

# 1. Обобщенное и реинтрантное распределения в радиобиологии

## 1.1. Обобщенное распределение Пуассона

Случайная величина  $\zeta$  имеет обобщенное Пуассоновское распределение  $\mathcal{P}(\lambda|f(x))$ , если она равна сумме  $\tau$  независимых одинаково распределенных случайных величин  $\xi_i$  с плотностью распределения  $f(x)$ , где  $\tau \sim \mathcal{P}(\lambda)$ ,

$$\zeta = \begin{cases} 0, & \text{при } \tau = 0 \text{ с вероятностью } p_0 = e^{-\lambda} \\ \xi_1 + \dots + \xi_\tau, & \text{при } \tau > 0 \text{ с вероятностью } 1 - e^{-\lambda} \end{cases}.$$

Обозначим через  $F_j(x) = P\{\xi_1 + \dots + \xi_j \leq x\}$  функцию распределения и через  $f_j(x)$  соответствующую плотность, через  $\mu, \sigma^2$  соответствующие МО и дисперсию. Применим формулу полной вероятности в

$$\begin{aligned} P\{\xi_1 + \dots + \xi_\tau \leq x | \tau > 0\} &= \frac{P\{\xi_1 + \dots + \xi_\tau \leq x, \tau > 0\}}{P\{\tau > 0\}} = \\ &= \frac{1}{1 - p_0} \sum_{j=1}^{\infty} p_j P\{\xi_1 + \dots + \xi_j \leq x\} = \frac{1}{1 - p_0} \sum_{j=1}^{\infty} p_j F_j(x), \end{aligned}$$

таким образом, условная плотность имеет вид  $\frac{1}{1-p_0} \sum_{j=1}^{\infty} p_j f_j(x)$ . Обозначим через  $g(t)$  ПФМ для СВ  $\xi_i$ . Вычислим ПФМ для СВ  $\zeta$ .

$$\begin{aligned} \phi(t) = \mathbb{E}e^{t\zeta} &= p_0 e^{t \cdot 0} + (1 - p_0) \int_{-\infty}^{\infty} \frac{1}{1 - p_0} \sum_{j=1}^{\infty} p_j f_j(x) e^{tx} dx = \\ &= p_0 + \sum_{j=1}^{\infty} p_j \int_{-\infty}^{\infty} f_j(x) e^{tx} dx = p_0 + \sum_{j=1}^{\infty} p_j (g(t))^j = \\ &= \sum_{j=0}^{\infty} p_j (g(t))^j = e^{-\lambda + \lambda g(t)}, \end{aligned}$$

так как  $p_j = \frac{\lambda^j}{j!} e^{-\lambda}$ , и ПФ распределения Пуассона имеет вид

$$\sum_{j=0}^{\infty} p_j \nu^j = \sum_{j=0}^{\infty} \frac{(\lambda \nu)^j}{j!} e^{-\lambda} = e^{-\lambda + \lambda \nu}.$$

Аналогично для производящей функции  $h(\nu) = \sum_{j=0}^{\infty} p_j \nu^j = \mathbb{E}\nu^\zeta$

$$\begin{aligned} h(\nu) &= \mathbb{E}\nu^\zeta = p_0 \nu^0 + (1 - p_0) \int_{-\infty}^{\infty} \frac{1}{1 - p_0} \sum_{j=1}^{\infty} p_j f_j(x) \nu^x dx = \\ &= p_0 + \sum_{j=1}^{\infty} p_j \int_{-\infty}^{\infty} f_j(x) \nu^x dx = p_0 + \sum_{j=1}^{\infty} p_j G_j(t) = \\ &= p_0 + \sum_{j=1}^{\infty} p_j (g(t))^j = \sum_{j=0}^{\infty} p_j (g(t))^j = f(g(t)), \end{aligned}$$

## 1.2. Метод моментов

Рассматривая количество моментов, равное количеству оцениваемых параметров, и приравнивая выборочные моменты к соответствующим моментам распределения, являющимся функциями от параметров, можно получить оценки параметров распределения.

