

## PENAKSIRAN PARAMETER ANALISIS REGRESI COX DAN ANALISIS SURVIVAL BAYESIAN

Rina Wijayanti

IKIP Budi Utomo Malang

rina.statistika12@gmail.com

### Abstrak

Teori estimasi/penaksiran memiliki dua pendekatan yaitu pendekatan statistika klasik dan pendekatan statistika global (*Bayesian*). Statistika klasik adalah statistika dimana tatacara pengambilan keputusannya didasarkan hanya pada data sampel yang diambil dari populasi. Sedangkan statistika *Bayesian* dalam pengambilan keputusannya didasarkan pada informasi baru dari data yang diamati (sampel) dan pengetahuan sebelumnya. Pada tulisan ini Analisis Regresi Cox akan diambil sebagai contoh penaksiran parameter dengan pendekatan statistika klasik dan Analisis Survival Bayesian sebagai contoh pendekatan statistika global (*Bayesian*). Estimasi parameter Survival Bayesial menggunakan algoritma MCMC karena model yang komplek/rumit sehingga sulit diselesaikan sedangkan model Regresi Cox menggunakan metode *partial likelihood*. Hasil yang diperoleh dari estimasi parameter tersebut tidak *close form* sehingga perlu dilakukan dengan metode iterasi Newton-Raphson.

Kata Kunci: penaksiran, klasik, bayes

### Abstract

The theory of estimation have two approaches, namely the classical statistical approach and global statistical approach (*Bayesian*). Classical statistics are statistics in which the procedure is the decision based only on the data samples taken from the population. While Bayesian statistics in making decisions based on new information from the observed data (sample) and prior knowledge. At this writing Cox Regression Analysis will be taken as an example of parameter estimation by the classical statistical approach Survival Analysis and Bayesian statistical approach as an example of global (*Bayesian*). Survival Bayesial parameter estimation using MCMC algorithms for model complex / complicated and difficult to resolve while the Cox regression models using the method of partial likelihood. Results of the parameter estimates do not close form that needs to be done by the method of Newton-Raphson iteration.

Key words: estimation, classic, Bayes

## PENDAHULUAN

Teori estimasi/penaksiran memiliki dua pendekatan yaitu pendekatan statistika klasik dan pendekatan statistika global (*Bayesian*). Statistika klasik adalah statistika dimana tatacara pengambilan keputusannya didasarkan hanya pada data sampel yang diambil dari populasi. Sedangkan statistika *Bayesian* dalam pengambilan keputusannya didasarkan pada informasi baru dari data yang diamati (sampel) dan pengetahuan sebelumnya. Inferensi statistik dengan pendekatan statistika *Bayesian* berbeda dengan pendekatan statistika klasik. Pendekatan statistika klasik memandang parameter  $\beta$  sebagai parameter bernilai tetap. Sedangkan pendekatan statistika *Bayesian* memandang parameter  $\beta$  sebagai variabel random yang memiliki distribusi, disebut distribusi *prior*. Dari distribusi *prior* selanjutnya dapat ditentukan distribusi posterior sehingga diperoleh estimator *Bayesian* yang merupakan mean atau modus dari distribusi posterior. Pada penelitian ini Analisis Regresi Cox akan diambil sebagai contoh penaksiran parameter dengan pendekatan statistika klasik dan Analisis Survival Bayesian sebagai contoh pendekatan statistika global (*Bayesian*). Pemilihan kajian penaksiran parameter ini karena

dua metode tersebut digunakan dalam *Anilisis Survival*.

Regresi Cox pertama kali dikenalkan oleh Cox dan respon yang digunakan adalah data yang diperoleh dari perhitungan waktu suatu peristiwa tertentu (waktu *survival*). Misalnya data tentang waktu pasien menderita penyakit tertentu, dimana perhitungannya dimulai dari awal sakit hingga terjadi pada kejadian khusus, yaitu seperti kematian, sembuh atau kejadian khusus lainnya. Pada Regresi Cox tidak ada asumsi bahwa harus berdistribusi tertentu, distribusi yang digunakan adalah sesuai dengan respon yang digunakan. Estimasi parameter pada regresi cox menggunakan *partial likelihood*.

