

# Проверка гипотез однородности при перепараметризации обобщенного гамма распределения с применением в фармакологии

Гольдвирт Серафима Константиновна, гр. 522

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

Научный руководитель: к.ф.-м.н., доцент, Алексеева Н.П.  
Рецензент: к.ф.-м.н., ассистент, Коробейников А.И.



Санкт-Петербург  
2012г.

# Постановка задачи

- **Модель:** обобщенное гамма распределения (ОГР) с плотностью

$$f(x|a, k, \nu) = \frac{\nu k^a}{\Gamma(a)} e^{-kx^\nu} x^{\nu a - 1} \quad x > 0. \quad (1)$$

- Частными случаями ОГР являются гамма распределение ( $\nu = 1$ ), и распределения Вейбулла ( $a = 1$ ).
- Для оценки параметров используется логарифмическое обобщенное гамма распределение (ЛОГР) с плотностью

$$f_1(x|a, k, \nu) = \frac{\nu k^a}{\Gamma(a)} e^{\nu a x - k e^{\nu x}} \quad x > 0, \quad (2)$$

которое имеет случайная величина  $\eta = \ln \xi$ , где  $\xi \sim f(\cdot|a, k, \nu)$ .

- На основе модели ОГР выяснить влияние химических препаратов простагландина и тиреолиберина на высоту секреторного эпителия у лактирующих мышей.

## Оценки по методу максимального правдоподобия для ЛОГР

ОМП:

- $x_1, x_2, \dots, x_n \sim f_1(\cdot | a, k, \nu)$ . Оценка параметра  $\tilde{\nu}$  является решением уравнения:

$$\log\left(\frac{1}{\nu C}\right) - \psi\left(\frac{M_2}{\nu C}\right) + \nu M_1 = 0, \quad (3)$$

где  $\psi(y) = (\ln \Gamma(y))' = \frac{\Gamma(y)'}{\Gamma(y)},$

$$M_1 = \frac{1}{n} \sum_{i=1}^n x_i, \quad M_2 = \frac{1}{n} \sum_{i=1}^n e^{\nu x_i},$$

$$M_3 = \frac{1}{n} \sum_{i=1}^n x_i e^{\nu x_i}, \quad C = M_3 - M_1 M_2.$$

- Оценки  $\tilde{k}, \tilde{a}$  выражаются через  $\tilde{\nu}$  :

$$\tilde{k} = \frac{1}{\tilde{\nu} C}, \quad \tilde{a} = \frac{M_2}{\tilde{\nu} C}. \quad (4)$$

- Дисперсии оценок  $\tilde{k}$ ,  $\tilde{a}$ ,  $\tilde{\nu}$  максимального правдоподобия для ЛОГР имеют вид:

$$D_a = \frac{a^2 \varphi(a) + a - 1}{n (a^2 \varphi^2(a) - \varphi(a) - 1)}, \quad (5)$$

$$D_k = \frac{k^2 (a \varphi^2(a) + \varphi(a) + a \varphi(a) Z^2 + 2 \varphi(a) Z - Z^2)}{n (a^2 \varphi^2(a) - \varphi(a) - 1)}, \quad (6)$$

$$D_\nu = \frac{\nu^2 (a \varphi(a) - 1)}{n (a^2 \varphi^2(a) - \varphi(a) - 1)}, \quad (7)$$

где

$$Z = \psi(a) - \ln k,$$

$$\varphi(y) = (\ln \Gamma(y))'' = \frac{\Gamma(y)''}{\Gamma(y)} - \psi(y).$$

# Тестирование метода на модельной выборке

- При моделировании случайной выборки (ЛОГР) объема  $n = 200$  с параметрами  $a = 2$ ,  $k = 0.5$ ,  $\nu = 0.8$

$$\tilde{a} = 2.5565 \pm 2.7420,$$

$$\tilde{k} = 0.8120 \pm 0.9607,$$

$$\tilde{\nu} = 0.9525 \pm 0.5670.$$

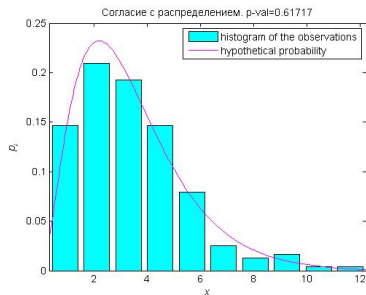


Рис. : Проверка согласия с ЛОГР модельной выборки, по критерию  $\chi^2$  ( $p = 0.62$ )

