

**WEIBULL, GAMMA ATAU EKSPONENSIAL SEBAGAI SEBARAN PELUANG MASA  
INKUBASI HIV UNTUK PENDUGAAN INFEKSI HIV DENGAN ALGORITMA EM**  
(Weibull, Gamma or Exponential as Probability Distributions of HIV Incubation Period for  
estimation of HIV infection cases with EM Algorithm)

Herlina Hanum

Jurusan Matematika FMIPA Universitas Sriwijaya

**ABSTRAK**

*This paper was summarised from two research carried out by the author. The aim was to analyze and determine the probability distribution of HIV incubation period so it can be used for estimation of HIV infection cases with EM Algorithm. The distribution was determined through the analysis of EM Algorithm resistance against the change of Weibull and Gamma parameters. Exponential distribution is a special form of Weibull and Gamma distributions. In conclusion, the best distribution of HIV incubation period to use with EM Algorithm is Exponential distribution.*

**PENDAHULUAN**

Algoritma EM (*Expectation Maximization*) sering digunakan dalam pendugaan secara Statistika. Penggunaan algoritma EM dipadukan dengan bentuk fungsi sebaran teoritis dari masalah dimana algoritma EM diterapkan. Salah satu penerapan algoritma EM adalah dalam pendugaan jumlah orang terinfeksi HIV (Bacchetti *et al.*, 1993). Pendugaan ini dilakukan berdasarkan data jumlah orang terdiagnosis menderita AIDS. Jumlah kasus AIDS diasumsikan bersebaran Poisson. Algoritma EM digunakan untuk

memaksimumkan fungsi kemungkinan (*likelihood function*) sebaran Poisson tersebut.

Faktor utama yang harus dipertimbangkan adalah masa inkubasi HIV yang diukur dari jangka waktu dari saat terjadinya infeksi sampai timbul gejala AIDS. Masa inkubasi ini sulit ditentukan karena tidak diketahui secara pasti kapan terjadinya infeksi, disamping HIV mempunyai masa inkubasi yang berbeda-beda dalam tubuh manusia, sehingga besar kemungkinan terjadi kesalahan dalam menentukan masa inkubasi. Untuk memudahkan pendugaan diasumsikan

bahwa masa inkubasi HIV mengikuti sebaran peluang tertentu sehingga dapat ditentukan peluang masa inkubasi untuk populasi terinfeksi HIV. Anderson dan Medley (1988) mengasumsikan masa inkubasi HIV mengikuti sebaran Gamma dan Weibull.

Penentuan peluang masa inkubasi menggunakan sebaran Weibull ( $\beta, \rho$ ) dipengaruhi oleh nilai  $\beta$  (Beta) dan masa inkubasi yang digunakan. Nilai  $\rho$  ditentukan oleh masa inkubasi dan  $\beta$ . Perubahan nilai Beta dan/atau masa inkubasi akan mengakibatkan perubahan nilai peluang masa inkubasi. Demikian juga bila digunakan sebaran Gamma ( $\alpha, \lambda$ ). Perubahan nilai  $\alpha$  dan/atau masa inkubasi akan mengakibatkan perubahan nilai peluang masa inkubasi. Perubahan peluang masa inkubasi akan mengakibatkan perubahan hasil pendugaan kasus infeksi HIV.

Untuk mengantisipasi perbedaan nilai dugaan yang disebabkan oleh penggunaan parameter sebaran dan/atau masa inkubasi yang berbeda, diperlukan bentuk sebaran yang tidak peka terhadap perubahan masa inkubasi. Disamping itu, diharapkan kedekatan antara nilai dugaan dengan nilai sebenarnya. Dalam tulisan ini dianalisis

bentuk sebaran peluang yang terbaik bagi masa inkubasi HIV untuk digunakan dalam pendugaan jumlah kasus infeksi HIV menggunakan Algoritma EM.

### METODE PENELITIAN

Kedua penelitian yang hasilnya dijadikan dasar kajian dalam tulisan ini adalah Hanum (1996) dan Hanum & Samsuryadi (1999). Kedua penelitian tersebut bersifat verifikasi terhadap pendugaan jumlah kasus infeksi HIV yang menggunakan, berturut-turut, sebaran Gamma dan Weibull dalam penentuan peluang masa inkubasi. Analisa dilakukan terhadap kedekatan nilai dugaan dengan nilai parameter atau jumlah kasus infeksi HIV sebenarnya. Langkah-langkah yang dilakukan dalam kedua penelitian tersebut adalah :

1. Menetapkan jumlah kasus infeksi HIV menggunakan model Eksponensial Kuadratik yaitu:  
$$\theta_t = d_0 \exp(d_1 t - d_2 t^2) \quad t = 1, 2, 3, \dots, 20$$
  
dengan  $t$  adalah waktu dalam semester. Nilai  $\theta$  dianggap sebagai nilai parameter jumlah kasus infeksi HIV.