Persamaan dari Regresi Cox Proportional Hazard adalah sebagai berikut:

$$h(t) = h_0(t) \exp(\beta'X) \quad (1)$$

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

Estimasi parameter pada regresi cox menggunakan *partial likelihood*. Apabila diketahui banyak sampel adalah  $n$ , dan diantara  $n$  sampel tersebut terdapat individu/objek yang mengalami *failure event* sebanyak  $k$ , kemudian dari  $k$  sampel diurutkan dari waktu yang terkecil sampai yang terbesar  $t(1) < t(2) < \dots < t(i)$  maka fungsi *partial likelihood* diberikan sebagai berikut [4]

$$L(\beta) = \prod_{i=1}^n \frac{\exp(\beta' X_i)}{\sum_{l \in R(t_i)} \exp(\beta' X_l)} \quad (2)$$

Dengan  $x_{(t)}$  adalah vektor variabel prediktor dari objek yang mengalami *failure event* pada saat ke- $i$  dengan urutan waktu  $t(l)$

Menurut Kleinbaum asumsi dari proportional hazard adalah variabel prediktor independen terhadap waktu dan hubungan antara kumulatif hazard konstan pada setiap waktu [5]. Asumsi proportional hazard tersebut didasarkan pada sebagai berikut:

$$S(t, X) = [S_0(t)] \exp\left(\sum_{j=1}^p \beta_j x_j\right) \quad (3)$$

$$\ln[S(t, X)] = \exp\left(\sum_{j=1}^p \beta_j x_j\right) \ln[S_0(t)]$$

Keterangan:

$h_0(t)$  = fungsi *bazeline hazard*

$S(t, X)$  = fungsi probabilitas survival yang bernilai

$$0 \leq S(t, X) \leq 1$$

Apabila data observasi dinyatakan sebagai  $x$  sedangkan parameter data dinyatakan sebagai  $\beta$ . Distribusi  $\beta$  dengan syarat  $x$  diberikan melalui teorema bayes sebagai berikut,

$$p(\beta | x) = \frac{l(x | \beta) p(\beta)}{p(x)} \quad (4)$$

Disebut juga metode yang melakukan *update* informasi prior parameter  $\beta$  ( $p(\beta)$ ) dari data sebelum pengamatan dilakukan menggunakan informasi sampel dalam likelihood data  $l(x | \beta)$  untuk mendapatkan

informasi posterior  $p(\beta | x)$  yang digunakan dalam pengambilan keputusan.  $p(x)$  merupakan *normalized constant*. Distribusi posterior merupakan likelihood dari distribusi prior, sehingga dapat dituliskan sebagai

$$p(\beta | x) \propto l(x | \beta) p(\beta) \quad (5)$$

Distribusi prior merupakan informasi awal yang dibutuhkan dalam membentuk distribusi posterior. Selain itu, juga dibutuhkan informasi dari sampel yang dinyatakan melalui *likelihood*.

Pada model yang kompleks, distribusi posterior terlalu sulit untuk memecahkan permasalahan, untuk memperbaharui (*update*) parameter dari sampel melalui distribusi posterior. Dengan Markov Chain Monte Carlo (MCMC) dapat menyelesaikan hal tersebut. MCMC dapat digunakan untuk membentuk model yang sangat kompleks, berdimensi tinggi atau sifat data yang berkorelasi tinggi. Ide dasar dari MCMC adalah membangkitkan sebuah Markov Chain dengan simulasi Monte Carlo yang beriterasi, sehingga didapatkan distribusi posterior yang stasioner [6].

Gibbs sampling adalah algoritma MCMC yang mencakup iterative sampling dari tiap distribusi bersyarat, dimana parameter  $\beta$  dipartisi menjadi beberapa bagian,

$\beta = (\beta_1, \beta_2, \dots, \beta_p)$  dan distribusi posterior full conditional adalah  $p(\beta_1 | x, \beta_2, \dots, \beta_p), \dots, p(\beta_p | x, \beta_1, \dots, \beta_{p-1})$  atau secara sederhana dapat ditulis  $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(p)}$  [7]. Gibbs sampling bekerja dengan langkah-langkah sebagai berikut.