# Оценки максимального правдоподобия ЛОГР на основе перепараметризации

- Плотность ЛОГР через параметры [J.F. Lawless, 1980]

$$\mu = \frac{1}{\nu} (\log(a) - \log(k)),$$

$$\sigma = \frac{1}{\nu \sqrt{a}}$$

приводится к виду:

$$f_2(w; a, \mu, \sigma) = \frac{a^{a-\frac{1}{2}}}{\Gamma(a)\sigma} \exp \left( \sqrt{a} w(\mu, \sigma) - a e^{\frac{w(\mu, \sigma)}{\sqrt{a}}} \right), \quad (8)$$

где  $w(\mu, \sigma) = \frac{x - \mu}{\sigma}$ .

- Обозначим через  $e_1 = \frac{1}{n} \sum_1^n e^{\frac{x}{\sigma \sqrt{a}}}$ ,  $c_1 = \frac{1}{n} \sum_1^n x_i e^{\frac{x_i}{\sigma \sqrt{a}}} - \frac{e_1}{n} \sum_1^n x_i$ .  
Оценка параметра  $\mu$ :

$$\tilde{\mu} = \tilde{\sigma} \sqrt{\tilde{a}} \ln e_1, \quad (9)$$

$\tilde{a}$ ,  $\tilde{\sigma}$  являются решением системы:

$$\begin{cases} \bar{w} = \sqrt{a} (\psi(a) - \log a), \\ c_1 = e_1 \frac{\sigma}{\sqrt{a}}. \end{cases} \quad (10)$$

- Дисперсиями оценок параметров  $\theta = (a, \sigma, \mu)$  являются величины:

$$D_{\theta_i} = \frac{(I^{-1})_{ii}}{n}, \quad (11)$$

где  $I$  - информационная матрица, с элементами

$$I_{11} = -\frac{3}{4a^2} - \frac{\log a}{2a^2} + \frac{\psi(a)}{2a^2} - \frac{1}{a} + \frac{\log^2 a}{4a} + \frac{\psi^2(a)}{4a} + \frac{\varphi(a)}{4a} + \varphi(a) - \frac{1}{2a} \log(a)\psi(a),$$

$$I_{12} = \frac{1}{2\sigma} \left( \frac{2}{a} \psi(a) + \varphi(a) + \psi^2(a) - \frac{2}{a} \log(a) + \log^2(a) - \frac{1}{a} - 2a \log(a)\psi(a) \right),$$

$$I_{13} = \frac{1}{2\sqrt{a}\sigma} \left( \frac{1}{a} + \psi(a) - \log a \right),$$

$$I_{22} = \frac{1}{\sigma^2} (a \log^2 a - 2 \log a + a\psi^2(a) + 2\psi(a) + a\varphi(a) - 2a \log(a)\psi(a) + 1),$$

$$I_{23} = \frac{1}{\sigma^2 \sqrt{a}} (1 + a\psi(a) - a \log a),$$

$$I_{33} = \frac{1}{\sigma^2}.$$

# Тестирование метода с перепараметризацией на модельной выборке

Выборка объема  $n = 200$ , подчиняется ЛОГР с параметрами  $a = 2$ ,  $\mu = 1.7329$ ,  $\sigma = 0.8839$ , для перепараметризованной плотности. Оценки параметров:

$$\tilde{a} = 2.5276 \pm 2.6962,$$

$$\tilde{\mu} = 1.0935 \pm 0.1479,$$

$$\tilde{\sigma} = 0.5955 \pm 0.0759.$$

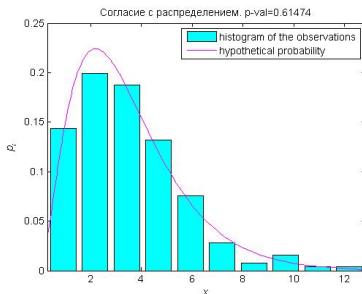


Рис. : Проверка согласия с ЛОГР модельной выборки, по критерию  $\chi^2$  ( $p = 0.61$ )



