

# О редукации параметров в моделях сложных распределений с применением в радиобиологии

Олейник М. В., СПбГУ, Санкт-Петербург [st087252@student.spbu.ru](mailto:st087252@student.spbu.ru),  
Алексеева Н.П., СПбГУ, Санкт-Петербург [nina.alekseeva@spbu.ru](mailto:nina.alekseeva@spbu.ru)

## Аннотация

В данной статье рассматривается возможность применения двухпараметрической модели сложного логарифмически-пуассоновского распределения для описания динамики роста числа аномалий на ядрах клеток рабдомиосаркомы в зависимости от облучения. Приведены вероятности распределения с применением чисел Стирлинга второго рода и Эйлера. Приведено обоснование возникновения этих специальных чисел через ряды по упорядоченным разбиениям.

## Введение

Для радиобиологических данных числа аномалий в ядрах клеток рабдомиосаркомы у крыс при облучении разной степени интенсивности в экспериментах *in vivo* и *in vitro* ранее была исследована модель реинтрантного бинома с двумя парами неизвестных параметров. В поисках адекватных моделей с меньшим числом параметров рассматривались разные варианты, сначала обобщение отрицательного биномиального распределения за счет перехода от пуассоновского распределения числа слагаемых к биномиальному — биномиально-логарифмического, затем его антипод — логарифмически-биномиальное распределение. Предварительная оценка параметров указала на преимущество последнего в предельном случае, соответствующем пуассоновскому распределению случайных компонент. В данной статье изучаются модели пуассон-логарифмического (отрицательно-биномиальное) и логарифмически-пуассоновское (ЛПР). Эти двухпараметрические модели показали удовлетворительное согласование с эмпирическими данными.

Наиболее адекватной для интерпретации оказалась модель ЛПР. Соотношение между оценками параметров логарифмического и пуассоновского распределений свидетельствуют об интенсивном характере образования аномалий в эксперименте *in vivo* и экстенсивном в эксперименте *in vitro*.

## Логарифмически-пуассоновское распределение

Логарифмически-пуассоновское распределение вводится как распределение случайной суммы независимых случайных величин

$$\zeta_\tau = \xi_1 + \dots + \xi_\tau,$$

где  $\xi_i$  имеют распределение Пуассона с параметром  $\lambda$ , и производящей функцией  $g(t) = e^{\lambda(t-1)}$ , а  $\tau$  — логарифмическое с вероятностями

$$P(\tau = j) = \frac{\alpha q^j}{j}, \text{ где } \alpha = -(\ln(1 - q))^{-1}, \quad j = 1, 2, \dots$$

с производящей функцией  $f(t) = -\alpha \ln(1 - qt)$ . Согласно [2], производящая функция случайной величины  $\zeta$  имеет вид суперпозиции производящих функций логарифмического и пуассоновского распределений,

$$h(t) = -\alpha \ln(1 - qe^{\lambda(t-1)}). \quad (1)$$

**Теорема 1.** Обозначим через  $S(k, j)$  числа Стирлинга второго рода [3], имеющие следующую рекуррентную формулу:

$$S(k, j) = S(k - 1, j - 1) + j \cdot S(k - 1, j), \quad 0 < j \leq k,$$

$S(0, 0) = 1, S(k, 0) = 0$  при  $k > 0, S(k, j) = 0$  при  $j > k$ . Пусть  $Q(\lambda, q) = \frac{qe^{-\lambda}}{1 - qe^{-\lambda}}$ , где  $\lambda > 0$  и  $q \in (0, 1)$ . Тогда вероятности логарифмически-пуассоновского распределения с параметрами  $\lambda, q$  и с производящей функцией (1) имеют вид:

$$P(\zeta_\tau = k) = \frac{\alpha}{k!} \lambda^k \sum_{j=1}^k (j-1)! S(k, j) Q(\lambda, q)^j \text{ при } k = 1, 2, \dots, \quad (2)$$