1. Mengambil nilai  $m=0$  dan menentukan nilai inisial (*initial value*) dari

$$\beta^{(0)} = \{\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)}\}$$

2. Membangkitkan tiap komponen dari

$$\beta^{(m+1)} = \{\beta_1^{(m+1)}, \beta_2^{(m+1)}, \dots, \beta_p^{(m+1)}\}$$

Dimana

Nilai  $\beta_1^{(m+1)}$  berasal dari

$$p(\beta_1 | x, \beta_2^{(m)}, \dots, \beta_p^{(m)})$$

Nilai  $\beta_2^{(m+1)}$  berasal dari

$$p(\beta_2 | x, \beta_1^{(m+1)}, \beta_3^{(m)}, \dots, \beta_p^{(m)})$$

⋮

Nilai  $\beta_p^{(m+1)}$  berasal dari

$$p(\beta_p | x, \beta_1^{(m+1)}, \beta_2^{(m+1)}, \dots, \beta_{p-1}^{(m+1)})$$

3. Mengambil nilai  $m_1=m+1, m_2=m+2, \dots, m_x=m+x$  dan ulangi langkah 1 dan 2
4. Anggap  $\{\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(x)}\}$  sebagai sampel untuk analisis posterior
5. Membuat plot distribusi posterior
6. Mendapatkan mean, median, deviasi standar dari distribusi posterior

## METODE PENELITIAN

1. Langkah-langkah mengkaji penaksiran parameter regresi cox

- a. Menentukan fungsi *likelihood*, yaitu

$$L(\beta) = \prod_{i=1}^n \frac{\exp(\beta' X_i)}{\sum_{l \in R(t_i)} \exp(\beta' X_l)}$$

(6)

- b. Mencari turunan pertama dan kedua dari fungsi *ln-likelihood*
- c. Jika parameter yang diperoleh tidak *close form* atau implisit, maka digunakan metode iterasi numerik Newton-Raphson [8].

2. Langkah-langkah mengkaji penaksir parameter dari Analisis Survival Bayesian

- a. Menentukan/menduga fungsi kepadatan peluang dari waktu ketahanan hidup penderita kanker serviks. Waktu ketahanan hidup penderita kanker serviks diduga mengikuti Distribusi Weibull dengan fungsi kepadatan peluang sebagai berikut

$$f(t | \lambda, k) = \frac{k}{\lambda} \left( \frac{t}{\lambda} \right)^{k-1} \exp \left[ - \left( \frac{t}{\lambda} \right)^k \right]$$

(7)

dimana  $t \geq 0, \lambda > 0, k > 0$ .  $\lambda$  merupakan parameter skala dan  $k$  adalah parameter

bentuk dan  $\lambda, k$  dianggap sebagai variabel random.

- b. Menentukan fungsi likelihood dari waktu survival/ketahanan hidup penderita kanker serviks. Distribusi dari waktu ketahanan hidup pasien kanker serviks diduga mengikuti Distribusi Weibull 2 parameter, dengan  $k$  adalah parameter bentuk dan  $\lambda$  parameter skala. Maka fungsi likelihood adalah

$$L(t_1, t_2, \dots, t_n | \lambda, k) = \prod_{i=1}^n f(t_i | \lambda, k) \tag{8}$$

- c. Menentukan distribusi prior
- d. Mencari fungsi posterior dengan MCMC dan Gibbs sampling
- e. Membentuk fungsi survival dari hasil taksiran parameter

**HASIL DAN PEMBAHASAN**

**1. Penaksiran Parameter Regresi Cox**

Estimasi parameter pada regresi cox menggunakan *partial likelihood*. Apabila data terdiri dari variabel independen  $X_1, X_2, \dots, X_p$  dan terdiri atas  $n$  pengamatan waktu *survival* tidak tersensor, terdapat individu yang mengalami *failure event* sebanyak  $k$ , kemudian dari  $k$  sampel diurutkan dari waktu yang terkecil sampai yang terbesar  $t(1) < t(2) < \dots < t(i)$ . Maka

fungsi *likelihood* diberikan sebagai berikut [4]

$$L(\beta) = \prod_{i=1}^n \frac{\exp(\beta' X_i)}{\sum_{l \in R(t_i)} \exp(\beta' X_l)} \tag{9}$$

$$\begin{aligned} \ln L(\beta) &= \sum_{i=1}^n (\beta' X_i) - \sum_{i=1}^n \ln \left[ \sum_{l \in R(t_i)} \exp(\beta' X_l) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^p (\beta_j x_{ji}) - \sum_{i=1}^n \ln \left[ \sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl}) \right] \\ &= \sum_{i=1}^n \left( \sum_{j=1}^p (\beta_j x_{ji}) - \ln \left[ \sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl}) \right] \right) \end{aligned} \tag{10}$$

