Age-sex specific EMS backcalculation model

Suwanee Surasiengsunk[1]

1. Introduction

Backcalculation has been widely used to estimate past rates of HIV infection from the numbers of incident AIDS cases and knowledge of the distribution of times between HIV infection and AIDS onset, called the incubation distribution. The need for backcalculation in this context arises from the fact that HIV incidence is not directly observable.

Backcalculation has been used extensively in the literature and a number of different approaches have been suggested. These range from methods utilizing various degrees of parametric assumptions (e.g. Isham, 1989, Rosenberg & Gail, 1990, Brookmeyer & Liao, 1990) to non-parametric approaches incorporating smoothing (Becker et al., 1991 and Brookmeyer, 1991). Becker et al (1991) combined the EM algorithm and a smoothing step, it is thus referred to as the EMS algorithm, and applied in the AIDS context. The basic idea of adding a smoothing step to the EM algorithm was originally suggested by Silverman et al (1990) and Wilson (1989), in the context of image reconstruction and other indirect estimation problems.

Here we adapt the method proposed by Becker et al (1991) for estimation age-sex specific relative risk of HIV infection. This report describes only a general model. Application to AIDS incidence data for Thai populaton will be presented in greater detail in a future report.

---

[1] Department of Statistics, Faculty of Commerce and Accountancy

Chulalongkorn University

2. Model

We use a multiplicative model to represent the expected number of age-sex-specific HIV infections and the EM-algorithm with smoothing, to estimate HIV/AIDS age-sex-specific infections in a backcalculation model.

We use the following notation:

Y_{gat} = number of AIDS cases age a sex g at time t , the time of their diagnosis, Observed.

N_{gat} = number of individuals age a sex g infected with HIV at time t . Unobserved

f_{sd} = the probability that an individual infected at time s will develop AIDS by duration d .

Suppose that $N_{gat}\ ; g = 1,2;\ a = 1,\dots,A;\ t = 1,\dots,T$ are mutually independent Poisson variates. The assumption that the incubation period of each individual infected in year t is independent leads to

$$E\left(Y_{gat} \mid \left\{N_{ga1},\dots,N_{gaT}\right\}\right) = \sum_{x=1}^{t} N_{g,a-t+x,x}\, f_{x,t-x}.$$

Whenever the age subscript of a quantity takes a nonpositive value, the quantity is interpreted to be zero. Then taking expectations

$$\mu_{gat} = \sum_{x=1}^{t} V_{g,a-t+x,x}\, f_{x,t-x}, \tag{1}$$

where $\mu_{gat} = E(Y_{gat})$ and $V_{gat} = E(N_{gat})$.

Model Assumptions

Let us assume the N_{gat} to be independent Poisson variates. The dependence on age and sex is incorporated via the multiplicative model

$$V_{gat} = \pi_{ga}\beta_{ga}\lambda_t, \tag{2}$$

Where π_{ga} is the proportion of individuals in the population who are of age a and sex g. We consider that π_{ga} are known from census tables. The π_{ga} allow β_{ga} to reflect the relative susceptibility of individuals of age a and sex g. The β_{ga} are estimated only for ages smaller than some large value, A, because the age specific relative risk will be negligible for very old individuals. Thus we can set $\beta_{ga} = 0$ for all $a > A$.

If we take the summation for all a and g in equation (2) we will get

$$\sum_{g=1}^{G}\sum_{a=1}^{A} V_{gat} = \lambda_t \sum_{g=1}^{G}\sum_{a=1}^{A} \pi_{ga}\beta_{ga}.$$

If we impose the constraint

$$\sum_{g=1}^{G}\sum_{a=1}^{A} \pi_{ga}\beta_{ga} = 1 , \tag{3}$$

then λ_t is the overall HIV infection intensity without regard to age and sex, and its estimate can be compared with estimates by backcalculation methods that do not use age and sex as a covariate.

