ISBN: 978-979-792-552-9

Aplikasi Bayesian Tanpa Informasi Distribusi Prior dalam Mengestimasi Parameter Distribusi Weibull

Rado Yendra, Ari Pani Despina, Rahmadeni

Department of Mathematics, Faculty of Science and Technology
Islamic State University of Riau, 28293, Pekanbaru, Riau, Indonesia

Abstrak

Estimasi parameter suatu distribusi biasanya dihasilkan dengan menggunakan metode maksimum *likelihood*. Metoda ini sangat bergantung pada proses numerik yang cukup rumit. Metoda Bayesian tanpa informasi pada distribusi prior telah dipergunakan dalam penelitian ini untuk mengestimasi parameter distribusi weibull. Data waktu survival untuk 30 pasien penyakit kanker dan Distribusi Normal~N (0.4,100) dan N(0.1,10) akan digunakan sebagai distribusi prior pada penelitian ini. Distribusi posterior yang dibangkitkan sebanyak 30000 kali iterasi telah berhasil mengestimasi dua parameter distribusi Weibull. Disribusi posterior yang dihasilkan telah konvergen pada iterasi yang ke 500, hal ini telah membuktikan bahwa metode ini sangat mudah untuk digunakan dalam mengestimasi parameter suatu distribusi tertentu.

Kata kunci: bayesian tanpa informasi prior, distribusi posterior, distribusi weibull

1 Pendahuluan

Pemodelan suatu statistik secara garis besarnya dapat dibagi dua, yaitu mendapatkan hubungan antara variabel bebas dan tidak bebas (analisa regresi) dan menentukan besar peluang terjadinya suatu peristiwa tertentu (distribusi peluang). Penentuan distribusi peluang terbaik dalam suatu peristiwa tertentu didahului oleh mendapatkan parameter-parameter distribusi peluang tersebut. Teknik Momen, L Momen, dan Maksimum *Likelihood* adalah sebagian teknik yang sering digunakan untuk mengestimasi parameter. Teknik di atas dikenal sangat bergantung kepada suatu proses numerik yang cukup rumit. Pada dewasa ini Metoda Bayesian telah menarik minat peneliti lain untuk menggunakannya dalam mengestimasi parameter [1]. Metoda Metoda Bayesian tidak memerlukan suatu proses numerik untuk tujuan ini. Perbandingan teknik estimasi parameter terhadap teknik maksimum *likelihood* dan Metoda Bayesian [2] telah membuktikan bahwa Metoda Bayesian sangat mudah untuk dilakukan.

Pada penelitian ini Metoda Bayesian MCMC (*Markov Chain Monte Carlo*) tanpa informasi untuk distribusi prior akan digunakan untuk mengestimasi parameter Distribusi Weibull dua parameter. Oleh sebab itu distribusi prior diberikan dengan mengasumsikan parameter distribusi Weibull terdistribusi secara normal dengan rata-rata yang kecil dan variasi yang besar. Distribusi Normal dengan parameter seperti di atas akan menjamin bahwa distribusi posterior akan cepat konvergen kepada suatu nilai tertentu [3].

Distribusi Weibull sangat baik digunakan dalam menganalisis data survival dan kegagalan suatu proses peramalan terhadap suatu penyakit [4]. Diantara peneliti yang menggunakan Metoda Bayesian untuk mengestimasi parameter Weibull [5] dimana penelitian yang dilakukan menggunakan dengan data sensor untuk memodelkan suatu fungsi bahaya.