Dari persamaan (9) maka dapat dicari turunan pertamanya. Hasil turunan pertama dari fungsi *ln-likelihood* dapat dilihat dan dijabarkan lebih lanjut seperti persamaan (10).

$$\begin{aligned} \frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_1} &= \sum_{i=1}^n \left[ x_{1i} - \frac{\sum_{l \in R(t_i)} x_{1l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right] \\ \frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_2} &= \sum_{i=1}^n \left[ x_{2i} - \frac{\sum_{l \in R(t_i)} x_{2l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right] \\ &\vdots \\ \frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_p} &= \sum_{i=1}^n \left[ x_{pi} - \frac{\sum_{l \in R(t_i)} x_{pl} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right] \end{aligned} \tag{11}$$

Turunan kedua dari fungsi *ln-likelihood* berasal dari persamaan (11), dimana persamaan tersebut berbentuk  $\frac{u}{v}$  sehingga turunan

kedua dicari menggunakan rumus

$$\frac{u'}{v'} = \frac{u'v - v'u}{v^2} \quad \text{Hasil turunan}$$

pertama dari fungsi *ln-likelihood* adalah persamaan (12)

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta_k \partial \beta_{k^*}} = - \sum_{i=1}^n \left\{ \frac{\sum_{l \in R(t_i)} x_{kl} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} x_{k^*l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right\}$$

$$= - \sum_{i=1}^n \frac{\sum_{l \in R(t_i)} x_{kl} x_{k^*l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})}$$

$$+ \sum_{i=1}^n \left( \frac{\sum_{l \in R(t_i)} x_{kl} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} x_{k^*l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right) \mathbf{H}(\beta) =$$

(12)

dengan  $j, j^* = 1, 2, \dots, p$

Karena estimasi parameter yang diperoleh tidak *close form* atau implisit, maka digunakan metode iterasi numerik Newton-Raphson [8]. Jika  $\mathbf{g}(\beta)$  adalah vektor berukuran  $p \times 1$  yang merupakan turunan pertama fungsi  $\ln L(\beta)$  terhadap parameter  $\beta$ .  $\mathbf{H}(\beta)$  adalah matrik hessian berukuran  $p \times p$  yang berisi turunan kedua dari fungsi *likelihood* yang ditransformasikan ke *ln-likelihood*, maka estimasi parameter pada

iterasi ke  $(n+1)$  adalah sebagai berikut.

$$\beta^{(n+1)} = \beta^{(n)} - H^{-1}(\beta^{(n)}) \mathbf{g}(\beta^{(n)}) \tag{13}$$

Iterasi akan berhenti jika,  $\|\beta^{(n+1)} - \beta^{(n)}\| \leq \varepsilon$ , dimana  $\varepsilon$  adalah suatu bilangan yang sangat kecil dab  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$

$$\mathbf{g}(\beta) = \begin{bmatrix} \frac{\partial \ln L(\beta)}{\partial \beta_1} \\ \frac{\partial \ln L(\beta)}{\partial \beta_2} \\ \vdots \\ \frac{\partial \ln L(\beta)}{\partial \beta_p} \end{bmatrix}$$

$$\mathbf{H}(\beta) = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta_1^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_p} \\ \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_2^2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \text{simetris} & \dots & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_p^2} \end{bmatrix}$$

Penelitian ini menggunakan data tersensor kanan, sehingga fungsi *likelihood* dinyatakan sebagai berikut

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp(\beta' \mathbf{X}_i)}{\sum_{l \in R(t_i)} \exp(\beta' \mathbf{X}_l)} \right]^{\delta_m}$$

(14)

