturan Bayes akan diperoleh sebaran posterior sebagai berikut:

$$\begin{aligned} f(\theta_i|\mathbf{y}_i) &= \frac{f(\mathbf{y}_i|\theta_i)f(\theta_i)}{\int f(\mathbf{y}_i|\theta_i)f(\theta_i)d\theta_i} \\ &= \frac{\frac{e^{-n_i\theta_i} \theta_i^{\sum_{j=1}^{n_i} y_{ij}}}{\prod_{j=1}^{n_i} y_{ij}!} \frac{b^a}{\Gamma(a)} \theta_i^{a-1} e^{-b\theta_i}}{\int \frac{e^{-n_i\theta_i} \theta_i^{\sum_{j=1}^{n_i} y_{ij}}}{\prod_{j=1}^{n_i} y_{ij}!} \frac{b^a}{\Gamma(a)} \theta_i^{a-1} e^{-b\theta_i} d\theta_i} \end{aligned}$$

$$= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-(n_i + b)\theta_i}}{\int \theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-(n_i + b)\theta_i} d\theta_i}$$

Misalkan  $w_i = (n_i + b)\theta_i$ , maka diperoleh  $dw_i = (n_i + b)d\theta_i$ , serta misalkan pula bahwa  $\theta_i = \frac{1}{(n_i + b)}w_i$ , sehingga diperoleh  $d\theta_i = \frac{1}{(n_i + b)}dw_i$ . Dengan demikian diperoleh

$$\begin{aligned} f(\theta_i | \mathbf{y}_i) &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-(n_i + b)\theta_i}}{\int \left(\frac{1}{n_i + b} w_i\right)^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-w_i} \frac{1}{n_i + b} dw_i} \\ &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-(n_i + b)\theta_i}}{\left(\frac{1}{n_i + b}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \int w_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-w_i} dw_i} \end{aligned}$$

Diketahui bahwa  $\int_0^\infty w_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-w_i} dw_i = \Gamma(\sum y_i + a)$ , sehingga diperoleh

$$f(\theta_i | \mathbf{y}_i) = \frac{(n_i + b)^{\sum_{j=1}^{n_i} y_{ij} + a}}{\Gamma(\sum y_i + a)} \theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-(n_i + b)\theta_i}$$

Artinya bahwa  $(\theta_i | \mathbf{y}_i) \sim \text{gamma}(\sum y_i + a, n_i + b)$  atau dapat dinyatakan juga sebagai  $(\theta_i | \mathbf{y}_i) \sim \text{gamma}(n_i \bar{y} + a, n_i + b)$ , dimana  $\bar{y}$  adalah rata-rata contoh.

Selanjutnya, diketahui pula bahwa rata-rata dan ragam untuk peubah acak  $x$  yang bersebaran gamma masing-masing diberikan oleh  $E(X) = \frac{\alpha}{\beta}$  dan  $Var(X) = \frac{\alpha}{\beta^2}$ . Di sini, diketahui bahwa  $\alpha = n_i \bar{y} + a$  dan  $\beta = n_i + b$ , sehingga diperoleh rata-rata dan ragam untuk sebaran posterior  $\theta$  masing-masing diberikan oleh

$$E(\theta_i | \mathbf{y}_i) = \frac{n_i \bar{y} + a}{n_i + b}$$

dan

$$\text{var}(\theta_i | \mathbf{y}_i) = \frac{n_i \bar{y} + a}{(n_i + b)^2}$$

Dengan demikian terbukti bahwa sebaran gamma merupakan sebaran prior yang bersifat *conjugate* bagi sebaran Poisson. □

Perhatikan bahwa model Poisson-gamma bersifat *conjugate* yang akan membawa pada inferensi posterior mengenai parameter eksak  $\theta_i$  bersyarat pada  $\tau$  dan  $\beta$ , dimana sebaran posteriornya dapat dinyatakan sebagai berikut:

$$\theta_i | \beta, \tau, \mathbf{y} \sim \text{Gamma}\left(y_i + \tau, n_i + \frac{\tau}{\mu_i}\right)$$

Dalam penelitian ini juga akan dipertimbangkan suatu sebaran prior yang tidak harus bersifat *conjugate*. Gelman (2006) memperkenalkan suatu sebaran prior yang bersifat *conditionally conjugate*. Suatu keluarga sebaran prior  $p(\theta)$  merupakan *conditionally conjugate* untuk parameter  $\theta$  apabila sebaran posterior bersyarat,  $p(\theta | \mathbf{y})$  juga berada

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

dalam kelas tersebut. Untuk keperluan komputasi, *conditionally conjugate* mempunyai makna bahwa apabila memungkinkan untuk mengambil  $\theta$  yang berasal kelas sebaran prior tertentu, maka hal ini juga memungkinkan untuk membentuk Gibbs sampler bagi  $\theta$  dalam sebaran posterior.

Menurut Gelman (2006) bahwa sebaran prior yang bersifat *conditionally conjugate* merupakan konsep yang sangat bermanfaat apabila diterapkan pada model Bayes yang berhirarki. Dalam penelitian ini akan dipertimbangkan suatu sebaran prior yang lain bagi parameter  $\theta$  yang bersifat *conditionally conjugate*. Sebaran prior itu adalah invers-gamma, yang dinyatakan dalam

$$\theta_i | \tau, \beta \sim \text{IGamma}(a, b)$$

dimana  $a = y_i + \tau$  dan  $b = n_i + \tau/\mu_i$ .

Sekali lagi bahwa pemilihan sebaran invers gamma dilakukan dengan pertimbangan bahwa sifat sifat sebaran dari invers gamma tidak jauh berbeda dengan sebaran gamma. Untuk menemukan sebaran posterior berdasarkan pada sebaran prior yang bersifat *conditionally conjugate* secara analitik akan sulit. Namun demikian dalam penelitian ini menemukan bentuk dari sebaran posterior untuk peubah acak  $\theta_i$  yang bersebaran Poisson dengan mengambil prior yang bersebaran invers gamma. Hasil temuan ini berupa sebaran posterior yang bersifat pendekatan, atau tidak berbentuk eksak, diringkas dalam suatu teorema sebagai berikut:

**Teorema 2.** Misalkan peubah acak  $y$  mengikuti sebaran Poisson dengan parameter  $\theta_i$  yang bersifat tetap (*fixed*),  $y_{ij} | \theta_i \sim \text{Poisson}(n_i \theta_i)$ , untuk  $i = 1, \dots, m$  dan  $j = 1, 2, \dots, n_i$ . Kemudian diasumsikan bahwa sebaran prior bagi parameter  $\theta_i$  adalah sebaran invers gamma dengan parameter  $a$  dan  $b$ , maka sebaran posterior bagi parameter  $\theta_i$  adalah

$$f(\theta_i | \mathbf{y}_i) \approx \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - (\frac{1}{b} \theta_i)}}{\left(\frac{1}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \sum_{j=1}^{n_i} w_j e^{1/b \theta_i} (n_i \theta_i)^{\sum_{j=1}^{n_i} y_{ij} + a - 1}}$$

**Bukti:**

Perhatikan bahwa apabila  $\theta_i$  merupakan peubah acak yang bersebaran invers gamma dengan parameter  $a$  dan  $b$ , dimana fungsi kepekatan peluangnya diberikan oleh

$$f(\theta_i) = \frac{\left(\frac{1}{b}\right)^a}{\Gamma(a)} \theta_i^{-a-1} e^{-1/b \theta_i}, \text{ untuk } \theta_i > 0.$$

Dengan menggunakan aturan Bayes akan diperoleh sebaran posterior sebagai berikut:

$$f(\theta_i | \mathbf{y}_i) = \frac{f(\mathbf{y}_i | \theta_i) f(\theta_i)}{\int f(\mathbf{y}_i | \theta_i) f(\theta_i) d\theta_i}$$

$$f(\theta_i | \mathbf{y}_i) = \frac{\frac{e^{-n_i \theta_i} \theta_i^{\sum_{j=1}^{n_i} y_{ij}} \left(\frac{1}{b}\right)^a}{\prod_{j=1}^{n_i} y_{ij}!} \frac{\Gamma(a)}{\Gamma(a)} \theta_i^{-a-1} e^{-1/b \theta_i}}{\int_0^\infty \frac{e^{-n_i \theta_i} \theta_i^{\sum_{j=1}^{n_i} y_{ij}} \left(\frac{1}{b}\right)^a}{\prod_{j=1}^{n_i} y_{ij}!} \frac{\Gamma(a)}{\Gamma(a)} \theta_i^{-a-1} e^{-1/b \theta_i} d\theta_i}$$



$$\begin{aligned}
 &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\int_0^\infty \theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)} d\theta_i} \\
 &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\int_0^\infty e^{-n_i \theta_i} e^{-1/(b \theta_i)} \theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} d\theta_i}
 \end{aligned}$$

Dimisalkan bahwa  $t_i = n_i \theta_i$  sehingga diperoleh  $dt_i = n_i d\theta_i$ , kemudian dimisalkan pula bahwa  $\theta_i = t_i/n_i$ , sehingga diperoleh  $d\theta_i = \frac{dt_i}{n_i}$ . Dengan demikian diperoleh

$$\begin{aligned}
 f(\theta_i | \mathbf{y}_i) &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\int_0^\infty e^{-t_i} e^{-n_i/(bt_i)} \left(\frac{t_i}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a - 1} \frac{1}{n_i} dt_i} \\
 &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\left(\frac{1}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \int_0^\infty e^{-t_i} e^{-n_i/(bt_i)} t_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} dt_i} \\
 f(\theta_i | \mathbf{y}_i) &= \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\left(\frac{1}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \int_0^\infty e^{-t_i} f(t_i) dt_i}
 \end{aligned}$$

dimana:  $f(t_i) = e^{-n_i/(bt_i)} t_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1}$ . Untuk menyelesaikan bentuk integrasi di atas akan menggunakan Laguerre *integration*, sehingga diperoleh:

$$f(\theta_i | \mathbf{y}_i) \approx \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - 1/(b \theta_i)}}{\left(\frac{1}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \sum_{j=1}^{n_i} w_i e^{-n_i/bt_i} (t_i)^{\sum_{j=1}^{n_i} y_{ij} + a - 1}}$$

atau dapat dinyatakan sebagai berikut:

$$f(\theta_i | \mathbf{y}_i) \approx \frac{\theta_i^{\sum_{j=1}^{n_i} y_{ij} + a - 1} e^{-n_i \theta_i - (\frac{1}{b} \theta_i)}}{\left(\frac{1}{n_i}\right)^{\sum_{j=1}^{n_i} y_{ij} + a} \sum_{j=1}^{n_i} w_i e^{1/b \theta_i} (n_i \theta_i)^{\sum_{j=1}^{n_i} y_{ij} + a - 1}}$$

Perhatikan bahwa bentuk sebaran posterior yang diperoleh merupakan sebaran posterior yang bersifat pendekatan. □

### Masalah Komputasi

Pada dasarnya akan sangat sulit untuk menghitung besaran yang sedang dikaji dalam masalah parametrik yang bersifat nonlinear, sehingga perlu dilakukan penyederhanaan pendekatan masalah komputasi yang biasa digunakan, misalnya

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

metode rantai Markov Monte Carlo (MCMC). Pada bagian ini dibahas mengenai metode Bayes dengan sebaran prior dua-tahap yang akan menghasilkan sebaran posterior bagi dua buah *hyperparameter*. Perlu diketahui bahwa metode yang saat ini berkembang biasanya menghasilkan sebaran posterior bersyarat yang tidak eksplisit sehingga mengakibatkan contoh Gibbs (Gelfand dan Smith, 1990) agak sulit untuk digunakan. Untuk mengatasi masalah tersebut kemudian digunakan algoritma Metropolis-Hasting. Namun perlu dicatat bahwa jika sebaran bersyarat posterior nonstandar berbentuk log-konkaf, maka penarikan contoh Gibb dapat digunakan dengan menggunakan algoritma Gilks-Wild (Nandram, 2000). Nandram (2000) menyatakan bahwa sebaran posterior bersama bagi parameter yang diamati akan bersifat proper untuk sembarang model. Dalam penelitian ini, masalah komputasi didasarkan pada apa yang dilakukan oleh Ghosh et al. (1998) mengenai penerapan model linear terampat pada pendugaan area kecil. Proses komputasi dilakukan pada  $m$  buah area lokal atau  $m$  strata.

Misalkan  $Y_i$  menyatakan statistik cukup minimum (diskrit atau kontinu) yang berhubungan dengan unit pengamatan dalam strata ke- $i$  ( $i = 1, \dots, m$ ). Peubah acak  $Y_i$  diasumsikan sebagai peubah acak yang saling bebas yang merupakan anggota dari keluarga eksponensial. Untuk memudahkan mengkaitkan bentuk model keluarga eksponensial yang dituliskan dalam Persamaan (2.13) ke dalam masalah pemodelan Bayes berhirarki, maka fungsi kepekatan peluangnya dapat dituliskan lagi sebagai berikut

$$f(y_i|\theta_i, \phi_i) = \exp[\phi_i^{-1}(y_i\theta_i - \psi(\theta_i)) + \rho(y_i; \phi_i)] \quad \dots (3.5)$$

dimana  $i = 1, \dots, m$ . Fungsi kepekatan yang diberikan dalam (2.29) diparameterisasi terhadap parameter kanonik  $\theta_i$  dan parameter skala  $\phi_i > 0$ . Dalam hal ini parameter skala  $\phi_i$  diasumsikan diketahui nilainya.

Dalam penelitian ini peubah acak respons yang diperhatikan adalah yang bersebaran Poisson,  $Y_i \sim \text{Poisson}(\theta_i)$ , dimana fungsi peluangnya diberikan oleh

$$f(y, \theta_i) = \frac{\theta_i^y e^{-\theta_i}}{y!}, \text{ untuk } y = 0, 1, 2, \dots$$

Fungsi peluang di atas kemudian dituliskan dalam bentuk keluarga eksponensial sebagai berikut:

$$f(y_i|\theta_i, \phi_i) = \exp\left[\frac{y_i \log \theta_i - \theta_i}{a(\phi)} - \log y_i!\right]$$

Dalam hal ini parameter alamiah  $\lambda_i = \log \theta_i$ ,  $a(\phi) = \phi_i = 1$ ,  $\psi(\theta_i) = \theta_i$ , serta  $\rho(y_i; \phi_i) = \log y_i!$ . Hal ini menunjukkan bahwa sebaran Poisson merupakan anggota dari keluarga eksponensial.

Selanjutnya, parameter alamiah  $\theta_i$  terlebih dahulu dimodelkan sebagai

$$h(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + u_i + \varepsilon_i \quad (i = 1, \dots, m) \quad \dots (3.6)$$

dimana  $h$  merupakan fungsi naik;  $\mathbf{x}_{ik}$  adalah vektor rancangan berukuran  $(p \times 1)$ ,  $\boldsymbol{\beta}$  adalah vektor koefisien regresi berukuran  $(p \times 1)$ ,  $u_i$  merupakan efek acak, dan  $\varepsilon_i$  adalah galat. Disini diasumsikan bahwa  $u_i$  dan  $\varepsilon_i$  adalah saling bebas dengan  $u_i \sim N(0, \sigma_u^2)$  dan  $\varepsilon_i \sim N(0, \sigma^2)$ . Apabila diperhatikan lebih jauh, model yang diberikan dalam Persamaan (3.6) merupakan model Fay-Herriot yang dijadikan sebagai model dasar dalam pendugaan area kecil sebagaimana yang ditunjukkan dalam Persamaan (2.4). Dengan

demikian dapat dikatakan bahwa model linear terampat dapat dihubungkan ke masalah pendugaan area kecil melalui hubungan antara model dalam Persamaan (3.5) dan (3.6).

Apabila diperhatika lebih jauh persamaan yang dinyatakan dalam (3.5) dan (3.6) ternyata tidak membentuk model Bayes berhirarki. Akan tetapi model tersebut akan distandarkan sedemikian rupa sehingga mewakili suatu model dalam kerangka kerja metode Bayes sebagaimana yang telah dilakukan oleh Ghosh et al. (1998). Misalkan  $R_u = \sigma_u^{-2}$  dan  $R = \sigma^{-2}$ . Dimisalkan bahwa vektor parameter dan vektor galat efek acak masing-masing adalah  $\Theta = (\theta_1, \dots, \theta_m)^T$  dan  $\mathbf{u} = (u_1, \dots, u_m)^T$ .

Kemudian model hirarki yang dipertimbangkan adalah

- I. Bersyarat pada  $\Theta, \beta, \mathbf{u}, \mathbf{R}_u = r_u$  dan  $R = r$ , dimana peubah acak  $Y_{ik}$  adalah saling bebas dengan fungsi kepekatan yang diberikan dalam (3.5);
- II. Bersyarat pada  $\beta, \mathbf{u}, \mathbf{R}_u = r_u$  dan  $R = r$ , dan  $h(\theta_{ik}) \sim N(\mathbf{x}_i^T \beta + u_i, r^{-1})$ ;
- III. Bersyarat pada  $\beta, \mathbf{R}_u = r_u$  dan  $R = r$ , dan  $u_i \sim N(0, r_u^{-1})$ .
- IV. Untuk melengkapi model Bayes berhirarki, Ghosh et al. (1998) menentukan sebaran prior untuk  $\beta, \mathbf{R}_u = r_u$  dan  $R = r$ .
- IV. Besaran  $\beta, \mathbf{R}_u = r_u$  dan  $R = r$  adalah saling bebas dengan  $\beta \sim \text{uniform}(\mathbf{R}^p)$ , untuk  $p < m$ ,  $R_u \sim \text{gamma}(\frac{1}{2}a, \frac{1}{2}b)$ , dan  $R \sim \text{gamma}(\frac{1}{2}c, \frac{1}{2}d)$ .

Pada bagian (IV) suatu peubah acak  $Z \sim \text{gamma}(\alpha, \beta)$  apabila  $Z$  mempunyai fungsi kepekatan peluang sebagai berikut:

$$f(z) = \frac{\beta^\alpha \exp(-\beta z) z^{\alpha-1}}{\Gamma(\alpha)}, \text{ untuk } z > 0$$

Ketidakpastian yang dinyatakan dalam model I – IV berisi dua buah komponen, yaitu (i) efek atau pengaruh dari area lokal, dan (ii) komponen galat, yang berarti bahwa model tersebut mempertimbangkan masalah kelebihan dispersi (overdispersion) dengan cara menyertakan suatu komponen keragaman ekstra.

Perhatian utama dalam penelitian ini adalah menemukan sebaran posterior bagi  $g(\theta_i)$  yang bersyarat pada data  $\mathbf{y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{m1}, \dots, y_{mn_m})^T$ , dimana  $g$  adalah fungsi naik dan secara khusus untuk menemukan rata-rata, ragam dan koragam posterior dari parameter yang berada dalam model. Dalam aplikasi tertentu diketahui bahwa  $g(\theta_i) = \psi(\theta_i) = E(\mathbf{Y}_i | \theta_i)$ .

Untuk dapat menemukan rata-rata, ragam dan koragam, maka perlu diyakinkan terlebih dahulu bahwa sebaran posterior bersama dari  $\Theta_i$  dengan syarat  $\mathbf{y}$  adalah bersifat *proper*. Misalkan  $\theta_i$  dibatasi pada selang terbuka  $(\theta_i^{BB}, \theta_i^{BA})$ , dimana batas bawah (BB) dari interval dapat berupa  $-\infty$ , batas atas (BA) dari interval dapat berupa  $+\infty$ . Teorema untuk mendukung masalah ini diberikan sebagai berikut: diasumsikan bahwa  $a > 0, c > 0, \sum_i n_i - p + d > 0$ , dan  $m + b > 0$ . Kemudian jika

$$\int_{\theta_i^{BB}}^{\theta_i^{BA}} \exp\{[\theta_i - \phi(\Theta)]/\phi_i\} h(\theta) d\theta < \infty \quad \dots (3.7)$$

Untuk semua  $y_i$  dan  $\phi_i (> 0)$ , maka fungsi kepekatan posterior bersama bagi  $\theta_i$  dengan syarat  $\mathbf{y}$  adalah bersifat *proper* (Ghosh, et al., 1998).

Untuk menunjukkan bahwa bahwa fungsi kepekatan posterior bersama bagi  $\theta_i$  dengan syarat  $\mathbf{y}$  adalah bersifat *proper*, maka perhatikan bentuk fungsi kepekatan bersama berikut ini:

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$\begin{aligned}
 f(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{v}, \sigma_v^2, \sigma_e^2 | \boldsymbol{y}) &= \prod_i \exp[\phi_i^{-1}(y_i \theta_i - \psi(\theta_i))] (\sigma_e^2)^{1/2 \sum_{i=1}^m n_i} \\
 &= \prod_i \exp \left[ -\frac{\sigma_e^2}{2} \{h(\theta_i) - \mathbf{x}_i^T \boldsymbol{\beta} - \boldsymbol{v}_i\}^2 \right] \\
 &= \left( \prod_i h(\theta_i) \right) (\sigma_v^2)^{m/2} \exp \left( -\frac{\sigma_v^2}{2} \sum_{i=1}^m v_i^2 \right) \\
 &= \exp \left( -\frac{a \sigma_v^2}{2} \right) (\sigma_v^2)^{\frac{1}{2b}-1} \exp \left( -\frac{c \sigma_e^2}{2} \right) (\sigma_e^2)^{\frac{1}{2d}-1}
 \end{aligned}$$

Dengan cara mengintegalkan masing-masing terhadap  $\boldsymbol{\beta}$ ,  $\sigma_v^2$ , dan  $\sigma_e^2$  maka akan diperoleh

$$f(\boldsymbol{\theta}, \boldsymbol{v} | \boldsymbol{y}) = C \prod_i \exp(\phi_i^{-1}(y_i \theta_i - \psi(\theta_i))) \left( a + \sum_{i=1}^m v_i^2 \right) h(\theta_i)$$

dimana  $C$  merupakan suatu konstanta tertentu yang tidak bergantung pada  $\boldsymbol{\theta}$  dan  $\boldsymbol{v}$ . Kemudian bentuk di atas diintegalkan kembali terhadap  $\boldsymbol{v}$  sehingga diperoleh

$$f(\boldsymbol{\theta} | \boldsymbol{y}) = C \prod_i \exp(\phi_i^{-1}(y_i \theta_i - \psi(\theta_i))) h(\theta_i)$$

Dengan demikian terbukti bahwa fungsi kepekatan posterior bersama bagi  $\theta_i$  dengan syarat  $\boldsymbol{y}$  adalah bersifat *proper*.

Selanjutnya, untuk kasus dimana peubah respons berbentuk data cacahan yang mengikuti sebaran Poisson,

$$Y_i | \theta_i \sim \text{Poisson}(\exp(\theta_i))$$

Kemudian, jika  $h$  merupakan fungsi hubung kanonik, dan  $g(\theta_i) = \psi(\theta_i) = \exp(\theta_i)$ , maka kondisi dalam Persamaan (3.7) akan menghasilkan

$$\int_0^\infty \tau_i^{y_i-1} \exp(-\tau_i) d\tau_i < \infty$$

akan terpenuhi pada saat  $y_i = 1, 2, \dots$

Evaluasi langsung untuk sebaran posterior bersama bagi  $g(\theta_i)$  dengan syarat  $\boldsymbol{y}$  melibatkan integrasi numerik berdimensi tinggi, dan masalah ini dapat diselesaikan melalui metode rantai Markov Monte Carlo (MCMC), seperti menggunakan Gibbs penarikan contoh. Implementasi dari Gibbs penarikan contoh memerlukan contoh yang dibangkitkan dari sebaran posterior bersyarat. Misalkan

$$\mathbf{h}(\boldsymbol{\Theta}) = (h(\theta_{11}), \dots, h(\theta_{1n_1}), \dots, h(\theta_{m1}), \dots, h(\theta_{mn_m}))$$

dan

$$\mathbf{X} = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \dots, \mathbf{x}_{m1}, \dots, \mathbf{x}_{mn_m})$$

Dan  $\mathbf{X}$  adalah nonsingular. Kemudian sebaran posterior bersyarat yang diperlukan berdasarkan pada model Bayes berhirarki yang diberikan dalam (I) – (IV) adalah

- i.  $\boldsymbol{\beta} | \boldsymbol{\Theta}, \mathbf{u}, \mathbf{r}_u, \mathbf{r}, \mathbf{y} \sim N((\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{h}(\boldsymbol{\Theta}) - \sum_i u_i \sum_i \mathbf{x}_i), r^{-1}(\mathbf{X}^T \mathbf{X})^{-1});$
- ii.  $u_i | \boldsymbol{\Theta}, \boldsymbol{\beta}, \mathbf{r}_u, \mathbf{r}, \mathbf{y} \sim N((rn_i + r_u)^{-1} \sum_i (h(\theta_i) - \mathbf{x}_i^T \boldsymbol{\beta}); (rn_i + r_u)^{-1});$
- iii.  $R | \boldsymbol{\Theta}, \boldsymbol{\beta}, \mathbf{r}_u, \mathbf{u}, \mathbf{y} \sim \text{Gamma}\left(\frac{1}{2}(c + \sum_i (h(\theta_i) - \mathbf{x}_i^T \boldsymbol{\beta} - u_i)^2); \frac{1}{2}(d + \sum_{i=1}^m n_i)\right);$
- iv.  $R_u | \boldsymbol{\Theta}, \boldsymbol{\beta}, \mathbf{r}, \mathbf{u}, \mathbf{y} \sim \text{Gamma}\left(\frac{1}{2}(a + \sum_i u_i^2); \frac{1}{2}(b + \sum_{i=1}^m n_i)\right);$
- v.  $\theta_{ik} | \boldsymbol{\beta}, \mathbf{u}, \mathbf{r}_u, \mathbf{r}, \mathbf{y} \sim \pi(\theta_{ij} | \boldsymbol{\beta}, \mathbf{u}, \mathbf{r}_u, \mathbf{r}, \mathbf{y}) \propto \exp\left[\left(y_i \theta_i - \psi(\theta_i)\right) \phi_i^{-1} - \frac{r}{2}(h(\theta_i) - \mathbf{x}_i^T \boldsymbol{\beta} - u_i)^2\right] h(\theta_i)$

Contoh dapat dibangkitkan dengan mudah dari sebaran normal dan gamma sebagaimana yang diberikan dalam (i) – (iv). Namun demikian, sebaran posterior bersyarat dalam (v), dalam hal ini parameter  $\theta_{ij}$  bersyarat pada  $\boldsymbol{\beta}, \mathbf{u}, \mathbf{r}_u, \mathbf{r}, \mathbf{y}$  diketahui hanya pada konstanta multiplikatif, sehingga hal ini menjadi kesulitan dalam mengambil contoh dari sebaran posterior bersyarat seperti itu. Dalam kasus khusus dimana  $h(z) = z$  untuk seluruh  $z$ , Ghosh et al. (1998) telah menunjukkan bahwa  $\log \pi(\theta_i | \boldsymbol{\beta}, \mathbf{u}, \mathbf{r}_u, \mathbf{r}, \mathbf{y})$  merupakan fungsi konkaf bagi  $\theta_{ij}$ .

Inferensi mengenai  $\boldsymbol{\Theta}$  berdasarkan pada yang diberikan dalam (i) – (v) secara langsung dapat diperoleh dengan cara membentuk hasil analisis output dari Gibbs penarikan contoh. Artinya,  $E(\theta_i | \mathbf{y})$ ,  $V(\theta_i | \mathbf{y})$ , dan  $\text{cov}(\theta_i, \theta_i | \mathbf{y}) i \neq i'$  dapat dengan mudah diperoleh dari rumusan untuk nilai harapan dan ragam bersyarat yang diiterasikan. Hal ini merupakan penduga Rao-Blackwell sebagaimana yang ditunjukkan oleh Gelfand dan Smith (1990).

### Evaluasi Kekonvergenan Rantai Markov

Dalam penelitian ini dikaji suatu model yang memadukan antara konsep Bayes, khususnya Bayes berhirarki, dengan konsep model linear terampat yang diterapkan pada masalah pendugaan area kecil. Oleh karena model Bayes berhirarki ini memerlukan teknik integrasi berdimensi tinggi, maka masalah komputasinya perlu dilakukan melalui metode MCMC. Dengan demikian, proses diagnostik yang perlu dilakukan sehubungan dengan pemodelan ini menyangkut dua hal, yaitu diagnostik konvergensi rantai Markov, serta diagnostik untuk mengevaluasi kecocokan model sebagaimana yang biasa dilakukan dalam model linear terampat. Diagnostik kecocokan model ini dilakukan melalui berbagai ukuran statistik (seperti devians) dan analisis sisaan.

Selanjutnya, sebagaimana yang telah diketahui bahwa pada dasarnya akan sangat sulit untuk menghitung besaran yang sedang dikaji dalam masalah parametrik yang bersifat nonlinear, sehingga perlu dilakukan penyederhanaan pendekatan masalah komputasi yang biasa digunakan, misalnya seperti metode rantai Markov Monte Carlo (MCMC). Pendekatan umum untuk memperoleh penduga dari parameter dalam model area kecil melalui pendekatan Bayes berhirarki ini bersifat langsung. Dengan menggunakan representasi dalam model Poisson Bayes berhirarki dua-level sebagaimana yang ditunjukkan pada Persamaan (3.10) dan (3.12) secara langsung akan membawa pada sebaran posterior dengan sejumlah parameter yang diamati untuk vektor pengamatan  $\mathbf{y}$  tertentu. Penarikan contoh berulang dari sebaran posterior ini menghasilkan  $K$  buah nilai bagi segugus parameter yang akan diduga dan  $K$  buah nilai dari  $\tau_i$  untuk  $i = 1, \dots, m$ . Nilai dari  $\tau_i$  ini selanjutnya digunakan untuk menduga

beberapa ringkasan statistik, seperti rata-rata dan simpangan baku, sebaran posterior dari  $\tau_i$ .

Untuk mengambil contoh dari sebaran posterior yang diperlukan, akan digunakan metode rantai Markov Monte Carlo, khususnya dengan menggunakan algoritma Metropolis-Hastings (M-H) dengan setiap tahap Metropolis diperoleh dengan menggunakan rantai yang saling bebas. Pendekatan lain yang dapat digunakan selain algoritma Metropolis-Hastings (M-H) adalah algoritma Gibb sampler. Kedua pendekatan ini harus digunakan secara hati-hati, karena dalam prakteknya seringkali ditemui beberapa masalah. Pendekatan model Bayes berhirarki dalam pendugaan area kecil yang didasarkan pada metode rantai Markov Monte Carlo (MCMC) mempunyai beberapa keterbatasan, yaitu pemilihan sebaran prior yang tepat, proses komputasi dalam bentuk single run atau multiple run, serta penentuan periode *burn-in length*.

