

Poisson-gamma モデルに基づく 市区町村別死亡危険度の推定

大瀧 慈¹, 川崎 裕美², 佐藤 健一¹,
中山 晃志³, 柳原 宏和⁴, 山口 直人⁵

¹広島大学原爆放射能医学研究所, ²広島大学医学部保健学科,

³原子力研究所, ⁴統計数理研究所, ⁵国立がんセンター研究所

1. はじめに

本研究は、任意年次時点での各種死因別死亡危険度地図作製およびその動画表示のためのコンピュータシステムを開発することを目的とする。データは、厚生省から目的外使用を許可された人口動態統計票のうち 1974 年から 1994 年までの 20 年間における全国の市区町村別死亡数、および、1975 年から 1995 年までの期間に行われた 5 回の国勢調査による市区町村別性別年齢階級別人口数である。

2. Poisson-gamma モデル

第 i 市区町村での第 j 年次での死亡数を Y_{ij} と、それに対する基準集団の年齢階級別死亡率に基づいて算出された期待死亡数を ξ_{ij} とするとき、下記のような Poisson モデルを想定する。

$$Pr(Y_{ij} = y) = \frac{(\xi_{ij}\eta_j\alpha_i)^y}{y!} e^{-\xi_{ij}\eta_j\alpha_i}, \quad j = 1, \dots, k, \quad i = 1, \dots, n, \quad (1)$$

ここで、 ξ_{ij} は全期間期間全市区町村平均に基づく期待死亡数であり、 η_j は固定効果母数であり、 α_i は、平均が 1 で分散が $\frac{1}{\nu}$ のガンマ分布に従うランダム効果母数で、互いに独立であるとする。即ち、

$$\alpha_i \sim_{i.i.d.} G(\nu, \nu), \quad i = 1, \dots, n,$$

とする。

3. 母数および相対危険度の推定

いま、 $Y_{\cdot j} = \sum_i Y_{ij}$, $\xi_{\cdot j} = \sum_i \xi_{ij}$ $j = 1, \dots, k$, および、 $Y_{i\cdot} = \sum_j Y_{ij}$, $\xi_{i\cdot} = \sum_j \xi_{ij}$ $i = 1, \dots, n$, とすると、式(1)より

$$Pr(Y_{\cdot j} = y) = \frac{(\eta_j \xi_{\cdot j} R_j)^y}{y!} e^{-\eta_j \xi_{\cdot j} R_j}, \quad R_j = \frac{\sum_i \xi_{ij} \alpha_i}{\xi_{\cdot j}},$$

$$Pr(Y_{i\cdot} = y) = \frac{(\alpha_i \sum_j \xi_{ij} \eta_j)^y}{y!} e^{-\alpha_i \sum_j \xi_{ij} \eta_j}$$

となる。ここで、 $E(R_j) = 1$, $Var(R_j) = \frac{\sum_{i=1}^n \xi_{ij}^2}{(\sum_{i=1}^n \xi_{ij})^2} \cong 0$ であることより、データ y_{ij} , $j = 1, \dots, k$, $i = 1, \dots, n$ が与えられたときの η_j の推定値、

$$\hat{\eta}_j = \frac{y_{\cdot j}}{\xi_{\cdot j}}, \quad j = 1, \dots, k, \quad (2)$$

を得る。さらに、 α_i の分散 ν の推定値 $\hat{\nu}$ を、事後確率最大化法により求める。すなわち、 $\hat{\nu}$ は母数 ν に関する尤度関数

$$L(\nu | (y_{i\cdot})) = \prod_{i=1}^n \frac{\Gamma(y_{i\cdot} + \nu)}{y_{i\cdot}! \Gamma(\nu)} \left(\frac{\sum_j \xi_{ij} \hat{\eta}_j}{\sum_j \xi_{ij} \hat{\eta}_j + \nu} \right)^{y_{i\cdot}} \left(\frac{\nu}{\sum_j \xi_{ij} \hat{\eta}_j + \nu} \right)^\nu$$

を最大化する解である。このとき、 α_i の事後分布に基づくベイズ推定値、

$$\hat{\alpha}_i = \frac{y_{i\cdot} + \hat{\nu}}{\sum_j \xi_{ij} \hat{\eta}_j + \hat{\nu}}, \quad i = 1, \dots, n, \quad (3)$$

を得る。よって、(1), (2) および (3) より (i, j) における相対危険度の推定値として、

$$RR_{ij} = \hat{\alpha}_i \hat{\eta}_j = \frac{y_{i\cdot} + \hat{\nu}}{\sum_\ell \xi_{i\ell} \hat{\eta}_\ell + \hat{\nu}} \hat{\eta}_j, \quad j = 1, \dots, k, \quad i = 1, \dots, n, \quad (4)$$

が得られる。

4. データ解析

式(4)を適用して、「肺がん」、「胃がん」、「結腸がん」、「肝がん」、「膵臓がん」、「乳がん」および「舌がん」について、1975年～1994年の20年間の各年次毎の性別市区町村別相対危険度を算出してSMR地図を作製し、アニメーション化を行った。