Dimana  $\delta_m \begin{cases} 0 = \text{tidak tersensor} \\ 1 = \text{tersensor} \end{cases}$

$$\ln L(\beta) = \sum_{i=1}^n \delta_m \left\{ \left( \beta' \mathbf{X}_j - \ln \left[ \sum_{l \in R(t_i)} \exp(\beta' \mathbf{X}_l) \right] \right) \right\}$$

(15)

Langkah-langkah estimasi parameter data tersensor dan tidak tersensor sama. Sehingga diperoleh turunan pertama dan kedua dalam persamaan (14) dan (15)

$$\frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_p} = \sum_{i=1}^n \delta_m \left[ x_{pi} - \frac{\sum_{l \in R(t_i)} x_{pl} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right]$$

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta_k \partial \beta_{k^*}} = \sum_{i=1}^n \delta_m \left[ - \frac{\sum_{l \in R(t_i)} x_{kl} x_{k^*l} \exp \sum_{j=1}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} + \frac{\sum_{l \in R(t_i)} x_{kl} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} \beta_{k^*l} \exp \sum_{i=j}^p (\beta_j x_{jl})}{\sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl}) \sum_{l \in R(t_i)} \exp \sum_{j=1}^p (\beta_j x_{jl})} \right]$$

(17)

Karena estimasi parameter yang diperoleh tidak *close form* atau implisit, maka juga digunakan metode iterasi numerik Newton-Raphson.

$$\beta^{(n+1)} = \beta^{(n)} - H^{-1}(\beta^{(n)})g(\beta^{(n)})$$

(18)

Iterasi akan berhenti jika,  $\|\beta^{(n+1)} - \beta^{(n)}\| \leq \varepsilon$ , dimana  $\varepsilon$  adalah suatu bilangan yang sangat kecil dan  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$

$$g(\beta) = \begin{bmatrix} \frac{\partial \ln L(\beta)}{\partial \beta_1} \\ \frac{\partial \ln L(\beta)}{\partial \beta_2} \\ \vdots \\ \frac{\partial \ln L(\beta)}{\partial \beta_p} \end{bmatrix}$$

$$H(\beta) = \begin{bmatrix} \frac{\partial^2 \ln L(\beta)}{\partial \beta_1^2} & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_1 \partial \beta_p} \\ \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_2^2} & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \text{simetris} & \dots & \dots & \frac{\partial^2 \ln L(\beta)}{\partial \beta_p^2} \end{bmatrix}$$

(16)

## 2. Pengujian Parameter Model pada Regresi Cox

Uji signifikansi parameter model pada regresi cox adalah sebagai berikut.

### a) Uji Serentak

Pengujian ini digunakan untuk mengetahui apakah variabel prediktor yang digunakan pada model berpengaruh signifikan secara bersama-sama.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

H1: paling sedikit ada satu  $\beta_k \neq 0$  dengan  $k=1,2,\dots,p$

Statistik Uji:

$$G^2 = -2 \ln \left( \frac{\max_{\omega} \ln L(\omega)}{\max_{\Omega} \ln L(\Omega)} \right) = -2 \ln \left( \frac{L(\hat{\omega})}{L(\hat{\Omega})} \right)$$

dengan

$$L(\hat{\Omega}) = \prod_{i=1}^n \left[ \frac{\exp(\beta' X_i)}{\sum_{l \in R(t_i)} \exp(\beta' X_l)} \right]^{\delta_m}$$

$L(\hat{\omega}) = 1$  (karena tidak mengandung variabel)

dimana  $L(\hat{\Omega}) =$  nilai likelihood dengan semua variabel bebas

$L(\hat{\omega}) =$  nilai likelihood variabel bebas tereduksi

Daerah penolakan: tolak  $H_0$  jika nilai statistik uji  $G^2 > \chi^2_{p,\alpha}$

b) Uji Parsial

Uji Parsial dilakukan untuk mengetahui variabel mana yang berpengaruh signifikan terhadap model .