## Сравнение методов по эффективности оценок параметра формы

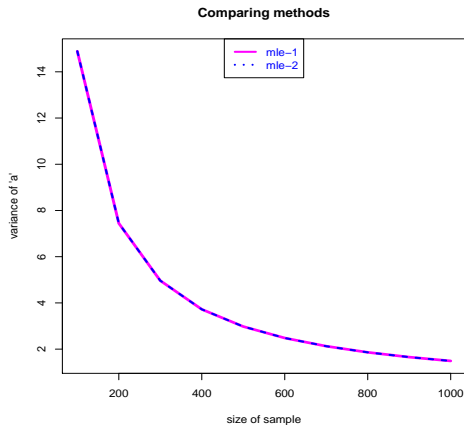


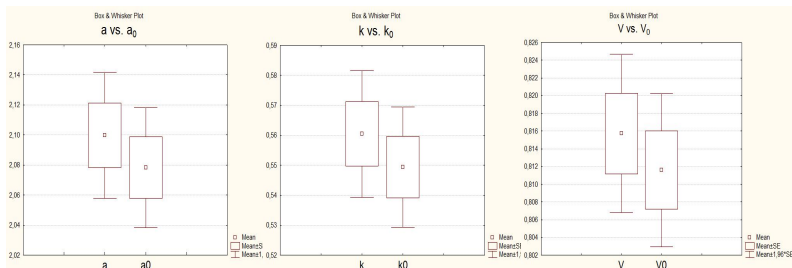
Рис. : Изменение  $D_a^{f_1}$  и  $D_a^{f_2}$  в зависимости от объема выборки ( $a = 4$ ).

# Сравнение методов по эффективности на модельных выборках

- Параметры  $f_2(\cdot; a, \mu, \sigma)$  и  $f_1(\cdot; a, k, \nu)$  связаны следующим образом:

$$k_0 = k(a, \mu, \sigma) = e^{\ln a - \frac{\mu}{\sigma\sqrt{a}}}, \quad \nu_0 = \nu(a, \sigma) = \frac{1}{\sigma\sqrt{a}}. \quad (12)$$

- Т-тест. Сравнение средних оценок по двум методам ( $p\text{-val} = 0$ ).



**Рис. :** Оценки  $a$  и  $a_0$ ,  $k$  и  $k_0$ ,  $\nu$  и  $\nu_0$  на 500 модельных выборках объемом 1000 с параметрами  $a = 2$ ,  $k = 0.5$ ,  $\nu = 0.8$ .

# Критерий однородности для параметров

- **Предложение.** Пусть  $x_1, \dots, x_n$  и  $y_1, \dots, y_m$  две независимые выборки,  $\tilde{\theta}_n^x$  и  $\tilde{\theta}_m^y$  - оценки максимума правдоподобия одного из параметров логарифмического обобщенного гамма распределения  $f_1(\cdot; a, k, \nu)$  или  $f_2(\cdot; a, \mu, \sigma)$ ,  $\frac{\sigma_x^2}{n}$  и  $\frac{\sigma_y^2}{m}$  соответственно дисперсии оценок. Тогда при справедливости гипотезы

$$H_0 : \theta^x = \theta^y$$

статистика

$$Z = \frac{\tilde{\theta}_n^x - \tilde{\theta}_m^y}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \quad (13)$$

имеет асимптотически стандартное нормальное распределение.

# Описание эксперимента

- Исследование влияния химических веществ тиреолиберина и простагландина на лактацию у 38 белых мышей.
- Экспериментальные данные:
  - ◆ Привесы;
  - ◆ Высота секреторного эпителия молочной железы на основе морфологического анализа гистологических препаратов.

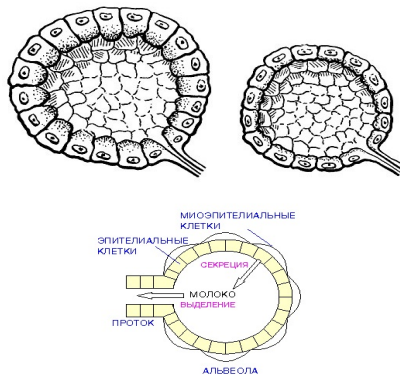


Рис. : Альвеола молочной железы.