$$P(\zeta_\tau = 0) = -\alpha \ln(1 - qe^{-\lambda}).$$

Вероятности (2) получены через формулу для производящих функций:  $P(S_\tau = k) = \frac{1}{k!} h^{(k)}(0)$ ,  $k = 0, 1, 2, \dots$ , а доказательство теоремы осуществлено по методу математической индукции. Для чисел Стирлинга второго рода известна явная формула  $S(k, j) = \frac{1}{j!} \sum_{i=0}^j (-1)^{j+i} C_j^i i^k$ , поэтому не видно никаких ограничений для использования (2). Однако вероятности (2) можно получить при помощи специальных чисел другого вида. Также методом математической индукции доказано следующее утверждение.

**Теорема 2.** Обозначим через  $E(k, j)$  числа Эйлера первого рода [3] для которых справедлива рекуррентная формула:

$$E(k, j) = (j + 1) \cdot E(k - 1, j) + (k - j) \cdot E(k - 1, j - 1), 0 < j < k - 1,$$

$$E(k, 0) = 1 \text{ при } k \geq 0, E(k, j) = 0 \text{ при } j \geq k > 0.$$

Вероятности логарифмически-пуассоновского распределения для  $k > 0$  могут быть выражены следующим образом:

$$P(\zeta_\tau = k) = \frac{\alpha}{k!} \frac{\lambda^k q e^{-\lambda}}{(1 - q e^{-\lambda})^k} \sum_{j=0}^{k-2} E(k - 1, j) (q e^{-\lambda})^j. \quad (3)$$

Числа Эйлера (табл.1) представляют собой количество перестановок порядка  $k$  с  $j$  подъёмами, то есть в перестановке  $\pi = (\pi_1, \dots, \pi_k)$  всего  $j$  индексов  $i$  таких, что  $\pi_i < \pi_{i+1}$ .

$k \setminus j$	0	1	2	3	4	5
0	1					
1	1					
2	1	1				
3	1	4	1			
4	1	11	11	1		
5	1	26	66	26	1	
6	1	57	302	302	57	1

Таблица 1: Числа Эйлера.

Эти числа, явный вид которых  $E(k, j) = \sum_{i=0}^j C_{k+1}^i (-1)^i (j + 1 - i)^k$ , возникают также в функциональном ряде

$$\sum_{i=1}^{\infty} i^k x^i = \frac{x}{(1 - x)^{k+1}} \sum_{j=0}^{k-1} E(k, j) x^j,$$

который в нашем случае возникает при использовании формулы полной вероятности

$$\begin{aligned} P(\zeta_\tau = k) &= \sum_{j=1}^{\infty} P(\zeta_\tau = k | \tau = j) \cdot P(\tau = j) = \\ &= \sum_{j=1}^{\infty} -\frac{1}{\ln(1 - q)} \frac{q^j}{j} \frac{(j\lambda)^k}{k!} e^{-j\lambda} = \frac{\alpha}{k!} \lambda^k \sum_{j=1}^{\infty} j^{k-1} (q e^{-\lambda})^j. \end{aligned}$$

## Оценка параметров и пример из радиобиологии

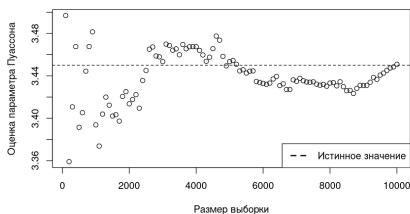


Рис. 1: Сходимость оценки параметра  $\lambda$  к истинному значению для логарифмически-пуассоновского распределения.

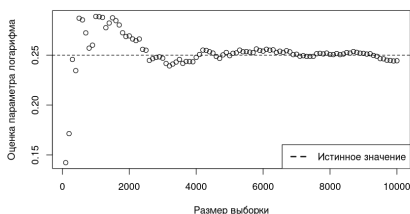


Рис. 2: Сходимость оценки параметра  $q$  к истинному значению для логарифмически-пуассоновского распределения.

