Модель двумерного гамма–распределения в исследовании повторных статистических наблюдений с приложением в иммунологии

Алиева Наталия Дмитриевна, гр. 422

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., д. Алексеева Н. П.
Рецензент: к. т. н., д. Белякова Л. А.

СтатМод

Санкт-Петербург, 2015г.

Виды двумерного гамма–распределения

- Furman E. (2008): $\xi_i \sim \gamma(\cdot|b_i, \lambda_i)$ — нез., $i = 0, 1, 2$. $\zeta = (\eta_1, \eta_2)$ имеет двумерное гамма–распределение.

$$\begin{cases} \eta_1 = \frac{b_0}{b_1}\xi_0 + \xi_1 \\ \eta_2 = \frac{b_0}{b_1}\xi_0 + \frac{b_1}{b_2}\xi_1 + \xi_2. \end{cases} \tag{1}$$

Виды двумерного гамма–распределения

- Furman E. (2008): $\xi_i \sim \gamma(\cdot|b_i, \lambda_i)$ — нез., $i = 0, 1, 2$. $\zeta = (\eta_1, \eta_2)$ имеет двумерное гамма–распределение.

$$\begin{cases} \eta_1 = \frac{b_0}{b_1}\xi_0 + \xi_1 \\ \eta_2 = \frac{b_0}{b_1}\xi_0 + \frac{b_1}{b_2}\xi_1 + \xi_2. \end{cases} \tag{1}$$

- Sen S., Lamichhane R., Diawara N. (2014): $\xi_0 \sim \gamma(\cdot|b_0, \lambda_0), \eta_i \sim \gamma(\cdot|b_i, \lambda_i, \xi_0), i = 1, 2$. η_1 и η_2 — нез. $f(\xi_0, \eta_1, \eta_2) = f_1(\xi_0) f_2(\eta_1) f_3(\eta_2)$ — плотность двумерного гамма–распределения.

Виды двумерного гамма–распределения

- Furman E. (2008): $\xi_i \sim \gamma(\cdot|b_i, \lambda_i)$ — нез., $i = 0, 1, 2$. $\zeta = (\eta_1, \eta_2)$ имеет двумерное гамма–распределение.

$$\begin{cases} \eta_1 = \frac{b_0}{b_1}\xi_0 + \xi_1 \\ \eta_2 = \frac{b_0}{b_1}\xi_0 + \frac{b_1}{b_2}\xi_1 + \xi_2. \end{cases} \tag{1}$$

- Sen S., Lamichhane R., Diawara N. (2014): $\xi_0 \sim \gamma(\cdot|b_0, \lambda_0), \eta_i \sim \gamma(\cdot|b_i, \lambda_i, \xi_0), i = 1, 2$. η_1 и η_2 — нез. $f(\xi_0, \eta_1, \eta_2) = f_1(\xi_0)f_2(\eta_1)f_3(\eta_2)$ — плотность двумерного гамма–распределения.
- Алексеева Н. П., Щербакова И. С., Щекина Л. А. (2009): $\xi_i \sim \gamma(\cdot|b = 1, \lambda_i), i = 0, 1, 2$ — независимые. $\zeta = (\eta_1, \eta_2)$ имеет двумерное гамма–распределение.

$$\begin{cases} \eta_1 = \xi_0 + \xi_1 \\ \eta_2 = \xi_0 + \xi_2. \end{cases} \tag{2}$$

Структура реальных данных

Повторные наблюдения иммунологических показателей до и после лечения.

- Источники: НИИ онкологии им. проф. Н. Н. Петрова, НИИ Фтизиопульмонологии.
- Иммунологические показатели: иммуноглобулины класса A и G, матриксные металлопротеиназы и их ингибиторы.

Структура реальных данных

Повторные наблюдения иммунологических показателей до и после лечения.

- Источники: НИИ онкологии им. проф. Н. Н. Петрова, НИИ Фтизиопульмонологии.
- Иммунологические показатели: иммуноглобулины класса A и G, матриксные металлопротеиназы и их ингибиторы.

Прикладная задача: влияние иммунитета на эффективность лечения.

Основные понятия

Определение

$\xi \sim \gamma(\cdot|b,\lambda)$, $\Gamma(\lambda)$ — *гамма–функция Эйлера. Плотность гамма–распределения:*

$$f_\xi(x) = x^{\lambda-1}\frac{e^{-x/b}}{b^\lambda\Gamma(\lambda)}, x \geq 0$$

Утверждение (Алексеева, Щербакова, 2007)

$\xi_i \sim \gamma(\cdot|1,\lambda_i), i = 0,1,2$ — *независимые.* $\zeta = (\eta_1,\eta_2)$ *имеет двумерное гамма–распределение с параметрами* $\Lambda_i = \lambda_0 + \lambda_i, i = 1,2$ *и* ρ — *коэф. корреляции:*

$$\begin{cases} \eta_1 = \xi_0 + \xi_1 \\ \eta_2 = \xi_0 + \xi_2 \end{cases} \implies \begin{cases} \lambda_0 = \rho\sqrt{\Lambda_1\Lambda_2} \\ \lambda_1 = \Lambda_1 - \rho\sqrt{\Lambda_1\Lambda_2} \\ \lambda_2 = \Lambda_2 - \rho\sqrt{\Lambda_1\Lambda_2}. \end{cases}$$

Обозначение: $(\eta_1,\eta_2) \sim \gamma_2(\cdot|\lambda_0,\lambda_1,\lambda_2)$.