Tetiga permasalahan yang sering ditemui dalam metode MCMC ini dapat dideteksi melalui diagnostik konvergensi. Diagnostik ini perlu dilakukan untuk mengetahui seberapa lama proses rantai Markov diperlukan agar supaya memperoleh (atau paling tidak mendekati) suatu gugus pengamatan dengan sebaran yang bersifat stasioner dari suatu rantai Markov. Beberapa metode untuk melakukan diagnostik konvergensi akan dibahas pada bab ini yang meliputi diagnostik melalui grafik maupun melalui statistik formal tertentu. Kemudian bagian akhir pada bab ini akan diberikan hasil-hasil studi simulasi terhadap model yang sedang dikembangkan.

Metode MCMC memerlukan proses diagnostik untuk menentukan apakah simulasi yang dilakukan secara iteratif telah mencapai kesetimbangan sebaran dari rantai Markovnya. Rantai yang tercontoh perlu dijalankan untuk suatu periode *burn-in* awal tertentu sampai diperoleh contoh yang tepat dari sebaran posterior yang sedang diamati. Periode *burn-in* ini dapat bervariasi bergantung pada masalah yang sedang dihadapi. Selain itu, sangat penting dalam metode MCMC ini untuk memastikan bahwa rantai Markov yang terbentuk berada dalam ruang parameter yang tepat sedemikian rupa sehingga tidak menghasilkan contoh yang bersifat maksima lokal dari permukaan sebarannya. Dengan demikian sangat penting untuk memastikan bahwa penentuan periode *burn-in* sudah memadai untuk masalah yang sedang diteliti.

Terdapat banyak metode yang ada saat ini untuk memperkirakan masalah konvergensi dari suatu rantai dalam MCMC. Robert dan Casella (2005) memberikan beberapa banyak ulasan tentang masalah ini. Metode yang ada saat ini pada umumnya berdasarkan pada pemeriksaan sifat-sifat sebaran contoh dari rantai. Secara umum didefinisikan aliran output untuk vektor parameter  $\theta$  sebagai

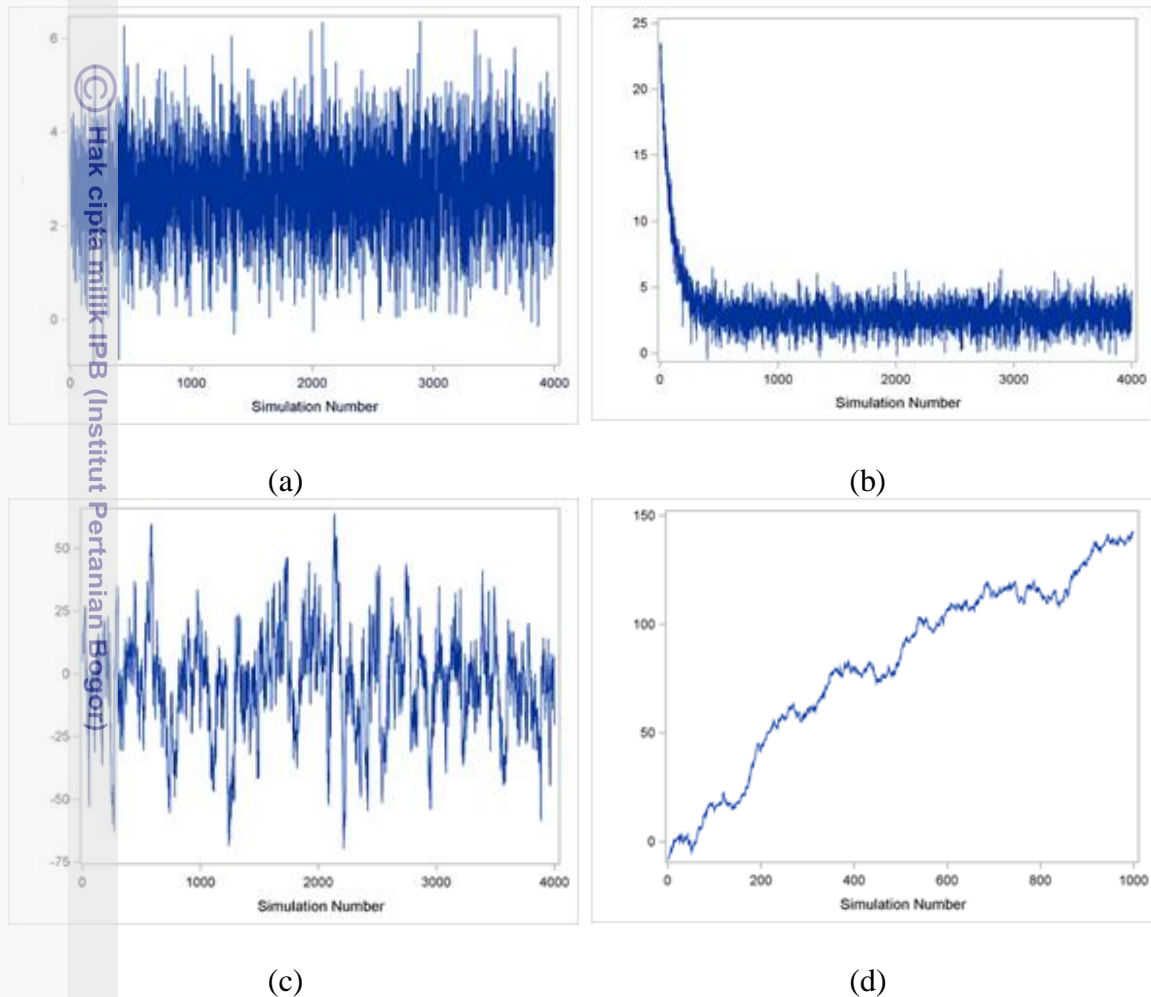
$$\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(k)}, \theta^{(k+1)}, \dots, \theta^{(k+k_K)}\}$$

Di sini nilai ke- $k$  adalah akhir dari periode *burn-in* pada ukuran contoh  $m_p$  yang sudah konvergen. Dengan demikian contoh yang sudah konvergen adalah  $\{\theta^{(k+1)}, \dots, \theta^{(k+k_K)}\}$ . Misalkan didefinisikan aliran outputnya sebagai  $\gamma = t(\theta)$ , sehingga  $\gamma^1 = t(\theta^1)$ . Pada bagian ini akan dibahas beberapa metode untuk menentukan kekonvergenan suatu rantai Markov dari prosedur MCMC, baik melalui analisis visual plot tras maupun melalui statistik uji formal.

### Metode Grafik

Pertama, untuk mengevaluasi kekonvergenan suatu rantai Markov dapat dilakukan secara grafis melalui plot tras contoh dengan indeks simulasi. Tras ini mendeskripsikan tentang apakah rantai sudah mencapai kekonvergenan terhadap suatu sebaran yang

stasioner atau belum. Apabila belum mencapai kekonvergenan, biasanya periode *burn-in* perlu diperpanjang. Suatu rantai Markov dikatakan sudah mencapai stasioner apabila sebaran dari titik-titik tidak berubah sepanjang rantai Markovnya. Konsep kestasioneran di sini dapat dilihat dengan dari plot tras apabila rata-rata dan ragamnya relatif konstan. Sebagai ilustrasi untuk menunjukkan bahwa suatu rantai Markov dikatakan sudah mencapai kekonvergenan disajikan pada Gambar 3.1.



Gambar 3.1 *Trace plot* untuk mengevaluasi kekonvergenan rantai Markov.

Gambar 3.1 (a) mengilustrasikan suatu plot tras yang ‘sempurna’, dalam arti bahwa proses rantai Markov sudah mencapai pada sebaran targetnya, atau dengan kata lain sudah mencapai kekonvergenan pada ke-stasioneran sebarannya. Gambar 3.1(b) mengilustrasikan bahwa plot tras untuk rantai yang dimulai pada suatu nilai awal tertentu menuju pada sebaran sasarannya. Beberapa ratus observasi pertama mungkin dapat diabaikan. Apabila hasil plot tras menghasilkan seperti Gambar 3.1(b) tersebut, maka perlu memperpanjang periode *burn-in*. Atau apabila menggunakan data tersebut untuk membuat inferensi, maka data yang digunakan adalah pada bagian akhir dari rantai Markovnya.

Selanjutnya, hasil tras plot yang disajikan pada Gambar 3.1(c) mengilustrasikan suatu rantai Markov yang bersifat campuran secara marginal. Hasil tras plot semacam ini biasanya berhubungan dengan adanya masalah otokorelasi yang tinggi antar contoh.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Untuk mengatasi masalah ini, biasanya perlu memperpanjang periode *burn-in*. Sementara itu, plot tras yang ditunjukkan pada Gambar 3.1(d) mengilustrasikan suatu rantai Markov yang mempunyai masalah yang serius, yang berarti bahwa rantai Markov tidak mencapai kekonvergenan. Untuk mengatasi masalah ini biasanya perlu melakukan reparameterisasi model pada skala logaritma, kemudian lakukan rantai Markov dengan periode *burn-in* yang jauh lebih panjang.

### Metode Rasio Ragam

Sebagaimana yang telah dibahas pada bab sebelumnya bahwa metode MCMC dapat dijalankan melalui proses tunggal ataupun ganda. Proses tunggal dapat memberikan penduga Monte Carlo bagi besaran posterior yang andal pada saat memilih  $D$  yang cukup besar. Akan tetapi hal ini masih menyisakan porsi yang signifikan dari ruang yang dibangkitkan oleh posterior  $f(\eta|y)$  yang secara keseluruhan belum terekplorasi. Untuk mengatasi masalah tersebut, Gelman dan Rubin (1992) menggunakan proses mutlipel-run dengan menggunakan beberapa nilai awal yang berbeda sehingga menghasilkan contoh paralel. Walaupun demikian, proses multiple run dapat sia-sia disebabkan oleh periode *burn-in* awal diabaikan dari setiap proses, walaupun hal ini bukan merupakan keterbatasan yang serius apabila menggunakan processor paralel untuk membangkitkan contoh paralel. Gelfand dan Smith (1990) menggunakan banyak proses short-run, dimana setiap proses berisi  $(l + 1)$  buah nilai  $\eta$ , dan hanya menyimpan observasi terakhir pada setiap proses.

Metode Gelman dan Rubin (1992) adalah menganalisis  $K$  buah barisan yang saling bebas untuk membentuk penduga sebaranonal yang diketahui mengenai beberapa peubah acak bersyarat pada data pengamatan yang disimulasikan. Metode ini dijadikan dasar untuk mengetahui bagaimana suatu penduga mendekati kekonvergenan. Proses ini dimulai dengan mensimulasikan  $K \geq 2$  barisan dengan panjang  $2n$ , pada setiap barisan dimulai dengan nilai awal yang berbeda yang bersifat overdispersi terhadap kestasioneran sebarannya. Pada proses ini  $n$  iterasi pertama dibuang dan sisanya adalah sebanyak  $n$ . Kemudian, untuk sembarang skalar fungsional yang diamati, sebut saja  $\theta(v)$ , akan dihitung  $B/n$ , yaitu ragam antar rata-rata  $K$  barisan yang dinotasikan sebagai  $\bar{\theta}_k$ . Kemudian didefinisikan

$$\frac{B}{n} = \frac{1}{K-1} \sum_{k=1}^K (\bar{\theta}_k - \bar{\theta})^2 \quad \dots (3.8)$$

dimana:

$\bar{\theta}_k = \frac{1}{n} \sum_{t=n+1}^{2n} \theta_k^{(t)}$ , dan  $\bar{\theta} = \frac{1}{K} \sum_{k=1}^K \bar{\theta}_k$ , dan  $\theta_k^{(t)} = \theta(v_k^{(t)})$  merupakan nilai  $\theta$  pengamatan ke- $t$  dari rantai  $k$ . Selanjutnya didefinisikan  $W$  sebagai rata-rata dari  $K$  buah ragam dalam barisan,  $s_k^2$ , yang masing-masing berdasarkan pada derajat bebas  $n - 1$ . Besaran  $W$  ini dihitung dengan menggunakan rumus:

$$W = \frac{1}{K} \sum_{k=1}^K s_k^2 \quad \dots (3.9)$$

dimana

$$s_k^2 = \frac{1}{n-1} \sum_{t=n+1}^{2n} (\theta_k^{(t)} - \bar{\theta}_k)^2$$

Kemudian, Gelman dan Rubin (1992) memberikan pendekatan sebaran  $t$  untuk menggambarkan keragaman dari  $\hat{\mu}$  dan  $\hat{\sigma}^2$ , dimana rata-rata  $\hat{\mu}$  dan ragam



$$\hat{V} = \frac{n-1}{n}W + \left(1 + \frac{1}{K}\right)\frac{B}{n} \quad \dots (3.10)$$

serta derajat bebasnya diduga oleh  $\widehat{db} = \frac{2\hat{V}^2}{\widehat{var}(\hat{V})}$ .

Selain itu, Gelman dan Rubin (1992) mengusulkan untuk mengevaluasi kekonvergenan rantai Markov dengan cara menduga suatu faktor yang merupakan penduga skala dari sebaran posterior  $\theta$  yang akan menciut pada saat  $n \rightarrow \infty$ . Besaran faktor tersebut dirumuskan sebagai

$$\hat{R}_c = \frac{d+3}{d+1}\frac{\hat{V}}{W} \quad \dots (3.11)$$

Nilai dari  $\hat{R}_c$  selanjutnya disebut sebagai faktor reduksi skala potensial (*potential scale reduction factor*, PSRF). Faktor ini diinterpretasikan sebagai suatu diagnostik kekonvergenan sebagai berikut. Jika  $\hat{R}_c$  bernilai besar, maka hal ini mengindikasikan bahwa penduga ragam  $\hat{\sigma}^2$  nilainya akan mengecil pada saat ditambah simulasinya, atau penambahan simulasi akan meningkatkan besaran  $W$ . Hal ini berarti bahwa metode MCMC belum mencapai kekonvergenan pada sebaran sasarannya. Sebaliknya, jika nilai PSRF ini mendekati 1, maka dapat disimpulkan bahwa setiap  $K$  gugus pengamatan yang disimulasikan akan mendekati sebaran sasarannya, atau dengan kata lain sudah mencapai kekonvergenan.

Sementara itu, Gelman et al. (2004) memberikan ukuran lain yang bisa digunakan untuk memonitor kekonvergenan tersebut diberikan oleh

$$R = \sqrt{\frac{n-1}{n} + \frac{1}{n}\frac{B}{W}} \quad \dots (3.12)$$

Suatu proses rantai Markov ini dikatakan konvergen apabila nilai  $R$  tersebut mendekati nilai 1. Menurut Gelman, et al (2004) apabila nilai  $R$  untuk seluruh parameter dan fungsi dari parameter itu sendiri berada antara 1.0 dan 1.1, maka keadaan kekonvergenan dapat diterima.

### Metode Spectral

Untuk mengevaluasi kekonvergenan dari rantai Markov yang dibentuk dari proses tunggal, statistik yang dapat digunakan disebut juga sebagai statistik Geweke. Statistik ini bekerja dengan cara untuk setiap barisan output dibagi ke dalam dua buah segmen berdasarkan suatu *burn-in* dengan panjang periode  $K$ . Misalkan didefinisikan panjang segmen pertama dan kedua masing-masing didefinisikan sebagai  $n_B$  dan  $n_A$ .

Menurut Brooks dan Gelman (1998) fungsi dari rantai Markov merupakan kasus khusus deret waktu yang stasioner. Dengan demikian metode deret waktu dapat digunakan untuk mengevaluasi kekonvergenan rantai Markov. Pada bagian ini dibahas sebuah metode diagnostik yang merupakan teknik yang sudah baku dari analisis spektral untuk memperoleh penduga ragam melalui kepekatan spektralnya, yaitu metode diagnostik melalui statistik Geweke.

Misalkan akan diduga nilai harapan dari beberapa fungsi,  $\theta(\mathbf{X})$ . Dimisalkan bahwa  $\theta^{(k)} = \theta(\mathbf{X}^{(k+n_0)})$ , untuk  $k = 1, 2, \dots, K$ , kemudian untuk barisan  $\{\theta^{(k)}\}$ , Geweke (1992) menyatakan bahwa jika suatu rantai Markov dikatakan konvergen menurut waktu  $n_0$ , maka seharusnya menerima suatu pengujian mengenai kesamaan lokasi untuk dua buah sub barisan  $\{\theta^{(k)}: k = 1, 2, \dots, n_A\}$  dan  $\{\theta^{(k)}: k = n^*, \dots, n\}$ , dimana  $1 < n_A < n^*$  dan  $n_B = n - n^* + 1$ . Kemudian didefinisikan

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$\theta_A = \frac{1}{n_A} \sum_{t=1}^{n_A} \theta^{(k)} \text{ dan } \theta_B = \frac{1}{n_B} \sum_{t=n^*}^n \theta^{(k)}$$

Dimisalkan pula  $\hat{S}_\theta^A(0)$  dan  $\hat{S}_\theta^B(0)$  masing-masing menyatakan penduga sensitivitas spektral yang konsisten untuk  $\{\theta^{(k)}: k = 1, 2, \dots, n_A\}$  dan  $\{\theta^{(k)}: k = n^*, \dots, n\}$ . Kemudian jika rasio  $n_A/n$  dan  $n_B/n$  adalah tetap, dengan ketentuan

$$\frac{(n_A + n_B)}{n} < 1$$

serta jika barisan dari  $\{\theta^{(k)}\}$  adalah stasioner, maka

$$Z_n = \frac{(\theta_A - \theta_B)}{\sqrt{\frac{1}{n_A} \hat{S}_\theta^A(0) + \frac{1}{n_B} \hat{S}_\theta^B(0)}} \sim N(0,1) \quad \dots (3.13)$$

Hal ini berlaku pada saat  $n \rightarrow \infty$ . Hasil ini dapat digunakan untuk menguji hipotesis nol mengenai kesamaan lokasi, dimana pada hipotesis nol ditolak, maka hal ini berarti bahwa rantai Markov tidak konvergen. Lebih jauh Geweke (1992) menyarankan untuk mengambil  $n_A = n/10$  dan  $n_B = n/2$ , dengan alasan bahwa pemilihan tersebut akan memenuhi asumsi di bawah Persamaan (3.13) dalam rangka memberikan kuasa diagnostik melawan kemungkinan bahwa barisan dari  $\{\theta^{(k)}\}$  tidak sepenuhnya mencapai kekonvergenan.

### Ukuran Kecocokan Model Bayes Berhierarchy

Kriteria kecocokan model bisa bervariasi bergantung pada sifat-sifat kriteria dan model yang dibentuk. Dalam pemodelan linear terampat dengan efek tetap, devians merupakan suatu ukuran kecocokan model yang penting (McCullagh dan Nelder, 1983). Devians sering digambarkan sebagai statistik kecocokan model (*goodness-of-fit statistics*). Statistik ini secara implisit merupakan perbandingan antara model yang diamati dengan model “maksimal” yang lebih kompleks. Model maksimal selalu mempunyai kecocokan yang lebih baik daripada model yang sedang diamati, tetapi perbedaan ini secara statistik dianggap signifikan. Dalam perhitungan devians untuk model logit, model maksimal sering dirujuk sebagai model penuh (*full model*). Model penuh ini mempunyai satu parameter untuk setiap peluang dugaan sehingga akan menghasilkan kecocokan yang sempurna terhadap data. Sebagai rasio likelihood, devians adalah sama dengan dua kali perbedaan absolut antara logkemungkinan untuk model taksiran, yang dinotasikan dengan  $l(y|\hat{\theta}_{fit})$ , dengan log-kemungkinan untuk model penuh, yang dinotasikan dengan  $l(y|\hat{\theta}_{full})$ .

Devians mengukur kelayakan model dengan cara membandingkan model dugaan terhadap model penuh (*saturated model*). Devians merupakan selisih antara log kemungkinan dari data untuk kedua model

$$D = -2[l(y|\hat{\theta}_{fit}) - l(y|\hat{\theta}_{full})] \quad \dots (3.14)$$

Suatu nilai  $D$  yang besar terjadi jika  $l(y|\hat{\theta}_{fit})$  mempunyai nilai yang relatif kecil terhadap  $l(y|\hat{\theta}_{full})$ , yang menunjukkan bahwa model prediksi adalah model yang kurang baik. Sebaliknya, suatu nilai  $D$  yang kecil diperoleh jika nilai  $l(y|\hat{\theta}_{fit})$  dan  $l(y|\hat{\theta}_{full})$  hampir sama, yang memberi indikasi bahwa model prediksi merupakan

model yang baik. Dengan demikian, dapat dikatakan bahwa statistik  $D$  atau devians ini digunakan untuk mengukur seberapa besar model prediksi menyimpang dari model penuh.

Model penuh mempunyai satu parameter tunggal per data pengamatan. Sering kali ukuran relatif dari kecocokan model digunakan sedemikian rupa sehingga devians dibandingkan dan perubahan dalam devians antara model 1 dan model 2 sebagai berikut:

$$\Delta D = -2[l(y|\hat{\theta}_1) - l(y|\hat{\theta}_2)] \quad \dots (3.15)$$

Dengan demikian untuk keperluan perbandingan model, maka fungsi likelihood untuk model penuh ini diabaikan. Ukuran devians biasa digunakan sebagai ukuran kecocokan model dalam metode Bayes, akan tetapi dalam pemodelan Bayes ini biasanya tanpa melibatkan model penuh.

Salah satu kelemahan dari penggunaan devians secara langsung adalah bahwa devians tidak mempertimbangkan derajat parameterisasi dalam model, artinya suatu model dapat dikatakan cocok terhadap data sebagaimana banyaknya parameter dalam model meningkat. Untuk mengatasi masalah tersebut kemudian dikembangkan beberapa ukuran untuk mengevaluasi kecocokan model, seperti ukuran *Akaike information criterion* (AIC) dan *Bayesian information criterion* (BIC). Statistik AIC didefinisikan sebagai

$$AIC = -2l(y|\hat{\theta}_{fit}) + 2p \quad \dots (3.16)$$

dimana  $p$  merupakan banyaknya parameter. Bentuk kedua dalam persamaan di atas bertindak sebagai faktor koreksi dalam masalah parameterisasi dari model. Konsepnya adalah semakin banyak parameter ditambahkan ke dalam model, maka model dianggap lebih cocok terhadap data. Sementara itu, BIC juga merupakan ukuran yang biasa digunakan sebagai kriteria pemilihan model. Ukuran ini banyak digunakan dalam model Bayes dan model berhirarki. BIC didefinisikan sebagai

$$BIC = -2l(y|\hat{\theta}_{fit}) + p \ln m \quad \dots (3.17)$$

Dalam suatu model dengan fungsi log-likelihood  $l(\theta)$  tertentu, nilai AIC dan BIC dapat diduga berdasarkan output dari algoritma MCMC melalui

$$AIC = -2l(y|\hat{\theta}) + 2p \quad \dots (3.18)$$

$$BIC = -2l(y|\hat{\theta}) + p \ln m \quad \dots (3.19)$$

dimana  $p$  adalah banyaknya parameter dan  $m$  menyatakan banyaknya titik data, serta dugaan fungsi log-likelihoodnya diberikan oleh

$$\hat{l}(\theta) = \frac{1}{K} \sum_{k=1}^K l(y|\theta^{(k)}) \quad \dots (3.20)$$

yang tidak lain dari fungsi log-likelihood yang dirata-ratakan pada  $G$  buah contoh posterior bagi  $\theta$ . Cara lain yang dapat ditempuh adalah penduga posterior  $\hat{\theta}$  sebelumnya dihitung terlebih dahulu, kemudian disubstitusikan ke dalam rumus AIC atau BIC. Menurut Lawson (2009) salah satu kelemahan dari ukuran AIC dan BIC ini adalah bahwa model dengan efek acak (*random effect*) akan sangat sulit untuk memutuskan berapa banyak parameter yang bisa disertakan ke dalam model.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak Cipta Dilindungi Undang-Undang

### Deviance Information Criterion

Salah satu ukuran kecocokan model yang dapat digunakan dalam mengevaluasi kecocokan model Bayes adalah *deviance information criterion* (DIC) sebagaimana yang diusulkan oleh Spiegelhalter et al, (2002). Kriteria ini didefinisikan sebagai

$$DIC = 2E_{\theta|y}(D) - D[E_{\theta|y}(\theta)] \quad \dots (3.20)$$

dimana  $D(\cdot)$  merupakan devians model dan  $y$  adalah data pengamatan. Perlu dicatat bahwa besaran DIC berdasarkan pada perbandingan rata-rata devians

$$\bar{D} = -2 \sum_{k=1}^K l(y|\theta^{(k)}) / G \quad \dots (3.21)$$

dan devians dari nilai harapan penduga parameter posterior,  $\hat{\theta}$ ,

$$\hat{D}(\hat{\theta}) = -2l(y|\hat{\theta}) \quad \dots (3.22)$$

Menurut Lawson (2009), untuk sembarang nilai parameter contoh  $\theta^{(k)}$ , maka devians ini menjadi  $\hat{D}(\theta^{(k)}) = -2l(y|\theta^{(k)})$ . Kemudian, banyaknya parameter yang efektif  $pD$ , diduga melalui  $p\hat{D} = \bar{D} - \hat{D}(\hat{\theta})$ , sehingga diperoleh  $DIC = \bar{D} + p\hat{D} = 2\bar{D} - \hat{D}(\hat{\theta})$ . Kelemahan dari ukuran DIC ini adalah dapat bernilai negatif, yaitu pada saat  $\hat{D}(\hat{\theta}) > \bar{D}$ . Ketidakstabilan dalam  $pD$  dapat membawa masalah pada penggunaan DIC sebagai ukuran untuk mengevaluasi model Bayes. Untuk mengatasi masalah ini disarankan untuk menghitung banyaknya parameter yang efektif menurut ragam posterior dari devians. Gelman et al (2004) mengusulkan suatu besaran untuk menduga  $pD$  yang didefinisikan sebagai berikut:

$$p\hat{D} = \frac{1}{2} \frac{1}{K-1} \sum_{k=1}^K (\hat{D}(\theta^{(k)}) - \bar{D})^2 \quad \dots (3.23)$$

Penduga ragam alternatif lainnya yang dapat digunakan untuk masalah ini adalah

$$v\hat{ar}(D) = \frac{1}{K-1} \sum_{k=1}^K (\hat{D}(\theta^{(k)}) - \bar{D})^2 = 2p\hat{D} \quad \dots (3.24)$$

Dengan demikian nilai DIC berdasarkan penduga ragam tersebut adalah

$$DIC = \bar{D} + v\hat{ar}(D) \quad \dots (3.25)$$

Perlu dicatat pula bahwa nilai harapan dari devians prediktif ( $\widehat{D}_{pr}$ ) merupakan ukuran kecocokan model alternatif dan nilai harapan ini ditaksir sebagai  $\widehat{D}_{pr} = 2\bar{D} - \hat{D}(\hat{\theta})$ .

### Posterior Predictive Loss

Gelfand dan Ghosh (1998) mengusulkan suatu fungsi rugi (*loss function*) berdasarkan pendekatan pada kecocokan model yang menggunakan sebaran prediktif. Pada dasarnya pendekatan ini membandingkan data pengamatan dengan data yang diprediksi dari model dugaan. Misalkan didefinisikan item data prediktif ke- $i$  dinyatakan sebagai  $y_i^{pr}$ . Perlu diketahui bahwa data prediktif dapat diperoleh melalui contoh posterior yang konvergen. Diberikan parameter akhir pada iterasi ke- $k$ ,  $\theta^{(k)}$ , maka

$$p(y_i^{pr} | \mathbf{y}) = \int p(y_i^{pr} | \theta^{(k)}) p(\theta^{(k)} | \mathbf{y}) d\theta^{(k)} \quad \dots (3.26)$$

Dengan demikian pada iterasi ke- $k$  dapat menghasilkan  $y_i^{pr}$  dari  $p(y_i^{pr} | \theta^{(k)})$ . Nilai prediktif ini mempunyai sebaran marginal  $p(\theta^{(k)} | \mathbf{y})$ . Dalam penelitian ini, dimana

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

peubah acak yang diamati diasumsikan mengikuti sebaran Poisson, maka bentuk item data prediktif ke- $i$  merupakan data cacahan sebagai berikut  $y_i^{pr} \leftarrow \text{Poisson}(e_i \theta_i^{(k)})$ .

Suatu fungsi rugi (*loss function*) mengasumsikan bahwa  $L_0(y, y^{pr}) = f(y, y^{pr})$ . Salah satu pilihan yang biasa digunakan dari fungsi rugi ini berupa kuadrat kerugian galat (*squared error loss*) yang didefinisikan sebagai:

$$L_0(y, y^{pr}) = (y - y^{pr})^2 \quad \dots (3.27)$$

Fungsi rugi lainnya yang yang biasa digunakan sebagai ukuran kecocokan model Bayes berhirarki dinyatakan dalam bentuk kerugian dalam data yang diberikan oleh rata-rata kerugian dari seluruh item. Ukuran ini biasa dinyatakan dalam bentuk kuadrat rata-rata galat prediktif (*mean squared predictive error*, MSPE) yang didefinisikan sebagai berikut:

$$MSPE_k = \sum_i (y_i - y_{ik}^{pr})^2 / m \quad \dots (3.28)$$

$$MSPE = \sum_i \sum_k (y_i - y_{ik}^{pr})^2 / (K \times m) \quad \dots (3.29)$$

dimana  $m$  merupakan banyaknya observasi dan  $K$  adalah ukuran contoh yang tercontoh. Ukuran lainnya yang dapat digunakan di sini dapat dinyatakan dalam bentuk absolutnya, yang disebut juga sebagai *mean absolute predictive error* (MAPE), yang didefinisikan sebagai berikut:

$$MAPE_k = \sum_i |y_i - y_{ik}^{pr}| / m \quad \dots (3.30)$$

$$MAPE = \sum_i \sum_k |y_i - y_{ik}^{pr}| / (K \times m) \quad \dots (3.31)$$

Gelfand dan Ghosh (1998) mengusulkan bentuk ukuran kecocokan model yang lebih kompleks berdasarkan fungsi rugi ini. Fungsi rugi yang diusulkan oleh Gelfand dan Ghosh (1998) ini diberikan oleh

$$D_g = \frac{g}{g+1} A + B \\ = \frac{g}{g+1} \sum_{i=1}^m (y_i - \overline{y_i^{pr}}) + \frac{1}{m_p} \sum_{i=1}^m \sum_{k=1}^{m_p} (y_{ik}^{pr} - \overline{y_i^{pr}})^2 \quad \dots (3.31)$$

dimana:  $\overline{y_i^{pr}} = \sum_{k=1}^{m_p} y_{ik}^{pr} / m_p$ , dan  $m_p$  merupakan ukuran contoh prediksi (dalam hal ini biasanya  $K = m_p$ ). Di sini, besaran  $g$  dapat dipilih sedemikian rupa sehingga memberikan bobot untuk komponen yang berbeda. Untuk  $k = \infty$ , maka  $D_g = A + B$ . Pemilihan besarnya  $g$  ini biasanya tidak mempengaruhi urutan kecocokan model. Masing-masing komponen mempunyai makna yang berbeda mengenai kecocokan model, dimana komponen  $A$  menyatakan ketidakcocokan model dan komponen  $B$  mengukur derajat kehalusannya (*smoothness*). Suatu model dikatakan cocok apabila mempunyai nilai  $D_k$  (atau MSPE atau MAPE) yang lebih kecil.