В случае сложно-пуассоновского распределения  $\mathcal{P}(\lambda | \mathcal{N}(\mu, \sigma))$  с ПФМ  $g(t)$  внутреннего нормального распределения имеет вид  $\phi(t) = e^{-\lambda + \lambda g(t)}$ . Используя то, что  $g'(0) = \mu$ ,  $g''(0) = \mu^2 + \sigma^2$ , можем вычислить  $\mathbb{E}\zeta$ ,  $\mathbb{D}\zeta$ .

$$\begin{aligned} \phi'(t) &= e^{-\lambda + \lambda g(t)} \lambda g'(t), \quad \phi'(0) = \lambda \mu \\ \phi''(t) &= e^{-\lambda + \lambda g(t)} (\lambda g'(t))^2 + e^{-\lambda + \lambda g(t)} \lambda g''(t), \quad \phi''(0) = \lambda^2 \mu^2 + \lambda(\mu^2 + \sigma^2), \\ \mathbb{E}\zeta &= \lambda \mu, \quad \mathbb{D}\zeta = \lambda(\mu^2 + \sigma^2). \end{aligned}$$

Если имеется выборка  $x_1, \dots, x_n$  и  $\bar{x}$  выборочное среднее,  $S^2$  несмещенная оценка дисперсии, то в  $\hat{p}_0 = \frac{n_0}{n}$ , где  $n_0$  число нулей в выборке,

$$\hat{\lambda} = -\ln(\hat{p}_0), \quad \hat{\mu} = \frac{\bar{x}}{\hat{\lambda}}, \quad \hat{\sigma}^2 = \frac{S^2}{\hat{\lambda}} - \hat{\mu}^2. \quad (1)$$

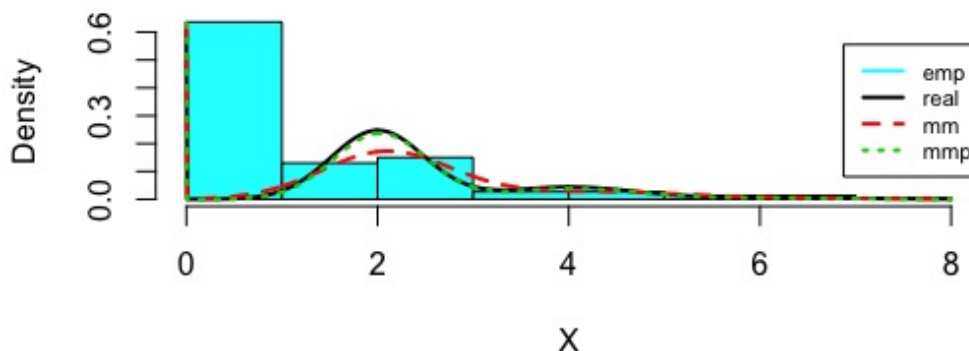


Рис. 1. Сложно-пуассоновское распределение, согласование гистограммы с плотностями при разных параметрах.

### 1.3. Проблема выбора адекватной модели

При исследовании изменения вероятности формирования интерфазных опухолевых клеток с одиночными или множественными аномалиями<sup>1</sup> на ядре под воздействием облучения была предпринята попытка найти подходящую статистическую модель.

В эксперименте рассматривались две популяции крыс. В одной серии опытов ядерные аномалии были исследованы в популяциях подкожных клеточных трансплантатов перевивной рабдомиосаркомы РА-23, фиксированных через 52 часа (достаточный срок для 2-3 митозов) с момента облучения X-ray *in vivo* дозами 5-45 Гр<sup>2</sup>. В другой серии опытов были исследованы ядерные аномалии у отдаленных потомков этих трансплантатов. Суспензированные клетки прививались путем инъекции в переднюю камеру глаза

<sup>1</sup> Аномалиями считались межъядерные мосты, ядерные протрузии и гантелевидные ядра.

<sup>2</sup> Грей (обозначение: Гр, Gy) — единица измерения поглощенной дозы ионизирующего излучения в системе СИ. Поглощенная доза равна одному грею, если в результате поглощения ионизирующего излучения вещество получило один джоуль энергии в расчёте на один килограмм массы.

другого животного сразу после облучения *in vitro*. Распределения числа аномалий представлены в табл. 1.