# Описание эксперимента

- Тиреолиберин — один из представителей класса релизинг-гормонов гипоталамуса, приводящий к гиперпролактинемии (избыточной секреции пролактина). Опытные группы:  
 $T_1$  - контрольная (5);  
 $T_2$  - физиологический раствор (4);  
 $T_3 - 10^{-6} \frac{\text{мМоль}}{\text{мл}}$  (10);  
 $T_4 - 10^{-10} \frac{\text{мМоль}}{\text{мл}}$  (4).
- Простагландин—окисленные производные полиненасыщенных жирных кислот. Функции простагландинов сводятся к изменению тонуса гладких мышц. Группы:  
 $P_0$  - контрольная группа (6);  
 $P_1$  - лактопоз (10-11-й день) (5);  
 $P_2$  - лактогенез (3-4-й день) (4).
- В эксперименте с простагландином выборки объемом по 200 наблюдений, с тиреолиберином — 50.

- Метод множественных сравнений, статистики Тамхейна, Даннетта и Геймс-Хоуэлла (при условии разных дисперсий).

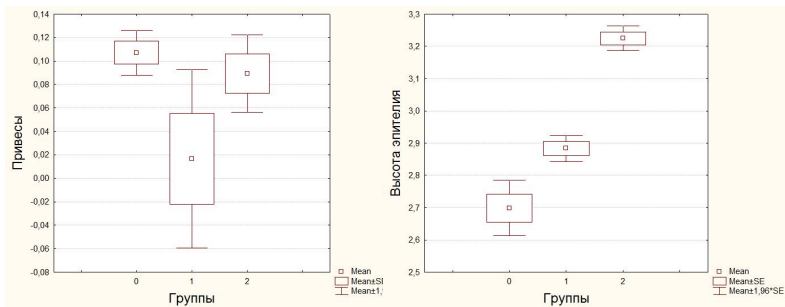


Рис. : Значимые изменения логарифма высоты эпителия и привесов в зависимости от фактора группы:  $P_0$ ,  $P_1$ ,  $P_2$  ( $p - \text{val} < 0.001$ ).

# Оценка параметров по ММП

- Примеры оценки параметров

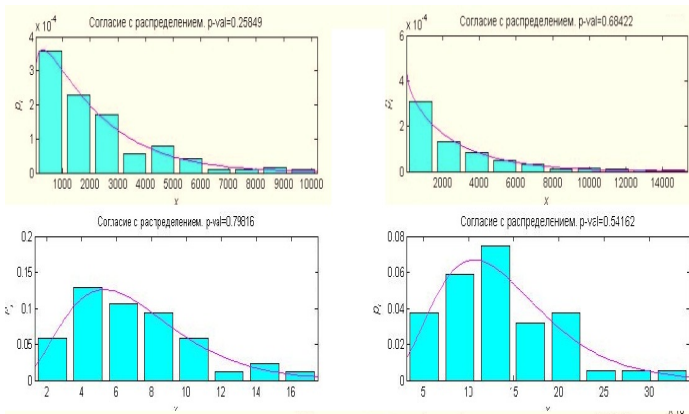


Рис. : Проверка согласия с распределением  $f_1(\cdot; a, k, \nu)$  для мышей ( $p > 0.25$ ).

# Влияние простагландина на параметры

- Группа контроля выделяется по параметрам формы и масштаба

	Лактопозз			Лактогенез				Контроль
	1	0,46	0,62	0,60	0,86	0,64	0,55	0,17
	0,46	1	0,48	0,78	0,46	0,47	0,48	0,42
Лактопозз	0,62	0,48	1	0,60	0,75	0,99	0,88	0,12
	0,60	0,78	0,60	1	0,60	0,60	0,61	0,58
	0,86	0,46	0,7	0,6	1	0,77	0,65	0,15
	0,64	0,47	1,0	0,6	0,77	1	0,87	0,13
Лактогенез	0,55	0,48	0,9	0,6	0,65	0,87	1	0,13
Контроль	0,17	0,42	0,1	0,6	0,15	0,13	0,13	1

Рис. : Пример значимостей отличия параметра формы у мышей из групп  $P_0, P_1, P_2$ .