### Analisis Sisaan pada Model Bayes Berhirarki

Dalam pemodelan Bayes berhirarki, setelah dicapai kekonvergenan algoritma MCMC tidak senantiasa menghasilkan model yang baik. Apabila model dinyatakan dengan kaliru, maka penggunaan model tersebut menjadi terbatas. Terdapat banyak hal yang berkenaan dengan uji kecocokan model yang perlu dipertimbangkan sebagaimana

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

yang telah dibahas sebelumnya. Pada bagian ini akan dikembangkan suatu analisis sisaan yang perlu diperhatikan pada saat mengevaluasi kelayakan suatu model.

### Sisaan Umum

Analisis sisaan dan ringkasan dari fungsi sisaan merupakan bagian paling mendasar dalam menentukan kecocokan model dalam berbagai aplikasi pemodelan statistika. Hal ini termasuk juga dalam model Poisson Bayes berhirarki dua-level yang biasa diterapkan untuk memodelkan angka mortalitas, walaupun analisis sisaan ini jarang dipublikasikan dalam berbagai literatur. Di sini didefinisikan sisaan individu sebagai selisih antara nilai pengamatan dan nilai taksiran model yang dibakukan. Pada umumnya proses pembakuan ini dilakukan berdasarkan pada ukuran keragaman dari perbedaan antara dua buah nilai.

Sisaan itu sendiri dinyatakan sebagai  $r_i = y_i - \hat{y}_i$  atau  $r_i^s = r_i / \sqrt{\widehat{var}(r_i)}$ , dimana  $\hat{y}_i$  merupakan nilai taksiran yang diperoleh dari model tertentu. Pada saat mengkaji model spasial, maka seringkali menghitung sisaan seperti  $r_i$  dengan menggunakan metode Monte Carlo. Akan tetapi perhitungan sisaan akan diperoleh secara langsung ketika menggunakan metode *parametric bootstrap* (PB). Untuk kasus dimana datanya berbentuk cacahan, khususnya ketika model likelihood Poisson diasumsikan dengan  $y_i \sim \text{Poisson}(e_i \theta_i)$ . Kemudian metode PB dapat secara langsung diimplementasikan dengan jalan membangkitkan segeugus data cacahan hasil simulasi, sebut saja  $\{y_{ik}^*\}$ , untuk  $k = 1, \dots, K$ , dari contoh acak yang bersebaran Poisson dengan rata-rata  $e_i \theta_i$ . Dengan cara seperti ini maka peringkat *tract-wise*, dan juga p-value, dapat dihitung dengan cara menentukan peringkat sisaan dalam gugus data gabungan dari  $K + 1$  buah sisaan, yang dinyatakan sebagai

$$\{y_i - e_i \hat{\theta}_i; \{y_{ik}^* - e_i \hat{\theta}_i\}, k = 1, \dots, K\}$$

Misalkan sisaan dibakukan pengamatan dinyatakan sebagai  $r_i^s$  dan sisaan dibakukan hasil simulasi dinyatakan sebagai  $r_{ik}^s$ . Dengan menyatakan keadaan seperti itu dimungkinkan untuk membandingkan fungsi sisaan untuk melakukan perbandingan secara langsung. Sebagai contoh, dalam konteks spasial, cukup tepat untuk menentukan otokorelasi spasial dari sisaan observasi yang akan memberikan informasi tentang ketidakcocokan model. Dengan demikian, untuk menentukan derajat atau besarnya otokorelasi sisaan yang diperoleh melalui prosedur Monte Carlo dapat dilakukan dengan cara membandingkan statistik Moran,  $M(\{r_i^s\})$ , untuk sisaan pengamatan dengan statistik Moran untuk sisaan hasil simulasi,  $M(\{r_{ik}^s\})$ , dimana:

$$M(\{u\}) = \frac{u^T W u}{u^T u}, \text{ serta } u_i = r_i / \sqrt{\widehat{var}(r_i)}; r_i = (y_i - e_i \hat{\theta}_i);$$

dan  $W$  merupakan matriks *adjacency*.

### Sisaan Bayes

Carlin dan Louis (2000) mengusulkan suatu bentuk sisaan yang lebih tepat ketika melakukan pemodelan melalui metode Bayes yang dikenal juga sebagai sisaan Bayes yang didefinisikan sebagai berikut:

$$r_i = y_i - \frac{1}{K} \sum_{k=1}^K E(y_i | \theta_i^{(k)}) \quad \dots (3.32)$$

dimana  $E(y_i|\theta_i)$  merupakan nilai harapan dari sebaran posterior prediktif, dan dalam konteks penarikan contoh MCMC, bentuk  $\theta_i^{(k)}$  merupakan segugus nilai parameter yang tercontoh dari sebaran posterior.

Dalam pemodelan data cacahan, dengan fungsi kemungkinan yang didasarkan pada sebaran Poisson dengan nilai harapan  $e_i\theta_i$ , maka sisaan yang diberikan dalam Persamaan (3.32) dapat didekati melalui persamaan

$$r_i = y_i - \frac{1}{K} \sum_{k=1}^K e_i \theta_i^{(k)} \quad \dots (3.33)$$

Sisaan di atas merupakan rata-rata dari contoh posterior. Suatu bentuk alternatif dalam memudahkan komputasi adalah dengan jalan mengambil rata-rata contoh  $\theta_i^{(k)}$ , yaitu  $\hat{\theta} = \frac{1}{K} \sum_{k=1}^K e_i \theta_i^{(k)}$  untuk menghasilkan nilai harapan posterior bagi  $y_i$ , atau  $\hat{y}_i = e_i \hat{\theta}_i$ , sehingga diperoleh  $r_i = y_i - \hat{y}_i$ . Sisaan ini dapat memberikan ukuran kecocokan dari setiap titik data pengamatan, sebagaimana ukuran kecocokan model secara keseluruhan (seperti kuadrat tengah galat,  $KTG = \frac{1}{m} \sum_{i=1}^m r_i^2$ ), serta dapat dihitung dengan menggunakan metode Monte Carlo.

Untuk keperluan eksplorasi maka sisaan tersebut akan sangat bermanfaat apabila dilakukan pembakuan terlebih dahulu, walaupun hal ini menurut Lawson (2009) tidak begitu penting terutama dalam metode Monte Carlo. Untuk memberikan pendugaan Monte Carlo dari perilaku sisaan secara individu, maka perlu dilakukan pengulangan pada simulasi Monte Carlo dari contoh yang saling bebas dari sebaran prediktifnya. Hal ini dapat dicapai dengan cara mengambil  $K$  contoh dari hasil prosedur MCMC yang sudah konvergen.

## 4 STUDI SIMULASI

### Pendahuluan

Bab ini membahas tentang studi simulasi yang dilakukan dengan tujuan untuk mengevaluasi performa dari model Bayes berhirarki dua-level dengan peubah respons yang mengikuti sebaran Poisson atau dapat dinyatakan dalam bentuk:  $y_i|\theta_i \sim \text{Poisson}(n_i\theta_i)$ , untuk  $i = 1, \dots, m$ , dimana parameter  $\theta_i$  selanjutnya dianggap sebagai suatu peubah acak yang mengikuti sebaran tertentu. Pemilihan sebaran prior inilah yang menjadi isu utama dalam studi simulasi ini. Ada dua sebaran prior yang dipertimbangkan dalam studi simulasi, yaitu sebaran prior gamma dan sebaran prior invers gamma, yang masing-masing dapat ditulis dalam bentuk:  $f(\boldsymbol{\theta}|\mathbf{y}) \sim \text{gamma}(a, b)$  dan  $f(\boldsymbol{\theta}|\mathbf{y}) \sim \text{invers gamma}(a, b)$ . Sebagaimana yang telah diketahui bahwa sebaran gamma merupakan sebaran prior yang bersifat *conjugate* bagi sebaran Poisson. Sedangkan sebaran invers gamma merupakan sebaran prior yang bersifat *nonconjugate* bagi sebaran Poisson, tapi masih merupakan suatu sebaran yang bersifat *conditionally conjugate*.

Studi simulasi ini kemudian dilakukan dengan membandingkan performa dari model Poisson Bayes berhirarki dua-level yang berasal dari dua sebaran prior yang berbeda tersebut, yaitu sebaran gamma dan invers gamma. Pembahasan mengenai studi simulasi ini dimulai dengan menspesifikasikan proses simulasinya, proses

pembangkitan data, dan pembahasan hasil-hasil dari studi simulasi. Terdapat empat hal yang akan dikaji sehubungan dengan performa model yang akan menjadi perhatian utama dari studi simulasi ini, yaitu diagnostik kekonvergenan rantai Markov, dalam hal ini diagnostik difokuskan pada rantai Markov yang dijalankan melalui proses tunggal; diagnostik kecocokan model dengan cara menerapkan konsep pemodelan linear terampat dalam model Bayes berhirarki pada masalah pendugaan area kecil, termasuk didalamnya adalah melakukan analisis sisaan; sifat-sifat dari penduga parameter model Poisson Bayes berhirarki, terutama yang berkaitan dengan ketidakbiasan dan akurasi dari penduga parameter yang diamati, termasuk didalamnya adalah galatbaku dari sebaran posteriornya; serta nilai prediksi untuk sebaran posterior, dimana ukuran statistik yang diamatinya adalah rata-rata dan simpangan baku prediksi, serta selang kepercayaannya.

### Spesifikasi Simulasi

Secara umum, simulasi ini dilakukan dengan spesifikasi bahwa contoh diambil berdasarkan pada sebaran bersyarat penuh. Untuk implementasinya simulasi ini akan dilakukan dengan spesifikasi sebagai berikut:

- a. Menetapkan proses ganda secara paralel yaitu sebanyak tiga yang masing-masing mempunyai panjang 'burn-in' sebesar  $L = 1000$ , serta ukuran penarikan contoh Gibbs sebesar  $K = 10.000$ .
- b. Menetapkan nilai parameter gamma  $a$  dan  $b$  yang relatif cukup kecil, dalam ini digunakan nilai  $a = b = 0.002$ .
- c. Menetapkan prior untuk *hyperparameter*  $a$  dan  $b$  yang masing-masing juga mengikuti sebaran gamma.

Simulasi berdasarkan model Bayes berhirarki dua-level untuk peubah respons yang mengikuti sebaran Poisson sebagaimana yang ditunjukkan pada Persamaan (3.1), atau dalam hal ini  $y_i | \theta_i \sim \text{Poisson}(n_i \theta_i)$ , untuk  $i = 1, 2, \dots, m$ . Dalam hal ini  $\theta_i$  merupakan parameter mengenai angka kematian (*mortality rate*) yang diasumsikan mengikuti suatu sebaran Poisson. Mengikuti kajian teori tentang model Poisson Bayes berhirarki yang dikembangkan pada bab sebelumnya, maka dalam hal terdapat dua model yang akan dipertimbangkan. Yang membedakan dari kedua model ini adalah terletak pada pemilihan sebaran prior bagi parameter  $\theta_i$ .

### Model dengan Prior Gamma

Model pertama yang dipertimbangkan dalam studi simulasi ini adalah model Poisson Bayes berhirarki, dimana parameter  $\theta_i$  merupakan suatu parameter yang berkenaan dengan angka kematian (*mortality rate*) yang diasumsikan mengikuti sebaran gamma. Perlu diketahui bahwa parameter  $\theta_i$  yang bersebaran gamma ini merupakan level pertama dari model Bayes berhirarki dua-level, sedangkan level kedua dari hirarki ini terletak pada parameter gamma  $a$  yang bersebaran hyperprior,  $h_a(v)$  dan parameter gamma  $b$  yang bersebaran *hyperprior*  $h_b(\rho)$ , dimana  $v$  dan  $\rho$  masing-masing menunjukkan parameter dari sebaran *hyperprior* tersebut. Untuk melengkapi konsep model linear terampat pada masalah pendugaan area kecil dalam model Bayes berhirarki ini, maka model dapat dituliskan secara lengkap sebagai berikut:



- i. Bersyarat pada  $\theta, \beta, \nu, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka peubah acak  $y_i \sim \text{Poisson}(n_i\theta_i)$ .
- ii. Bersyarat pada  $\beta, \nu, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka  $\theta_i \sim \text{gamma}(a, b)$
- iii. Bersyarat pada  $\beta, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka  $u_i \sim N(0, \sigma_u^2)$
- iv. Untuk melengkapi model Bayes berhirarki, maka diperlukan prior untuk parameter  $\beta$  dan  $\sigma_\varepsilon^2$ , dimana  $\beta, \sigma_\nu^2$ , dan  $\sigma_\varepsilon^2$  adalah saling bebas, dengan  $\beta \sim N(\mu, \sigma^2)$  dan  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ .

**Model dengan Prior Invers Gamma**

Spesifikasi simulasi kedua yang dipertimbangkan dalam penelitian ini adalah dengan menggunakan sebaran prior untuk parameter  $\theta_i$  berdasarkan pada sebaran invers gamma atau  $\theta_i \sim \text{Invers Gamma}(a, b)$ , dengan mengambil nilai  $a$  dan  $b$  sama seperti pada spesifikasi simulasi pertama, yaitu nilai  $a = b = 20$ . Sedangkan prior untuk *hyperparameter*  $a$  dan  $b$  yang masing-masing juga mengikuti sebaran invers gamma. Model 2 ini dapat dituliskan sebagai berikut:

- i. Bersyarat pada  $\theta, \beta, \nu, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka peubah acak  $y_i \sim \text{Poisson}(n_i\theta_i)$ .
- ii. Bersyarat pada  $\beta, \nu, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka  $\theta_i \sim \text{invers gamma}(a, b)$
- iii. Bersyarat pada  $\beta, \sigma_u^2$ , dan  $\sigma_\varepsilon^2$ , maka  $u_i \sim N(0, \sigma_u^2)$
- iv. Untuk melengkapi model Bayes berhirarki, maka diperlukan prior untuk parameter  $\beta$  dan  $\sigma_\varepsilon^2$ , dimana  $\beta, \sigma_\nu^2$ , dan  $\sigma_\varepsilon^2$  adalah saling bebas, dengan  $\beta \sim N(\mu, \sigma^2)$  dan  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ .

**Pembangkitan Data**

Untuk keperluan studi simulasi ini, peubah respons,  $y_i$ , dibangkitkan dengan menggunakan terminologi yang ada dalam model linear terampat. Dalam hal ini peubah respons,  $y_i$ , diasumsikan mengikuti sebaran Poisson dan menggunakan fungsi hubungan log. Oleh karena dalam penelitian ini memerlukan peubah  $n_i$ , yang di dalam terminologi model linear terampat dianggap sebagai peubah *exposure*, maka model yang dipertimbangkan untuk keperluan pembangkitan data adalah:

$$\log \mu_i = \eta_i = o_i + \mathbf{x}_i^T \beta \quad \dots (4.1)$$

Dalam model ini peubah respons diasumsikan sebagai  $y_i \sim \text{Poisson}(\mu_i)$ , dimana  $\mu_i = n_i\theta_i$ . Untuk kasus dalam penelitian ini parameter  $\theta_i$  merupakan angka mortalitasnya. Besaran  $o_i$  dalam Persamaan (4.1), dalam terminologi model linear terampat disebut juga sebagai *offset*, dan dirumuskan sebagai  $o_i = \log(n_i)$ . Peubah *offset* ini merupakan kovariat dalam prediktor linear dimana koefisiennya tidak ditaksir, akan tetapi diasumsikan sama dengan satu.

Dalam studi simulasi ini akan dilakukan untuk model regresi Poisson dengan dua buah peubah prediktor,  $x_1$  dan  $x_2$ , sehingga model yang dipertimbangkan adalah

$$\log \mu_i = \eta_i = o_i + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad \dots (4.2)$$

Tahap-langkah proses pembangkitan data dilakukan sebagai berikut:

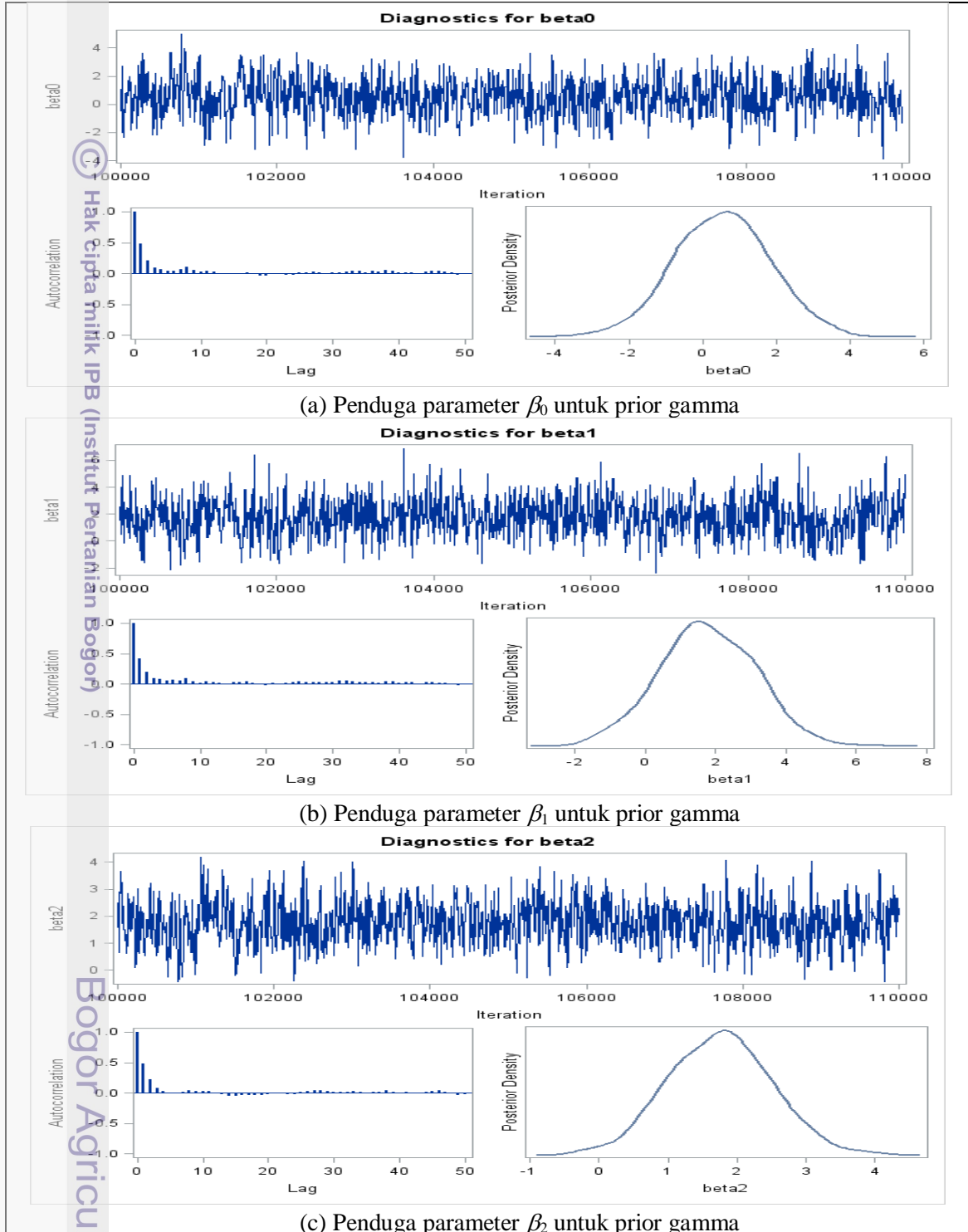
- a. Menetapkan banyaknya pengamatan, yaitu  $m = 10$ .
- b. Menetapkan nilai-nilai koefisien regresi, yaitu  $\beta_0 = 0.5, \beta_1 = 1.5$ , dan  $\beta_2 = 2.0$ .
- c. Menetapkan nilai-nilai untuk peubah *offset*,  $o_i$ , untuk  $i = 1, 2, \dots, m$ .
- d. Membangkitkan data untuk peubah prediktor  $x_1$  dan  $x_2$  dengan ketentuan sebagai berikut:  $x_1 \sim \text{uniform}(0, 100)$  dan  $x_2 \sim \text{uniform}(15, 40)$ .

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Himpunan Ilmiah IPB (Institut Pertanian Bogor) Bogor Agricultural University

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

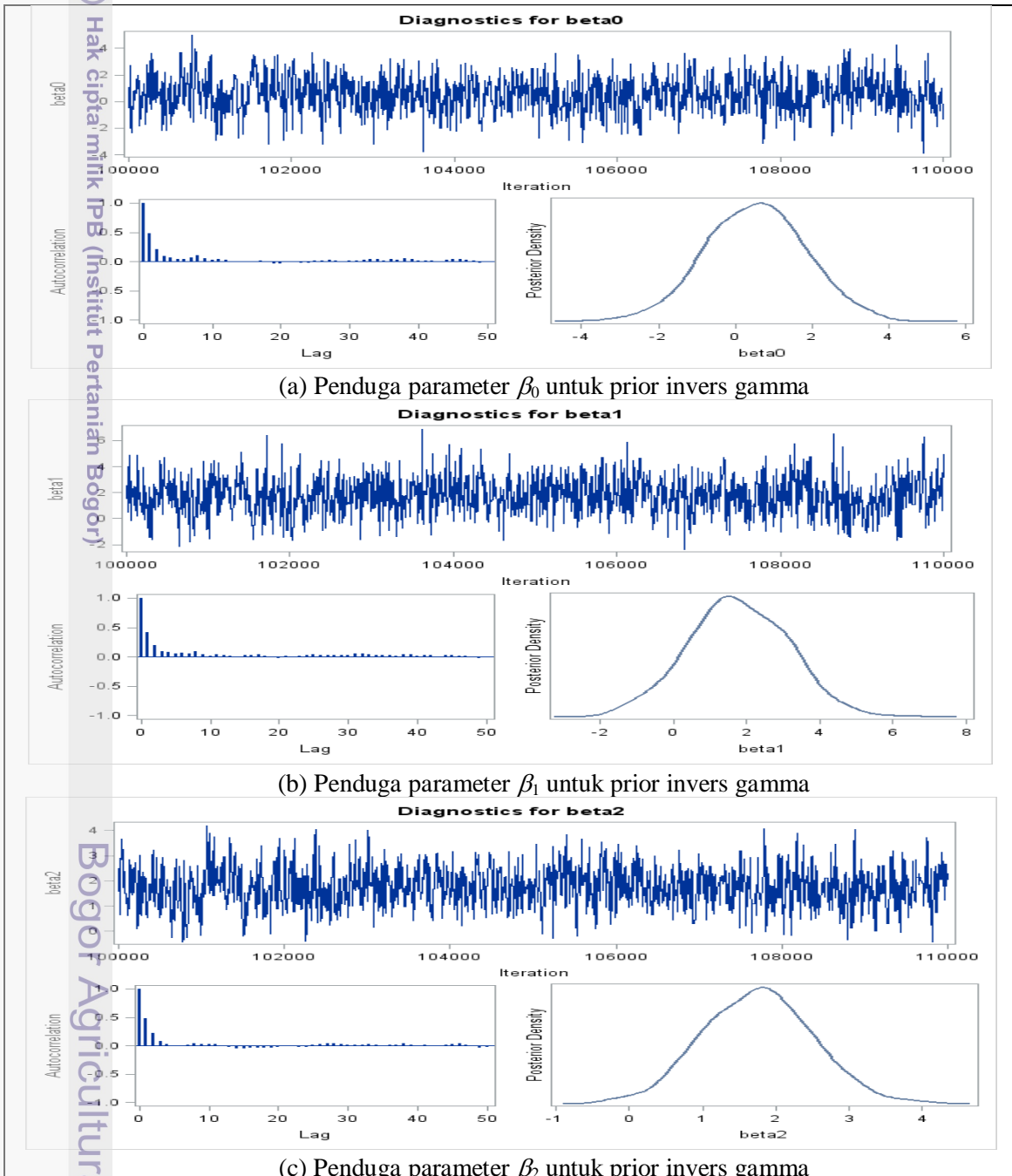
- e. Menghitung besaran  $\mu_i$ , dengan rumus:  $\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})$
- f. Membangkitkan data untuk peubah respons yang mengikuti sebaran Poisson dengan parameter  $\mu_i$  yang diperoleh pada bagian sebelumnya.



Gambar 4.1 Trace plot untuk studi simulasi yang berasal dari sebaran prior Gamma.

### Hasil-hasil Simulasi

Sebagaimana yang telah disebutkan pada bagian sebelumnya, hasil-hasil dari studi simulasi yang dilakukan pada penelitian ini adalah untuk melihat performa model terutama yang berkenaan ketidakbiasan dan akurasi dari penduga parameter dan galat bakunya. Selain itu, inferensi Bayes berdasarkan studi simulasi ini perlu dievaluasi apakah rantai Markov telah mencapai kestasioneran dari sebaran posterior yang diinginkan atau tidak. Proses ini dilakukan melalui diagnostik kekonvergenan dengan menggunakan *trace plot* atau dengan menggunakan beberapa ukuran statistik yang relevan



(a) Penduga parameter  $\beta_0$  untuk prior invers gamma

(b) Penduga parameter  $\beta_1$  untuk prior invers gamma

(c) Penduga parameter  $\beta_2$  untuk prior invers gamma

Gambar 4.2 *Trace plot* untuk studi simulasi yang berasal dari sebaran prior Invers Gamma.

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

### Diagnostik Kekonvergenan Rantai Markov dalam Proses Tunggal

Hasil dari diagnostik kekonvergenan dari studi simulasi ini dilakukan melalui *trace plot* sebagaimana yang ditunjukkan pada Gambar 4.1 dan Gambar 4.2. *Trace plot* adalah memplotkan antara contoh dengan indeks simulasi yang dapat digunakan sebagai alat untuk memperkirakan apakah kekonvergenan suatu rantai Markov terhadap kestasioneran sebarannya sudah tercapai atau belum. Apabila kekonvergenan terhadap kestasioneran sebaran ini belum tercapai, maka perlu menambah periode ‘*burn-in*’ dalam proses simulasinya. Suatu rantai Markov dikatakan mencapai kestasioneran apabila sebaran dari titik-titiknya tidak berubah sebagaimana perkembangan rantai Markovnya. Dalam hal ini dapat dilihat melalui *trace plot* yang relatif konstan antara rata-rata dan ragamnya melalui plot fungsi otokorelasi antar contoh dan plot kepekatan peluang.

Berdasarkan hasil *trace plot* yang ditunjukkan pada Gambar 4.1 terlihat bahwa nilai dugaan parameter untuk setiap parameter ( $\beta_0$ ,  $\beta_1$ , dan  $\beta_2$ ) dari sebaran prior gamma dipusatkan di sekitar nilai yang asal yang dibangkitkan, yaitu  $\beta_0 = 0.5$ ,  $\beta_1 = 1.5$ , dan  $\beta_2 = 2.0$  (lihat Gambar 4.1a, 4.1b, dan 4.1c). Hal ini menunjukkan bahwa proses simulasi MCMC yang didasarkan pada sebaran prior gamma sudah mencapai kekonvergenan terhadap kestasioneran sebaran posteriornya.

Lebih jauh, berdasarkan plot otokorelasi dan plot kepekatan peluang yang ditunjukkan pada Gambar 4.1 memberikan indikasi yang sejalan dengan hasil *trace plot*. Menurut hasil plot otokorelasi untuk setiap parameter yang diamati menunjukkan pola yang hampir sama, yaitu terjadi penurunan yang cepat setelah lag kedua. Demikian juga halnya menurut plot kepekatan peluang menunjukkan pola bahwa sebaran dari penduga parameter yang diamati adalah cenderung simetris.

Sementara itu, hasil plot otokorelasi dan plot kepekatan peluang yang ditunjukkan Gambar 4.2 yang dihasilkan dari sebaran prior invers gamma untuk setiap parameter yang diamati menampilkan pola yang tidak jauh berbeda dengan apa yang dihasilkan dari sebaran prior gamma. Namun demikian, berdasarkan hasil dari *trace plot*nya menunjukkan bahwa nilai dugaan parameter yang diamati ( $\beta_0$ ,  $\beta_1$ , dan  $\beta_2$ ) tidak jauh berbeda dengan nilai aslinya, akan tetapi mempunyai fluktuasi relatif cukup besar. Hal ini ditunjukkan oleh nilai simpangan baku untuk sebaran prior invers gamma relatif lebih besar dibandingkan dengan simpangan baku untuk prior gamma.

Tabel 4.1 Diagnostik kekonvergenan rantai Markov melalui statistik uji Geweke

Parameter	Prior Gamma		Prior Invers Gamma	
	z	Pr >  z	z	Pr >  z
$\beta_0$	0.6637	0.5069	0.6929	0.4884
$\beta_1$	1.2032	0.2289	1.6075	0.1079
$\beta_2$	-0.8148	0.4152	-0.0461	0.9632

Hasil-hasil *trace plot* di atas juga didukung oleh hasil dari statistik uji Geweke. Pada dasarnya statistik uji Geweke ini dilakukan untuk menguji apakah rata-rata penduga telah mencapai kekonvergenan dengan cara membandingkan rantai Markov yang telah terbagi ke dalam dua bagian. Selain itu, statistik ini juga merupakan uji dua-pihak yang berdasarkan pada statistik skor-Z, dimana suatu nilai Z absolut yang besar

mengindikasikan bahwa hasil pengujian adalah signifikan. Artinya bahwa secara rata-rata penduga parameter yang diamati belum mencapai kekonvergenan. Hasil pengujian kekonvergenan untuk masing-masing parameter  $\beta_0$ ,  $\beta_1$ , dan  $\beta_2$  baik yang diperoleh berdasarkan prior gamma maupun sebaran prior invers gamma disajikan pada Tabel 4.1.

Berdasarkan hasil pengujian melalui statistik uji Geweke terlihat bahwa rata-rata penduga parameter  $\beta_0$ ,  $\beta_1$ , dan  $\beta_2$  yang berasal dari sebaran prior invers gamma sudah mencapai kekonvergenan. Hal ini terlihat dari nilai statistik Geweke untuk masing-masing parameter tersebut sebesar 0.6929, 1.6075, dan -0.0461, dengan nilai p-value masing-masing sebesar 0.4884, 1.0079, dan 0.9632. Sebaliknya hasil statistik uji Geweke untuk contoh yang berasal dari sebaran prior gamma semuanya juga menunjukkan hasil pengujian yang non signifikan, yang mengindikasikan bahwa rata-rata penduga parameter yang diamati semuanya sudah mencapai kekonvergenan. Hal ini terlihat dari nilai statistik Geweke untuk masing-masing parameter yang diperoleh sebesar 0.6637, 1.2032, dan -0.8148, dengan nilai p-value masing-masing adalah 0.5069, 0.2289, dan 0.4152.