2. Berdasarkan nilai parameter tersebut kemudian ditentukan jumlah kasus AIDS menggunakan peluang masa inkubasi dari Sebaran Weibull  $(\beta, \rho)$  dan Gamma  $(\alpha, \lambda)$ . Fungsi kepekatan peluang sebaran Weibull  $(\beta, \rho)$  adalah:

$$f(i) = \beta \rho (\rho i)^{\beta-1} \exp(-(\rho i)^\beta) \quad (i \geq 0)$$

nilai tengahnya adalah :

$$E(i) = \frac{1}{\rho} \Gamma(1 + \frac{1}{\beta}) \quad (i \geq 0)$$

Bentuk fungsi kepekatan Gamma  $(\alpha, \lambda)$  dan nilai tengahnya

$$f(t) = \lambda (\lambda t)^{\alpha-1} \exp(-\lambda t) / \Gamma(\alpha) \quad (t \geq 0)$$

dan  $E(t) = \frac{\alpha}{\lambda}$

3. Selanjutnya dilakukan proses sebaliknya yaitu pendugaan jumlah kasus infeksi HIV berdasarkan jumlah kasus AIDS tersebut dengan algoritma EM.
4. Evaluasi dilakukan terhadap kedekatan nilai  $\theta$ , dan  $\hat{\theta}$ , dan kepekaan pendugaan terhadap perubahan parameter sebaran Gamma dan Weibull dan/atau masa inkubasi

### Pendugaan dengan Algoritma EM

Algoritma EM memiliki dua tahap yaitu tahap nilai harapan dan tahap pemaksimuman. Untuk tahap nilai harapan diasumsikan ada data lengkap  $x_{ij}$  yaitu jumlah individu yang terinfeksi pada waktu ke- $t$  terdiagnosis pada waktu ke- $j$ , dimana  $t < j$ , yang memiliki sebaran Poisson dengan nilai tengah  $a_{ij} \theta_t$ . Algoritma EM dimulai dengan tahap nilai harapan yaitu menentukan nilai awal  $(\theta_j)^0$  dilanjutkan dengan penghitungan nilai harapan dari fungsi kemungkinan  $x_{ij}$  yaitu  $E[\ln L(x_{ij}) | y, (\theta_t)^0]$  dengan

$$\ln L(x_{ij}) = \sum_{i=1}^j [x_i \ln(\theta_i) - \theta_i \sum_{j=1}^n a_{ij}]$$

dengan  $x_i = \sum_{j=1}^n x_{ij}$  dan

$$E[x_i | y, (\theta_t)^0] = \sum_{j=1}^n y_j \frac{\theta_t a_{ij}}{\sum_{r=1}^j \theta_r a_{rj}}$$

(Green, 1990; Bacchetti *et al.*, 1993).

Tahap pemaksimuman digunakan untuk mendapatkan nilai  $(\theta_j)^{baru}$  dengan cara memaksimumkan  $E[\ln L(x_{ij}) | y, (\theta_t)^0]$ . Dugaan

n  $\theta_i$  didapat dengan membuat turunan pertama nilai harapan tersebut terhadap  $\theta_i$  sama dengan nol yaitu:

$$(\theta_i)^{baru} = (\theta_i)^{lama} \frac{\sum_{j=1}^n y_j a_{ij}}{\sum_{r=1}^j (\theta_r)^{lama} a_{rj}}$$

dimana  $(\theta_i)^{lama}$  adalah dugaan  $\theta_i$  saat ini dan  $(\theta_i)^{baru}$  adalah dugaan baru bagi  $\theta_i$  (Shepp dan Vardi, 1982; Green, 1990). Penentuan nilai  $\theta_i$  berikutnya dilakukan secara iteratif menggunakan persamaan (3). Iterasi berakhir setelah proses konvergen ke suatu titik, yaitu setelah tercapai

$$l_{m+1}(\theta_i) - l_m(\theta_i) \leq 10^{(-4)} l_m(\theta_i)$$

dengan  $l_m$  dan  $l_{m+1}$  adalah ln kemungkinan maksimum pada iterasi ke-m dan ke-(m+1). Dalam setiap iterasi, dugaan  $(\theta_i)^{baru}$  merupakan perbaikan dari dugaan  $(\theta_i)^{lama}$  karena  $L((\theta_i)^{baru}) \geq L((\theta_i)^{lama})$  dengan kesamaan hanya berlaku jika dan hanya jika  $L((\theta_i)^{lama}) = \max_{\theta} L(\theta)$ . Dengan demikian semua maksimum dari  $L(\theta)$  merupakan maksimum global dan iterasi persamaan (3) menghasilkan suatu barisan  $(\theta_i)^0, (\theta_i)^1, (\theta_i)^2, \dots$  dimana 0,1,2,... adalah iterasi yang dilakukan.