Оценка параметров  $\lambda$  и  $q$  проводилась методом максимального правдоподобия численно на компьютере с использованием функции `optim` на языке R. Моделируя выборку из  $n$  индивидов для известных параметров можно проверить состоятельность такой оценки (рис. 1, 2).

В таблице 2 представлены оценки параметров для радиобиологических данных из статьи [1], а также показано согласие с эмпирическим распределением. При уровне значимости  $\alpha = 0.05$  получаем согласие для  $16/19 \cdot 100\% = 84\%$  случаев.

Наблюдаемое количество аномалий определяется в целом двумя факторами: их исходной распространенностью и интенсивностью их образования в процессе митоза. За увеличение исходной распространенности отвечает параметр  $q$  логарифмического распределения, а за интенсивность образования аномалий в процессе митоза параметр  $\lambda$  пуассоновского распределения. Поскольку распределения суммы и самих случайных величин однопараметри-

In vivo				In vitro			
Гр	$\lambda$	$q$	<b>p-v</b>	Гр	$\lambda$	$q$	<b>p-v</b>
0	0.38	5.9e-7	<b>0.13</b>	0	0.39	3.7e-6	<b>0.59</b>
5	0.67	3.3e-7	<b>0.81</b>	5	0.14	0.80	<b>0.73</b>
10	0.83	6.7e-7	<b>0.21</b>	10	0.59	1.1e-7	<b>0.03</b>
15	1.15	5.3e-7	<b>0.01</b>	15	0.41	0.76	<b>0.36</b>
20	1.71	1e-5	<b>0.03</b>	20	0.37	0.56	<b>0.45</b>
25	1.04	0.07	<b>0.55</b>	25	0.88	0.37	<b>0.22</b>
30	1.48	0.33	<b>0.33</b>	30	0.78	0.53	<b>0.26</b>
35	1.75	0.19	<b>0.11</b>	35	1.10	0.43	<b>0.43</b>
40	2.05	0.22	<b>0.16</b>	40	1.46	0.29	<b>0.38</b>
45	2.36	0.18	<b>0.08</b>				

Таблица 2: Оценки параметров и значимости критерия хи-квадрат по данным in vivo и in vitro для логарифмически-пуассоновского распределения.

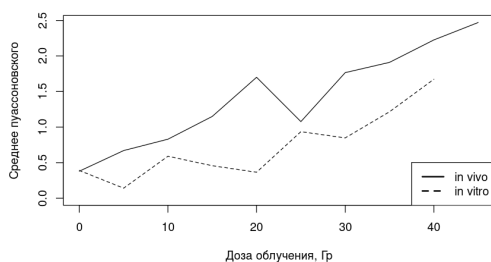


Рис. 3: Оценка параметра  $\lambda$  в зависимости от дозы облучения

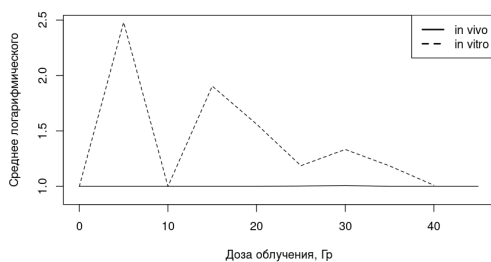


Рис. 4: Среднее лог. распределения в зависимости от дозы облучения

ческие, при интерпретации параметров можно опираться на их средние значения.

Динамика оценок параметра  $\lambda$ , соответственно средних значений, свидетельствует о положительной линейной зависимости их от дозы облучения, что очевидно, и о значимо меньших значениях в эксперименте *in vitro* (рис. 3), так как выжившие клетки обладают большим иммунитетом.

Что касается исходной распространенности аномалий, то, согласно динамике оценок параметра  $q$  в зависимости от дозы облучения (рис. 4), зависимости от дозы облучения нет, а в эксперименте *in vivo* базовая распространенность практически не выражена и существенно меньше, чем в эксперименте *in vitro*. Это объясняется тем, что от дозы облучения зависит только количество выживших клеток, а не распространенность их аномалий, которая, судя по графику, очень вариабельна, но существенно выше базовой распространенности до начала эксперимента.