Условия применимости модели (для проекций):

- $b > 0$, $\lambda_0 + \lambda_i > 0$,
- согласие с $\gamma(\cdot|1, \lambda_0 + \lambda_i)$,
- $\rho \geq 0$.

Профили правдоподобия параметров распределения при $n = 1000$.

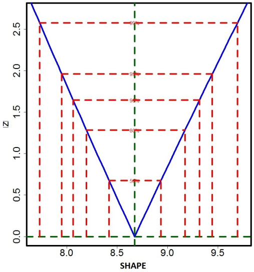

Рис. : Параметр формы λ.

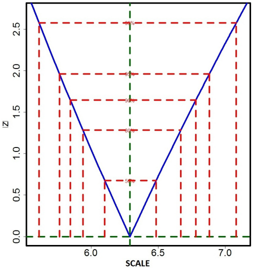

Рис. : Параметр масштаба b.

Плотность двумерного гамма–распределения

Утверждение

$$\gamma_2(u,v|\lambda_0,\lambda_1,\lambda_2) = S(u,v)\frac{e^{-(u+v)}}{\Gamma(\lambda_0)\Gamma(\lambda_1)\Gamma(\lambda_2)},$$ *где*

$$S(u,v) = \begin{cases} \sum\limits_{k=0}^{\infty} \frac{u^{c-1}v^{-a}\Gamma(b)\Gamma(c-b)}{\Gamma(c)} {}_2F_1(a,b;c,z), & u<v \\ \sum\limits_{k=0}^{\infty} \frac{v^{c-1}u^{-a}\Gamma(b)\Gamma(c-b)}{\Gamma(c)} {}_2F_1(a,b;c,z), & v<u \\ \sum\limits_{k=0}^{\infty} \frac{v^{c-a-1}\Gamma(b)\Gamma(c-a-b)}{\Gamma(c-a)}, & u=v, \end{cases}$$

$$(a,b,c,z) = \begin{cases} (1-\lambda_2, k+\lambda_0, k+\lambda_0+\lambda_1, u/v), & u<v \\ (1-\lambda_1, k+\lambda_0, k+\lambda_0+\lambda_2, v/u), & v<u, \end{cases}$$

$${}_2F_1(a,b;c,z) = \sum_{\ell=0}^{\infty} \frac{(a)_\ell (b)_\ell}{(c)_\ell}\frac{z^\ell}{\ell!}$$ *— гипергеометрическая функция,*

$(a)_\ell = \frac{\Gamma(a+\ell)}{\Gamma(a)}$ *и* $c-a-b>0$.

Утверждение (О производной гипергеометрической функции)

Пусть $\psi(x) = \Gamma'(x)/\Gamma(x)$ *— дигамма–функция.*

$$\frac{\partial {}_2F_1(a,b;c,z)}{\partial \lambda_0} = \sum_{\ell=0}^{\infty} \frac{(a)_\ell (b)_\ell}{(c)_\ell} [\psi(b+\ell) + \psi(c) - \psi(b) - \psi(c+\ell)] \frac{z^\ell}{\ell!},$$

где $a = 1 - \lambda_2$, $b = k + \lambda_0$, $c = k + \lambda_0 + \lambda_1$, $z = u/v$ *(по методу из (L. U. Arcani, G. Gasaneo, 2009)).*

$$D_i = \sum_{k=0}^{\infty} \frac{u_i^{c-1} v_i^{-a}}{k!} \frac{\Gamma(b)\Gamma(c-b)}{\Gamma(c)} {}_2F_1\Big(a,b;c,z_i\Big).$$

$$\begin{cases} \frac{\partial \ln \gamma_2(u,v|\lambda_0,\lambda_1,\lambda_2)}{\partial \lambda_0} = \sum_{i=1}^{n} \frac{D_i'}{D_i} - n\psi(\lambda_0) = 0 \\ \frac{\partial \ln \gamma_2(u,v|\lambda_0,\lambda_1,\lambda_2)}{\partial \lambda_1} = \sum_{i=1}^{n} \frac{D_i'}{D_i} - n\psi(\lambda_1) = 0 \\ \frac{\partial \ln \gamma_2(u,v|\lambda_0,\lambda_1,\lambda_2)}{\partial \lambda_2} = \sum_{i=1}^{n} \frac{D_i'}{D_i} - n\psi(\lambda_2) = 0. \end{cases}$$