2 Distribusi Normal dan Weibull

Dua distribusi peluang yaitu Distribusi Normal dan Weibull sangat diperlukan dalam melakukan estimasi parameter dengan menggunakan Metoda Bayesian. Distribusi Normal akan digunakan sebagai distribusi prior dan Distribusi Weibull akan diperlukan sebagai fungsi *likelihood*. Untuk itu fungsi densitas peluang distribusi Normal dan Weibull secara berturut-turut akan diberikan seperti pada dua persamaan berikut

$$f\left(y;\mu,\sigma^2\right)=\left\{\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-\left(y-\mu\right)^2}{2\sigma^2}\right]\right\}, -\infty\le y\le\infty \qquad (1)$$

$$f\left(y;\lambda,\gamma\right)=\gamma\lambda^{\gamma}y^{\gamma-1}\exp\left(-\left(\lambda y\right)^{\gamma}\right), 0\le y\le\infty \qquad (2)$$

Likelihood untuk distribusi peluang Weibull dengan fungsi densitas peluang seperti pada persamaan (2) dinotasikan sebagai $L(\lambda,\gamma|y)$, dan dihasilkan dengan menggunakan persamaan berikut

$$L(\lambda,\gamma|y)=\prod_{i=1}^{n}f(y;\lambda,\gamma) \qquad (3)$$

3 Bayesian MCMC Tanpa Informasi Distribusi Prior

Pada bagian ini akan diperkenalkan ide penggunaan Metoda Bayesian MCMC tanpa informasi distribusi prior. Andaikan distribusi prior diasumsikan terdistribusi secara normal dengan fungsi densitas peluang seperti yang diberikan pada persamaan (1) dan diberi notasi $\pi(\theta)$, sedangkan *likelihood* seperti yang ditampilkan pada persamaan (2) dan diberi notasi $L(\theta|y)$. Ide Bayesian bermula dari ditentukannya distribusi posterior dengan notasi $\pi(\theta|y)$ yang berasal dari perkalian distribusi prior dan *likelihood*, seperti yang ditampilkan pada persamaan berikut

$$\pi(\theta|y)=\frac{\pi(\theta)L(\theta|y)}{\int\pi(\theta)L(\theta|y)d\theta}. \qquad (4)$$

Persamaan (4) di atas dapat disingkat dengan

$$\pi(\theta|y) \propto \pi(\theta)L(\theta|y). \tag{5}$$

Kesulitan yang diakibatkan oleh integral dalam membangkitkan distribusi posterior akan diselesaikan dengan menggunakan teknik simulasi MCMC. Algoritma simulasi distribusi posterior MCMC untuk mengestimasi parameter distribusi Weibull pada penelitian ini adalah sebagai berikut.

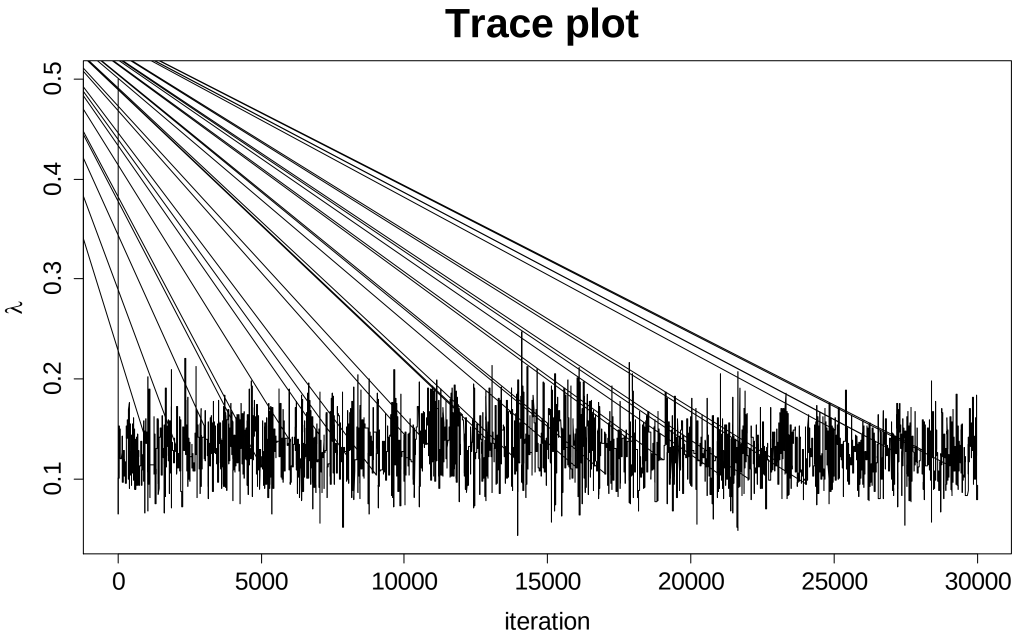
Algoritma simulasi distribusi posterior MCMC