Parameter Estimation

Let T be the last year for which the AIDS count is considered reliable. Due to the Poisson assumption on the $N's$ the AIDS counts $Y = \{Y_{gat};\ g = 1,2; a = 1,\ldots, A+t-1;\ t = 1,\ldots,T\}$ are mutually independent Poisson variates. Since is potentially positive for a $\in \{1,\ldots, A+t-1\}$, the age subscript of the AIDS counts takes values up to $A+t-1$. Observing $Y = y$ yields the likelihood function

$$L(\beta,\lambda|y) = \prod_{t=1}^{T}\prod_{g=1}^{G}\prod_{a=1}^{A+t-1} \mu_{gat}^{y_{gat}} e^{-\mu_{gat}},$$

where μ_{gat} is given by (1) and (2). To estimate and we maximize the likelihood $L(\beta,\lambda|y)$ with respect to these parameters subject to the constraint (3). By adapting the EMS-algorithm of Silverman et al (1990) we obtain smooth estimated curves of β_{ga} and λ_t.

3. The Algorithm

The complete data are observations $N_{gatd}, (d = 0,\dots,T-t)$, the number of individuals of age a sex g infected at time t and having incubation period of duration d They include the infection time for all individuals who develop AIDS by time T

The log likelihood function for the "complete" data $N_{gatd} = n_{gatd}$ is

$$log\, L(\beta,\lambda|\{N_{gatd}\}) = \sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{a=1}^{A}\sum_{d=0}^{T-t}\left[n_{gatd}\, log\left(\pi_{ga}\beta_{ga}\lambda_t f_{td}\right) - \pi_{ga}\beta_{ga}\lambda_t f_{td}\right]. \qquad (4)$$

E-step

This step involves replacing in (4) by $\hat{n}_{gat} = E\left(N_{gatd}|y;\beta^{old},\lambda^{old}\right)$, the conditional expectation of N_{gatd} given $Y = y$ and parameter values $\left(\beta^{old},\lambda^{old}\right)$.

Here

$$n_{gatd} = E\left(N_{gatd}|y;\beta,\lambda\right) = y_{g,a+d,t+d}\frac{\pi_{ga}\beta_{ga}\lambda_t f_{td}}{\sum_{i=1}^{t+d}\pi_{g,a+i-t}\beta_{g,a+i-t}\lambda_i f_{i,t+d-i}} \qquad (5)$$

M-step

At this step we maximize log $L\left(\beta,\lambda|\{\hat{n}_{gatd}\}\right)$ with respect to β and λ ,subject to the constraint (3). The method of Lagrange multipliers gives

$$\beta^{*}_{ga} = \pi^{-1}_{ga}\hat{n}_{ga\bullet\bullet}/\hat{n}_{g\bullet\bullet\bullet} \qquad (g = 1,2,\ a = 1,\dots,A), \qquad (6)$$

$$\lambda_t^* = \hat{n}_{\bullet\bullet t\bullet}/F_{t,T-t} \qquad (t = 1,\ldots,T), \tag{7}$$

where $F_{t,T-t} = \sum_{d=0}^{T-t} f_{td}$ and $\bullet$ in place of a subscript denotes summation over that subscript.

S-step

The smoothing step is incorporated after each M-step. After smoothing,

$$\beta_{ga}^{new} = \sum_{j=0}^{k_1} w_{1j}\beta^*_{g,a+j-\frac{1}{2}k_1}, \tag{8}$$

$$\lambda_t^{new} = \sum_{j=0}^{k_2} w_{2j}\lambda^*_{t+j-\frac{1}{2}k_2} \tag{9}$$