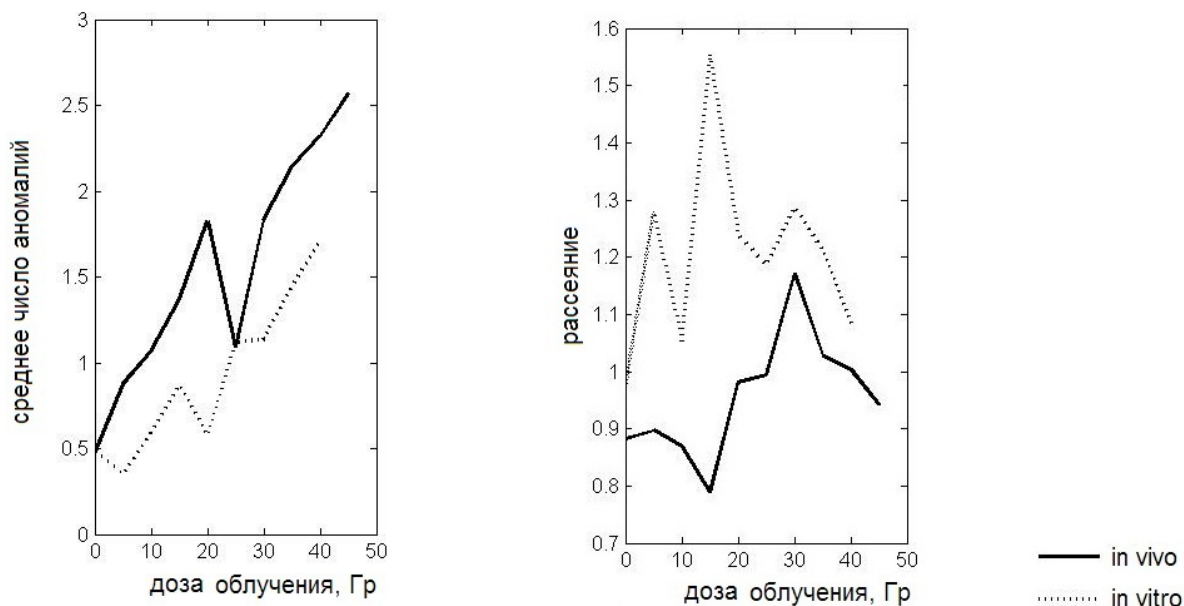


Рис. 2. Динамика среднего числа аномалий и рассеяния.

Результаты первичной статистики, во-первых, свидетельствуют о том, что существует значимая положительная корреляция между средним числом аномалий на ядре и дозой облучения (рис. 2), во-вторых, распределение меняет свой характер рассеяния<sup>3</sup> – при увеличении дозы облучения распределение из недорассеянного становится перерассеянным (рис. 2). В связи с этим недорассеянное положительно биномиальное или перерассеянное отрицательно биномиальное распределения не могут быть использованы в качестве модели описания роста числа аномалий при увеличении дозы облучения.

<sup>3</sup> Рассеяние  $e$  – это отношение дисперсии к среднему. Пуассоновское распределение имеет рассеяние, равное единице. Распределение с  $e < 1$  называется недорассеянным, с  $e > 1$  перерассеянным.

Таблица 1. Распределения числа аномалий на ядре.

Доза, Гр	in vitro						in vivo					
	число аномалий						число аномалий					
	0	1	2	3	4	> 5	0	1	2	3	4	> 5
0	68	25	7				66	31	2	1		
5	74	19	5	2			50	35	13	2		
10	59	24	16	1			41	39	17	2	1	
15	48	33	11	3	2	3	27	39	29	3	1	1
20	59	31	5	4	1		22	22	32	15	6	3
25	37	31	22	6	2	2	33	39	18	8	1	1
30	35	37	17	5	3	3	21	29	21	14	10	5
35	26	36	19	10	5	4	17	24	32	11	10	6
40	19	33	25	11	7	5	15	17	29	18	11	10
45							13	14	24	24	12	13

#### 1.4. Реинтрантный бином

На уровне характеристики рассеяния выясним возможность использования модели сложного распределения для описания роста числа ядерных аномалий при увеличении дозы облучения.

**Определение 1.** *Сложным* называется распределение случайной суммы независимых одинаково распределенных случайных величин

$$\xi = \eta_1 + \dots + \eta_\tau,$$

где  $\tau$  и  $\eta_j$  также независимые случайные величины [? ].

Производящая функция  $h(\nu) = \sum_{j=0}^{\infty} p_j \nu^j$ , где  $p_j = P\{\xi = j\}$ , сложно распределенной случайной величины  $\xi$  имеет вид суперпозиции производящих функций

$$h(\nu) = f_1(f_0(\nu)), \quad (2)$$

где  $f_1(\nu)$  и  $f_0(\nu)$  – производящие функции распределений  $\tau$  и  $\eta_i$  соответственно.