## Обобщение на случай произвольных распределений

Рассмотрим модель сложного распределения  $\zeta_\tau = \xi_1 + \dots + \xi_\tau$ , в котором случайное число слагаемых  $\tau$  распределено по  $(P_0, P_1, P_2, \dots)$ , а независимые компоненты по  $(p_0, p_1, p_2, \dots)$ . Вероятности сложного закона распределения можно выразить через конечную суммы

$$P\{\zeta_\tau = n\} = \frac{1}{n!} \sum_{k=1}^n \Theta_k G_n^k, \quad \text{где } \Theta_k = \sum_{i=k}^{\infty} P_i C_i^{i-k} p_0^{i-k},$$

$$G_n^k = n! \sum_{\sum_i^k n_i = n} \prod_{i=1}^k p_{n_i}$$

и  $n = (n_1, \dots, n_k), n_i > 0$ , упорядоченные разбиения натурального числа  $n$ . Сравнивая выражения  $G_n^k$  в случае логарифмических и пуассоновских независимых величин, приходим к двум видам суммирования над упорядоченными разбиениями:

$$u_n^k = \sum_{\sum_i^k n_i = n, n_i > 0} \frac{n!}{n_1 \dots n_k}, \quad v_n^k = \sum_{\sum_i^k n_i = n, n_i > 0} \frac{n!}{n_1! \dots n_k!}$$

с рекуррентными соотношениями

$$u_n^k = k u_{n-1}^{k-1} + (n-1) u_{n-1}^k, \quad v_n^k = k (v_{n-1}^{k-1} + v_{n-1}^k),$$

которые приводят к существенным частным случаям в виде чисел Стирлинга первого и второго рода  $s(n, k) = \frac{1}{k!} u_n^k$  и  $S(n, k) = \frac{1}{k!} v_n^k$ . В случае пуассоновского распределения независимых компонент сумма по упорядоченным разбиениям имеет тип  $v_n^k$ , а в случае логарифмического тип  $u_n^k$ .

При логарифмическом распределении (2) числа компонент  $\tau$  имеем  $\Theta_k = \frac{\alpha}{k} \left( \frac{q}{1-qp_0} \right)^k$ , а при пуассоновском  $\Theta_k = \frac{\lambda^k}{k!} e^{-\lambda + \lambda p_0}$ . Отличие в константах  $\frac{1}{k}$  или  $\frac{1}{k!}$  приводит к разному виду коэффициентов. Например, пуассоновское  $\tau$  с внутренним логарифмическим дает сочетание  $\frac{1}{k!} u_n^k$ , приводящее напрямую к степенному ряду с коэффициентами в виде чисел Стирлинга первого рода, соответственно к убывающему факториалу и к вероятностям отрицательного биномиального распределения. Реинтрантный пуассон приводит к вероятностям с весами в виде чисел Стирлинга второго рода, реинтрантный логарифм или логарифм-пуассон дадут соответственно взвешенные числа Стирлинга первого и второго рода вида  $(k-1)!s(n, k)$  и  $(k-1)!S(n, k)$ .

## Заключение

Конечные суммы над упорядоченными разбиениями, связанные с числами Стирлинга, позволяют систематизировать разные виды сложных распределений и тем самым ускорить процесс выбора наиболее адекватной модели.

## Список литературы

- [1] Динамика роста числа ядерных аномалий рабдомиосаркомы RA-23 при увеличении дозы острого редкоизирующего облучения. Исследование на основе модели реинтрантно-биномиального распределения / Алексеева Н. П., Алексеев А. О., Вахтин Ю. Б., Кравцов В. Ю., Кузоватов С. Н. и Скорикова Т. И. // Цитология. 2008. — С. 528–534.
- [2] Феллер В. Введение в теорию вероятностей и её приложения. В 2 т. Москва : Мир, 1952. Т. 1.
- [3] Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. Москва : Мир, 1998. ISBN: 5-03-001793-3.