# Влияние простагландина на параметры

- Сравнение групп простагландина по параметрам при использовании множественных сравнений на основе объединенной дисперсии.

	Контроль	Лактопоз	Лактогенез	Значимость	Объемы групп	
	a_0	a_1	a_2	p	n	m
1	1,04	2,9328		0,11	212	597
2		2,9328	3,2717	0,77	597	788
3	1,04		3,2717	0,06	212	788

Рис. : Проверка равенства параметров формы по группам.

	Контроль	Лактопоз	Лактогенез	Значимость	Объемы групп	
	k_0	k_1	k_2	p	n	m
1	0,0032	0,0637		0,09	212	597
2		0,0637	0,0861	0,54	597	788
3	0,0032		0,0861	0,02	212	788

Рис. : Проверка равенства параметров масштаба по группам.

- Морфологическая интерпретация: простагландин стимулирует увеличение высоты эпителия за счет образования внутриклеточных фагосом, а не из-за увеличения секреторной активности.

# Влияние тиреолиберина на параметры

- Критерий Краскала-Уоллиса для  $\mu, \sigma$  среди групп из эксперимента с тиреолиберином.

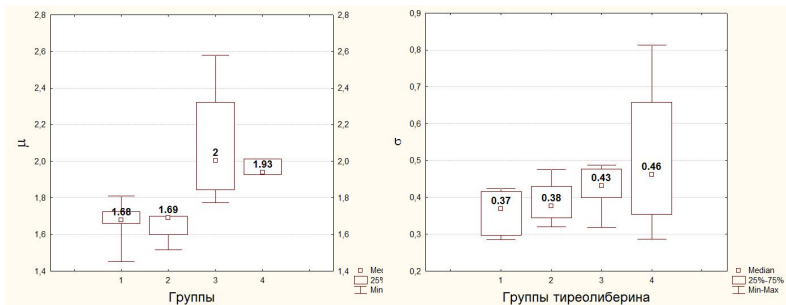


Рис. : Влияние фактора препарата на оценки параметров  $\mu, \sigma$  по критерию К-В ( $p = 0.0014$ ,  $p = 0.043$  соответственно).

- Коэффициенты корреляции Спирмена между  $\tilde{\mu}, \tilde{\sigma}$  и привесами соответственно  $R = 0.44$ ,  $p = 0.021$  и  $R = 0.43$ ,  $p = 0.024$ .
- Морфологическая интерпретация: большую секреторную активность характеризует не только  $\mu$ , но и  $\sigma$ .

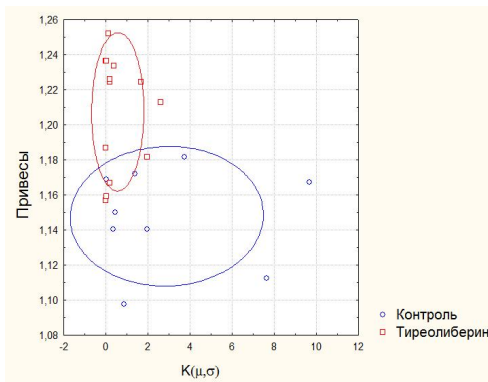


Рис. : Двумерная диаграмма между параметром  $k(\mu, \sigma)$  и привесами.

- Группы  $T_3, T_4$  отличаются от  $T_1, T_2$  значимо ( $p = 0.018$ , критерий М-У) по параметру  $k(\mu, \sigma)$ , медианы равны соответственно 0.18 и 1.38.

# Заключение

- Построены ОМП для ЛОГР и его перепараметризации.
- Вычислены дисперсии оценок. Показана инвариантность дисперсии оценок при перепараметризации.
- Программы ОМП с проверкой гипотез согласия протестированы на модельных выборках.
- Построены критерии однородности для параметров ЛОГР.
- На примере реальных данных исследовано влияние химических препаратов на параметры ЛОГР высоты секреторного эпителия у лактирующих мышей.