The value of the $k_i's$ determine the "window width" for the weighted average, and should be an even integer. We must choose a value for k_i and the weights $w_{ij}, (i = 1,2;\ j = 0,1,\ldots,k_i)$ such that $\sum_j w_{ij} = 1$. Symmetric binomial weights as used by Silverman et al (1990) are one example that can be used for the weights w_{ij}:

$$w_{ij} = \binom{k_i}{j} / 2^{k_i},\ i = 1,2;\ j = 0,1,\ldots,k_j.$$

Convergence

For $T' < T, A' < A$ and small value of ε_1 and ε_2 the iteration stops when both

$$\frac{\left|\sum_{g=1}^{2}\sum_{a=1}^{A'}\beta_{ga}^{new} - \sum_{g=1}^{2}\sum_{a=1}^{A'}\beta_{ga}^{old}\right|}{\sum_{g=1}^{2}\sum_{a=1}^{A'}\beta_{ga}^{old}} < \varepsilon_1,$$

$$\frac{\left|\sum_{t=1}^{T'}\lambda_t^{new} - \sum_{t=1}^{T'}\lambda_t^{old}\right|}{\sum_{t=1}^{T'}\lambda_t^{old}} < \varepsilon_2,$$

are first satisfied.

องค์ประกอบของต้นทุนการผลิต

ต้นทุนผลิตภัณฑ์ที่ผลิต เป็นต้นทุนของการผลิตทั้งหมดที่เกี่ยวข้องโดยทางตรงและทางอ้อมกับผลิตภัณฑ์นั้น โดยทั่วไปแบ่งองค์ประกอบของต้นทุนการผลิตออกเป็น 3 ประเภท คือต้นทุนวัตถุดิบทางตรง (Direct Material Cost) ต้นทุนแรงงานทางตรง (Direct labor Cost) และโสหุ้ยการผลิต (Factory Overhead) สำหรับโรงงาน

ตัวอย่าง ต้นทุนการผลิตสามารถแยกตามองค์ประกอบดังกล่าวข้างต้นได้ดังนี้

1. **ต้นทุนวัตถุดิบทางตรง**

ต้นทุนวัตถุดิบทางตรงของการผลิตสติกเกอร์ ประกอบด้วย ค่าใช้จ่ายในการจัดหาวัตถุดิบหลักในการผลิต อันได้แก่ สติกเกอร์ หรือกระดาษ สีและผงสีน้ำมันต่าง ๆ

2. **ต้นทุนแรงงานทางตรง**

ต้นทุนแรงงานทางตรง เป็นต้นทุนค่าแรงที่จ่ายให้แก่คนงานที่ทำงานในเวลาปกติและทำงานล่วงเวลา ต้นทุนในส่วนนี้ได้แก่ ค่าแรงงานของในหน่วยพิมพ์ซิลสกรีน หน่วยพิมพ์ออฟเซ็ทและหน่วยพิมพ์สติกเกอร์ม้วนระบบอัตโนมัติ

3. **โสหุ้ยการผลิต**

โสหุ้ยการผลิต เป็นต้นทุนค่าใช้จ่ายที่นอกเหนือไปจากต้นทุนวัตถุดิบทางตรงและค่าแรงทางตรง ต้นทุนในส่วนที่เป็นโสหุ้ยการผลิต ได้แก่ วัตถุดิบทางอ้อม แรงงานทางอ้อม ค่าไฟ ค่าน้ำ เป็นต้น

เมื่อพิจารณาหาวิธีการเก็บข้อมูลแล้วจะต้องจัดทำเอกสาร เพื่อช่วยให้การเก็บข้อมูลให้สมบูรณ์ครบถ้วนยิ่งขึ้น จึงต้องจัดทำระบบการส่งเอกสารหรือการรายงานจากแหล่งข้อมูลเพื่อนำมาประมวล ซึ่งเรียกว่า ระบบการไหลของเอกสาร (Document Flow System)

ผลของการประมวลผลข้อมูลนี้ ทำให้ทราบต้นทุนการผลิตผลิตภัณฑ์แต่ละชนิด โดยอาศัยข้อมูลที่เก็บรวบรวม และจากเอกสารและการไหลของเอกสารทำให้ข้อมูลต่าง ๆ ที่เกี่ยวกับต้นทุนส่งผ่านมายังแผนกบัญชี แผนกบัญชีจะนำข้อมูลนี้มาสรุป และเป็นแนวทางในการนำข้อมูลมาประมวลผลด้วยคอมพิวเตอร์ต่อไป