1. Rantai pertama dibangkitkan dengan nilai awal pada parameter $\theta^0 = (\lambda^0, \gamma^0)$ untuk j = 1
2. Bangkitkan $\omega_\lambda \sim N(0{,}4)$
3. Bangkitkan rantai $\lambda^* = \lambda^{j-1} + \omega_\lambda$
4. Terima $\lambda^j = \lambda^*$ dengan peluang = min $\left(1, \frac{\pi(\lambda^*|\gamma^{j-1})}{\pi(\lambda^{j-1}|\gamma^{j-1})}\right)$ dan $\lambda^j = \lambda^{j-1}$ untuk yang lainnya
5. Bangkitkan $\omega_\gamma \sim N(0{,}0.3)$
6. Bangkitkan rantai $\gamma^* = \gamma^{j-1} + \omega_\gamma$
7. Terima $\gamma^j = \gamma^*$ dengan peluang = min $\left(1, \frac{\pi(\gamma^*|\gamma^j)}{\pi(\lambda^{j-1}|\gamma^j)}\right)$ dan $\gamma^j = \gamma^{j-1}$ untuk yang lainnya
8. Bangkitkan rantai dari j hingga j+1 dengan mengulangi tahap ke 2

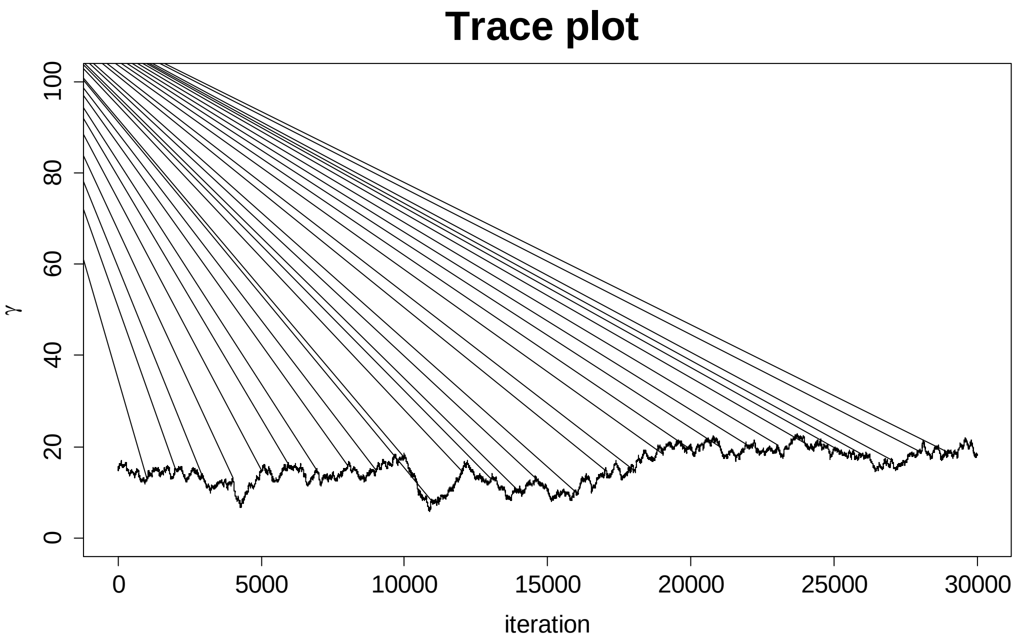
4 Hasil Penelitian

Data waktu survival untuk 30 orang penderita penyakit kanker, akan dipergunakan dalam penelitian ini. Algoritma Bayesian MCMC tanpa informasi distribusi prior, seperti yang telah dibahas pada bagian 3 akan digunakan pada bagian ini. Distribusi posterior untuk dua parameter distribusi Weibull akan dibangkitkan rantai sebanyak 30000 kali. Kekonvergenan parameter yang dibangkitkan tersebut akan dianalisa. Untuk memeriksa kekonvergenan dari rantai paremeter yang dibangkitkan tersebut akan dipergunakan dua jenis grafik, yaitu grafik trace plot dan grafik densitas posterior. Grafik trace plot akan sangat berguna dalam memeriksa kekovergenan parameter, grafik yang cendrung menuju ke suatu bilangan tertentu akan digunakan sebagai pertanda bahwa parameter yang dibangkitkan hingga ke-k rantai telah konvergen. Seterusnya untuk memastikan nilai parameter tersebut, grafik densitas posterior dapat digunakan.