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

Statistik Uji :

$$W_k = \left[ \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right]^2$$

Keterangan :  $\hat{\beta}_k$  adalah penduga  $\beta_k$

$SE(\hat{\beta}_k)$  adalah *standard error* dari  $\beta_k$

Daerah penolakan :

Tolak  $H_0$  jika  $W_k > \chi^2_{\alpha,1}$

**3. Estimasi Parameter Model Analisis Survival Bayesian Menggunakan MCMC dan Gibbs Sampling**

Estimasi parameter model survival dalam penelitian ini mengkaji estimasi parameter analisis survival bayesian menggunakan MCMC dan *Gibbs sampling*. Langkah-langkah estimasi parameter model pada

analisis survival bayesian adalah sebagai berikut:

1. Menentukan fungsi probabilitas dari waktu ketahanan hidup penderita kanker serviks.

Misalkan waktu ketahanan hidup penderita kanker serviks mengikuti distribusi Weibull dua parameter, maka bentuk distribusinya adalah sebagai berikut:

$$f(t|\lambda, k) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{t}{\lambda}\right)^k\right]$$

(19)

dimana  $t \geq 0, \lambda > 0, k > 0$ .  $\lambda$  merupakan parameter skala dan  $k$  adalah parameter bentuk dari  $\lambda$ , dan  $k$  dianggap sebagai variabel random.

2. Menentukan fungsi likelihood dari waktu survival/ketahanan hidup penderita kanker serviks

$$L(t_1, t_2, \dots, t_n | \lambda, k) = \prod_{i=1}^n f(t_i | \lambda, k)$$

$$= \prod_{i=1}^n \left[ \frac{k}{\lambda} \left(\frac{t_i}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{t_i}{\lambda}\right)^k\right] \right]$$

$$= \left(\frac{k}{\lambda}\right)^n \left(\frac{\sum_{i=1}^n t_i}{\lambda^n}\right)^{k-1} \exp\left[-\left(\frac{\sum_{i=1}^n t_i}{\lambda}\right)^k\right]$$

(20)



3. Menentukan distribusi prior

Penelitian ini menggunakan distribusi prior distribusi normal ( $g(\mu, \sigma)$ ), sehingga fungsi padat probabilitas bersama yang diperoleh adalah

$$H(t_1, t_2, \dots, t_n, \lambda, k, \mu, \sigma) = \prod_{i=1}^n f(t_i | \lambda, k) g(\mu, \sigma) = L(t_1, t_2, \dots, t_n | \lambda, k) g(\mu, \sigma)$$

$$= \left(\frac{k}{\lambda}\right)^n \left(\frac{\sum_{i=1}^n t_i}{\lambda^n}\right)^{k-1} \exp\left[-\left(\frac{\sum_{i=1}^n t_i}{\lambda}\right)^k\right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\lambda - \mu}{\sigma}\right)^2\right)$$

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{k - \mu}{\sigma}\right)^2\right)$$

$$= \left(\frac{k}{\lambda}\right)^n \left(\frac{\sum_{i=1}^n t_i}{\lambda^n}\right)^{k-1} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2 \exp\left[-\left(\frac{\sum_{i=1}^n t_i}{\lambda}\right)^k - \left(\frac{\lambda - \mu}{2\sigma}\right)^2 - \left(\frac{k - \mu}{2\sigma}\right)^2\right]$$

(21)

Sehingga distribusi posterior

$$f(\lambda, k, \mu, \sigma | t_1, t_2, \dots, t_n) = \frac{L(t_1, t_2, \dots, t_n | \lambda, k) g(\mu, \sigma)}{\iint_{\Omega} H(t_1, t_2, \dots, t_n, \lambda, k, \mu, \sigma) dk d\lambda d\mu d\sigma}$$

(22)

Persamaan (22) di atas kompleks sehingga sulit untuk memecahkan permasalahan, untuk memperbaharui update parameter dari sampel melalui distribusi posterior selanjutnya menggunakan *Markov Chain Monte Carlo* (MCMC).

4. Mencari fungsi posterior dengan MCMC dan Gibbs sampling

*Gibbs sampling* adalah algoritma MCMC yang mencakup iterative sampling

dari tiap distribusi bersyarat, dimana parameter  $\beta$  dipartisi menjadi beberapa bagian,  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  dan distribusi posterior full conditional adalah  $p(\beta_1 | x, \beta_2, \dots, \beta_p), \dots, p(\beta_p | x, \beta_1, \dots, \beta_{p-1})$  atau secara sederhana dapat ditulis  $\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(p)}$

(Congdon, 2003). Gibbs sampling bekerja dengan langkah-langkah sebagai berikut.