Barisan tersebut akan konvergen ke  $(\theta_i)^\infty$  yaitu penduga kemungkinan maksimum global (Shepp dan Vardi, 1982).

Nilai awal  $\theta$  yang digunakan dalam algoritma EM adalah nilai awal seragam yaitu 100 untuk semua. Nilai alpha dan Beta yang digunakan adalah 1 dan 3. Dengan masa inkubasi 5 dan 10 tahun. Kombinasi antara nilai Alpha atau Beta dengan masa inkubasi menghasilkan 4 bentuk sebaran Gamma ( $\alpha, \lambda$ ) yaitu Gamma (1, 0.1), Gamma (1, 0.05), Gamma (3, 0.3) dan Gamma (3, 0.15). Kombinasi tersebut juga digunakan dalam sebaran Weibull.

#### Analisis Data

Kebaikan hasil pendugaan ditentukan dengan nilai Kuadrat Tengah Galat (KTG). Hasil pendugaan disebut baik jika nilai KTG kecil. KTG dihitung dengan rumus:

$$KTG = \sum_{t=1}^N (\theta_t - \hat{\theta}_t)^2 / N \quad t = 1, 2, 3, \dots; \text{ dan}$$

$$N = 20$$

Kepekaan hasil pendugaan ditentukan melalui ukuran jarak antara nilai-nilai dugaan pada nilai Beta yang sama dengan masa inkubasi yang berbeda. Ukuran jarak tersebut dihitung dengan rumus:

$$D = \sum_{i=1}^N (\theta_i - \theta_{i^*})^2 \quad t = 1, 2, 3, \dots;$$

dan  $N=20$  dengan  $i$  dan  $i^*$  adalah masa inkubasi 5 tahun dan 10 tahun. Hasil pendugaan disebut lebih peka bila nilai  $D$  lebih besar.

## HASIL DAN PEMBAHASAN

Pertumbuhan jumlah kasus infeksi HIV dapat terjadi pada tingkat yang rendah maupun tingkat yang sangat tinggi tergantung pada kecepatan penyebaran virus HIV. Untuk mengantisipasi hal tersebut, dalam penelitian ini digunakan dua bentuk pertumbuhan jumlah kasus infeksi HIV. Bentuk pertama adalah pertumbuhan cukup rendah yaitu 100 kasus per semester. Sedangkan bentuk kedua adalah pertumbuhan tinggi yang diwakili oleh persamaan  $\theta_t = \exp(t - 0,01 t^2)$ , dengan  $t$  adalah waktu dalam semester.

### 1. Pengaruh Penggunaan Alpha dan Beta

Alpha dan Beta adalah, berturut-turut, adalah parameter lokasi dari sebaran Gamma dan Weibull. Dalam pendugaan dengan algoritma EM nilai Alpha dan Beta secara terpisah digunakan dalam menentukan peluang masa inkubasi. Peluang masa

inkubasi ini sebelumnya digunakan dalam menentukan jumlah kasus AIDS. Jadi digunakan sebaran peluang masa inkubasi yang sama, baik dalam penentuan jumlah kasus AIDS berdasarkan nilai parameter jumlah kasus infeksi HIV maupun dalam langkah baliknya yaitu pendugaan jumlah kasus infeksi HIV berdasarkan jumlah kasus AIDS.

Untuk menganalisa kebaikan pendugaan tersebut, digunakan nilai Kuadrat Tengah Galat (KTG) yang merupakan jumlah kuadrat dari jarak atau selisih antara nilai parameter dengan nilai dugaannya. Semakin kecil nilai KTG berarti semakin baik pendugaan tersebut karena hal tersebut menunjukkan jarak nilai dugaan dan nilai parameter semakin kecil.