Gambar 1 dan 2 secara berturut-turut ditampilkan untuk melihat kekonvergenan dua parameter distribusi weibull λ dan γ. Gambar 1 merupakan distribusi posterior untuk parameter λ yang dibangkitkan sebanyak 30000 iterasi. Dari gambar tersebut dapat dilihat bahwa nilai awal yang diberikan sebesar 0.5 telah berhasil konvergen pada bilangan tertentu diantara 0.1 hingga 0.2. Sementara untuk melihat kekonvergenan parameter γ, Gambar 2 turut ditampilkan, dari gambar tersebut juga dapat dilihat parameter tersebut cukup sulit untuk konvergen pada suatu bilangan tertentu, hal ini dapat dilihat dari fluktuasi trace plot hingga pada 30000 iterasi.



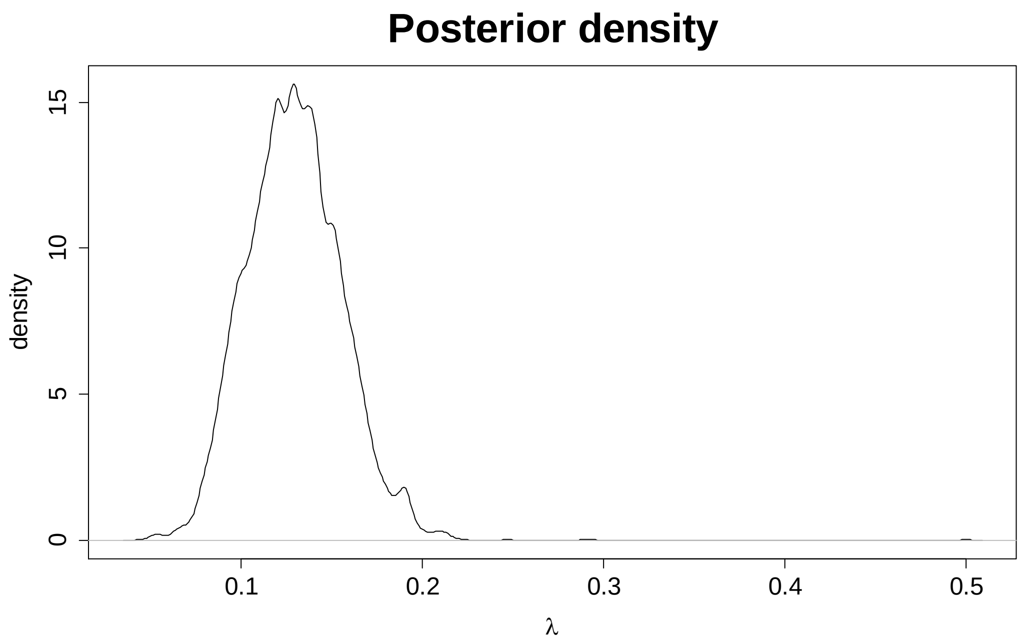
Gambar 1: Trace plot parameter λ



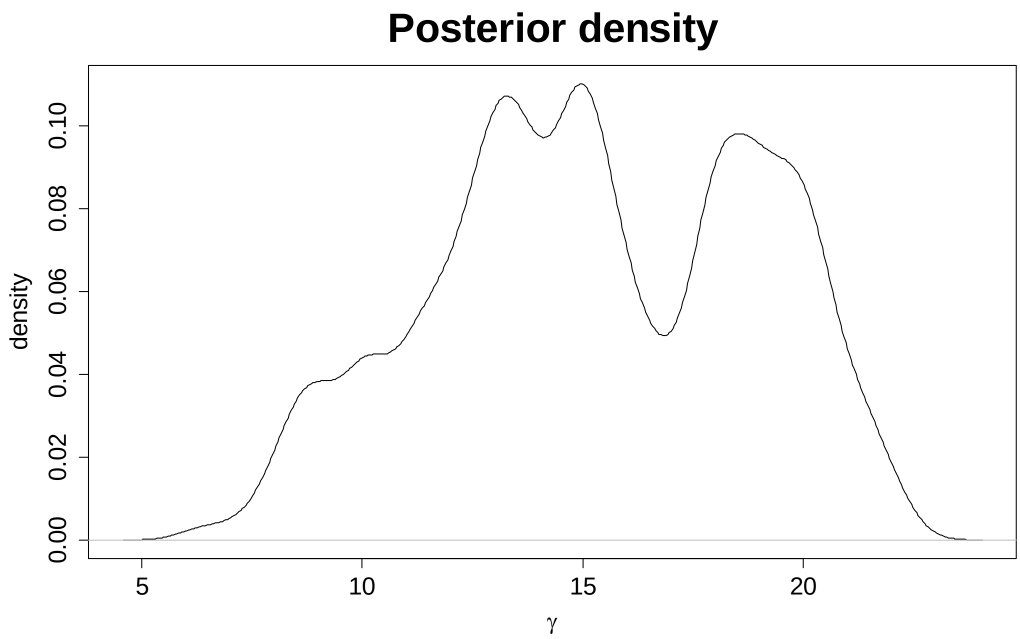
Gambar 2: Trace plot parameter γ

Untuk memastikan kedua nilai parameter distribusi ini, grafik densitas posterior turut ditampilkan sebagai berikut.

ISBN: 978-979-792-552-9



Gambar 3: Densitas posterior parameter λ



Gambar 4: Densitas posterior parameter γ

Gambar 3 dan 4 secara berturtu-turut adalah nilai densitas posterior parameter distribusi weibull λ dan γ yang dibangkitkan sebanyak 30000 iterasi. Dari Gambar 3 dapat dilihat bahwa nilai parameter λ dapat dipilih pada nilai peluang (*density*) yang tertinggi yaitu pada interval 0.12 hingga 0.15. Selanjutnya pada Gambar 4 juga dapat dilihat bahwa terdapat kemungkinan dua nilai untuk parameter γ, yaitu pada interval 12-15 dan interval 15-17. Dari gambar tersebut juga dapat dipastikan bahwa nilai peluang (*density*) yang tertinggi pada interval 15-17.