Свойства ОМП на модельных данных

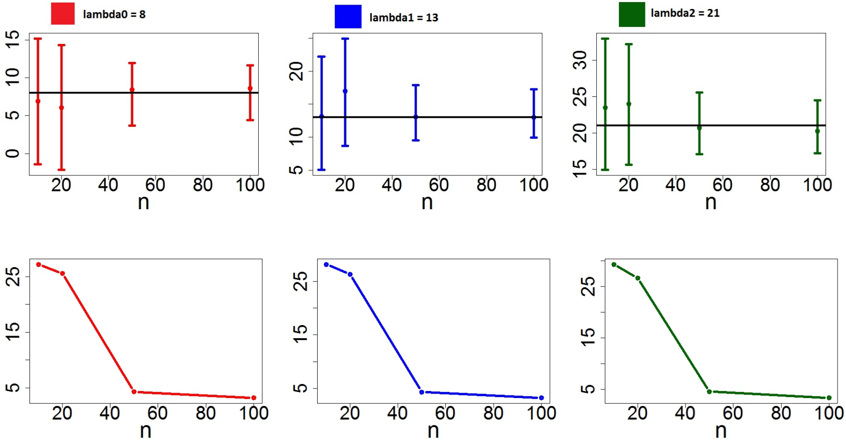

Рис. : Оценки параметров распределения и их оценки дисперсий.

$$T = n\sum_{i=1}^{k}\frac{(n_i/n - P_i(\theta))^2}{P_i(\theta)},$$

n — объем выборки,
k — число разбиений,
n_i — кол–во точек в i–й ячейке,
$P_i(\theta)$ — теор. вероятность попадания в i–ю ячейку,
$\theta = (\theta_1, \dots, \theta_s)$ — параметр.

$$T \sim \begin{cases}\chi^2(k-1), \\ \chi^2(k-s-1), \quad \theta = \hat{\theta}.\end{cases}$$

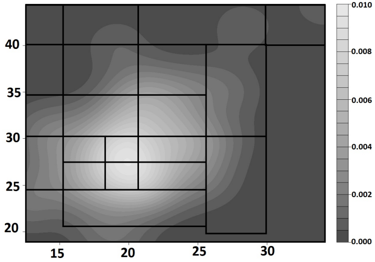

Рис. : Пример разбиения области определения $\gamma_2(\cdot|\lambda_0, \lambda_1, \lambda_2)$ на ячейки.

Проверка согласия с $\gamma_2(\cdot|\theta)$ на модельных данных

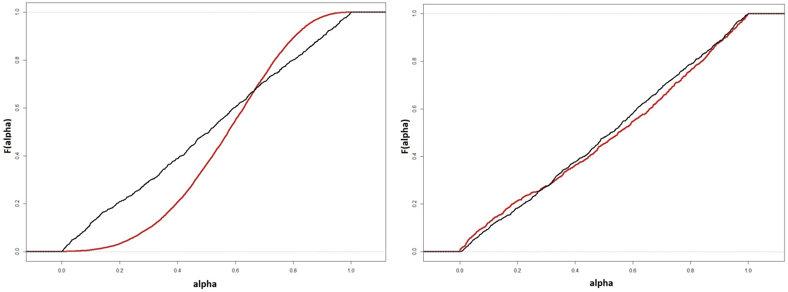

Рис. : Распределения $p - value$ до и после bootstrap'а при $n = 1000, N = 10000$.

Пример: матриксные металлопротеиназы MMP1

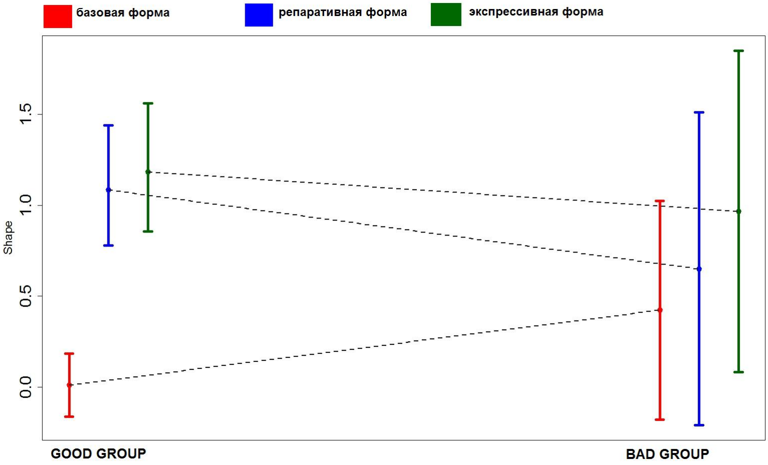

Рис. : Оценка параметров по MMP1.

Планы на будущее

- Сравнение оценок по методу максимального правдоподобия через решение системы и максимизацию функции правдоподобия.
- Проверка согласия с моделью при помощи различных критериев.
- Апробация метода на других данных.