### Diagnostik Kecocokan Model dan Analisis Sisaan

Tabel 4.2 menyajikan hasil-hasil mengenai berbagai kriteria informasi devians yang biasa dilakukan untuk mengevaluasi kecocokan model dalam metode MCMC. Beberapa kriteria yang dihitung di sini adalah rata-rata devians posterior ( $Dbar$ ), nilai devians yang dievaluasi pada rata-rata posterior ( $Dmean$ ), banyaknya parameter yang efektif ( $pD$ ), serta devians information criteria ( $DIC$ ) itu sendiri.

Perlu diketahui bahwa statistik  $DIC$  merupakan ukuran kecocokan model yang biasa digunakan dalam analisis Bayes, sebagai alternatif dari statistik  $AIC$  dan  $BIC$ . Dalam hal ini  $DIC$  menggunakan fungsi kepekatan posterior, yang berarti bahwa ukuran ini memperhitungkan informasi yang berasal dari sebaran priornya. Lebih jauh, suatu  $DIC$  yang nilainya lebih kecil menunjukkan bahwa model tersebut model yang cocok terhadap data. Berdasarkan hasil perhitungan  $DIC$  yang ditunjukkan pada Tabel 4.2 terlihat bahwa besaran  $DIC$  yang berasal dari sebaran prior gamma ( $DIC_G = 61.900$ ) lebih kecil daripada  $DIC$  untuk sebaran prior invers gamma ( $DIC_{IG} = 63.923$ ). Hal ini menunjukkan bahwa model Poisson Bayes berhirarki dengan sebaran prior gamma mempunyai derajat kecocokan yang lebih tinggi dibandingkan dengan model yang berasal dari sebaran prior invers gamma. Selisih antara kedua nilai  $DIC$  tersebut adalah sebesar 2.023, dan apabila nilai selisih  $DIC$  ini dibandingkan dengan nilai chi-kuadrat dengan derajat bebas 1, akan diperoleh nilai p-value sebesar 0.1549. Hal ini menunjukkan bahwa perbedaan atau selisih antara kedua nilai  $DIC$  yang diperoleh dari model prior gamma dengan prior invers gamma secara statistik adalah sama.

Tabel 4.2 Kriteria informasi devians untuk percobaan simulasi

Kriteria	Gamma	Invers Gamma
Rata-rata devians posterior ( $Dbar$ )	53.440	55.929
Nilai devians yang dievaluasi pada rata-rata posterior ( $Dmean$ )	44.980	47.935
Banyaknya parameter yang efektif ( $pD$ )	6.460	7.994
<i>Deviance Informations Criterion</i> ( $DIC$ )	61.900	63.923

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hasil yang ditunjukkan oleh DIC di atas juga sejalan dengan apa yang diberikan oleh rata-rata devians posterior (Dbar) dan devians yang dievaluasi pada rata-rata posterior (Dmean). Perhatikan bahwa nilai Dbar dan Dmean yang berasal dari sebaran prior gamma masing-masing adalah 53.440 dan 44.980. Sedangkan nilai Dbar dan Dmean yang berasal dari sebaran prior invers gamma masing-masing adalah 55.929 dan 47.935. Jadi nilai Dbar dan Dmean yang berasal dari sebaran prior gamma lebih kecil daripada nilai Dbar dan Dmean yang berasal dari sebaran prior invers gamma. Hasil ini masuk akal karena hasil perhitungan DIC didasarkan pada kedua statistik tersebut.

Sementara itu, masih dalam Tabel 4.2, besaran pD atau banyaknya parameter yang efektif di dalam model untuk sebaran prior gamma adalah sebesar  $pD = 6.460$ , dan ini relatif lebih besar dibandingkan dengan nilai pD untuk sebaran prior invers gamma yang sebesar  $pD = 7.994$ . Perlu diketahui bahwa besaran pD adalah mengukur selisih antara rata-rata devians posterior (Dbar) dan devians yang dievaluasi pada rata-rata posterior (Dmean). Ukuran pD ini menggambarkan kompleksitas dari suatu model, dimana model dengan parameter lebih banyak (atau nilai pD yang lebih besar) menunjukkan ketepatan dari propensitas devians. Dengan demikian berdasarkan hasil yang ditunjukkan pada Tabel 4.2, maka dapat dikatakan bahwa model Poisson Bayes berhirarki yang lebih cocok terhadap data adalah model dengan sebaran prior gamma.

Selanjutnya, hasil dari analisis sisaan dari model Poisson Bayes berhirarki dua-level untuk kedua sebaran prior yang digunakan dalam studi simulasi ini disajikan pada Tabel 4.3. Berdasarkan hasil dari analisis sisaan ini tampak bahwa model Poisson Bayes berhirarki dengan sebaran prior gamma memberikan nilai-nilai sisaan (sisaan Bayes, MSPE dan MAPE) sedikit lebih kecil dibandingkan dengan model dengan sebaran prior invers gamma. Berdasarkan hasil dari analisis sisaan ini dapat dikatakan bahwa model Poisson Bayes berhirarki dengan prior gamma lebih cocok terhadap data dibandingkan model dengan prior invers gamma.

Tabel 4.3 Hasil analisis sisaan

Ukuran	Gamma	Invers Gamma
Sisaan Bayes	48.9440	53.9848
MSPE	56.2788	58.8265
MAPE	5.3648	5.4140

Perlu diketahui bahwa nilai sisaan yang diberikan pada Tabel 4.3, pada dasarnya sisaan dalam analisis Bayes tidak jauh berbeda dengan sisaan dalam model linear statistika pada umumnya, yaitu selisih antara data aktual dengan data dugaan dari model. Di sini yang menjadi data aktual adalah peubah respons yang dibangkitkan dalam proses simulasi yang diasumsikan mengikuti sebaran Poisson,  $y_i$ , sedangkan data dugaan adalah peubah respons prediksi yang dihasilkan dari model Poisson Bayes berhirarki,  $y_i^{pr}$ . Walaupun nilai sisaan untuk model Poisson Bayes berhirarki untuk sebaran prior gamma secara umum lebih kecil daripada model dengan prior invers gamma, akan tetapi sebaran rata-rata dan simpangan baku dari peubah respons prediksi untuk kedua model tersebut mempunyai pola yang tidak jauh berbeda.

### Sifat-sifat Penduga Parameter Model

Tabel 4.4 menyajikan ringkasan statistik berupa rata-rata, galat baku, serta besaran bias untuk sebaran posteriornya berdasarkan pada dua penggunaan sebaran prior yang berbeda. Rata-rata dan simpangan baku (yang dinyatakan dalam bentuk ragam) untuk sebaran posterior ini masing-masing dihitung dengan menggunakan rumus:

$$E(\beta|\mathbf{y}) = \frac{1}{K} \sum_{k=1}^K \beta_k$$

dan

$$V(\beta|\mathbf{y}) = \frac{1}{n-1} \sum_{k=1}^K (\beta_k - E(\beta|\mathbf{y}))^2$$

Dalam hal ini  $K$  adalah banyaknya contoh yang dibangkitkan dalam proses komputasi melalui metode MCMC. Sementara itu besaran bias dari penduga parameter  $\beta$  didefinisikan sebagai berikut:

$$\text{Bias}(\hat{\beta}|\mathbf{y}) = E(\beta|\mathbf{y}) - \beta$$

Berdasarkan hasil dari simulasi ini menunjukkan bahwa ringkasan statistik sebaran posterior (yang ditunjukkan melalui rata-rata dan simpangan bakunya) yang berasal dari sebaran prior gamma maupun invers gamma memberikan hasil yang tidak jauh berbeda dengan nilai asli dari parameternya. Misalnya besaran rata-rata penduga untuk parameter  $\beta_0$ ,  $\beta_1$ , dan  $\beta_2$  yang dihasilkan dari sebaran prior gamma masing-masing adalah 0.5501, 1.5775, dan 1.9490; sedangkan yang dihasilkan dari sebaran prior gamma masing-masing adalah 0.4526, 1.4036, dan 2.0615. Perlu diketahui bahwa nilai parameter sebenarnya dalam studi simulasi ini untuk setiap parameter tersebut masing-masing adalah 0.5, 1.5, dan 2.0.

Tabel 4.4 Ringkasan statistik untuk sebaran posterior berdasarkan sebaran prior gamma dan invers gamma

Parameter	Prior Gamma			Prior Invers Gamma		
	Rata-rata	Galat Baku	Rata-rata Bias	Rata-rata	Galat Baku	Rata-rata Bias
$\beta_0$	0.5501	0.2696	0.0523	0.4526	0.3355	0.0499
$\beta_1$	1.5775	0.2138	0.0767	1.4036	0.6892	0.0861
$\beta_2$	1.9490	0.1093	0.0504	2.0615	0.2051	0.0654

Sementara itu, simpangan baku untuk setiap parameter yang diamati pada sebaran prior gamma semuanya lebih kecil dibandingkan dengan sebaran prior invers gamma. Sedangkan besarnya bias untuk parameter  $\beta_1$ , dan  $\beta_2$  yang dihasilkan dari sebaran prior gamma relatif lebih kecil dibandingkan dengan besarnya bias yang diperoleh dari model dengan prior invers gamma. Sebaliknya, besarnya bias untuk parameter  $\beta_0$  untuk model yang berasal dari prior gamma relatif lebih kecil dibandingkan dengan model yang berasal dari model dengan prior invers gamma.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber;

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Tabel 4.5 Galat baku Monte Carlo dan Simpangan baku posterior

Parameter	Sebaran Prior Gamma			Sebaran Prior InversGamma		
	GBMC	SBP	GBMC/SBP	GBMC	SBP	GBMC/SBP
$\beta_0$	0.0421	0.2696	0.1562	0.0189	0.3355	0.0563
$\beta_1$	0.0351	0.2138	0.1640	0.0213	0.6892	0.0309
$\beta_2$	0.0159	0.1093	0.1454	0.0066	0.2051	0.0324

Walaupun demikian apabila diperhatikan hasil-hasil dari galat baku Monte Carlo yang disajikan pada Tabel 4.5 menunjukkan bahwa rasio antara galat baku Monte Carlo dan simpangan bakunya untuk rantai Markov yang diperoleh dari sebaran prior invers gamma untuk setiap parameter yang diamati jauh lebih kecil dibandingkan dengan apa yang dihasilkan dari sebaran prior gamma. Hal ini menunjukkan bahwa model Poisson Bayes berhirarki dengan prior invers gamma mempunyai derajat presisi yang lebih tinggi dibandingkan model dengan prior gamma.

Selanjutnya, hasil-hasil dari nilai prediksi untuk sebaran posterior, dimana ukuran statistik yang diamatinya adalah rata-rata dan simpangan baku prediksi, serta selang kepercayaannya disajikan pada Tabel 4.6 dan Tabel 4.7. Pada kedua tabel tersebut berisi rata-rata, simpangan baku, dan interval HPD dari peubah acak respons yang diamati,  $Y_i$ , bersamaan dengan nilai penduga dari parameter  $\theta_i$ . Tabel 4.6 berisi ringkasan statistik posterior yang dihasilkan dari model dengan sebaran prior gamma, sedangkan pada Tabel 4.7 menampilkan ringkasan statistik posterior yang dihasilkan dari model dengan sebaran prior invers gamma.

Tabel 4.6 Ringkasan statistik posterior dari sebaran prior gamma

Area	Rata-rata	Simpangan Baku	Selang HPD		$\hat{\theta}_i^{(HB)}$	$\hat{\theta}_i^{(PL)}$
			BB	BA		
1	1.2995	1.1401	0	3	1.34	2.06
2	49.0002	7.0030	35	62	3.97	4.86
3	39.4267	6.3347	27	51	3.04	3.24
4	6.1579	2.4574	2	11	4.27	6.24
5	154.0000	12.5235	128	176	7.98	7.56
6	4.5168	2.1461	0	8	2.04	3.16
7	0.9897	0.9947	0	3	0.43	0.44
8	27.4210	5.2489	17	37	2.60	2.46
9	18.1505	4.2748	10	26	2.52	2.78
10	34.6166	6.7522	20	46	2.69	3.11

Hasil-hasil ringkasan statistik posterior yang diberikan pada Tabel 4.6 dan Tabel 4.7 menunjukkan bahwa penduga Bayes berhirarki baik yang dihasilkan dari prior gamma maupun prior invers gamma mempunyai sifat-sifat yang tidak jauh berbeda. Rata-rata dan simpangan baku untuk kedua sebaran prior yang digunakan dalam studi simulasi ini memberikan hasil yang tidak jauh berbeda. Demikian juga halnya dengan besar selang kepercayaan HPD untuk penduga prediksi,  $\hat{y}_i^{pr}$ , keduanya memberikan



cakupan selang yang juga tidak jauh berbeda. Sementara itu, penduga parameter yang dihasilkan dari model Bayes berhirarki,  $\hat{\theta}_i^{(HB)}$ , yang berasal dari kedua sebaran prior, memberikan hasil prediksi yang tidak jauh berbeda.

Tabel 4.7 Ringkasan statistik posterior dari sebaran prior invers gamma

Area	Rata-rata	Simpangan Baku	Selang HPD		$\hat{\theta}_i^{(HB)}$	$\hat{\theta}_i^{(PL)}$
			BB	BA		
1	1.0868	1.0342	0	3	1.12	2.06
2	58.4753	10.4192	40	79	4.73	4.86
3	38.6249	8.6893	22	55	2.98	3.24
4	8.9500	4.3984	1	17	6.21	6.24
5	145.4000	16.8901	111	177	7.53	7.56
6	6.0762	3.5344	1	13	2.75	3.16
7	1.0368	1.0184	0	3	0.45	0.44
8	26.7510	7.2959	12	40	2.53	2.46
9	18.8200	6.0594	7	30	2.61	2.78
10	41.3866	9.0294	24	58	3.22	3.11

Sebagai bahan perbandingan, pada kolom terakhir dari masing-masing Tabel 4.6 dan Tabel 4.7 diberikan penduga langsung (PL) untuk gugus data simulasi yang digunakan dalam penelitian ini. Hasilnya menunjukkan bahwa kedua penduga parameter yang diperoleh dari model Bayes berhirarki dan penduga langsung juga memberikan hasil yang tidak jauh berbeda. Namun perlu diingat bahwa ragam penduga yang diperoleh metode pendugaan langsung ini lebih besar dibandingkan dengan ragam dari model Bayes berhirarki.

### Pembahasan

Dalam studi simulasi ini telah ditunjukkan hasil-hasil dari model Poisson Bayes berhirarki dua-level dengan menggunakan dua buah sebaran prior yang berbeda, yaitu prior gamma dan prior invers gamma. Secara umum performa model dugaan yang ditunjukkan oleh kedua sebaran prior tersebut adalah hampir sama. Perlu diketahui bahwa selain pemilihan sebaran prior, keterbatasan pendekatan Bayes berhirarki dengan menggunakan metode MCMC adalah penentuan panjang atau periode *burn-in* dan apakah proses simulasi MCMC ini dijalankan secara tunggal atau ganda. Kedua masalah ini dapat dilihat melalui diagnostik kekonvergenan rantai Markovnya. Berdasarkan hasil diagnostik kekonvergenan rantai Markov pada proses tunggal diketahui bahwa model Poisson Bayes berhirarki dengan sebaran prior gamma cenderung memiliki derajat kekonvergenan lebih tinggi dibandingkan dengan model yang didasarkan pada prior invers gamma. Hasil ini didasarkan pada penggunaan panjang atau periode *burn-in* yang cukup besar, yaitu  $B = 1000$ , sebagaimana yang disarankan dalam banyak literatur tentang metode MCMC ini.

Berdasarkan hasil diagnostik kekonvergenan untuk model Poisson Bayes berhirarki dengan menggunakan sebaran prior invers gamma belum sepenuhnya dikatakan konvergen, terutama yang berkaitan dengan parameter  $\beta_1$  (dilihat dari hasil

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak cipta milik IPB (Institut Perikanan Bogor) | Bogor Agricultural University

statistik uji Geweke). Apabila belum mencapai kekonvergenan, biasanya panjang atau periode *burn-in* perlu diperpanjang. Akan tetapi hal ini masih menyisakan porsi yang signifikan dari ruang yang dibangkitkan oleh sebaran posterior yang secara keseluruhan belum tereksplorasi.

Selanjutnya, berdasarkan hasil diagnostik mengenai kecocokan model dan analisis sisaan menunjukkan bahwa model Poisson Bayes berhirarki dengan sebaran prior gamma cenderung memberikan derajat kecocokan yang lebih baik dibandingkan dengan model dengan sebaran prior invers gamma. Untuk mengevaluasi kecocokan model, khususnya dalam model linear terampat, ukuran yang biasa digunakan adalah devians. Namun statistik devians ini mempunyai kelemahan, yaitu devians tidak mempertimbangkan derajat parameterisasi dalam model, artinya suatu model dapat dikatakan cocok terhadap data sebagaimana banyaknya parameter dalam model meningkat.

Ukuran atau statistik lain yang dapat digunakan untuk mengevaluasi kecocokan model adalah statistik AIC dan BIC. Namun penggunaan AIC dan BIC dalam pendekatan Bayes berhirarki juga mempunyai kelemahan, yaitu bahwa model dengan efek acak (*random effect*) akan sangat sulit untuk memutuskan berapa banyak parameter yang bisa disertakan ke dalam model. Padahal dalam pendugaan area kecil yang dibahas dalam penelitian ini mengandung efek acak dalam modelnya. Oleh karena itu penggunaan DIC dalam pendekatan model Bayes berhirarki menjadi lebih tepat. Dalam hal ini DIC menggunakan fungsi kepekatan posterior, yang berarti bahwa ukuran ini memperhitungkan informasi yang berasal dari sebaran priornya. Lebih jauh, suatu DIC yang nilainya lebih kecil menunjukkan bahwa model tersebut mempunyai derajat atau tingkatan kecocokan yang lebih baik terhadap data.

Pengertian kecocokan model terhadap dalam konteks analisis Bayes adalah suatu model yang sesuai dengan sebaran sasarannya berdasarkan informasi yang berasal dari sebaran prior yang digunakan. Dengan demikian dalam hal ini pemilihan sebaran prior memegang peranan penting dalam memperoleh suatu dugaan dari peubah acak yang diamati. Dengan menggunakan representasi dalam model Poisson Bayes berhirarki dua-level secara langsung akan membawa pada sebaran posterior dengan sejumlah parameter yang diamati untuk vektor pengamatan  $\mathbf{y}$  tertentu. Penarikan contoh berulang dari sebaran posterior ini menghasilkan  $M$  buah nilai bagi segugus parameter yang akan diduga dan  $M$  buah nilai dari  $\theta_i$ , untuk  $i = 1, \dots, m$ . Nilai dari  $\theta_i$  ini selanjutnya digunakan untuk menduga beberapa ringkasan statistik, seperti rata-rata dan simpangan baku, sebaran posterior dari  $\theta_i$ . Konsep ini nantinya digunakan dalam keperluan diagnostik kecocokan model dan analisis sisaan melalui beberapa ukuran, seperti DIC, sisaan Bayes, MSPE, dan MAPE.

Berdasarkan hasil dari diagnostik kecocokan model dan analisis sisaan ini tampak bahwa model Poisson Bayes berhirarki dengan sebaran prior gamma memberikan berbagai ukuran statistik (DIC, sisaan Bayes, MSPE, dan MAPE) sedikit lebih kecil dibandingkan dengan model dengan sebaran prior invers gamma. Berdasarkan hasil dari diagnostik kecocokan model dan analisis sisaan ini dapat dikatakan bahwa model Poisson Bayes berhirarki dengan prior invers gamma mempunyai derajat kecocokan yang tidak jauh berbeda dengan model yang berasal dari sebaran prior invers gamma. Pernyataan ini terutama diperoleh dari selisih nilai DIC diantara kedua model tersebut adalah kurang dari 5, sebagaimana yang direkomendasikan oleh Spiegelhalter, et al (2002). Walaupun demikian kesimpulan ini masih bersifat relatif karena dari berbagai ukuran statistik yang dikaji menunjukkan performa yang tidak jauh berbeda.

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Berdasar hasil studi simulasi juga telah ditunjukkan bahwa ringkasan statistik sebaran posterior (yang ditunjukkan melalui rata-rata, galat baku, serta besarnya bias) yang berasal dari sebaran prior gamma maupun invers gamma memberikan hasil yang tidak jauh berbeda dengan nilai asli dari parameternya yang diperoleh dari proses pembangkitan data. Sementara itu, galat baku untuk setiap parameter yang diamati pada sebaran prior gamma semuanya lebih kecil dibandingkan dengan sebaran prior invers gamma. Hal ini masuk akal karena walaupun secara rata-rata nilai penduga parameter yang diperoleh dari prior invers gamma tidak jauh berbeda dengan nilai aslinya, akan tetapi dari hasil *trace plot* menunjukkan bahwa nilai penduganya tersebut tidak terpusat pada nilai yang ditetapkan. Bahkan cenderung berfluktuasi cukup tinggi, dan hal ini ditunjukkan oleh nilai simpangan baku yang lebih besar dibandingkan dengan hasil yang diperoleh dari prior gamma. Sementara itu, besarnya bias untuk parameter  $\beta_1$ , dan  $\beta_2$  yang dihasilkan dari sebaran prior gamma relatif lebih kecil dibandingkan dengan besarnya bias yang diperoleh dari model dengan prior invers gamma. Sebaliknya, besarnya bias untuk parameter  $\beta_0$  untuk model yang berasal dari prior gamma relatif lebih kecil dibandingkan dengan model yang berasal dari model dengan prior invers gamma.

Apabila diperhatikan hasil-hasil dari galat baku Monte Carlo menunjukkan bahwa rasio antara galat baku Monte Carlo dan simpangan bakunya untuk rantai Markov yang diperoleh dari sebaran prior invers gamma untuk setiap parameter yang diamati jauh lebih kecil dibandingkan dengan apa yang dihasilkan dari sebaran prior gamma. Perlu diketahui bahwa ragam dari fungsi kepekatan posterior (kadang-kadang cukup dirujuk juga sebagai ragam posterior) menggambarkan ketidakpastian dalam parameter, yang merupakan suatu peubah acak dalam paradigma Bayes. Analisis Bayes biasanya menggunakan ragam atau simpangan baku posterior untuk melihat karakteristik tentang dispersi parameter. Seluruh prosedur Bayes bergantung pada metode MCMC untuk memperoleh seluruh penduga posteriornya. Dengan hanya menggunakan banyaknya contoh yang terbatas, maka simulasi dalam analisis Bayes mengukur derajat ketidakpastian untuk melihat akurasi dari suatu penduga. Dalam hal ini menggunakan galat baku Monte Carlo (Monte Carlo *standard error*, MCSE), yang merupakan galat baku dari rata-rata penduga posterior, yang mengukur akurasi dari hasil simulasi.

Perlu ditambahkan pula bahwa simpangan baku posterior dan MCSE merupakan dua konsep yang sangat berbeda. Simpangan baku posterior menggambarkan ketidakpastian dalam parameter, sedangkan MCSE hanya menggambarkan ketidakpastian dalam penduga parameter sebagai hasil dari simulasi MCMC. Simpangan baku posterior merupakan fungsi dari ukuran contoh dalam suatu gugus data, sedangkan MCSE merupakan fungsi dari banyaknya iterasi dalam proses simulasi MCMC. Dengan demikian, maka dapat dikatakan bahwa penduga parameter yang diperoleh melalui prior gamma mempunyai derajat ketakbiasan dan akurasi yang relatif lebih tinggi dibandingkan dengan penduga parameter yang diperoleh dari prior invers gamma.

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## 5 APLIKASI MODEL POISSON BAYES BERHIRARKI UNTUK MENDUGA ANGKA KEMATIAN BAYI LEVEL KECAMATAN

### Pendahuluan

Bab ini berisi bahasan tentang aplikasi model Poisson Bayes berhirarki dua-level untuk menduga angka kematian bayi (AKB) level kecamatan pada salah Kabupaten Bojonegoro Provinsi Jawa Timur. Berdasarkan hasil studi simulasi yang dibahas pada bab sebelumnya diketahui bahwa model Poisson Bayes berhirarki dua-level dengan sebaran prior gamma memberikan performa yang relatif lebih baik dibandingkan dengan model dengan sebaran prior invers gamma. Hasil tersebut tentu saja masuk akal karena sebaran gamma merupakan sebaran prior yang bersifat *conjugate* bagi sebaran Poisson. Sebaliknya, sebaran invers gamma merupakan sebaran prior yang bersifat *non-conjugate* bagi sebaran Poisson. Merujuk pada teori yang dikembangkan oleh Gelman (2006) bahwa sebaran prior invers gamma tersebut diklasifikasikan sebagai sebaran prior yang bersifat *conditionally conjugate*.

Dalam pemodelan Bayes berhirarki, lebih jauh Gelman (2006) menyatakan bahwa penggunaan sebaran prior yang bersifat *conditionally conjugate* ini tidak akan memberikan pengaruh yang begitu besar pada sifat-sifat dari sebaran posteriornya. Hal ini berlaku apabila sebaran prior yang bersifat *conditionally conjugate* ini tidak digunakan pada level pertama atau kedua dari model Bayes berhirarki yang dirumuskan. Pernyataan tersebut telah ditunjukkan dalam studi simulasi, dimana walaupun model dengan prior gamma mempunyai sifat-sifat statistik yang relatif lebih baik daripada model dengan prior invers gamma, akan tetapi secara umum perbedaan diantara kedua penggunaan sebaran prior tersebut tidak terlalu nyata, terutama yang berkaitan dengan sifat-sifat dari sebaran posteriornya. Oleh karena itu, dalam aplikasi untuk menduga angka kematian bayi level kecamatan di salah satu kabupaten/kota di Provinsi Jawa Timur ini akan menggunakan kedua sebaran prior tersebut.

Selain itu, perlu diketahui bahwa peubah acak respons yang diperhatikan dalam studi ini adalah diasumsikan bersebaran Poisson. Model yang biasa digunakan untuk menganalisis data cacahan yang bersebaran Poisson adalah model linear terampat. Namun perlu diperhatikan bahwa model linear terampat ini tidak mempertimbangkan efek acak dari area yang diamati. Sedangkan model yang dikembangkan dalam penelitian ini adalah perpaduan antara model linear terampat dan model Bayes berhirarki untuk menangani masalah pendugaan area kecil, yang tentu saja didalamnya sudah memperhitungkan efek acak dari area yang diamati. Oleh karena itu, sebagai bahan perbandingan model yang tanpa memperhitungkan efek acak (model regresi Poisson biasa) juga akan digunakan dalam aplikasi ini.

Pada bab ini akan dibahas terlebih dahulu konsep tentang angka kematian bayi serta faktor-faktor yang berpengaruh pada kematian bayi. Pembahasan ini nantinya dapat dijadikan landasan dalam menentukan peubah tambahan (*auxilliary variable*) untuk disertakan dalam model Poisson Bayes berhirarki dua-level. Adapun sumber data yang digunakan dalam penelitian ini ada dua, yaitu Survei Sosial dan Ekonomi Nasional Tahun 2010 (SUSENAS2010) dan Survei Demografi dan Kesehatan Indonesia Tahun 2007 (SDKI2007).

SUSENAS2010 mengumpulkan data yang menyangkut bidang pendidikan, kesehatan/gizi, perumahan, sosial ekonomi lainnya, kegiatan sosial budaya,

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

konsumsi/pengeluaran dan pendapatan rumah tangga, dan perjalanan. Sejak tahun 1992, BPS melalui Susenas mengumpulkan data kor (data dasar) dan data modul (data sasaran) setiap tahun. Data modul dikumpulkan secara bergiliran setiap 3 tahun sekali. Modul Susenas dibagi atas 3 kelompok besar, yaitu modul sosial budaya dan pendidikan, perumahan dan kesehatan serta modul konsumsi dan pengeluaran rumah tangga.

Sumber data lain yang dapat digunakan dalam perhitungan Angka Kematian Bayi adalah Survei Demografi dan Kesehatan Indonesia (SDKI), yang merupakan salah satu survei sosial kependudukan yang secara berkala diselenggarakan BPS. Hingga saat ini, BPS telah menyelenggarakan kegiatan SDKI sebanyak lima kali yaitu pada tahun 1991, 1994, 1997, 2002, dan 2007. SDKI khusus dirancang untuk mengumpulkan berbagai informasi mengenai tingkat kelahiran, kematian, prevalensi keluarga berencana dan kesehatan khususnya kesehatan reproduksi. Sebagai bagian dari program internasional, survei serupa yang dikenal dengan nama Demographic and Health Surveys (DHS) juga dilaksanakan di negara-negara Amerika Latin, Asia, Afrika dan Timur Tengah. Sejalan dengan itu, pertanyaan yang dicakup dalam SDKI merujuk pada DHS yang telah berskala internasional.

### Angka Kematian Bayi

Tujuan Pembangunan Kesehatan adalah untuk meningkatkan kesadaran, kemauan dan kemampuan untuk hidup sehat bagi setiap orang agar terwujud derajat kesehatan masyarakat yang optimal sesuai dengan Undang-Undang No. 23 tahun 1992. Untuk mengetahui gambaran derajat kesehatan masyarakat dapat diukur dari indikator-indikator yang digunakan antara lain angka kematian, angka kesakitan serta status gizi. Indikator tersebut dapat diperoleh melalui laporan dari fasilitas kesehatan (*facility based*) dan dari masyarakat (*community based*). Perkembangan derajat kesehatan masyarakat dapat dilihat dari kejadian kematian dan kesakitan dalam masyarakat dari waktu ke waktu. Disamping itu kejadian kematian juga dapat digunakan sebagai indikator dalam penilaian keberhasilan pelayanan kesehatan dan program pembangunan kesehatan lainnya. Angka kematian pada umumnya dapat dihitung dengan melakukan berbagai survei dan penelitian.

Data kematian yang terdapat pada suatu komunitas dapat diperoleh melalui survei, karena sebagian besar kematian terjadi di rumah, sedangkan data kematian pada fasilitas pelayanan kesehatan hanya memperlihatkan kasus rujukan. Angka Kematian Bayi (AKB) di Indonesia berasal dari berbagai sumber yaitu Sensus Penduduk, Surkesnas/Susenas dan Survei Demografi dan Kesehatan Indonesia (SDKI). Ada banyak faktor yang mempengaruhi tingkat AKB tetapi tidak mudah untuk menemukan faktor yang paling dominan. Tersedianya berbagai fasilitas atau faktor aksesibilitas dan pelayanan kesehatan dari tenaga medis yang terampil, serta kesediaan masyarakat untuk merubah kehidupan tradisional ke norma kehidupan modern dalam bidang kesehatan merupakan faktor yang sangat berpengaruh terhadap tingkat AKB.