Menggunakan sebaran Gamma, didapat bahwa nilai KTG untuk  $\alpha=1$  selalu lebih kecil daripada nilai KTG untuk  $\alpha=3$  pada semua kombinasi masa inkubasi dan bentuk pertumbuhan. Sebagai contoh, untuk masa inkubasi 5 tahun dan bentuk pertumbuhan rendah, KTG  $\alpha=1$  adalah 31494 sedangkan KTG  $\alpha=3$  adalah 35977. Simulasi untuk  $\alpha=2$  juga menghasilkan KTG yang lebih besar daripada

KTG  $\alpha=1$ . Hal ini menunjukkan bahwa dalam pendugaan jumlah kasus infeksi HIV menggunakan algoritma EM dengan sebaran Gamma, penggunaan  $\alpha=1$  lebih baik daripada penggunaan  $\alpha=3$ .

Penggunaan sebaran Weibull juga memberikan hasil yang sama, yaitu penggunaan Beta =1 memberikan nilai dugaan yang lebih baik daripada hasil pendugaan dengan Beta=3.

## 2. Kepekaan Terhadap Kesalahan Masa Inkubasi

Masa inkubasi virus HIV adalah jangka waktu terjadinya infeksi sampai terlihatnya gejala penyakit AIDS pada orang yang terinfeksi. Oleh karena waktu terjadinya infeksi HIV sangat sulit diketahui, maka besar kemungkinan akan terjadi kesalahan dalam penentuan masa inkubasi yang sebenarnya. Ada dua jenis kesalahan yang dapat terjadi dalam penentuan masa inkubasi yaitu kesalahan 'Under estimate' dan 'Over estimate'. Kesalahan "Under estimate" adalah bila masa inkubasi yang ditetapkan lebih singkat dari yang sebenarnya, sebaliknya kesalahan 'Over estimate' terjadi bila ditetapkan masa inkubasi yang lebih lama dari masa inkubasi sebenarnya.

Dalam simulasi ini dianggap bahwa rata-rata masa inkubasi sebenarnya adalah 10 tahun. Jadi rata-rata masa inkubasi 10 tahun digunakan dalam penentuan peluang masa inkubasi untuk menentukan jumlah kasus AIDS untuk ketiga masa inkubasi yang digunakan dan dalam penentuan nilai dugaan jumlah kasus infeksi HIV yang dianggap sebagai nilai parameter. Nilai dugaan ini akan digunakan sebagai pembanding terhadap hasil pendugaan yang menggunakan masa inkubasi yang salah. Sedangkan kesalahan 'Under estimate' menggunakan masa inkubasi 5 tahun dan kesalahan 'Over estimate' dengan masa inkubasi 15 tahun. Kedua rata-rata masa inkubasi ini hanya digunakan dalam pendugaan jumlah kasus infeksi HIV setelah ditentukan jumlah kasus AIDS dengan masa inkubasi 10 tahun.

Hasil simulasi terhadap kesalahan masa inkubasi menunjukkan bahwa penggunaan masa inkubasi yang lebih singkat dari masa inkubasi sebenarnya akan menghasilkan pendugaan yang lebih rendah dari nilai sebenarnya pada awal periode pendugaan kemudian pada periode selanjutnya lebih tinggi dari nilai dari nilai sebenarnya. Dengan kata lain, penggunaan

masa inkubasi yang 'Under estimate' menghasilkan pendugaan yang 'Under estimate' pada awal periode pendugaan tetapi pada periode selanjutnya menjadi 'Over estimate'. Sebaliknya, penggunaan masa inkubasi yang lebih lama dari yang sebenarnya menghasilkan nilai dugaan yang 'Over estimate' di awal periode, pada periode selanjutnya nilai dugaan menjadi 'Under estimate'. Hal tersebut berlaku baik untuk  $\alpha=1$  maupun  $\alpha=3$ .

Akan tetapi jika ditinjau dari jarak antara nilai dugaan dari masa inkubasi yang salah dan masa inkubasi sebenarnya akan tampak perbedaan antara  $\alpha=1$  dan  $\alpha=3$ . Jarak tersebut diukur dengan nilai D yang merupakan jumlah dari kuadrat selisih antara nilai dugaan pada masa inkubasi sebenarnya dengan nilai dugaan pada masa inkubasi yang salah. Nilai D tersebut menunjukkan kepekaan pendugaan terhadap kesalahan penetapan masa inkubasi. Semakin besar nilai D berarti semakin peka pendugaan yang digunakan karena kesalahan masa inkubasi menyebabkan hasil pendugaan jauh dari yang seharusnya didapat bila digunakan masa inkubasi yang benar. Nilai D untuk  $\alpha=1$  selalu lebih kecil daripada

nilai D untuk  $\alpha=3$ . Hal ini berlaku untuk semua kombinasi masa inkubasi dan bentuk pertumbuhan jumlah kasus infeksi HIV. Sebagai contoh, untuk pertumbuhan jumlah kasus infeksi HIV rendah dengan kesalahan 'Under estimate', nilai D  $\alpha=1$  adalah 359 sedangkan nilai D  $\alpha=3$  adalah 1741.