Kesimpulan

Metoda Bayesian sangat baik digunakan untuk mengestimasi parameter distribusi peluang tertentu, terutama jika digunakan untuk mengestimasi parameter suatu parameter distribusi yang cukup banyak (melebihi 3 parameter). Metoda ini dikenal cukup ampuh dalam mengestimasi parameter dengan tingkat kecepatan kovergennya suatu parameter cukup cepat dan dapat mengelakkan proses numerik yang cukup rumit.

Daftar Pustaka

[1] M. A. Al Omari and N. A. Ibrahim, "Bayesian Survival Estimation for Weibull Distribution with Censored Data," *Journal of Applied Sciences*, 11 (2011), 393–396, 2011.

[2] B. N. Pandey, N. Dwividi, and B. Pulastya, "Comparison Between Bayesian and Maximum Likelihood Estimation of the scale parameter in Weibull distribution with known shape under linex lossfunction," *Journal of Scientific Research*, 55 (2011), 163–172.

[3] A. F. M. Smith and G. O. Roberts, Bayesian Computation via Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal Statistics Society*, 55(1993), 3-23.

[4] R. B. Abernethy, *The New Weibull Handbook*, 5th edition, 2006.

[5] S. K. Sinha, "Bayes Estimation of the Reliability Function and Hazard Rate of a Weibull Failure Time Distribution," *Tranbajos de Estadistica*, 1 (1986), 47–56.