### Konsep Angka Kematian Bayi

Kematian bayi itu sendiri adalah kematian yang terjadi antara saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun. Banyak faktor yang dikaitkan dengan kematian bayi. Secara garis besar, dari sisi penyebabnya, kematian bayi ada dua macam

yaitu endogen dan eksogen. Kematian bayi endogen atau yang umum disebut dengan kematian neonatal; adalah kematian bayi yang terjadi pada bulan pertama setelah dilahirkan, dan umumnya disebabkan oleh faktor-faktor yang dibawa anak sejak lahir, yang diperoleh dari orang tuanya pada saat konsepsi atau didapat selama kehamilan. Kematian bayi eksogen atau kematian post neo-natal, adalah kematian bayi yang terjadi setelah usia satu bulan sampai menjelang usia satu tahun yang disebabkan oleh faktor-faktor yang bertalian dengan pengaruh lingkungan luar.

Angka Kematian Bayi dapat menggambarkan keadaan sosial ekonomi masyarakat dimana angka kematian itu dihitung. Kegunaan Angka Kematian Bayi untuk pengembangan perencanaan berbeda antara kematian neo-natal dan kematian bayi yang lain. Karena kematian neo-natal disebabkan oleh faktor endogen yang berhubungan dengan kehamilan maka program-program untuk mengurangi angka kematian neo-natal adalah yang bersangkutan dengan program pelayanan kesehatan Ibu hamil, misalnya program pemberian pil besi dan suntikan anti tetanus. Sedangkan Angka Kematian Post-NeoNatal dan Angka Kematian Anak serta Kematian Balita dapat berguna untuk mengembangkan program imunisasi, serta program-program pencegahan penyakit menular terutama pada anak-anak, program penerangan tentang gizi dan pemberian makanan sehat untuk anak dibawah usia 5 tahun.

Angka Kematian Bayi (AKB) adalah banyaknya kematian bayi berusia dibawah satu tahun, per 1000 kelahiran hidup pada satu tahun tertentu. Selanjutnya, AKB dihitung dengan menggunakan rumus:

$$AKB = \frac{Y_{0-<1th}}{\sum \text{lahir hidup}} \times 1000$$

dimana:

- $AKB$  = Angka Kematian Bayi atau *Infant Mortality Rate* (IMR)
- $Y_{0,1th}$  = Jumlah Kematian Bayi (berumur kurang 1 tahun) pada satu tahun tertentu di daerah tertentu.
- $\sum$  lahir hidup = jumlah kelahiran hidup pada satu tahun tertentu di daerah tertentu.

Perlu diketahui bahwa konsep fertilitas hanya menghitung jumlah bayi yang lahir hidup. Organisasi Kesehatan Dunia (World Health Organization/WHO) mendefinisikan kelahiran hidup sebagai peristiwa kelahiran bayi, tanpa memperhitungkan lamanya berada dalam kandungan, dimana si bayi menunjukkan tanda-tanda kehidupan pada saat dilahirkan; misalnya bernafas, ada denyut jantung, atau denyut tali pusat, atau gerakan-gerakan otot. Dengan demikian, peristiwa bayi yang lahir dalam keadaan tidak hidup/meninggal (*still birth*) tidak dimasukkan dalam perhitungan jumlah kelahiran. Untuk bayi yang lahir hidup tetapi kemudian meninggal, beberapa saat setelah lahir atau dikemudian hari, kelahiran hidup ini tetap dimasukkan dalam perhitungan jumlah kelahiran. Tidak termasuk sebagai kelahiran hidup adalah peristiwa keguguran atau bayi yang lahir dalam keadaan meninggal (lahir mati).

### Faktor-faktor yang Mempengaruhi Kematian Bayi

Tingginya tingkat kematian balita sangat tergantung dari perhatian dan perawatan yang diberikan oleh kelompok penduduk dewasa dalam hal ini orang tua. Ini dikarenakan kematian balita lebih banyak disebabkan oleh penyakit-penyakit infeksi akibat pencemaran lingkungan karena perilaku orang dewasa termasuk didalamnya adalah para orang tua. Untuk itu perlu dilakukan suatu penelaahan terhadap beberapa faktor yang diperkirakan erat hubungannya dengan kematian balita berdasarka beberapa

penelitian yang pernah dilakukan yaitu faktor sosio-demografi seperti tingkat pendidikan ibu, usia ibu saat melahirkan, jarak antar kelahiran, urutan kelahiran, serta faktor perilaku pra dan pasca persalinan yang berhubungan dengan kondisi biologis dari balita tersebut sebelum dan sesudah kelahirannya.

Status sosial ekonomi menggambarkan tingkat penghidupan seseorang atau keluarga yang ditentukan oleh unsur pendidikan, pekerjaan dan penghasilan. Status ekonomi juga berkaitan dengan konsumsi (pengeluaran) dan produksi (pendapatan). Indikator status ekonomi bisa diukur melalui berbagai cara antara lain dengan menghitung tingkat pengeluaran perkapita. Status ekonomi mempengaruhi kebutuhan seseorang karena menentukan kemampuan keluarga untuk memperoleh makanan, karena pemenuhan kebutuhan hidupnya tergantung dari penghasilannya. Juga berpengaruh terhadap penyediaan bahan pangan, baik kuantitas maupun kualitas. Keluarga dengan status ekonomi rendah kemampuan untuk mempengaruhi konsumsi makanan keluarga yang berkaitan erat dengan status gizi keluarga.

Angka kematian bayi (AKB) merupakan salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat. Keberhasilan pembangunan kesehatan berperan penting dalam meningkatkan mutu kualitas sumber daya manusia suatu negara. Pembangunan bidang kesehatan diarahkan untuk mencapai komitmen internasional yang dituangkan dalam Millennium Development Goals (MDGs). Salah satu tujuan MDGs adalah menurunkan angka kematian anak. Di Jawa Timur, Badan Pusat Statistik (BPS) mencatat angka kematian bayi (AKB) menurun dari 39.6 per 1000 kelahiran hidup pada tahun 2004 menjadi 31.41 per 1000 kelahiran hidup tahun 2008 (Depkes, 2008). Angka kematian ini turun menjadi 31.28 pada tahun 2010. Namun AKB ini masih tergolong tinggi karena masih jauh dari target MDGs tahun 2015 sebesar 23 per 1.000 sehingga bisa dikatakan bahwa upaya penurunan angka kematian bayi di Jawa Timur belum maksimal. Penurunan AKB mengindikasikan peningkatan derajat kesehatan masyarakat sebagai salah satu wujud keberhasilan pembangunan di bidang kesehatan.

Beberapa penelitian menunjukkan bahwa tingginya angka kematian bayi bukan hanya karena masalah medis, melainkan non medis yang sifatnya sangat fundamental (mendasar dan besar), seperti minimnya sarana dan prasarana kesehatan, jauhnya akses masyarakat menuju puskesmas, minimnya tenaga bidan, lemahnya ekonomi dan juga bisa karena *human error* dari faktor ibu sendiri (Bappenas, 2009). Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2007 mengumpulkan berbagai informasi, diantaranya tentang kematian bayi yang dikelompokkan berdasarkan karakteristik sosial ekonomi dan biodemografis (BPS, 2007). Peubah sosial ekonomi tersebut mencakup tempat tinggal, pendidikan, serta indeks kekayaan kuantil. Peubah biodemografis meliputi umur ibu, paritas dan jarak kelahiran. Beberapa peubah lain yang berpengaruh terhadap kematian bayi antara lain berat bayi saat lahir, pemeriksaan kehamilan dan penolong kelahiran, serta komplikasi saat persalinan, semuanya telah tercakup didalamnya.

Tingginya AKB tidak bisa dibiarkan begitu saja mengingat kelangsungan hidup anak menentukan kualitas sumber daya manusia dimasa yang akan datang. Oleh karena itu diperlukan intervensi untuk mengurangi angka kematian bayi. Intervensi yang efektif dapat dilakukan jika faktor-faktor signifikan yang mempengaruhi kematian bayi dapat diketahui. Penelitian mengenai faktor-faktor yang mempengaruhi angka kematian bayi di Jawa Timur telah banyak dilakukan oleh berbagai pihak. Intisari dari berbagai hasil

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

penelitian tersebut menyatakan bahwa faktor-faktor yang mempengaruhi AKB ini disajikan pada Tabel 5.1.

Tabel 5.1 Definisi peubah penelitian

Kode	Definisi
BBLR	Persentase berat badan bayi lahir rendah, yaitu berat badan bayi yang kurang 2500 gram.
K1	Persentase ibu hamil yang tidak melakukan kunjungan untuk mendapatkan pelayanan kesehatan oleh tenaga kesehatan profesional (dokter spesialis kandungan dan kebidanan, dokter umum, bidan dan perawat) selama masa kehamilannya.
KN1	Persentase ibu yang tidak melakukan kunjungan bayi. Kunjungan anak usia kurang dari satu tahun(29 hari-11 bulan) untuk mendapatkan pelayanan
PSLN	Persentase persalinan di suatu wilayah dalam kurun waktu tertentu yang ditolong bukan oleh tenaga kesehatan profesional, seperti: dokter spesialis kebidanan, dokter umum, bidan, pembantu bidan, pembantu bidan dan perawat bidan.
RIST	Persentase ibu hamil risti, dimana ibu hamil risti adalah ibu hamil dengan keadaan penyimpangan dari normal yang secara langsung menyebabkan kesakitan dan kematian bagi ibu maupun bayinya.
ASI	Persentase bayi yang tidak mendapat ASI eksklusif, dimana ASI eksklusif adalah pemberian ASI saja tanpa makanan dan minuman lain sampai bayi berusia 6 bulan.
POSY	Persentase rasio ketersediaan POSYANDU terhadap penduduk.
RSHT	Persentase rumah tidak sehat. Bangunan rumah tinggal yang memenuhi syarat kesehatannya itu memiliki jamban sehat, tempat pembuangan sampah, sarana air bersih, sarana pembuangan air limbah, ventilasi baik, kepadatan hunian rumah sesuai dan lantai rumah tidak dari tanah.
PEND	Pendidikan terakhir yang ditempuh oleh ibu (dihitung dalam tahun).
KMSK	Persentase penduduk miskin.
RTKS	Rasio ketersediaan tenaga kesehatan profesional dan tenaga kesehatan masyarakat terhadap penduduk.

### Evaluasi Data Dasar

Sebagaimana yang telah disebutkan sebelumnya data yang digunakan dalam penelitian ini adalah data yang bersumber dari SUSENAS Tahun 2010 dan SDKI Tahun 2007. Model Poisson Bayes berhirarki dua-level yang dikembangkan pada bab sebelumnya akan digunakan untuk menduga angka kematian level kecamatan di salah satu kabupaten/kota di Provinsi Jawa Timur. Adapun kabupaten/kota yang dijadikan objek penelitian dipilih secara acak, dan yang terpilih adalah Kabupaten Bojonegoro.



Tabel 5.2 Data kesehatan tingkat kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur

No.	Kecamatan	POP	LH	BM	BBLR	K1	KN2	PSLN	RISTI	ASI	POSY	RSHT	PEND	KMSK	RTKS
1	Balen	65550	946	14	2.01	5.45	5.60	5.45	5.05	78.16	67.57	17.76	7.30	46.08	21.65
2	Baureno	81540	1206	15	1.82	3.37	11.28	3.37	4.60	28.83	37.80	17.30	6.78	46.99	17.40
3	Bojonegoro	83685	1434	6	0.21	4.57	6.04	4.57	7.88	69.79	56.07	12.62	5.51	20.47	16.96
4	Bubulan	15523	159	2	0.63	31.74	31.28	31.74	0.00	42.17	90.00	53.00	7.66	28.42	91.41
5	Dander	67962	1022	4	1.17	10.16	11.95	10.16	1.43	72.94	67.10	52.79	6.86	42.18	20.88
6	Kalitidu	60659	1020	5	2.16	8.01	9.11	8.01	4.21	34.80	53.51	51.45	6.18	57.03	23.39
7	Kanor	59410	782	6	1.66	17.11	16.33	17.11	0.74	32.56	85.71	52.25	5.97	15.92	23.88
8	Kapas	48784	793	6	1.89	1.43	4.07	2.73	5.32	62.33	85.46	51.35	9.35	76.12	29.09
9	Kasiman	27532	434	0	0.69	2.05	0.23	2.05	2.28	38.57	33.33	57.97	8.23	43.32	51.54
10	Kedungadem	84388	1102	6	1.18	17.50	19.13	17.50	4.15	7.68	65.33	52.63	7.79	23.11	16.82
11	Kepohbaru	61899	883	2	3.62	15.05	16.76	15.05	3.45	31.66	59.54	55.80	6.20	55.41	22.92
12	Malo	33036	473	4	2.11	9.78	5.65	9.78	3.99	96.14	42.86	56.92	5.59	56.12	42.95
13	Margomulyo	22037	341	1	0.59	7.48	9.78	7.48	1.66	10.91	10.71	55.23	6.69	68.83	64.39
14	Ngambon	12020	146	1	0.00	26.32	25.00	26.32	6.84	22.79	70.83	58.82	7.72	54.28	118.05
15	Ngasem	76525	1062	9	1.13	14.96	12.21	14.96	0.84	66.11	56.07	57.37	7.54	17.77	18.54
16	Ngraho	44655	675	4	0.89	5.29	2.02	5.29	2.29	95.24	44.44	13.87	7.42	41.04	31.78
17	Padangan	40619	670	3	0.75	1.33	0.45	1.33	7.39	68.76	28.57	54.92	7.22	49.06	34.93
18	Purwosari	28600	519	3	0.19	1.10	6.24	22.52	4.64	56.25	96.83	52.28	6.54	70.44	49.62
19	Sugihwaras	44492	710	5	0.28	4.14	8.09	4.14	5.61	87.02	90.00	55.56	8.45	23.34	31.89
20	Sumberejo	72014	1069	3	1.12	8.69	17.74	13.48	1.15	59.02	30.96	54.83	6.81	26.80	19.70
21	Tambakrejo	51793	779	2	0.51	13.52	19.67	13.52	11.19	13.00	8.00	63.42	6.94	49.50	27.40
22	Temayang	34009	438	4	1.37	23.39	22.20	23.39	2.86	81.28	65.62	61.89	5.65	57.07	41.72
23	Trucuk	35038	555	5	0.54	2.64	9.77	2.64	8.08	63.57	72.97	57.57	7.24	45.67	40.50
24	Sukosewu	43283	596	3	1.85	7.43	9.70	13.47	4.49	90.94	70.69	57.37	6.12	73.41	32.78
25	Gondang	24518	305	5	2.95	33.42	24.08	33.42	4.15	59.72	66.67	60.53	6.55	59.79	57.88
26	Sekar	25078	320	16	0.94	41.00	36.75	41.00	0.95	95.12	82.86	63.20	8.87	57.19	56.58
27	Kedewan	12147	178	0	2.25	4.81	12.97	4.81	6.42	37.76	28.57	55.18	9.12	95.06	116.82

Sumber: Hasil olah dari SUSENAS2010 dan SDKI2007



Kabupaten Bojonegoro merupakan bagian dari wilayah Provinsi Jawa Timur dengan jarak  $\pm 110$  Km dari ibukota Propinsi Jawa Timur yang memiliki luas 2.307,06 km<sup>2</sup> dengan jumlah penduduk tahun 2008 sebesar 1.256.796 jiwa (Proyeksi Dinkes Provinsi Jawa Timur dan BPS Propinsi Jawa Timur). Kabupaten Bojonegoro dibagi 27 wilayah Kecamatan dan 430 Desa/Kelurahan. Adapun rincian mengenai data kecamatan dan sejumlah peubah yang akan dianalisis dalam penelitian disajikan pada Tabel 5.2. Data yang tersaji pada tabel tersebut akan dievaluasi terutama yang berkaitan dengan kecocokan sebaran Poisson untuk peubah respons dan untuk pemilihan peubah tambahan (*auxiliary variables*).

**Pengujian Kecocokan Sebaran Poisson**

Dalam penelitian ini yang menjadi peubah respons adalah peubah BM, yaitu jumlah kematian bayi (berumur kurang 1 tahun) pada satu tahun 2008 pada setiap kecamatan. Peubah BM ini diasumsikan mengikuti sebaran Poisson. Untuk menguji asumsi tersebut akan dilakukan dengan menggunakan statistik uji chi-kuadrat dengan hipotesis sebagai berikut:

- H<sub>0</sub>: data mengikuti sebaran Poisson
- H<sub>1</sub>: data tidak mengikuti sebaran Poisson

Ringkasan hasil perhitungan statistik uji chi-kuadrat untuk menguji asumsi di atas disajikan pada Tabel 5.3. Kolom pertama dan kedua dari tabel tersebut menyatakan kategori dan jumlah kematian bayi yang berumur kurang dari 1 tahun pada masing-masing kategori tersebut. Misalnya pada baris pertama pada kolom kategori adalah ' $\leq 3$ ' menyatakan bahwa jumlah bayi yang mati antara 0 sampai dengan 3 terjadi pada 11 kecamatan.

Tabel 5.3 Perhitungan statistik chi-kuadrat untuk uji kecocokan sebaran Poisson

Kategori	Nilai Pengamatan (O)	Peluang Poisson (P)	Nilai Harapan (E)	$(O_i - E_i)^2 / E_i$
$\leq 3$	11	0.2703	7.2971	1.8790
4 – 5	8	0.3522	9.5093	0.2395
$\geq 6$	8	0.3775	10.1936	0.4720
				2.5906

Terdapat dua kecamatan yang melaporkan tidak ada bayi yang berumur kurang dari satu tahun mengalami kematian, yaitu Kecamatan Kasiman dan Kedewan. Sementara itu, jumlah kematian bayi tertinggi di Kabupaten Bojonegoro pada tahun 2008 adalah di Kecamatan Sekar, yaitu terjadi sebanyak 16 kematian bayi dari 320 banyaknya bayi yang lahir hidup. Adapun rata-rata jumlah kematian bayi yang berumur kurang dari 1 tahun untuk setiap kecamatan pada tahun 2008 di Kabupaten Bojonegoro adalah sebesar 4.9630. Artinya, untuk masing-masing kecamatan di Kabupaten Bojonegoro pada tahun 2008 terjadi kematian bayi yang berumur kurang dari satu tahun hampir sebanyak 5 kematian bayi.

Peluang dari sebaran Poisson yang ditampilkan pada Kolom ketiga dari Tabel 5.3 dihitung dengan cara sebagai berikut:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$P(Y \leq 3) = P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\ = \frac{e^{-4.9630}(4.9630)^0}{0!} + \dots + \frac{e^{-4.9630}(4.9630)^3}{3!} = 0.2703$$

Dengan cara yang sama dapat diperoleh peluang dari sebaran Poisson untuk kategori yang kedua dan ketiga, yang masing-masing diperoleh sebesar 0.3522 dan 0.3775. Perlu diketahui bahwa banyaknya kecamatan yang diamati dalam penelitian ini adalah sebesar  $m = 27$ . Dengan demikian, nilai harapan yang disajikan pada kolom keempat pada Tabel 5.3 dihitung dengan cara mengalikan antara banyaknya kecamatan yang diamati dengan peluang sebaran Poisson untuk setiap kategori. Misalnya untuk kategori ' $\leq 3$ ' diperoleh nilai harapan sebesar  $0.2703 \times 27 = 7.2971$ .

Selanjutnya, nilai statistik chi-kuadrat yang diberikan pada kolom terakhir pada Tabel 5.3 dihitung dengan menggunakan rumus:

$$\chi^2 = \sum_{l=1}^3 \frac{(O_l - E_l)^2}{E_l}$$

Hasil perhitungan yang disajikan pada Tabel 5.3 dapat dihitung nilai statistik chi-kuadrat sebesar 2.5906, dan dengan derajat bebas 2 (banyaknya kategori dikurangi 1) diperoleh p-value sebesar 0.2738. Dengan menggunakan taraf signifikansi  $\alpha = 0.05$ , maka hal ini menunjukkan hasil pengujian yang non signifikan. Berdasarkan hasil ini maka tidak terdapat cukup bukti untuk menyatakan bahwa data mengenai jumlah kematian bayi (berumur kurang 1 tahun) pada satu tahun 2008 di Kabupaten Bojonegoro, Jawa Timur, adalah bukan berasal dari sebaran Poisson.

### Pemilihan Peubah Tambahan

Selanjutnya adalah tahapan pemilihan peubah yang akan dijadikan sebagai peubah tambahan dalam pemodelan Bayes berhirarki dua-level. Pemilihan peubah tambahan ini menjadi penting, karena terdapat beberapa konsekuensi apabila terjadi kesalahan dalam menspesifikasikan model, diantaranya yaitu:

- a. Penduga kuadrat terkecil dari  $p$  buah peubah prediktor bersifat bias;
- b. Ragam dari penduga kuadrat terkecil bagi parameter dalam model penuh akan lebih besar atau sama dengan ragam dari parameter dalam model subset. Dengan demikian penghilangan peubah tidak akan meningkatkan ragam penduga dari parameter yang tersisa;
- c. Penduga kuadrat terkecil dalam model subset mempunyai kuadrat tengah galat yang lebih kecil dari penduga parameter dalam model penuh jika peubah yang dihilangkan mempunyai koefisien regresi yang lebih kecil dari galat baku penduga dalam model penuh;
- d. Penduga ragam dari peubah respons merupakan penduga yang bersifat bias.

Motivasi untuk pemilihan peubah tambahan yang dilakukan dalam penelitian ini dapat diringkas sebagai berikut: penghapusan beberapa peubah dari model, diharapkan dapat meningkatkan presisi dari penduga parameter dari peubah yang tersisa walaupun beberapa dari peubah yang dihilangkan itu tidak dapat diabaikan. Hal ini juga berlaku untuk ragam peubah respons. Penghilangan peubah berpotensi untuk menimbulkan bias pada penduga koefisien dari peubah tersisa dan peubah respons. Tetapi jika penghilangan peubah itu mempunyai efek yang kecil, maka kuadrat tengah galat dari

penduga yang bias akan lebih kecil dari ragam dari penduga tak bias. Artinya besarnya bias yang muncul lebih kecil dari reduksi ragamnya.

Tabel 5.4 Hasil pemilihan peubah tambahan melalui kriteria pemilihan himpunan bagian model regresi terbaik

$p$	$R^2$	$R^2_{adj}$	$C_p$	KTG	Peubah dalam model
1	0.2892	0.2583	12.7317	0.3802	{RTKS}
1	0.1562	0.1196	19.0432	0.4513	{RSHT}
2	0.4443	0.3938	7.3733	0.3108	{POSY, RTKS}
2	0.3942	0.3392	9.7492	0.3388	{K1, RTKS}
3	0.5122	0.4425	6.1508	0.2858	{K1, POSY, RTKS}
3	0.5117	0.4420	6.1730	0.2861	{K1, RSHT, RTKS}
4	0.6269	0.5523	2.7051	0.2295	{K1, POSY, RSHT, RTKS}
<b>4</b>	<b>0.7121</b>	<b>0.6345</b>	<b>4.1040</b>	<b>0.2386</b>	<b>{PSLN, RSHT, PEND, RTKS}</b>
5	0.6815	0.5977	2.1150	0.2062	{K1, POSY, RSHT, PEND, RTKS}
5	0.6781	0.5934	2.2773	0.2084	{K1, ASI, RSHT, PEND, RTKS}
6	0.7147	0.6197	2.5379	0.1950	{K1, ASI, POSY, RSHT, PEND, RTKS}
6	0.6942	0.5922	3.5140	0.2090	{K1, POSY, RSHT, PEND, KMSK, RTKS}
7	0.7212	0.6064	4.2315	0.2018	{K1, ASI, POSY, RSHT, PEND, KMSK, RTKS}
7	0.7197	0.6043	4.3014	0.2028	{K1, RISTI, ASI, POSY, RSHT, PEND, RTKS}
8	0.7250	0.5875	6.0520	0.2115	{K1, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}
8	0.7221	0.5831	6.1890	0.2137	{BBLR, K1, ASI, POSY, RSHT, PEND, KMSK, RTKS}
9	0.7253	0.5604	8.0380	0.2253	{K1, PSLN, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}
9	0.7252	0.5603	8.0411	0.2254	{BBLR, K1, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}
10	0.7261	0.5304	10.0006	0.2407	{BBLR, K1, PSLN, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}
10	0.7253	0.5291	10.0373	0.2414	{K1, KN2, PSLN, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}
11	0.7261	0.4943	12.0000	0.2592	{BBLR, K1, KN2, PSLN, RISTI, ASI, POSY, RSHT, PEND, KMSK, RTKS}

Berdasarkan pertimbangan tersebut, maka proses pemilihan peubah tambahan dilakukan dengan cara membangkitkan model subset dan memutuskan apakah model subset itu lebih baik daripada model subset yang lainnya. Ada empat kriteria yang biasa digunakan dalam pemilihan subset dari model regresi, yaitu koefisien determinasi ( $R^2$ ), koefisien determinasi terkoreksi ( $R^2_{adj}$ ), statistik  $C_p$ -Mallow ( $C_p$ ), serta kuadrat tengah galat (KTG). Dalam hal ini, pemilihan peubah tambahan difokuskan dengan menggunakan statistik  $C_p$ -Mallow, dimana peubah respons yang diamati terlebih dahulu

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

ditransformasikan ke dalam bentuk logaritma. Dalam penelitian ini pemilihan peubah tambahan lebih difokuskan dengan menggunakan statistik  $C_p$ -Mallow. Statistik  $C_p$ -Mallow merupakan fungsi dari jumlah kuadrat galat (JKG) untuk model penuh dan untuk model reduksi. Statistik  $C_p$ -Mallow dihitung dengan menggunakan rumus sebagai berikut:

$$C_p = \frac{JKG_p}{S^2} - n + 2p$$

dimana  $JKG_p$  adalah jumlah kuadrat galat untuk model yang berisi  $p$  buah peubah bebas,  $S^2$  adalah penduga bagi kuadrat tengah galat untuk model penuh, dan  $n$  adalah banyaknya pengamatan. Ketika model benar, maka nilai  $C_p$  akan mendekati banyaknya peubah bebas, sebaliknya ketika nilai  $C_p$  lebih besar dari  $p$ , maka hal ini mencerminkan adanya bias dalam penduga parameter dalam persamaan regresi.

Hasil-hasil pemilihan peubah tambahan melalui kriteria pemilihan himpunan bagian model regresi terbaik disajikan pada Tabel 5.4. Kolom pertama pada tabel tersebut merupakan banyak peubah tambahan yang disertakan ke dalam model. Walaupun fokus utama kriteria pemilihan ini adalah menggunakan statistik  $C_p$ -Mallow, akan tetapi sebagai bahan perbandingan ukuran atau kriteria lainnya (seperti koefisien determinasi ( $R^2$ ), koefisien determinasi terkoreksi ( $R_{adj}^2$ ), serta kuadrat tengah galat (KTG)) juga ditampilkan ke dalam tabel tersebut. Kemudian, pada kolom terakhir yang tersaji pada Tabel 5.4 adalah peubah-peubah yang berada di dalam model.

Perlu diketahui bahwa hasil pemilihan peubah tambahan yang diberikan pada Tabel 5.4 hanya menampilkan dua buah model yang dianggap 'terbaik' menurut kriteria tersebut untuk setiap kombinasi peubah yang berada di dalam model. Sebagai contoh misalnya, untuk kombinasi satu peubah, dua buah model yang dianggap 'terbaik' adalah model yang masing-masing mengandung peubah RTKS dan RSHT. Sedangkan untuk kombinasi dua peubah, dua buah model yang dianggap 'terbaik' adalah model yang masing-masing mengandung peubah {POSY, RTKS} dan {K1, RTKS}.

Berdasarkan hasil pemilihan peubah tambahan berdasarkan kriteria statistik  $C_p$ -Mallow adalah model yang mengandung empat buah peubah bebas, yaitu PSLN, RSHT, PEND, dan RTKS. Model ini mempunyai nilai statistik  $C_p$ -Mallow sebesar 4.1040, dimana nilai sebesar ini adalah nilai  $C_p$  yang paling mendekati dengan banyaknya peubah bebas yang berada di dalam model. Model ini juga mempunyai nilai koefisien determinasi dan koefisien determinasi terkoreksi yang relatif cukup besar dibandingkan model lainnya. Walaupun model ini mempunyai nilai KTG yang bukan paling kecil, akan tetapi secara keseluruhan model tersebut dianggap sebagai model yang 'terbaik', terutama karena mempunyai nilai statistik  $C_p$  yang paling mendekati dengan banyak peubah bebas yang berada dalam model. Dengan demikian, berdasarkan hasil tersebut peubah bebas yang dipertimbangkan untuk menjadi peubah tambahan dalam pemodelan Poisson Bayes berhirarki dua-level adalah sebagai berikut:

- PSLN, yaitu peubah yang menyatakan persentase persalinan di suatu wilayah dalam kurun waktu tertentu yang ditolong bukan oleh tenaga kesehatan profesional, seperti: dokter spesialis kebidanan, dokter umum, bidan, pembantu bidan, pembantu bidan dan perawat bidan.
- RSHT, yaitu peubah yang menyatakan persentase rumah tidak sehat.
- PEND, yaitu peubah yang menyakan pendidikan terakhir yang ditempuh oleh ibu (dihitung dalam tahun).

- d. RTKS, yaitu peubah yang menyatakan rasio ketersediaan tenaga kesehatan profesional dan tenaga kesehatan masyarakat terhadap penduduk.

### Hasil-hasil Pendugaan Angka Mortalitas

Target utama dari aplikasi pemodelan Bayes berhirarki dua-level untuk menangani masalah pendugaan area kecil ini adalah untuk memprediksi banyaknya bayi yang mati dan menduga angka kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur. Model yang dipertimbangkan adalah model Poisson Bayes berhirarki dua-level dengan menggunakan sebaran prior gamma dan invers gamma. Sebagaimana yang telah diketahui bahwa kerangka kerja dari model yang dikembangkan dalam penelitian ini merupakan perpaduan konsep model linear terampat dan model Bayes berhirarki untuk menangani masalah pendugaan area kecil. Hal-hal yang dikaji dalam aplikasi ini adalah sebagai berikut:

- a. Diagnostik kekonvergenan rantai Markov;
- b. Diagnostik kecocokan model dengan cara menerapkan konsep pemodelan linear terampat dalam model Bayes berhirarki pada masalah pendugaan area kecil, termasuk didalamnya adalah melakukan analisis sisaan;
- c. Sifat-sifat dari penduga parameter model Poisson Bayes berhirarki, terutama yang berkaitan dengan ketidakbiasan dan akurasi dari penduga parameter yang diamati, termasuk didalamnya adalah galat baku dari sebaran posteriornya; serta
- d. Nilai prediksi untuk sebaran posterior, dimana ukuran statistik yang diamatinya adalah rata-rata dan simpangan baku prediksi.