a. Mengambil nilai  $m=0$  dan menentukan nilai inisial (*initial value*) dari

$$\beta^{(0)} = \{\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)}\}$$

b. Membangkitkan tiap komponen dari

$$\beta^{(m+1)} = \{\beta_1^{(m+1)}, \beta_2^{(m+1)}, \dots, \beta_p^{(m+1)}\}$$

Dimana

Nilai  $\beta_1^{(m+1)}$  berasal dari

$$p(\beta_1 | x, \beta_2^{(m)}, \dots, \beta_p^{(m)})$$

Nilai  $\beta_2^{(m+1)}$  berasal dari

$$p(\beta_2 | x, \beta_2^{(m+1)}, \beta_3^{(m)}, \dots, \beta_p^{(m)})$$

⋮

Nilai  $\beta_p^{(m+1)}$  berasal dari

$$p(\beta_p | x, \beta_1^{(m+1)}, \beta_2^{(m+1)}, \dots, \beta_{p-1}^{(m+1)})$$

c. Mengambil nilai  $m_1=m + 1, m_2=m + 2, \dots, m_x=m+x$  dan ulangi langkah 1 dan 2

- d. Anggap  $\{\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(x)}\}$  sebagai sampel untuk analisis posterior
- e. Mendapatkan mean, median, deviasi standar dari distribusi posterior.
5. Membentuk fungsi survival dari hasil taksiran parameter
- $$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

dimana nilai  $\beta_1, \beta_2, \dots, \beta_p$  diperoleh dari mean posterior proses MCMC yang telah konvergen. Kejadian konvergen ketika nilai  $\beta_1, \beta_2, \dots, \beta_p$  stabil berada disekitar nilai tertentu saja.

#### 4. Selang Kepercayaan Analisis Survival Bayesian

Selang kepercayaan Bayes  $(1-\alpha)100\%$  untuk  $\beta$  Analisis Bayesian dapat diperoleh dengan menghitung selang yang berpusat pada *mean posterior*, dengan menghitung selang [9]

$$\hat{\beta} - z_{\frac{\alpha}{2}} SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + z_{\frac{\alpha}{2}} SE(\hat{\beta}) \quad (23)$$

dimana  $SE(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})}$

yang mengandung  $(1-\alpha)100\%$  peluang posteriornya.

#### KESIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari penelitian ini adalah estimasi parameter Survival Bayesial

menggunakan algoritma MCMC karena model yang kompleks/rumit sehingga sulit diselesaikan sedangkan model Regresi Cox menggunakan metode *partial likelihood*. Hasil yang diperoleh dari estimasi parameter tersebut tidak *close form* sehingga perlu dilakukan dengan metode iterasi Newton-Raphson

#### DAFTAR PUSTAKA

- Wong, M.C. M., Lam, K. F., & Lo, E. C. M. 2005. Bayesian analysis of clustered interval-censored data. *Journal of Dental Research*, 84(9), 817-821.
- Cox, D. R. 1972. *Regression Model and Life Table*. J Roy Stat Soc B, 34, 187-202.
- Ahmed, F. E., Vos, P. W., dan Holbert, D. 2007. Modeling Survival in Colon Cancer : A Methodological Review. *Molecular Cancer*, 6, 15.
- Omurlu, I.K., Ozdamar, K., Ture, M. 2009. Comparison of Bayesian survival analysis and Cox Regression Analysis in simulated and breast cancer data sets. *Journal International of Expert Systems with Applications*, 36, 11341-11346.
- Kleinbaum, D.G, and Klein, M. 2005. *Survival Analysis : A Self-Learning Text*, Second Edition, Springer, New York.
- Sorensen, D. Dan Gianola, D. 2002. *Likelihood, Bayesian, and MCMC Methods in*

**Wijayanti, Rina**

*Penaksiran Parameter Analisis Regresi Cox dan Analisis Survival Bayesian*

*Quantitative Genetics,*  
Springer.

Congdon, P. 2003. *Applied Bayesian Modelling*, London, John Wiley & Sons, Ltd.

Collet, D. 1994. *Modelling Survival Data in Medical Reseach*. London, Chapman and Hall.

Walpole, R.E. 1995. *Pengantar Metode Statistika*. Jakarta: Gramedia Pustaka Utama.