Uraian mengenai jarak antara hasil pendugaan dengan masa inkubasi salah dan inkubasi sebenarnya menunjukkan bahwa pendugaan dengan  $\alpha=3$  lebih peka terhadap kesalahan dalam penetapan masa inkubasi. Jadi dalam pendugaan jumlah kasus infeksi HIV sebaiknya digunakan  $\alpha=1$ .

Perilaku nilai dugaan tersebut juga dimiliki oleh hasil pendugaan menggunakan sebaran Weibull. Jarak D untuk Beta = 1 lebih rendah dari nilai D untuk Beta=3. Hal ini berlaku untuk semua kombinasi masa inkubasi dan bentuk pertumbuhan jumlah kasus infeksi HIV. Dari sini dapat disimpulkan bahwa penggunaan Beta=1 lebih resisten terhadap kesalahan masa inkubasi yang digunakan.

Berdasarkan kedua penelitian tersebut, bentuk sebaran yang menghasilkan nilai dugaan yang lebih baik dan tidak peka

terhadap kesalahan masa inkubasi adalah Sebaran Gamma(1,λ) dan Sebaran Weibull (1,ρ).

Fungsi kepekatan peluang Sebaran Gamma(1,λ) adalah

$$f(t) = \lambda (\lambda t)^{1-1} \exp(-\lambda t) / \Gamma(1) \quad (t \geq 0)$$

atau

$$f(t) = \lambda \exp(-\lambda t) \quad (t \geq 0) \quad (1)$$

Bentuk terakhir ini (persamaan (1)) adalah fungsi kepekatan peluang bagi sebaran Eksponensial (λ). Sejalan dengan itu, fungsi kepekatan peluang Sebaran Weibull (1,ρ) adalah

$$f(i) = \rho (\rho i)^{1-1} \exp(-(\rho i)^1)$$

$$f(i) = \rho \exp(-(\rho i)) \quad i \geq 0 \quad (2)$$

Seperti halnya persamaan (2), persamaan (1) ini juga merupakan fungsi kepekatan peluang sebaran Eksponensial (ρ). Parameter λ dan ρ merupakan parameter yang sama untuk sebaran Eksponensial. Keduanya dapat diganti dengan lambang yang sama, misalkan λ, sehingga bentuk Sebaran Gamma(1,λ) dan Weibull (1,ρ) adalah sebaran Eksponensial(λ).

Jadi penggunaan Sebaran Gamma(1,λ) atau Weibull (1,ρ) sama saja

dengan penggunaan sebaran Eksponensial(λ). Dengan demikian, sebaran terbaik untuk digunakan sebagai sebaran peluang masa inkubasi HIV adalah sebaran Eksponensial.

## KESIMPULAN

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa sebaran peluang masa inkubasi yang baik digunakan untuk pendugaan jumlah kasus infeksi HIV dengan Algoritma EM adalah sebaran Eksponensial karena hasil pendugaannya lebih baik dan tidak peka terhadap kesalahan penetapan masa inkubasi.

## DAFTAR PUSTAKA

- Anderson, R.M. and G.F. Medley. 1988. Epidemiology, HIV infection and AIDS: The Incubation and infectious periods, survival and vertical transmission, AIDS 2, S57-S63
- Bacchetti, P., M.R. Segal, and N.P. Jewell. 1993. Backcalculation of HIV infection rates. Statistical Science, Vol.8 No 6 82-119.
- Green, P.J. 1990. On use of EM algorithm for penalized likelihood estimation. J.Roy.Statist.Soc, Ser B 52: 443-452.
- Hanum, H. 1996. Studi simulasi Algoritma EM dalam pendugaan kasus infeksi HIV. Tesis. Program Pascasarjana



Institut Pertanian Bogor (tidak dipublikasikan).

Hanum, H dan Samsuryadi, 1999, Penggunaan Sebaran Gamma untuk pendugaan jumlah Kasus infeksi HIV dengan Algoritma EM.(penelitian dosen Muda DIKTI, belum dipublikasikan)

Shepp, L.A. and Y. Vardi. 1982. Maximum likelihood reconstruction in positron emission tomography, IEEE Trans. Med. Imaging, 1 113-122.