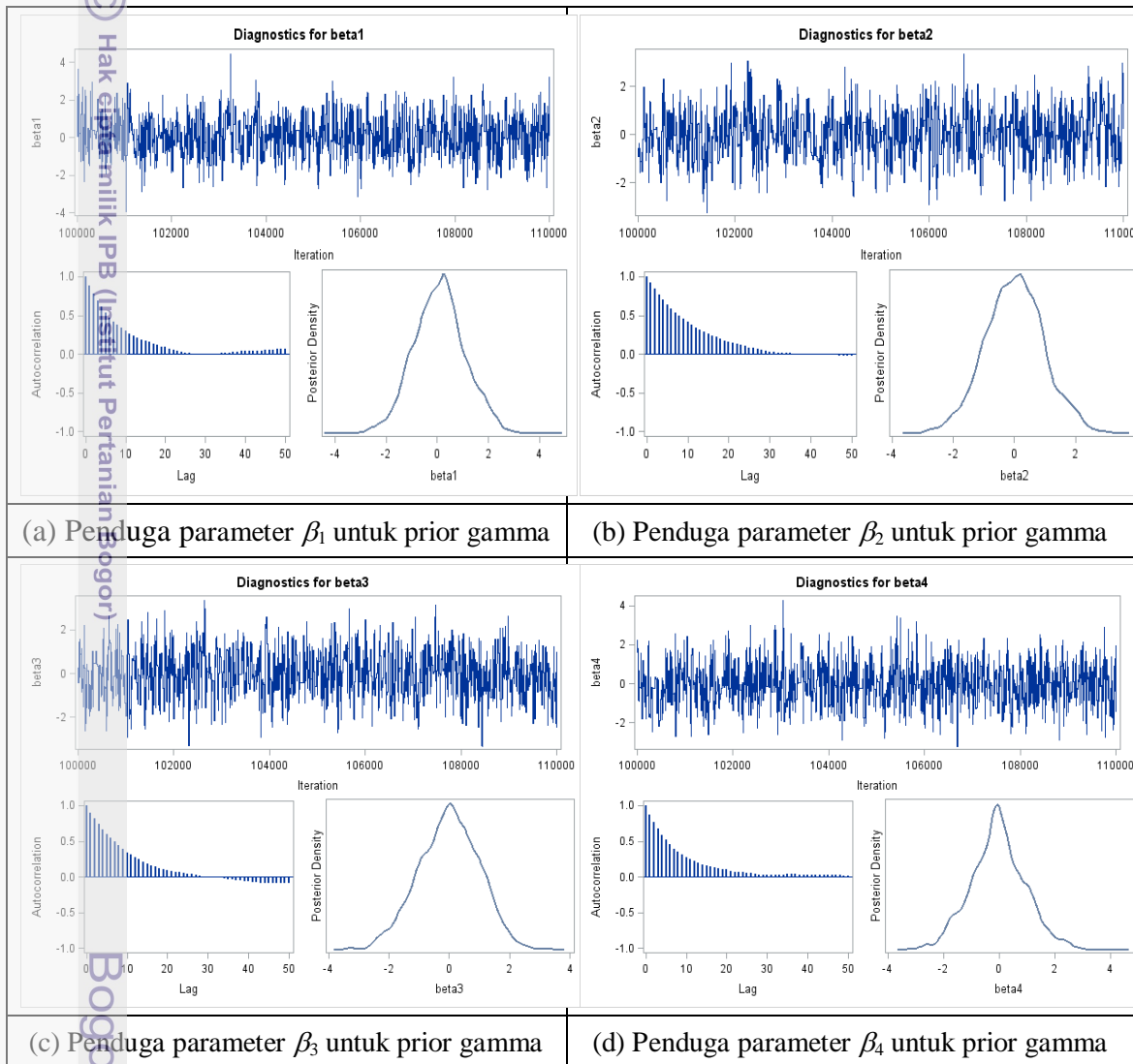
#### Diagnostik Kekonvergenan Rantai Markov

Hasil dari diagnostik kekonvergenan dari aplikasi model Poisson Bayes berhirarki dua-level ini dilakukan melalui *trace plot* sebagaimana yang ditunjukkan pada Gambar 5.1 dan Gambar 5.2. Dalam hal ini dapat dilihat melalui *trace plot* yang relatif konstan antara rata-rata dan ragamnya melalui plot fungsi otokorelasi antar contoh dan plot densitas peluang.

Berdasarkan hasil *trace plot* yang ditunjukkan pada Gambar 5.1 terlihat bahwa nilai dugaan parameter untuk setiap parameter ( $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , dan  $\beta_4$ ) dari sebaran prior gamma dipusatkan di sekitar nilai penduga parameter dari sebaran posteriornya, yaitu,  $\beta_1 = 0.0848$ ,  $\beta_2 = 0.0082$ ,  $\beta_3 = -0.0090$ , dan  $\beta_4 = -0.0339$  (lihat Gambar 5.1a, 5.1b, 5.1c, dan 5.1d). Hal ini menunjukkan bahwa proses pendugaan parameter melalui metode MCMC yang didasarkan pada sebaran prior gamma sudah mencapai kekonvergenan terhadap kestasioneran sebaran posteriornya. Lebih jauh, berdasarkan plot otokorelasi dan plot kepekatan peluang yang ditunjukkan pada Gambar 5.1 memberikan indikasi yang sejalan dengan hasil *trace plot*. Menurut hasil plot otokorelasi untuk setiap parameter yang diamati menunjukkan pola yang hampir sama, yaitu terjadi penurunan yang lambat. Demikian juga halnya menurut plot kepekatan peluang menunjukkan pola bahwa sebaran dari penduga parameter yang diamati adalah cenderung simetris.

Selanjutnya, hasil *trace plot* yang ditunjukkan pada Gambar 5.2 terlihat bahwa nilai dugaan parameter untuk setiap parameter ( $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , dan  $\beta_4$ ) dari sebaran prior invers gamma juga dipusatkan di sekitar nilai penduga parameter dari sebaran

posteriorinya, yaitu  $\beta_1 = 0.0892$ ,  $\beta_2 = 0.0142$ ,  $\beta_3 = -0.0023$ , serta  $\beta_4 = -0.0012$  (lihat Gambar 5.2a, 5.2b, 5.2c, dan 5.2d). Sedangkan, hasil plot otokorelasi dan plot kepekatan peluang yang ditunjukkan Gambar 5.2 yang dihasilkan dari sebaran prior invers gamma untuk setiap parameter yang diamati menampilkan pola yang tidak jauh berbeda dengan apa yang dihasilkan dari sebaran prior gamma. Artinya, terjadi penurunan yang lambat pada hasil plot otokorelasinya dan menurut plot kepekatan peluang menunjukkan pola bahwa sebaran dari penduga parameter yang diamati adalah cenderung simetris.



Gambar 5.1 Trace plot untuk model Poisson Bayes berhirarki level kecamatan di Kabupaten Bojonegoro yang berasal dari sebaran prior Gamma.

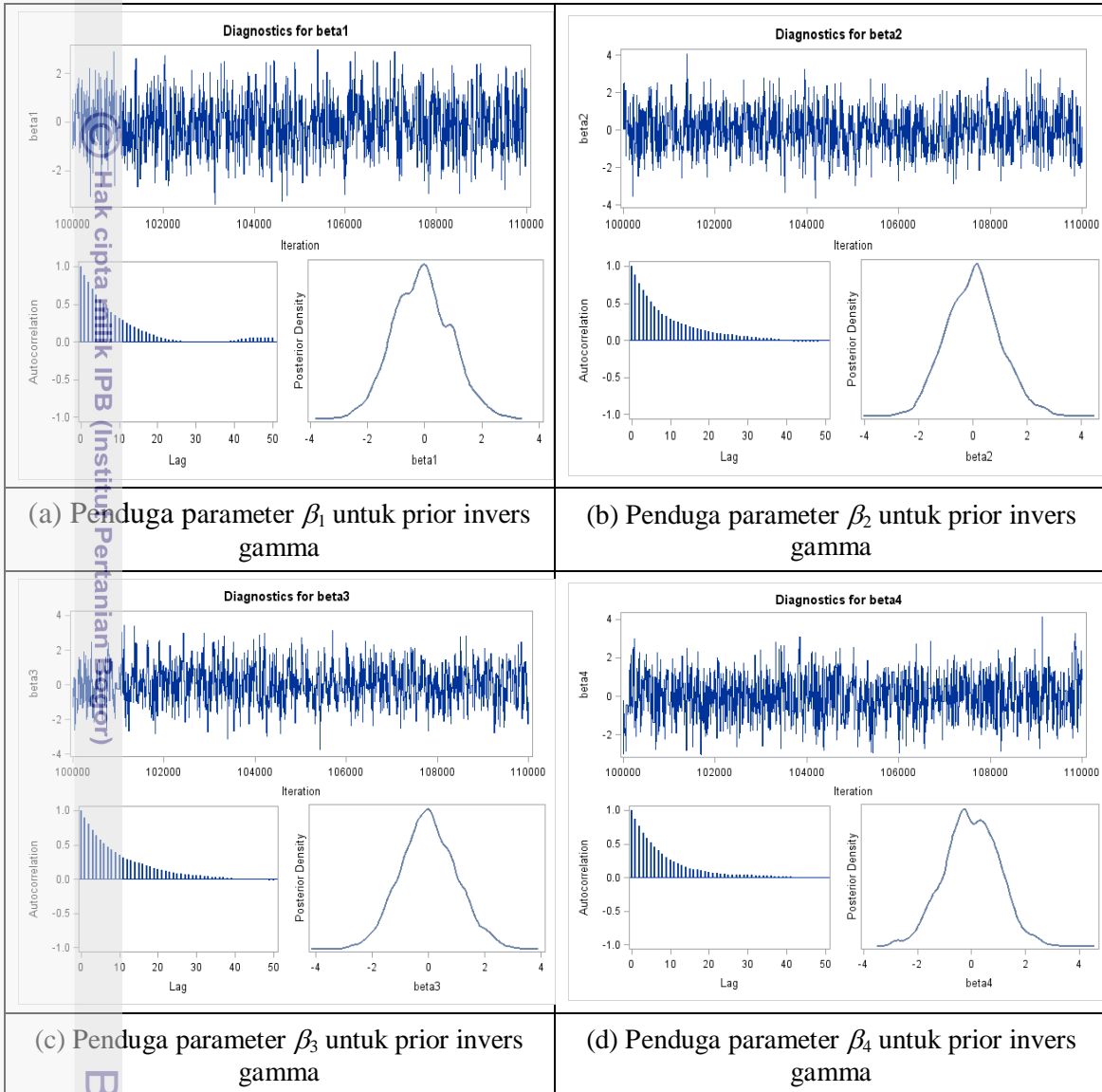
Hasil-hasil trace plot di atas juga didukung oleh hasil dari statistik uji Geweke yang disajikan pada Tabel 5.5. Perlu diketahui bahwa statistik Geweke digunakan untuk mendiagnostik kekonvergenan untuk rantai Markov berdasarkan pada pengujian kesamaan rata-rata bagian pertama dan akhir dari rantai Markov yang terbentuk (Geweke, 1992). Dalam hal ini bagian awal yang digunakan dari rantai Markov adalah sebesar 10%, sedangkan bagian akhir yang digunakan untuk pengujian adalah sebesar

Hak Cipta Dilindungi Undang-Undang

Hak Cipta Milik IPB (Institut Pertanian Bogor)

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

50%. Apabila contoh yang terambil berasal dari sebaran yang stasioner (atau dalam hal sudah mencapai kekonvergenan), maka kedua rata-rata tersebut adalah sama atau hasil pengujiannya adalah non-signifikan, serta statistik uji Geweke ini akan menyebar normal baku.



Gambar 5.2 *Trace plot* untuk model Poisson Bayes berhirarki level kecamatan di Kabupaten Bojonegoro yang berasal dari sebaran prior invers gamma.

Berdasarkan hasil diagnostik kekonvergenan penduga parameter rantai Markov melalui statistik uji Geweke untuk data level kecamatan di Kabupaten Bojonegoro, baik yang berasal dari sebaran prior gamma maupun invers gamma seluruhnya sudah mencapai kekonvergenan. Misalnya, untuk penduga parameter  $\beta_1$  yang berasal dari sebaran prior gamma diperoleh nilai statistik  $Z$  sebesar 1.4241 dengan  $p$ -value sebesar 0.1544, yang menunjukkan hasil pengujian yang non signifikan secara statistik. Hal ini berarti bahwa rata-rata penduga parameter  $\beta_1$  untuk 1000 contoh yang terambil pada bagian awal sama dengan rata-rata penduga parameter  $\beta_1$  untuk 5000 contoh yang

1. Diarangkan mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangkan mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



terambil pada bagian terakhir. Sementara itu, untuk penduga parameter  $\beta_1$  yang berasal dari sebaran prior invers gamma diperoleh nilai statistik  $Z$  sebesar 0.9859 dengan  $p$ -value sebesar 0.3242, yang juga menunjukkan hasil pengujian yang non signifikan secara statistik. Hasil yang sama juga untuk parameter lainnya yang berada di dalam model, baik yang berasal dari sebaran prior gamma maupun invers gamma.

Tabel 5.5 Diagnostik kekonvergenan rantai Markov melalui statistik uji Geweke untuk data level kecamatan di Kabupaten Bojonegoro

Parameter	Prior Gamma		Prior Invers Gamma	
	$z$	$Pr >  z $	$z$	$Pr >  z $
$\beta_0$	0.2043	0.8381	0.4596	0.6458
$\beta_1$	1.4241	0.1544	0.9859	0.3242
$\beta_2$	-1.2041	0.2286	0.5908	0.5546
$\beta_3$	-0.2721	0.7855	-0.9296	0.3526
$\beta_4$	-0.5084	0.6112	0.5096	0.6104

**Diagnostik Kecocokan Model dan Analisis Sisaan**

Tabel 5.6 menyajikan hasil-hasil mengenai berbagai kriteria informasi devians yang biasa dilakukan untuk mengevaluasi kecocokan model dalam metode MCMC. Kriteria yang digunakan untuk melakukan diagnostik model Poisson Bayes berhirarki ini adalah kriteria informasi devians (*Deviance Information Criterion*, DIC). Sebagaimana yang telah disebutkan bahwa dalam hal ini DIC menggunakan fungsi kepekatan posterior, yang berarti bahwa ukuran ini memperhitungkan informasi yang berasal dari sebaran priornya. Lebih jauh, suatu DIC yang nilainya lebih kecil menunjukkan bahwa model tersebut model yang cocok terhadap data. Berdasarkan hasil perhitungan DIC yang ditunjukkan pada Tabel 5.6 terlihat bahwa besaran DIC yang berasal dari sebaran prior gamma ( $DIC_G = 156.108$ ) lebih besar daripada DIC untuk sebaran prior invers gamma ( $DIC_{IG} = 133.054$ ). Hal ini menunjukkan bahwa model Poisson Bayes berhirarki dua-level dengan sebaran prior invers gamma mempunyai derajat kecocokan terhadap data yang lebih baik dibandingkan dengan model yang berasal dari sebaran prior gamma. Lebih jauh, selisih antara nilai DIC untuk model yang berasal dari prior gamma dengan DIC untuk model prior invers gamma adalah sebesar  $156.108 - 133.054 = 24.054$ . Selisih antara kedua nilai DIC tersebut adalah signifikan pada taraf sebesar 10%. Dengan demikian dapat dikatakan bahwa model Poisson Bayes berhirarki dua-level dengan prior invers gamma relatif lebih cocok untuk diaplikasikan dalam menduga angka kematian bayi level kecamatan di Kabupaten Bojonegoro.

Hasil tersebut juga sejalan dengan apa yang disampaikan oleh Spiegelhalter, et al. (2002) bahwa apabila selisih antara nilai DIC yang diperoleh dari dua buah model yang berbeda adalah lebih dari 10 atau antara 5 sampai dengan 10, maka dapat dikatakan dengan pasti bahwa model dengan nilai DIC lebih kecil adalah model yang lebih cocok terhadap data. Sedangkan apabila selisih antara nilai DIC yang diperoleh dari dua buah model yang berbeda adalah kurang dari 5, maka kesimpulan mengenai model yang lebih

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Hak cipta milik IPB (Institut Pertanian Bogor)

cocok terhadap data adalah model dengan nilai DIC yang paling kecil menjadi menyesatkan. Berdasarkan pernyataan tersebut, maka dikatakan bahwa model Poisson Bayes berhirarki dua-level dengan prior invers gamma relatif lebih cocok untuk diaplikasikan dalam menduga angka kematian bayi level kecamatan di Kabupaten Bojonegoro.

Tabel 5.6 Kriteria informasi devians untuk data level kecamatan di Kabupaten Bojonegoro

Kriteria	Gamma	Invers Gamma
Dbar (posterior mean of deviance)	152.145	127.950
Dmean (deviance evaluated at posterior mean)	148.182	123.846
pD (effective number of parameters)	3.963	4.104
Deviance Informations Criterion (DIC)	156.108	132.054

Hasil yang ditunjukkan oleh DIC di atas juga sejalan dengan apa yang diberikan oleh rata-rata devians posterior (Dbar) dan devians yang dievaluasi pada rata-rata posterior (Dmean). Perhatikan bahwa nilai Dbar dan Dmean yang berasal dari sebaran prior gamma masing-masing adalah 152.145 dan 148.182. Sedangkan nilai Dbar dan Dmean yang berasal dari sebaran prior invers gamma masing-masing adalah 127.950 dan 123.846. Jadi nilai Dbar dan Dmean yang berasal dari sebaran prior invers gamma lebih kecil daripada nilai Dbar dan Dmean yang berasal dari sebaran prior gamma. Hasil ini masuk akal karena hasil perhitungan DIC didasarkan pada kedua statistik tersebut.

Sementara itu, masih dalam Tabel 5.6, besaran pD atau banyaknya parameter yang efektif di dalam model untuk sebaran prior gamma adalah sebesar  $pD = 3.963$ , dan ini relatif lebih besar dibandingkan dengan nilai pD untuk sebaran prior invers gamma yang sebesar  $pD = 4.104$ . Sekali lagi bahwa besaran pD diukur melalui selisih antara rata-rata devians posterior (Dbar) dan devians yang dievaluasi pada rata-rata posterior (Dmean). Ukuran pD ini menggambarkan kompleksitas dari suatu model, dimana menurut Spiegelhalter, et al (2002) menyatakan bahwa model yang mempunyai nilai pD mendekati banyaknya parameter di dalam model menunjukkan model yang lebih 'baik'. Dalam kasus ini, model Poisson Bayes berhirarki untuk sebaran prior gamma dan invers gamma mempunyai nilai pD yang tidak jauh berbeda, Walaupun demikian, berdasarkan hasil diperoleh melalui besaran DIC, maka dapat dikatakan bahwa model Poisson Bayes berhirarki yang lebih cocok terhadap data mengenai angka kematian bayi level kecamatan di Kabupaten Bojonegoro pada Tahun 2008 adalah model dengan sebaran prior invers gamma.

Selanjutnya, hasil dari analisis sisaan dari model Poisson Bayes berhirarki dua-level untuk kedua sebaran prior yang digunakan dalam studi simulasi ini disajikan pada Tabel 5.7. Terdapat tiga buah ukuran sisaan yang digunakan dalam penelitian ini, yaitu sisaan Bayes, kuadrat tengah galat prediktif (*mean square predictive error*, MSPE), serta absolut tengah galat prediktif (*mean absolute predictive error*, MAPE). Perlu dijelaskan lagi di sini bahwa sisaan Bayes pada dasarnya digunakan untuk mendeteksi struktur hirarki yang tidak diketahui, sebagaimana yang biasa ditemukan dalam pemodelan Bayes berhirarki (Yan dan Sedransk, 2010). Sementara itu, besaran MSPE dan MAPE merupakan suatu ukuran untuk membandingkan data pengamatan dengan

data yang diprediksi dari model dugaan Poisson Bayes berhirarki, baik untuk sebaran prior gamma maupun invers gamma.

Tabel 5.7 Hasil analisis sisaan untuk data level kecamatan di Kabupaten Bojonegoro

Ukuran	Gamma	Invers Gamma
Sisaan Bayes	10.4753	6.0755
MSPE	14.1339	12.7704
MAPE	2.7820	2.6374

Berdasarkan hasil dari analisis sisaan ini tampak bahwa model Poisson Bayes berhirarki dengan sebaran prior invers gamma memberikan nilai-nilai sisaan (sisaan Bayes, MSPE, dan MAPE) sedikit lebih kecil dibandingkan dengan model dengan sebaran prior gamma. Misalnya, besarnya sisaan Bayes untuk model yang berasal dari sebaran prior invers gamma adalah sebesar 6.0755 adalah lebih kecil dibandingkan dengan sisaan Bayes untuk model dengan prior gamma yang sebesar 10.4753. Hasil yang sama juga ditunjukkan dari besaran MSPE dan MAPE untuk model dengan sebaran prior invers gamma (masing-masing sebesar 12.7704 dan 2.6374) relatif lebih kecil dibandingkan dengan nilai MSPE dan MAPE untuk model dengan sebaran prior gamma (masing-masing sebesar 14.1339 dan 2.7820). Berdasarkan hasil dari analisis sisaan ini dapat dikatakan bahwa model Poisson Bayes berhirarki dengan sebaran prior invers gamma lebih cocok terhadap data mengenai angka kematian bayi level kecamatan di Kabupaten Bojonegoro, Jawa Timur, dibandingkan model dengan sebaran prior gamma.

### Hasil Prediksi Jumlah Kematian Bayi dan Pendugaan Angka Kematian Bayi

Sebelum melakukan prediksi jumlah kematian bayi dan pendugaan angka kematian bayi level kecamatan akan diperhatikan terlebih dahulu derajat akurasi dari hasil pemodelan Poisson Bayes berhirarki dengan menggunakan dua buah sebaran prior yang berbeda. Dalam hal ini parameter yang diperhatikan adalah koefisien regresi dari model regresi Poisson Bayes berhirarki, yaitu  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , dan  $\beta_4$ , dimana ukuran statistik yang digunakan adalah rasio antara galat baku Monte Carlo (GBMC) dengan simpangan baku posterior (SBP) yang hasilnya disajikan pada Tabel 5.9.

Tabel 5.8 Galat baku Monte Carlo (GBMC) dan Simpangan baku posterior (SBP) untuk data level kecamatan Kabupaten Bojonegoro

Parameter	Sebaran Prior Gamma			Sebaran Prior InversGamma		
	GBMC	SBP	GBMC/SBP	GBMC	SBP	GBMC/SBP
$\beta_0$	0.0440	0.9662	0.0456	0.0387	0.9389	0.0412
$\beta_1$	0.0394	0.9984	0.0394	0.0395	0.9940	0.0398
$\beta_2$	0.0436	0.9601	0.0454	0.0414	1.0002	0.0414
$\beta_3$	0.0416	0.9997	0.0416	0.0438	0.9979	0.0439
$\beta_4$	0.0406	0.9870	0.0412	0.0397	1.0029	0.0396

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Berdasarkan hasil diberikan pada Tabel 5.8 nampak bahwa rasio antara galat baku Monte Carlo dan simpangan bakunya untuk rantai Markov yang diperoleh dari sebaran prior gamma dan prior invers gamma untuk setiap parameter memberikan nilai yang tidak jauh berbeda. Hasil-hasil yang diberikan Tabel 5.8 menunjukkan bahwa performa pendugaan model Poisson Bayes berhirarki dua-level yang dihasilkan dari sebaran prior gamma dan invers gamma adalah tidak jauh berbeda. Artinya bahwa penduga parameter yang diamati mempunyai derajat akurasi yang tidak jauh berbeda.

Tabel 5.9 Hasil prediksi kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur.

No	Kecamatan	N	Y	Gamma		Invers Gamma		Poisson	
				Rata-rata	SB	Rata-rata	SB	Rata-rata	SB
1	Balen	946	14	8.22	2.85	17.95	5.15	9.24	3.04
2	Baureno	1206	15	7.82	2.81	16.62	4.49	8.68	2.95
3	Bojonegoro	1434	6	3.62	1.93	3.64	1.99	7.98	2.82
4	Bubulan	159	2	3.12	1.78	1.10	1.06	4.78	2.19
5	Dander	1022	4	4.19	2.06	8.98	3.24	4.29	2.07
6	Kalitidu	1020	5	2.28	1.52	2.26	1.55	4.01	2.00
7	Kanor	782	6	6.12	2.44	5.91	2.83	3.83	1.96
8	Kapas	793	6	3.09	1.75	4.76	2.33	6.26	2.50
9	Kasiman	434	0	1.53	1.23	1.01	1.00	4.69	2.17
10	Kedungadem	1102	6	3.55	1.87	10.27	3.71	4.91	2.22
11	Kepohbaru	883	2	1.57	1.26	1.98	1.45	3.69	1.92
12	Malo	473	4	1.06	1.03	4.99	2.34	3.31	1.82
13	Margomulyo	341	1	0.99	0.99	1.10	1.06	3.99	2.00
14	Ngambon	146	1	1.01	1.01	2.52	1.61	4.29	2.07
15	Ngasem	1062	9	7.32	2.72	3.85	2.10	4.31	2.08
16	Ngraho	675	4	2.70	1.64	7.12	2.95	10.17	3.19
17	Padangan	670	3	2.18	1.47	1.10	1.05	4.33	2.08
18	Purwosari	519	3	1.09	1.05	1.06	1.03	4.15	2.04
19	Sugihwaras	710	5	2.91	1.70	8.77	3.48	5.07	2.25
20	Sumberejo	1069	3	1.56	1.26	1.04	1.04	4.09	2.02
21	Tambakrejo	779	2	4.82	2.19	3.84	2.03	3.51	1.87
22	Temayang	438	4	1.61	1.27	2.23	1.52	3.02	1.74
23	Trucuk	555	5	5.27	2.27	2.86	1.75	4.11	2.03
24	Sukosewu	596	3	3.18	1.76	5.98	2.65	3.53	1.88
25	Gondang	305	5	1.92	1.39	4.50	2.21	3.52	1.88
26	Sekar	320	16	8.46	2.88	16.25	5.07	4.61	2.15
27	Kedewan	178	0	7.91	2.82	1.02	1.03	5.62	2.37

**Keterangan:** peubah acak Poisson  $Y$  adalah banyak bayi yang mati, dan  $N$  adalah banyaknya bayi yang lahir hidup.

Sebagaimana yang telah dijelaskan dalam bab sebelumnya bahwa simpangan baku posterior dan GBMC merupakan dua konsep yang sangat berbeda. Simpangan baku posterior menggambarkan ketidakpastian dalam parameter, sedangkan GBMC hanya menggambarkan ketidakpastian dalam penduga parameter sebagai hasil dari simulasi MCMC. Simpangan baku posterior merupakan fungsi dari ukuran contoh dalam suatu gugus data, sedangkan GBMC merupakan fungsi dari banyaknya iterasi dalam proses

simulasi MCMC. Dengan demikian, maka dapat dikatakan bahwa penduga parameter yang diperoleh melalui prior gamma mempunyai derajat ketakbiasan dan akurasi yang relatif sama dengan penduga parameter yang diperoleh dari prior invers gamma pada saat diaplikasikan untuk menduga angka kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur.

Tabel 5.9 menyajikan hasil prediksi jumlah kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur. Pada tabel tersebut berisi hasil prediksi kematian bayi level kecamatan dengan menggunakan tiga buah model yang berbeda, yaitu model Poisson Bayes berhirarki dengan prior gamma, model Poisson Bayes berhirarki dengan prior invers gamma, serta model regresi Poisson biasa. Dua model yang pertama tentu saja merupakan model kajian utama dalam penelitian ini yang diaplikasikan dalam masalah pendugaan area kecil. Sedangkan model regresi Poisson disajikan dengan tujuan untuk membandingkan performa model antara model yang memperhitungkan efek area (model Poisson Bayes berhirarki) dan model yang tidak memperhitungkan efek area (model Poisson biasa).

Berdasarkan hasil prediksi banyaknya bayi yang mati level kecamatan di Kabupaten Bojonegoro ini terlihat bahwa hasil prediksi dari model Poisson Bayes berhirarki dengan sebaran prior invers gamma mendekati data aktual. Yang dimaksud dengan data aktual di sini adalah data yang diperoleh dari hasil survei, dimana dalam hal ini merupakan data hasil olahan dari SDKI 2007 dan Susenas 2010. Perhatikan bahwa jumlah kematian bayi di Kecamatan Balen dan Baureno menurut data hasil survei masing-masing adalah sebanyak 14 dan 15. Kemudian hasil prediksi menurut model Poisson Bayes berhirarki dua-level dengan sebaran prior invers gamma untuk kedua kecamatan tersebut masing-masing adalah 17.9530 dan 16.6234. Hasil ini lebih mendekati data hasil survei dibandingkan dengan hasil prediksi dari model yang berasal dari sebaran prior gamma dengan hasil prediksi masing-masing sebesar 8.2156 dan 7.8184.

Hasil prediksi berdasarkan model Poisson Bayes berhirarki dengan sebaran prior invers gamma untuk kecamatan lainnya pada umumnya juga lebih mendekati data hasil survei dibandingkan dengan model dengan prior gamma. Bahkan untuk Kecamatan Sekar yang menurut hasil survei terjadi kematian bayi sebanyak 16, kemudian diprediksi melalui model dengan sebaran prior invers gamma sebesar 16.2547. Sedangkan hasil prediksi menurut model dengan sebaran prior gamma untuk Kecamatan Sekar hanya sebesar 8.4627. Hasil ini cukup masuk akal karena sejalan dengan hasil-hasil dari analisis sisaan yang disajikan pada Tabel 5.7, dimana model dengan prior invers gamma memberikan sisaan yang relatif lebih kecil dibandingkan model dengan prior gamma.

Sementara itu Tabel 5.10 menyajikan hasil-hasil pendugaan angka kematian bayi level kecamatan di Kabupaten Bojonegoro untuk tiga buah model yang dikaji bersamaan dengan angka kematian bayi yang dihitung langsung dari data aktual atau data hasil survei. Hasil pendugaan angka kematian bayi yang diperoleh melalui model Poisson Bayes berhirarki dengan prior invers gamma terlihat mendekati hasil pendugaan angka kematian bayi yang dihitung dari data aktual. Sedangkan hasil pendugaan dari model Poisson Bayes berhirarki dengan prior gamma pada umumnya memberikan nilai penduga yang lebih kecil dibandingkan data AKB hasil survei. Sebagai contoh, misalnya hasil pendugaan menurut hasil survei,  $\hat{\theta}_i^{(PL)}$ , untuk Kecamatan Balen dan Baureno masing-masing adalah sebesar 14.80 dan 12.44. Sementara itu, hasil dugaan angka kematian bayi yang diperoleh dari model Poisson Bayes berhirarki dengan

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

sebaran prior invers gamma,  $\hat{\theta}_i^{(IG)}$ , untuk kedua kecamatan tersebut masing-masing adalah sebesar 16.98 dan 13.78. Hasil dugaan angka kematian bayi untuk kedua kecamatan tersebut relatif lebih mendekati dengan hasil dugaan data survei dibandingkan dengan hasil dugaan dari model Poisson Bayes berhirarki dengan sebaran prior gamma,  $\hat{\theta}_i^{(G)}$ . Hasil dugaan angka kematian bayi untuk Kecamatan Balen dan Baureno menurut model dengan sebaran prior gamma masing-masing adalah sebesar 8.68 dan 6.48, jauh lebih kecil dibandingkan dengan hasil dugaan survei.

Tabel 5.10 Hasil pendugaan angka kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur.

No.	Kecamatan	N	Y	$\hat{\theta}_i^{(PL)}$	$\hat{\theta}_i^{(G)}$	$\hat{\theta}_i^{(IG)}$	$\hat{\theta}_i^{(P)}$
1	Balen	946	14	14.80	8.68	16.98	9.77
2	Baureno	1206	15	12.44	6.48	13.78	7.19
3	Bojonegoro	1434	6	4.18	2.53	2.54	5.56
4	Bubulan	159	2	12.58	19.59	6.90	30.08
5	Dander	1022	4	3.91	4.10	8.79	4.20
6	Kalitidu	1020	5	4.90	2.23	2.21	3.93
7	Kanor	782	6	7.67	7.82	7.55	4.90
8	Kapas	793	6	7.57	3.89	6.01	7.90
9	Kasiman	434	0	0.00	3.53	2.33	10.80
10	Kedungadem	1102	6	5.44	3.23	9.32	4.45
11	Kepohbaru	883	2	2.27	1.78	2.24	4.17
12	Malo	473	4	8.46	2.25	10.54	7.00
13	Margomulyo	341	1	2.93	2.91	3.23	11.71
14	Ngambon	146	1	6.85	6.95	17.28	29.39
15	Ngasem	1062	9	8.47	6.90	3.63	4.06
16	Ngraho	675	4	5.93	4.00	10.54	15.06
17	Padangan	670	3	4.48	3.26	1.64	6.46
18	Purwosari	519	3	5.78	2.09	2.05	7.99
19	Sugihwaras	710	5	7.04	4.09	12.35	7.15
20	Sumberejo	1069	3	2.81	1.46	0.97	3.83
21	Tambakrejo	779	2	2.57	6.18	4.93	4.50
22	Temayang	438	4	9.13	3.69	5.10	6.89
23	Trucuk	555	5	9.01	9.50	5.16	7.41
24	Sukosewu	596	3	5.03	5.34	10.03	5.93
25	Gondang	305	5	16.39	6.30	14.75	11.54
26	Sekar	320	16	50.00	26.45	50.80	14.42
27	Kedewan	178	0	0.00	44.42	5.76	31.55
AKB Kabupaten Bojonegoro				7.36	5.32	7.67	7.20

**Keterangan:** peubah acak Poisson  $Y$  adalah banyak bayi yang mati, dan  $N$  adalah banyaknya bayi yang lahir hidup,  $\hat{\theta}_i^{(G)}$  adalah penduga langsung,  $\hat{\theta}_i^{(G)}$  adalah penduga AKB dari prior gamma,  $\hat{\theta}_i^{(IG)}$  adalah penduga AKB dari prior invers gamma, serta  $\hat{\theta}_i^{(P)}$  adalah penduga dari model regresi Poisson biasa.

Sebagai gambaran angka kematian bayi untuk Kabupaten Bojonegoro pada Tahun 2008 adalah sebesar 7.36, sebagaimana yang dilaporkan dalam Laporan Profil Kesehatan Provinsi Jawa Timur Tahun 2008 (Dinas Kesehatan Provinsi Jawa Timur, 2008). Sementara itu, secara agregat hasil pendugaan angka kematian bayi level

kabupaten menurut model Poisson Bayes berhirarki dengan prior invers gamma adalah sebesar 7.67, sedangkan hasil pendugaan AKB menurut model dengan prior gamma adalah sebesar 5.32. Hal ini menunjukkan bahwa model Poisson Bayes berhirarki dengan prior invers gamma memberikan dugaan pada AKB yang tidak jauh berbeda dengan apa yang dilaporkan dalam profil kesehatan Provinsi Jawa Timur tersebut, sedangkan model dengan prior gamma justru memberikan dugaan AKB yang jauh lebih kecil. Namun demikian, satu hal yang perlu dicatat adalah bahwa AKB yang dimuat dalam Laporan Profil Kesehatan tersebut belum bisa menggambarkan AKB yang sebenarnya di populasi.

Untuk mengevaluasi kelayakan suatu model pada umumnya ukuran statistik yang diperhatikan adalah besaran KTG. Akan tetapi perlu diketahui bahwa dalam penelitian ini komputasi yang dipakai menggunakan algoritma MCMC yang merupakan metode yang efisien dari penarikan contoh dari sebaran posterior untuk parameter yang dinyatakan dalam model. Sebagaimana yang telah disebutkan sebelumnya di sini menggunakan contoh dari sebaran posterior untuk menduga ringkasan statistik yang diamati. Artinya untuk sembarang contoh  $\theta^1, \theta^2, \dots, \theta^n$ , dapat menduga nilai harapan  $E[g(\theta)]$  untuk suatu fungsi  $g$  melalui

$$E[g(\theta)] = \frac{1}{n} \sum_{j=1}^n g(\theta^j)$$

yang merupakan penduga Monte Carlo biasa. Perlu dicatat bahwa  $g(\theta^1), \dots, g(\theta^n)$  membentuk contoh dari sebaran posterior  $g(\theta)$ . Jadi, jika  $\hat{\mu}_n = \sum_{j=1}^n g(\theta^j)$  dan  $\mu = E[g(\theta)]$ , maka  $\hat{\mu}_n$  merupakan penduga Monte Carlo bagi  $\mu$  berdasarkan pada suatu contoh yang berukuran  $n$ . Akan tetapi perlu juga dipertimbangkan galat Monte Carlo, yaitu  $\hat{\mu}_n - \mu$ .

Pada saat contoh  $g(\theta^1), \dots, g(\theta^n)$  adalah saling bebas dan menyebar secara identik sebagai realisasi bagi  $\theta$ , maka melalui hukum bilangan besar (*law of large number*, LLN) diperoleh  $\hat{\mu}_n \xrightarrow{a.s} \mu$  sebagaimana  $n \rightarrow \infty$ , serta berdasarkan dalil limit pusat diperoleh  $\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$  sebagaimana  $n \rightarrow \infty$ , dimana  $\sigma^2 = var[g(\theta)]$  yang diasumsikan diketahui.

Pada umumnya besar ragam  $\sigma^2 = var[g(\theta)]$  tidak diketahui sehingga perlu diduga. Galat baku Monte Carlo (GBMC) merupakan penduga yang konsisten bagi  $\sigma$  sedemikian rupa sehingga

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (g(\theta^j) - \hat{\mu}_n)^2$$

Kemudian, diketahui bahwa  $\frac{\hat{\mu}_n - \mu}{\hat{\sigma}_n} \xrightarrow{D} N(0, 1)$  sebagaimana  $n \rightarrow \infty$ . Artinya bahwa dapat diperoleh

$$\hat{\mu}_n \approx N(\mu, \hat{\sigma}_n^2/n)$$

Akan tetapi teori di atas berlaku dengan asumsi contoh dari  $\theta$  adalah saling bebas dan menyebar identik. Padahal contoh yang termabil melalui prosedur MCMC tentu saja tidak saling bebas. Oleh karena itu, Gelman et al (2004) mengusulkan untuk menggunakan suatu ukuran KTG yang berlaku pada saat contoh yang diperoleh melalui prosedur MCMC dengan cara menghitung ragam  $\sigma^2 = var[g(\theta)]$  sebagai berikut:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memunculkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Fakultas Pertanian (Institut Pertanian Bogor) Bogor Agricultural University

$$\sigma^2 = \text{var}[g(\theta)] + 2 \sum_{j \neq k=1}^n \text{cov}[g(\theta^j), g(\theta^k)]$$

Yang membedakan antara ragam di atas dengan ragam yang biasa adalah menyertakan bentuk koragam dalam perhitungan  $\sigma^2$ .

Berdasarkan hasil perhitungan diperoleh galat baku Monte Carlo untuk prediksi yang diperoleh dari model Poisson Bayes berhirarki dua-level dengan sebaran prior gamma adalah sebesar 0.0278. Sedangkan galat baku Monte Carlo untuk prediksi yang diperoleh dari model Poisson Bayes berhirarki dua-level dengan sebaran prior invers gamma adalah sebesar 0.0149. Dengan demikian dapat dinyatakan bahwa model Poisson Bayes berhirarki dua-level yang berasal dari sebaran prior invers gamma merupakan model yang relatif lebih baik dibandingkan dengan model yang sebaran prior gamma dalam memprediksi angka kematian bayi level kecamatan di Kabupaten Bojonegoro, Jawa Timur.

Hal lain yang perlu diperhatikan adalah adanya beberapa hasil pendugaan AKB yang sepeertinya tidak 'masuk akal'. Sebagai contoh, misalnya hasil pendugaan AKB pada Kecamatan Kadewan menurut model Poisson Bayes berhirarki dengan prior gamma adalah sebesar 44.42. Padahal AKB yang dihitung dari data aktual adalah sebesar 0, karena pada kecamatan tersebut dilaporkan tidak ada bayi yang meninggal. Sementara itu, hasil pendugaan AKB menurut model Poisson Bayes berhirarki dengan prior invers gamma hanya memberikan dugaan sebesar 5.76. Hasil lain yang juga cukup menonjol adalah terjadi Kecamatan Sekar, dimana AKB menurut data aktual adalah sebesar 50.00, sedangkan menurut model dengan prior gamma dan invers gamma masing-masing adalah sebesar 26.45 dan 50.80. Sekali lagi, telah ditunjukkan bahwa model Poisson Bayes berhirarki dengan prior gamma memberikan dugaan AKB yang relatif jauh berbeda dengan AKB yang dihitung dari data aktual, sementara itu model dengan prior invers gamma relatif memberikan hasil dugaan AKB yang tidak jauh berbeda dengan data aktual.

Ada beberapa hal yang menyebabkan terjadinya masalah tersebut. Salah satu diantaranya adalah kemungkinan adanya data pencilan yang mungkin berpotensi sebagai data berpengaruh. Kemungkinan yang kedua adalah munculnya masalah heteroskedastisitas dalam data. Sebagaimana yang telah diketahui bahwa dalam pemodelan linear terampat dimana peubah respons yang diamatinya adalah berbentuk data cacahan yang bersebaran Poisson, seringkali dihadapkan pada suatu masalah yang disebut dengan masalah *overdispersion*. Perlu diketahui bahwa peubah acak respons yang diamati dalam penelitian ini adalah peubah acak yang saling bebas yang bersebaran Poisson. Dalam prakteknya asumsi bahwa peubah acak ini adalah saling bebas, terutama dalam analisis data kategorik, seringkali tidak terpenuhi. Kemungkinan terjadinya kedua masalah tersebut dalam penelitian ini belum dikaji lebih mendalam. Sebagai contoh, misalnya untuk mengatasi masalah data pencilan dapat dilakukan dengan cara membentuk suatu model Poisson Bayes berhirarki yang bersifat *robust*. Sementara itu, pada dasarnya model yang dikembangkan dalam penelitian ini sudah mempertimbangkan bahwa parameter angka kematian dianggap sebagai suatu peubah acak yang secara tidak langsung bisa digunakan dalam mengatasi masalah *overdispersion*. Namun dalam penelitian ini belum dikaji mengenai pendugaan dari parameter dispersinya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memungut dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



## 6 KESIMPULAN DAN SARAN

### Kesimpulan

Dalam penelitian ini telah dikembangkan suatu model Poisson Bayes berhirarki dua-level yang memadukan konsep pemodelan linear terampat dan pemodelan Bayes berhirarki dalam menangani masalah pendugaan area kecil. Peubah respons yang diamati dalam penelitian ini adalah berbentuk data cacahan yang bersebaran Poisson yang merupakan anggota dari keluarga eksponensial sebagai bagian dari konsep dalam pemodelan linear terampat. Kemudian, peubah acak Poisson dengan parameter  $\theta$  dianggap sebagai suatu peubah acak yang dimodelkan melalui model Fay-Herriot sebagai model dasar yang digunakan dalam pendugaan area kecil. Kedua konsep pemodelan tersebut kemudian diimplementasikan melalui pendekatan Bayes berhirarki sehingga membentuk Model Poisson Bayes berhirarki dua-level.

Permasalahan utama yang sering dihadapi dalam pendugaan area kecil melalui pendekatan Bayes berhirarki seperti ini adalah pemilihan sebaran prior. Pemilihan sebaran prior dalam hal ini menjadi penting karena apabila sebaran prior ini diketahui maka inferensi dapat dengan mudah dilakukan dengan cara meminimumkan galat posterior, menghitung daerah kepekatan sebaran posterior yang lebih tinggi dimensinya, atau mengintegrasikan parameter untuk menemukan sebaran prediktifnya. Sebaran prior merupakan kunci pada inferensi analisis Bayes, sehingga penetapan prior ini menjadi tahapan yang paling penting dalam menggambarkan inferensi ini. Dalam beberapa kasus pemilihan sebaran prior ini merupakan hal yang sangat sulit. Akan tetapi dalam praktiknya seringkali sudah tersedia informasi prior yang mempunyai presisi yang cukup baik sehingga akan membawa pada penetapan eksak dari suatu sebaran prior tertentu.

Dalam penelitian telah digunakan dua buah sebaran prior yang berbeda, yaitu sebaran gamma yang merupakan sebaran prior yang bersifat *conjugate* bagi sebaran Poisson, serta sebaran invers gamma yang merupakan sebaran prior yang bersifat *non-conjugate*. Pemilihan sebaran invers gamma dilakukan dengan pertimbangan bahwa sifat sifat sebaran dari invers gamma tidak jauh berbeda dengan sebaran gamma. Lebih jauh, sebaran invers gamma ini dirujuk juga sebagai sebaran prior yang bersifat *conditionally conjugate*, yaitu suatu keluarga sebaran prior  $p(\lambda)$  merupakan *conditionally conjugate* untuk parameter  $\lambda$  apabila sebaran posterior bersyarat,  $p(\lambda|y)$  juga berada dalam kelas tersebut. Untuk keperluan komputasi, *conditionally conjugate* mempunyai makna bahwa apabila memungkinkan untuk mengambil  $\lambda$  yang berasal kelas sebaran prior tertentu, maka hal ini juga memungkinkan untuk membentuk Gibbs sampler bagi  $\lambda$  dalam sebaran posterior. Dalam penelitian ini telah ditemukan secara analitik bentuk dari sebaran posterior bagi peubah acak Poisson dengan menggunakan sebaran prior invers gamma, walaupun bentuk sebaran posteriornya masih berbentuk pendekatan (tidak eksak).

Berdasarkan hasil-hasil kajian yang telah dilakukan dalam penelitian ini menunjukkan bahwa performa model Poisson Bayes berhirarki dua-level ini dengan menggunakan dua buah sebaran prior (sebaran prior gamma dan invers gamma) yang berbeda adalah tidak jauh berbeda. Hasil-hasil tersebut diperoleh baik melalui studi simulasi maupun studi dari aplikasi penerapan model untuk menduga angka kematian bayi level kecamatan di Kabupaten Bojonegoro Provinsi Jawa Timur. Beberapa kriteria

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

yang digunakan untuk membandingkan kedua model tersebut adalah derajat akurasi model, derajat kekonvergenan rantai Markov, serta derajat kecocokan model.

Tidak seperti pemodelan pendugaan area kecil pada umumnya dalam melihat derajat akurasi penduga parameter yang menggunakan KTG, dalam konteks pemodelan Bayes berhirarki menggunakan galat baku Monte Carlo (GBMC). Hasilnya menunjukkan bahwa model Poisson Bayes berhirarki dua-level dengan sebaran prior gamma dan invers gamma memberikan nilai GBMC yang relatif cukup kecil dan keduanya memberikan nilai yang tidak jauh berbeda.

Sementara itu, hasil-hasil mengenai derajat kekonvergenan rantai Markov diantara kedua model ini juga menunjukkan hasil yang tidak jauh berbeda. Proses komputasi dengan menggunakan metode MCMC untuk proses tunggal (berdasarkan hasil *trace plot*, *plot* otokorelasi, *plot* fungsi kepekatan peluang, serta statistik Geweke) menunjukkan bahwa kedua model tersebut mempunyai derajat kekonvergenan yang tidak jauh berbeda. Perlu diketahui bahwa algoritma MCMC yang digunakan dalam penelitian ini adalah algoritma Metropolis-Hasting, dimana algoritma ini mendefinisikan suatu rantai Markov dari peubah acak dimana sebarannya akan konvergen pada suatu sebaran yang diinginkan. Dalam terminologi rantai Markov, sebaran semacam itu disebut sebagai sebaran yang stasioner, dan dalam statistik Bayes, ini merupakan sebaran posterior bagi parameter model. Dengan demikian dapat dikatakan bahwa sebaran prior gamma dan juga invers gamma memberikan suatu proses rantai Markov yang sudah mencapai pada sebaran targetnya, atau dengan kata lain sudah mencapai kekonvergenan pada kestasioneran sebarannya.

Selanjutnya, berdasarkan hasil evaluasi kecocokan model dan analisis sisaan juga menunjukkan bahwa model Poisson Bayes berhirarki dua-level untuk sebaran prior gamma dan invers gamma adalah juga tidak jauh berbeda. Beberapa ukuran untuk mengevaluasi kecocokan model seperti DIC (*deviance information criterion*), sisaan Bayes, kuadrat tengah galat prediktif (MSPE), serta absolut tengah galat prediktif (MAPE) pada umumnya memberikan hasil-hasil yang tidak jauh berbeda. Berdasarkan hasil prediksi banyaknya bayi yang mati level kecamatan di Kabupaten Bojonegoro ini terlihat bahwa hasil prediksi dari model Poisson Bayes berhirarki dengan sebaran prior invers gamma mendekati data aktual. Sedangkan hasil pendugaan angka kematian bayi yang diperoleh melalui model Poisson Bayes berhirarki dengan prior invers gamma terlihat mendekati hasil pendugaan angka kematian bayi yang dihitung dari data aktual. Sedangkan hasil pendugaan dari model Poisson Bayes berhirarki dengan prior gamma pada umumnya memberikan nilai penduga yang lebih kecil dibandingkan AKB aktual. Akhirnya, secara umum kesimpulan yang dapat diambil dari penelitian ini menunjukkan bahwa model Poisson Bayes berhirarki dua-level sebaran prior gamma (yang bersifat *conjugate*) dan sebaran prior invers gamma (yang bersifat *conditionally conjugate*) memberikan hasil pendugaan angka kematian bayi yang tidak jauh berbeda.

## Saran

Ketepatan penggunaan model ini sangat bergantung pada berbagai kondisi, seperti ukuran area, ketersediaan peubah penjelas yang baik di tingkat area, akurasi dari penduga ragam penarikan contoh. Pengembangan dari model HB ini menjadi dimungkinkan karena ketersediaan perangkat lunak yang menyediakan fasilitas untuk melakukan simulasi MCMC, sehingga proses pendugaan model menjadi relatif lebih

mudah. Permasalahan yang masih relevan untuk dibahas adalah untuk melakukan inferensi terutama yang berkaitan dengan pemilihan spesifikasi model yang ‘terbaik’.

Dalam penelitian terdapat beberapa keterbatasan untuk dikaji lebih mendalam. Salah satu diantaranya adalah kemungkinan adanya data pencilan yang mungkin berpotensi sebagai data berpengaruh. Kemungkinan yang kedua adalah munculnya masalah heteroskedastisitas dalam data. Sebagaimana yang telah diketahui bahwa dalam pemodelan linear terampat dimana peubah respons yang diamatinya adalah berbentuk data cacahan yang bersebaran Poisson, seringkali dihadapkan pada suatu masalah yang disebut dengan masalah *overdispersion*. Perlu diketahui bahwa peubah acak respons yang diamati dalam penelitian ini adalah peubah acak yang saling bebas yang bersebaran Poisson. Dalam prakteknya asumsi bahwa peubah acak ini adalah saling bebas, terutama dalam analisis data kategorik, seringkali tidak terpenuhi. Kemungkinan terjadinya kedua masalah tersebut dalam penelitian ini belum dikaji lebih mendalam.

Oleh karena itu, perlu kiranya dilakukan kajian yang mendalam tentang bagaimana mengatasi masalah data pencilan dapat dilakukan dengan cara membentuk suatu model Poisson Bayes berhirarki yang bersifat *robust*. Sementara itu, pada dasarnya walaupun model yang dikembangkan dalam penelitian ini sudah mempertimbangkan bahwa parameter angka kematian dianggap sebagai suatu peubah acak yang secara tidak langsung bisa digunakan dalam mengatasi masalah *overdispersion*. Namun dalam penelitian ini belum dikaji mengenai pendugaan dari parameter dispersinya.

## DAFTAR PUSTAKA

- Badan Pusat Statistik. 2007. *Survei Demografi dan Kesehatan Indonesia*. Jakarta: BPS.
- Badan Pusat Statistik 2010. *Survei Sosial Ekonomi Nasional*. Jakarta: BPS.
- Badan Perencanaan Pembangunan Nasional. 2009. *Kajian Evaluasi Pembangunan Sektoral: Faktor-faktor yang Mempengaruhi Kelangsungan Hidup Anak*. Jakarta: Kedeputian Evaluasi Kinerja Pembangunan, Bappenas.
- Breslow NE dan Clayton DG. 1993. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9-25.
- Brooks SP dan Gelman A. 1998. Alternative Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, **7**, 434-455.
- Browne WJ dan Draper D. 2006. A Comparison of Bayesian and Likelihood-based Methods for Fitting Multilevel Models. *Bayesian Analysis*, **1**, 473-514.
- Carlin, BP dan Gelfand AE. 1991. An iterative Monte Carlo method for nonconjugate Bayesian analysis. *Statistics and Computing*, **1**(2), 119-128.
- Carlin BP dan Louis TA. 2000. *Bayes and empirical bayes methods for data analysis*. London: Chapman and Hall.
- Casella G dan George EI. 1992. Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-174.
- Christiansen CL dan Morris CN. 1997. Hierarchical Poisson regression modeling. *Journal of the American Statistical Association*, **92**, 618-632.
- Clayton D dan Kaldor J. 1987. Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*, **43**, 671-681.



- Congdon PD. 2010. *Applied Bayesian hierarchical methods*. London: Chapman and Hall.
- Daniels MJ dan Gatsonis C. 1999. Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American Statistical Association*, **94**, 29-42.
- Daniels MJ dan Kass RE. 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254-1263.
- Datta GS dan Sweeting TJ. 2005. Probability Matching Priors dalam Sen PK dan Rao CR (Eds.), *Handbook of Statistics*, Volume 25, Amsterdam: Elsevier Science B.V., pp. 93-114.
- Datta GS, Ghosh M, dan Waller LA. 2000. Hierarchical and Empirical Bayes Methods for Environmental Risk Assessment. dalam Sen PK dan Rao CR (Eds.), *Handbook of Statistics*, Volume 18, Amsterdam: Elsevier Science B.V., pp. 223-245.
- Datta GS, Lahiri P, dan Lu KL. 1999. Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S. *Journal of the American Statistical Association*, **94**, 1074-1082.
- Dempster AP dan Tomberlin TJ. 1980. The analysis of census undercount from a post-enumeration surveys. *Proceedings of the Conference on Census Undercount*, 88-94.
- Departemen Kesehatan. 2008. *Profil Kesehatan Indonesia 2008*. Jakarta: Depkes.
- Farrel PJ, MacGibbon B, dan Tomberlin TJ. 1997. Empirical Bayes estimators of small area proportion in multistage designs. *Statistica Sinica*, **7**, 1065-1083.
- Farrel PJ, MacGibbon B, dan Tomberlin TJ. 1997. Empirical Bayes Small Area Estimation Using Logistic Regression Models and Summary Statistics. *Journal of Business & Economic Statistics*, **15**, 101-108.
- Fay RE dan Herriot RA. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269- 277.
- Gelfand AE dan Ghosh S. 1998. Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**, 1-11.
- Gelfand AE dan Smith AFM. 1990. Sample-based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 972-985.
- Gelman A dan Rubin DB. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**, 457-472.
- Gelman A, Carlin JB, Stern HS, dan Rubin DB. 2004. *Bayesian data analysis*. Second Edition. London: Chapman and Hall.
- Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515-533.
- Geman S dan Geman D. 1984. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- Geweke J. 1992. Evaluating the accuracy of sampling-based approach to the calculation of posterior moments. Dalam Bernardo JM, Smith AFM, David AP, dan Berger JO (eds.), *Bayesian Statistics*. New York: Oxford University Press.
- Ghosh M dan Maiti T. 1999. Adjusted Bayes Estimators With Applications To Small Area Estimation. *Sankhya: The Indian Journal of Statistics*, **61**, 71-90.
- Ghosh M dan Maiti T. 2008. Empirical Bayes Confidence Intervals for Means of Natural Exponential Family-Quadratic Variance Function Distributions with

- Application to Small Area Estimation. *Scandinavian Journal of Statistics*, **35**, 484-495.
- Ghosh M dan Rao JNK. 1994. Small Area Estimation: An Appraisal. *Statistical Science*, **9**, 55-93.
- Ghosh M, Kim D, Sinha K, Maiti T, Katzoff M, dan Parsons VL. 2009. Hierarchical and empirical Bayes small domain estimation of the proportion of persons without health insurance for minority subpopulations. *Survei Methodology*, **35**, 53-66.
- Ghosh M, Maiti T, dan Roy A. 2008. Influence Functions and robust Bayes and Empirical Bayes Small Area Estimation. *Biometrika*, **95**, 573-585.
- Ghosh M, Natarajan K, Stroud TWF, dan Carlin BP. 1998. Generalized Linear Models for Small Area Estimation. *Journal of the American Statistical Association*, **93**, 273-282.
- Gilks W, Richardson S, dan Spiegelhalter D. (1996). Introducing Markov Chain Monte Carlo. In W Gilks, S Richardson, D Spiegelhalter (eds), *Markov chain Monte Carlo in practice*. London: Chapman and Hall.
- Hajarisman N, Hasanudin N, dan Padmadisastra S. (2011) *Pertimbangan Penting Yang Mendasari Penggunaan Metode Small Area Estimation*. Prosiding Seminar Nasional Statistika 2011, Volume 2 November 2011, ISSN: 2087-5290.
- Hajarisman N, Komarudin AK, dan Iswani AA. (2012) *Pemodelan Area Kecil untuk Data Cacahan dalam Pendugaan Angka Mortalitas melalui Pendekatan Model Bayes Berhierarchy (Studi Kasus Pendugaan Angka Mortalitas Bayi Level Kecamatan di Provinsi Jawa Barat)*. Laporan Penelitian Hibah Unggulan Perguruan Tinggi. Bandung: Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Islam Bandung.
- Hoberg JP dan Casella G. (1996) The effect of improper priors on Gibbs penarikan contoh in hierarchical linear mixed models. *Journal of the American Statistical Association*, **91**, 1461-1473.
- Jiang J dan Lahiri P. 2001. Empirical Best Prediction For Small Area Inference With Binary Data. *Ann. Inst. Statist. Math*, **53**, 217-243.
- Kismiantini, Kurnia A, dan Notodiputro KA. 2006. Risk of Dengue Haemorrhagic Fever in Bekasi Municipality with Small Area Approach. *Proceeding at The First International Conference on Mathematics and Statistics (ICoMS-1)*, June 19-21, 2006. Bandung.
- Kurnia A and Notodiputro KA. 2005. Aplikasi Metode Bayes pada Small Area Estimation. *Prosiding Seminar Nasional Statistika VII*, 26 November 2005. ITS Surabaya
- Kurnia A and Notodiputro KA. 2006. EB-EBLUP MSE Estimator on Small Area Estimation with Application to BPS Data. *Proceeding at The First International Conference on Mathematics and Statistics (ICoMS-1)*, June 19-21, 2006. Bandung.
- Kurnia A dan Notodiputro KA. 2008. Generalized Additive Mixed Models for Small Area Estimation. *Mathematics Journal Universiti Teknologi Malaysia*, p: 385.
- Kurnia A. dan Notodiputro KA. 2007. Pendekatan Generalized Additive Mixed Models dalam Pendugaan Parameter pada Kasus Small Area Estimation. *Jurnal Sains MIPA (ISSN 1978-1873)* 13(3), 145-151.
- Lawson AB. 2009. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. London: Chapman and Hall.
- MacGibbon B dan Tomberlin TJ. 1989. Small Area Estimation of Proportions Via Empirical Bayes Techniques. *Survei Methodology*, **15**, 237-252.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

- MacNab YC, Farrell PJ, Gustafon P, dan Wen, S. 2004. Estimation in Bayesian Disease Mapping. *Biometrics*, **60**, 865-873
- Malec D, Sedransk J, Moriarity CL, dan LeClere FB. 1997. Small Area Inference for Binary Variables in The National Health Interview Survei. *Journal of the American Statistical Association*, **92**, 815-826.
- McCullagh P dan Nelder JA. 1983. *Generalized Linear Models*. Second Edition. New York: Chapman and Hall.
- Munnich R, Burgard JP, dan Vogt M. 2009. Small area estimation for population counts in the German Census 2011. *Section on Survei Research Methods*, 181-190
- Nandram B dan Choi JW. 2002. Hierarchical Bayesian Nonresponse Models for a Binary Data From Small Areas With Uncertainty About Ignorability. *Journal of the American Statistical Association*, **97**, 381-388.
- Nandram B dan Sedransk J. 1993. Bayesian predictive inference for a finite population proportion: two-stage cluster penarikan contoh. *Journal of the Royal Statistical Society. Serie B*, **55**, 399-408.
- Nandram B, Sendransk J, dan Pickle L. 1999. Bayesian Analysis of Mortality Rates For U.S. Health Service Areas. *Sankhya: The Indian Journal of Statistics*, **61**, 146-165.
- Nandram B, Sendransk J, dan Pickle LW. 2000. Bayesian Analysis and Mapping of Mortality Rates for Chronic Obstructive Pulmonary Disease. *Journal of the American Statistical Association*, **95**, 1110-1118.
- Nandram B. 2000. Bayesian generalized linear models for inference about small area. In: Dey DK, Ghosh SK, dan Mallick BK (Eds.) *Generalized Linear Models: A Bayesian Perspective*. New York: Marcel Dekker, pp. 91-114.
- Natarajan R dan Kass RE. 2000. Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, **95**, 227-237.
- Natarajan R dan McCulloch CE. 1995. A Note on the Existence of the Posterior Distribution for a Class of Mixed Models for Binomial Responses. *Biometrika*, **82**, 639-643.
- Ntzoufras I. 2009. *Bayesian modeling using WinBUGS*. New York: John Wiley and Sons.
- Rao JNK. 2003a. *Small Area Estimation*. New York: Wiley.
- Rao JNK. 2003b. Some New Developments in Small Area Estimation. *Proceedings of the Survei Methods Section, SSC Annual Meeting, June 2003*.
- Robert CP dan Casella G. 2004. *Monte Carlo Statistical Methods*. Second Edition. New York: Springer-Verlag.
- Rumiati AT (2012) *Model Bayes untuk Pendugaan Area Kecil dengan Penarikan Contoh Berpeluang Tidak Sama pada Kasus Respons Binomial dan Multinomial (Aplikasi: Pendugaan Indeks Pendidikan Level Kecamatan di Jawa Timur)*. Disertasi tidak dipublikasikan. Program Pascasarjana, Program Studi Statistika, Institut Pertanian Bogor.
- Sadik K dan Notodiputro KA. 2007. Model State Space pada GLMM untuk Pendugaan Area Kecil (Small Area Estimation). *Prosiding pada Seminar Nasional Statistika: 24 Mei 2007*, Unisba – Bandung.
- Sadik K dan Notodiputro KA. 2009. Hierarchical Bayes Estimation Using Time Series and Cross-sectional Data: A Case of Per-capita Expenditure in Indonesia. *Conference of Small Area Estimation*, 29 Juni – 01 Juli 2009, Elche, Spanyol
- Souza DF, Moura FAS, dan Migon HS. 2009. Small area population prediction via hierarchical models. *Survei Methodology*, **35**, 203-214.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Spiegelhalter DJ, Best NG, Carlin BP, dan Van der Linde A. 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B*, **64**(4), 583–616.

Stroud TWF. 1991. Hierarchical Bayes predictive means and variance with application to sample survei inference. *Communication in Statistics, Part A-Theory and Methods*, **20**, 13-36.

Stroud TWF. 1994. Bayesian inference from categorical survei data. *Canadian Journal of Statistics*, **22**, 33-45.

Torabi M dan Rao JNK. 2008. Small area estimation under a two-level model. *Survei Methodology*, **34**, 11-17.

Torabi M, Datta GS, dan Rao JNK. 2009. Empirical Bayes Estimation of Small Area Means under a Nested Error Linear Regression Model with Measurement Errors in the Covariates. *Scandinavian Journal of Statistics*, **36**, 355-368.

Trevisani M dan Torelli N. 2007. Hierarchical Bayesian models for small area estimation with count data. *Working Paper: Dipartimento di Scienze Economiche e Statistiche*, Università degli Studi di Trieste, Trieste, Italy.

Waller L, Carlin B, Xia H, dan Gelfand A. 1997. Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, **92**, 607-617.

Yan G dan Sedransk J. 2010. A note on Bayesian residuals as a hierarchical model diagnostic technique. *Statistical Papers*, **51**(1), 1-10.

You Y dan Chapman B. 2006. Small Area Estimation Using Area Level Models and Estimated Penarikan contoh Variances. *Survei Methodology*, **32**, 97-103.

You Y dan Rao JNK. 2000. Small area estimation using unmatched penarikan contoh and linking models. *Proceedings of the Survei Method Section*, 191-196.

You Y dan Zhou QM. 2011. Hierarchical Bayes small area estimation under a spatial model with application to health survei data. *Survei Methodology*, **37**, 25-37.

## LAMPIRAN

### Lampiran 1. Sifat-sifat Sebaran Gamma dan Invers Gamma

#### Sebaran Gamma

Misalkan diketahui dua buah parameter  $\alpha$  dan  $\beta$ , dimana  $\alpha$  merupakan parameter bentuk dan  $\beta$  merupakan parameter skala. Fungsi kepekatan peluang untuk peubah acak  $X$  yang mengikuti sebaran gamma atau  $X \sim \text{Gamma}(\alpha, \beta)$  diberikan oleh

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \text{ untuk } x \geq 0.$$

dimana  $\Gamma(\alpha)$  merupakan fungsi gamma yang didefinisikan oleh

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Apabila kedua sisi persamaan di atas dibagi dengan  $\Gamma(\alpha)$  maka diperoleh

$$\int_0^\infty \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy = 1$$

Sedangkan fungsi sebaran kumulatifnya diberikan oleh

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang memurnikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$F(x; \alpha, \beta) = \frac{\Gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

### A. Moment

Moment ke- $n$  untuk sebaran gamma ini diberikan oleh

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+n)-1} \exp(-\beta x) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\beta^{\alpha+n}} \int_0^\infty \frac{\beta^{\alpha+n}}{\Gamma(\alpha+n)} x^{(\alpha+n)-1} \exp(-\beta x) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\beta^{\alpha+n}} = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\beta^n} = \frac{(\alpha+n-1)\Gamma(\alpha+n-1)}{\Gamma(\alpha)\beta^n} \\ &= \frac{(\alpha+n-1)(\alpha+n-2) \dots \alpha \Gamma(\alpha)}{\Gamma(\alpha)\beta^n} = \frac{(\alpha+n-1)(\alpha+n-2) \dots \alpha}{\beta^n} \end{aligned}$$

### B. Rata-rata dan Ragam

Rata-rata dan ragam untuk peubah acak yang bersebaran gamma masing-masing diberikan oleh

$$E(X) = \frac{\alpha}{\beta}$$

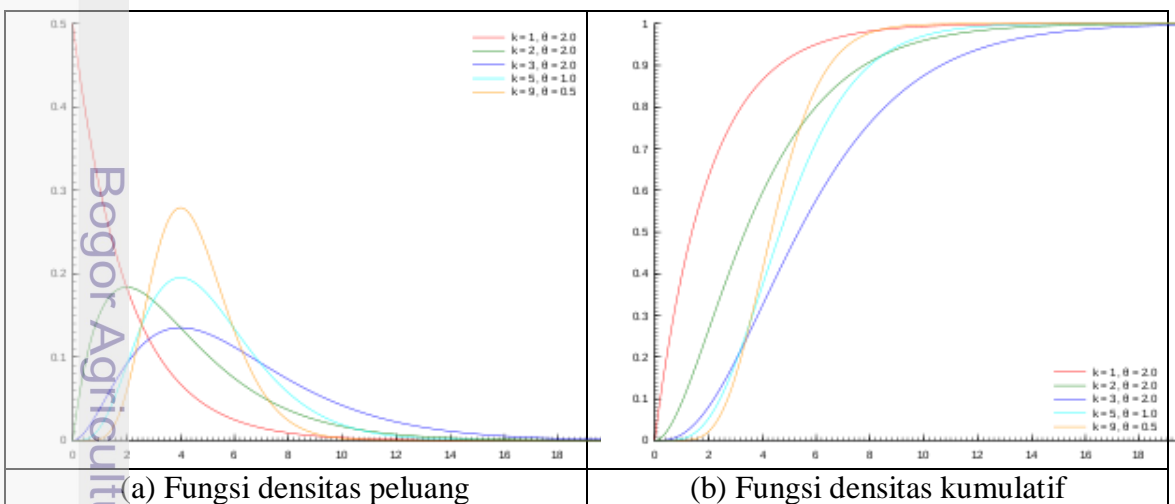
dan

$$E(X^2) = \frac{(\alpha+1)\alpha}{\beta^2}$$

Sehingga diperoleh:

$$Var(X) = E(X^2) - (E(X))^2 = \frac{(\alpha+1)\alpha}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}$$

Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatifnya disajikan dalam gambar berikut:



Gambar L1. Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatif sebaran gamma.



### Sebaran Inverse Gamma

Sebaran inverse gamma merupakan keluarga dari sebaran peluang kontinu dengan dua-parameter, yang merupakan sebaran dari kebalikan peubah acak yang bersebaran gamma. Pada umumnya, penggunaan sebaran inverse gamma dalam statistik Bayes, dimana sebaran ini merupakan *conjugate* prior bagi ragam dari sebaran normal.

Fungsi kepekatan peluang untuk sebaran inverse gamma diberikan oleh:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x), \text{ untuk } x > 0.$$

Parameter  $\alpha$  adalah parameter bentuk dan  $\beta$  adalah parameter skala. Fungsi sebaran kumulatifnya diberikan oleh

$$F(x; \alpha, \beta) = \frac{\Gamma(\alpha, \beta/x)}{\Gamma(\alpha)}$$

#### A. Penurunan Sebaran Invers Gamma dari Sebaran Gamma

Diberikan fungsi kepekatan peluang untuk peubah acak  $x$  yang bersebaran gamma. Kemudian, didefinisikan transformasi  $Y = g(X) = \frac{1}{x}$ , sehingga menghasilkan

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{1}{\theta^k \Gamma(\alpha)} \left(\frac{1}{y}\right)^{k-1} \exp\left(\frac{-1}{\theta y}\right) \frac{1}{y^2} \\ &= \frac{1}{\theta^k \Gamma(\alpha)} \left(\frac{1}{y}\right)^{k+1} \exp\left(\frac{-1}{\theta y}\right) \\ &= \frac{1}{\theta^k \Gamma(\alpha)} y^{-k-1} \exp\left(\frac{-1}{\theta y}\right) \end{aligned}$$

Dengan menggantikan  $k$  oleh  $\alpha$ ,  $\theta^1$  oleh  $\beta$ , serta  $y$  oleh  $x$ , maka diperoleh

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x), \text{ untuk } x > 0.$$

#### B. Perbandingan dengan Sebaran Gamma

Berdasarkan parameterisasi di atas, maka jika  $X$  adalah peubah acak yang bersebaran  $\text{gamma}(\alpha, \beta)$ , maka  $Y = 1/X$  akan mempunyai sebaran inverse gamma atau  $\text{IG}(\alpha, 1/\beta)$ . Hubungan ini dapat ditunjukkan sebagai berikut:

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{1}{y}\right) \left| \frac{d}{dy} (y^{-1}) \right| \\ &= \frac{1}{\Gamma(\alpha) \beta^\alpha} y^{-\alpha+1} \exp\left(-\frac{1}{\beta y}\right) y^{-2} \\ &= \frac{(1/\beta)^\alpha}{\Gamma(\alpha)} y^{-\alpha+1} \exp\left(\frac{-(1/\beta)}{y}\right) \end{aligned}$$

Selanjutnya moment ke- $n$  untuk sebaran inverse gamma ini diberikan oleh

$$\begin{aligned} E(X^n) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^n x^{-\alpha-1} \exp(-\beta/x) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{n-\alpha-1} \exp(-\beta/x) dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-n)}{\beta^{\alpha-n}} \\ &= \frac{\beta^n \Gamma(\alpha-n)}{(\alpha-1) \dots (\alpha-n) \Gamma(\alpha-n)} = \frac{\beta^n}{(\alpha-1) \dots (\alpha-n)} \end{aligned}$$

Untuk  $n = 1$  dan  $\alpha > 1$ , maka diperoleh nilai harapan untuk  $X$

Hak Cipta Dilindungi Undang-Undang  
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.  
 2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$E(X) = \frac{\beta}{(\alpha - 1)}$$

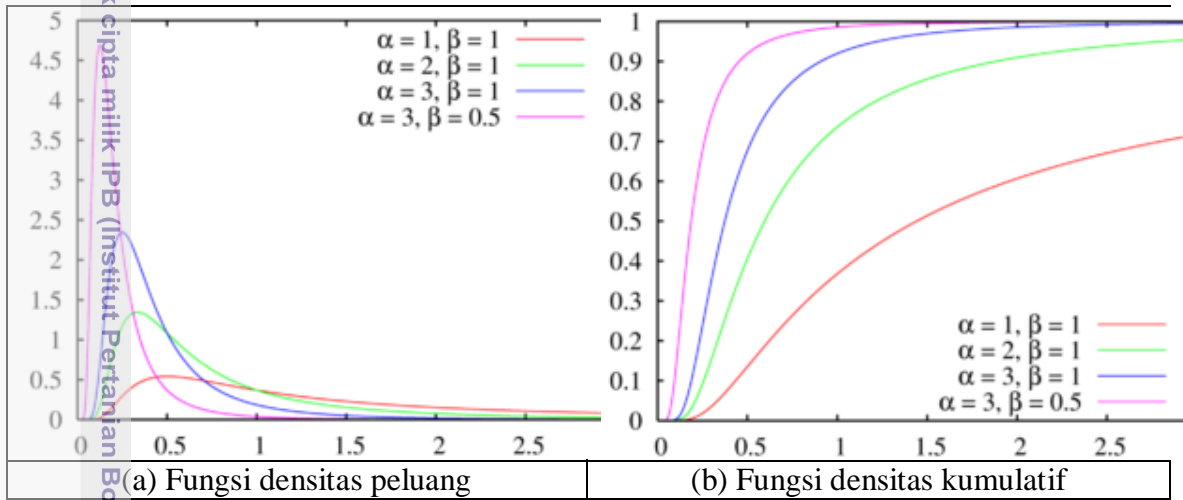
Untuk  $n = 2$  dan  $\alpha > 2$ , maka diperoleh

$$E(X^2) = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)}$$

Sehingga diperoleh ragam untuk  $X$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$$

Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatifnya disajikan dalam gambar berikut:



Gambar L2. Grafik fungsi kepekatan peluang dan fungsi sebaran kumulatif sebaran invers gamma.

## Lampiran 2

### Program Makro SAS untuk Pemodelan Poisson Bayes Berhierarchy

```

data simu;
b0=0.5;
b1=1.5;
b2=2.0;
al = 0;
be = 1;
do i 1 to 10;
x1 = rand("uniform");
x = rand("uniform");
x2 = al + (be-al)*x;
miu = exp(b0+(b1*x1)+(b2*x2));
z = rand("poisson",miu);
max = 20000;
m = ceil( max*x );
output;
end;
run;
data baru;

```

```
nobs = _n_;
set simu;
keep nobs m z x1 x2;
run;

proc print data=baru;
run;

ods graphics on;
proc mcmc data=baru outpost=postout1 seed=248601 nbi=100000 nmc=10000
monitor=(beta0 beta1 beta2) dic;
array ltheta[10];
parms beta0 0.5 beta1 1.5 beta2 2.0;
parms s2 1;
parms (ltheta:) 1;
prior beta: ~ normal(0,var=1000);
prior s2 ~ lognormal(0.01, sd=0.01);
w = beta0 + beta1*x1 + beta2*x2;
if nobs eq 1 then lp = 0;
lp = lp + lpdfnorm(ltheta[nobs], w, sqrt(s2));
hyperprior ltheta: ~ gamma(0.002, s=2);
theta = exp(ltheta[nobs]);
model z ~ poisson(theta);
mu = exp(w + s2/2);
sigma2 = mu + (exp(s2) - 1)*(mu**2);
if nobs eq 1 then Pearson = 0;
Pearson = Pearson + ((z - mu)**2/(sigma2));
preddest outpred=pout1;
run;
ods graphics off;

data dugaan1;
set pout1;
mean = mean(of z:);
sd = std(of z:);
max = max(of z:);
min = min(of z:);
run;

proc means data=baru noprint;
var z;
output out=stat1 mean=mean max=max min=min stddev=sd;
run;

data null_;
set stat1;
call symputx('mean',mean);
call symputx('sd',sd);
call symputx('min',min);
call symputx('max',max);
run;

data null_;
set dugaan1 end=eof nobs=nobs;
ctmean + (mean>&mean);
ctmin + (min>&min);
ctmax + (max>&max);
ctsd + (sd>&sd);
if eof then do;
```

Hak Cipta Dilindungi Undang-Undang

1. Diararang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diararang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

1. Diizinkan mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diizinkan mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

```
pmean = ctmean/nobs; call symputx('pmean',pmean);
pmin = ctmin/nobs; call symputx('pmin',pmin);
pmax = ctmax/nobs; call symputx('pmax',pmax);
psd = ctsd/nobs; call symputx('psd',psd);
end;
run;

proc template;
define statgraph twobytwo;
begingraph;
layout lattice / rows=1 columns=2;
layout overlay / yaxisopts=(display=none)
xaxisopts=(label="mean");
layout gridded / columns=2 border=false
autoalign=(topleft topright);
entry halign=right "p-value =";
entry halign=left eval(strip(put(&pmean, 12.2)));
endlayout;
histogram mean / binaxis=false;
lineparm x=&mean y=0 slope=. /
lineattrs=(color=red thickness=5);
endlayout;
layout overlay / yaxisopts=(display=none)
xaxisopts=(label="sd");
layout gridded / columns=2 border=false
autoalign=(topleft topright);
entry halign=right "p-value =";
entry halign=left eval(strip(put(&psd, 12.2)));
endlayout;
histogram sd / binaxis=false;
lineparm x=&sd y=0 slope=. /
lineattrs=(color=red thickness=5);
endlayout;
endlayout;
endgraph;
end;
run;

ods graphics on;
proc sgrender data=dugaan1 template=twobytwo;
run;
ods graphics off;

ods graphics on;
proc ncmc data=baru outpost=postout seed=248601 nbi=10000 nmc=10000
monitor=(beta0 beta1 beta2) dic;
array ltheta[10];
parms beta0 0.5 beta1 1.5 beta2 2.0;
parms s2 1;
parms \ltheta: 1;
prior beta: ~ normal(0,var=1000);
prior s2 ~ lognormal(0.01, sd=0.01);
w = beta0 + beta1*x1 + beta2*x2;
if nobs eq 1 then lp = 0;
lp = lp + lpdfnorm(ltheta[nobs], w, sqrt(s2));
hyperprior ltheta: ~ igamma(0.002, s=0.002);
theta = exp(ltheta[nobs]);
model z ~ poisson(theta);
mu = exp(w + s2/2);
```

```

sigma2 = mu + (exp(s2) - 1) * (mu**2);
if nobs eq 1 then Pearson = 0;
Pearson = Pearson + ((z - mu)**2/(sigma2));
preddist outpred=pout2;
run;
ods graphics off;

data dugaan2;
set pout2;
mean = mean(of z:);
sd = std(of z:);
max = max(of z:);
min = min(of z:);
run;

proc means data=baru noprint;
var z;
output out=stat2 mean=mean max=max min=min stddev=sd;
run;

data null_;
set stat2;
call symputx('mean',mean);
call symputx('sd',sd);
call symputx('min',min);
call symputx('max',max);
run;

data null_;
set dugaan2 end=eof nobs=nobs;
ctmean + (mean>&mean);
ctmin + (min>&min);
ctmax + (max>&max);
ctsd + (sd>&sd);
if eof then do;
pmean = ctmean/nobs; call symputx('pmean',pmean);
pmin = ctmin/nobs; call symputx('pmin',pmin);
pmax = ctmax/nobs; call symputx('pmax',pmax);
psd = ctsd/nobs; call symputx('psd',psd);
end;
run;

proc template;
define statgraph twobytwo;
begingraph;
layout lattice / rows=1 columns=2;
layout overlay / yaxisopts=(display=none)
xaxisopts=(label="mean");
layout gridded / columns=2 border=false
autoalign=(topleft topright);
entry halign=right "p-value =";
entry halign=left eval(strip(put(&pmean, 12.2)));
endlayout;
histogram mean / binaxis=false;
lineparm x=&mean y=0 slope=. /
lineattrs=(color=red thickness=5);
endlayout;
layout overlay / yaxisopts=(display=none)
xaxisopts=(label="sd");

```

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

```
layout gridded / columns=2 border=false
                    autoalign=(topleft topright);
                    entry halign=right "p-value =";
                    entry halign=left eval(strip(put(&psd, 12.2)));
endlayout;
histogram sd / binaxis=false;
lineparm x=&sd y=0 slope=. /
                    lineattrs=(color=red thickness=5);
endlayout;
endlayout;
endgraph;
end;
run;
ods graphics on;
proc sgrender data=dugaan2 template=twobytwo;
run;
ods graphics off;
```

### Lampiran 3

#### Program Makro SAS untuk Perhitungan Statistik Gelman-Rubin

```
data init1;
    input Chain beta0 beta1 beta2;
    datalines;
    1 -0.5 -1.5 -2
    2 0.5 1.5 2
    3 0 0 0
;

/* define constants */
%let nchain = 3;
%let nparm = 3;
%let nsim = 10000;
%let var = beta0 beta1 beta2;

%macro gcmc;
    %do i=1 %to &nchain;
        data _null_;
            set init1;
            if Chain=&i;
            %do j = 1 %to &nparm;
                call symputx("init1&j", %scan(&var, &j));
            %end;
            stop;
        %end;
    run;

ods graphics on;
proc mcmc data=baru outpost=postout1&i seed=248601 nbi=100000
nmc=&nsim
monitor=(beta0 beta1 beta2) dic;
array ltheta[10];
parms beta0 &init11 beta1 &init12 beta2 &init13;
parms s2 1;
```

```

parms (ltheta:) 1;
prior beta: ~ normal(0,var=1000);
prior s2 ~ lognormal(0.01, sd=0.01);
w = beta0 + beta1*x1 + beta2*x2;
if nobss eq 1 then lp = 0;
lp = lp + lpdfnorm(ltheta[nobs], w, sqrt(s2));
hyperprior ltheta: ~ gamma(0.002, s=2);
theta = exp(ltheta[nobs]);
model z ~ poisson(theta);
lambda = z/m;
mu = exp(w + s2/2);
sigma2 = mu + (exp(s2) - 1)*(mu**2);
if nobss eq 1 then Pearson = 0;
Pearson = Pearson + ((z - mu)**2/(sigma2));
preddist outpred=pout1;
run;
ods graphics off;

%end;
%mend;

ods listing close;
%gcmc;
ods listing;

data all;
  se postout11(in=in1) postout12(in=in2) postout13(in=in3);
  if in1 then Chain=1;
  if in2 then Chain=2;
  if in3 then Chain=3;
run;

%gelman(all, &nparm, &var, &nsim);

data GelmanRubin(label='Gelman-Rubin Diagnostics');
  merge _Gelman_Parms _Gelman_Ests;
run;

proc print data=GelmanRubin;
run;

/* plot the trace plots of three Markov chains. */
%macro trace;
  %do i = 1 %to &nparm;
    proc sgplot data=all cycleattrs;
      series x=Iteration y=%scan(&var, &i) / group=Chain;
    run;
  %end;
%mend;
%trace;

/* define sliding window size */
%let nwin = 200;
data PSRF;
run;

%macro PSRF(nsim);
  %do k = 1 %to %sysevalf(&nsim/&nwin, floor);
    %gelman(all, &nparm, &var, nsim=%sysevalf(&k*&nwin));
  %end;

```

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

```
data GelmanRubin;
    merge _Gelman_Parms _Gelman_Ests;
run;

data PSRF;
    set PSRF GelmanRubin;
run;
%end;
%mend PSRF;

options nonotes;
%PSRF(&nsim);
options notes;

data PSRF;
    set PSRF;
    if _n_ = 1 then delete;
run;

proc sort data=PSRF;
    by Parameter;
run;

%macro sepPSRF(nparm=, var=, nsim=);
    %do k = 1 %to &nparm;
        data save&k; set PSRF;
            if n > %sysevalf(&k*&nsim/&nwin, floor) then delete;
            if _n_ < %sysevalf((&k-1)*&nsim/&nwin + 1, floor) then
                delete;
            Iteration + &nwin;
        run;

        proc sgplot data=save&k(firstobs=10) cycleattrs;
            series x=Iteration y=Estimate;
            series x=Iteration y=upperbound;
            yaxis label="%scan(&var, &k)";
        run;
    %end;
%mend sepPSRF;

%sepPSRF(nparm=&nparm, var=&var, nsim=&nsim);

data init2;
    input Chain beta0 beta1 beta2;
    datalines;
1 -0.5 -1.5 -2
2 0.5 1.5 2
3 0 0 0
;

/* define constants */
%let nchain = 3;
%let nparm = 3;
%let nsim = 10000;
%let var = beta0 beta1 beta2;

%macro gmcmc;
    %do i=1 %to &nchain;
        data _null_;
```



Hak Cipta Dilindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

```

set init2;
if Chain=&i;
%do j = 1 %to &nparm;
    call symputx("init2&j", %scan(&var, &j));
%end;
stop;
run;

ods graphics on;
proc mcmc data=baru outpost=postout2&i seed=248601 nbi=100000
nmc=&nsim
monitor=(beta0 beta1 beta2) dic;
array ltheta[10];
parms beta0 &init21 beta1 &init122 beta2 &init23;
parms s2 1;
parms (ltheta:) 1;
prior beta: ~ normal(0, var=1000);
prior s2 ~ lognormal(0.01, sd=0.01);
w = beta0 + beta1*x1 + beta2*x2;
if nobs eq 1 then lp = 0;
lp = lp + lpdfnorm(ltheta[nobs], w, sqrt(s2));
hyperprior ltheta: ~ igamma(0.002, s=0.002);
theta= exp(ltheta[nobs]);
model z ~ poisson(theta);
lambda = z/m;
mu = exp(w + s2/2);
sigma = mu + (exp(s2) - 1)*(mu**2);
if nobs eq 1 then Pearson = 0;
Pearson = Pearson + ((z - mu)**2/(sigma2));
predist outpred=pout2;
run;
ods graphics off;

%end;
%mend;

ods listing close;
%gcmc;
ods listing;

data all;
    set postout21(in=in1) postout22(in=in2) postout23(in=in3);
    if in1 then Chain=1;
    if in2 then Chain=2;
    if in3 then Chain=3;
run;

%gelman(all, &nparm, &var, &nsim);

data GelmanRubin(label='Gelman-Rubin Diagnostics');
    merge _Gelman_Parms _Gelman_Ests;
run;

proc print data=GelmanRubin;
run;

/* plot the trace plots of three Markov chains. */
%macro trace;
%do i = 1 %to &nparm;

```



### Hak Cipta Dilindungi Undang-Undang

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Diarangi mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

```
proc sgplot data=all cycleattrs;
    series x=Iteration y=%scan(&var, &i) / group=Chain;
run;

%end;
%mend;
%trace;

/* define sliding window size */
%let nwin = 200;
data PSRF;
run;

%macro PSRF(nsim);
    %do k = 1 %to %sysevalf(&nsim/&nwin, floor);
        %gelman(all, &nparm, &var, nsim=%sysevalf(&k*&nwin));
        data GelmanRubin;
            merge _Gelman_Parms _Gelman_Ests;
        run;

        data PSRF;
            set PSRF GelmanRubin;
        run;
    %end;
%mend PSRF;

options nonotes;
%PSRF(&nsim);
options notes;

data PSRF;
    set PSRF;
    if _n_ = 1 then delete;
run;

proc sort data=PSRF;
    by Parameter;
run;

%macro sepPSRF(nparm=, var=, nsim=);
    %do k = 1 %to &nparm;
        data save&k; set PSRF;
            if n > %sysevalf(&k*&nsim/&nwin, floor) then delete;
            if _n_ < %sysevalf((&k-1)*&nsim/&nwin + 1, floor) then
delete;
                Iteration + &nwin;
            run;

proc sgplot data=save&k(firstobs=10) cycleattrs;
    series x=Iteration y=Estimate;
    series x=Iteration y=upperbound;
    yaxis label="%scan(&var, &k)";
run;
    %end;
%mend sepPSRF;

%sepPSRF(nparm=&nparm, var=&var, nsim=&nsim);
```

© Hak cipta milik IPB (Institut Pertanian Bogor) Bogor Agricultural University

## RIWAYAT HIDUP

Penulis dilahirkan di Bandung, Jawa Barat pada tanggal 4 April 1969 dari pasangan Bapak R. Djayusman Basar (alm) dan Ibu Yetty Rochayati. Penulis merupakan anak pertama dari empat bersaudara. Pada tahun 2008 Penulis menikah dengan Dewi Rosiana dan dikaruniai seorang putri bernama Syahmina Arkani Raihana. Pendidikan Sarjana ditempuh di Jurusan Statistika Universitas Islam Bandung pada tahun 1987-1993. Pada tahun 1998 Penulis memperoleh gelar Magister Sains dari Program Studi Statistika Sekolah Pascasarjana IPB dengan pembimbing thesis Dr. Ir. Asep Saefuddin, MSc, Prof. Dr. Ir. Khairil Anwar Notodiputro, dan Dr. Ir. Budi Susetyo, MSc. Sejak tahun 2008 Penulis menempuh Program Doktor pada Program Studi Statistika Sekolah Pascasarjana Institut Pertanian Bogor.

Selama mengikuti pendidikan Program Doktor, Penulis telah menghasilkan beberapa karya ilmiah yang telah dipublikasikan dalam seminar nasional dan jurnal ilmiah. Karya-karya ilmiah tersebut merupakan bagian dari Program Doktor Penulis. Karya-karya ilmiah tersebut adalah:

2009	Pemeriksaan Ketepatan Fungsi Hubungan dalam Analisis Data Biner	Jurnal Statistika: Forum Teori dan Aplikasi Statistika Volume 9 Nomor 1 Mei 2009, ISSN: 1411-5891
2009	Pemeriksaan Data Berpengaruh dalam Model Gamma	Jurnal Ilmiah Mat Stat Volume 10 No. 1 Januari 2010, ISSN: 1412-1220)
2010	Pendekatan Fungsi Quasi-Likelihood dan Implementasinya dalam Sistem SAS	Jurnal Statistika: Forum Teori dan Aplikasi Volume 10 Nomor 1 Mei 2010, ISSN: 1411-5891
2011	Algoritma Pendugaan Model Regresi Kekar melalui Penduga-M	Jurnal Ilmiah Mat Stat Volume 11 No. 1 Januari 2011, ISSN: 1412-1220
2011	Pertimbangan Penting yang Mendasari Penggunaan Metode Small Area Estimation	Prosiding Seminar Nasional Statistika 2011, Volume 2 November 2011, ISSN: 2087-5290
2011	Seleksi Model dan Validasi Silang dalam Model Efek Utama Aditif dan Interaksi Multiplikatif	Jurnal Statistika: Forum Teori dan Aplikasi Volume 11 Nomor 2 November 2011, ISSN: 1411-5891
2012	Pendugaan Parameter Model Regresi Beta untuk Memodelkan Data Proporsi	Jurnal Statistika: Forum Teori dan Aplikasi Volume 12 Nomor 1 Mei 2012, ISSN: 1411-5891
2012	Implementasi Pendugaan Area Kecil melalui Model Poisson Bayes Berhierarchy untuk Menduga Angka Kematian Bayi Level Kecamatan Provinsi Jawa Barat	Hibah Penelitian Unggulan Perguruan Tinggi (DIKTI)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.