Пусть  $\tau$  и  $\eta_i$  имеют биномиальные законы распределения с параметрами  $\beta(\cdot|n_1, p_1)$  и  $\beta(\cdot|n_0, p_0)$ . Производящая функция случайной величины  $\xi_1$  имеет вид:

$$h_1(\nu) = f_1(f_0(\nu)) = ((p_0\nu + q_0)^{n_0} p_1 + q_1)^{n_1}, \quad (3)$$

а соответствующий закон распределения  $\beta(\cdot|p_1, p_0; n_1, n_1)$  будем называть *реинтрантным*, поскольку в [?] сложно-биномиальным называется распределение с пуассоновской случайной величиной  $\tau$ .

Используя свойства производных производящих функций (3)

$$\begin{aligned} h'(1) &= \mathbf{E}\xi, \quad h''(1) + h'(1) - (h'(1))^2 = \mathbf{D}\xi, \\ h'(1) &= f_1'(1)f_0'(1), \quad h''(1) = f_1''(1)(f_0'(1))^2 + f_1'(1)f_0''(1), \\ f_i'(1) &= n_i p_i, \quad f_i''(1) = n_i(n_i - 1)p_i^2, \end{aligned}$$

получаем выражения для первых двух моментов и рассеяния:

$$\mathbf{E}\xi = n_0 p_0 n_1 p_1, \quad \mathbf{D}\xi = n_0 p_0 n_1 p_1 (n_0 p_0 q_1 + q_0), \quad e = n_0 p_0 q_1 + q_0.$$

Поскольку возможно как  $e > 1$  при  $q_1 > \frac{1}{n_0}$ , так и наоборот  $e < 1$ , то модель реинтрантного биномиального распределения не противоречит наблюдаемым данным.

Математические ожидания  $\mathbf{E}\eta_i = \mu_0$ ,  $\mathbf{E}\tau = \mu_1$  будем называть *средними реинтрантными компонентами*.

### Предложение 1.

$$P\{\xi_1 = k_0\} = \sum_{k_1=\lceil \frac{k_0}{n_0} \rceil}^{n_1} \beta(k_1|n_1; p_1)\beta(k_0|n_0 k_1; p_0).$$

**Доказательство.** Поскольку индекс  $i$  в параметрах  $p_i$  и  $n_i$  в формуле  $\beta(\cdot|n_i; p_i)$  один и тот же, то обозначим  $\beta(\cdot|n_i; p_i) = \beta(\cdot|n_i)$ .

$$\begin{aligned} h_1(\nu) &= ((p_0\nu + q_0)^{n_0} p_1 + q_1)^{n_1} = \\ &= \sum_{k_1=0}^{n_1} \beta(k_1|n_1) \sum_{k_0=0}^{n_0 k_1} \beta(k_0|n_0 k_1) \nu^{k_0}. \end{aligned}$$

Изменим в этом выражении порядок суммирования:

$$h_1(\nu) = \sum_{k_0=0}^{n_0 n_1} \sum_{k_1=t}^{n_1} \beta(k_1|n_1) \beta(k_0|n_0 k_1) \nu^{k_0} = \sum_{k_0=0}^{n_0 n_1} P\{\xi_1 = k_0\} \nu^{k_0}.$$

Для того чтобы вычислить  $t$  – начало суммирования по  $k_1$ , рассмотрим соответствие между  $k_0$  и  $k_1$ .

$k_0 \setminus k_1$	0	1	2	...	$n_1$
0	0	0	0	...	0
1		1	1	...	1
$\vdots$		$\vdots$	$\vdots$	...	$\vdots$
$n_0$		$n_0$	$n_0$	...	$n_0$
$n_0 + 1$			$n_0 + 1$	...	$n_0 + 1$
$\vdots$			$\vdots$	...	$\vdots$
$2n_0$			$2n_0$	...	$2n_0$
$\vdots$					$\vdots$
$(n_1 - 1)n_0 + 1$					$(n_1 - 1)n_0 + 1$
$\vdots$					$\vdots$
$n_1 n_0$					$n_1 n_0$

Заметим, что при  $k_0 = 0$  справедливо  $k_1 = 0$ ,

$$\begin{aligned} 0 < k_0 \leq n_0 &\iff 0 < \frac{k_0}{n_0} \leq 1 \implies 1 \leq k_1 \leq n_1, \\ n_0 < k_0 \leq 2n_0 &\iff 1 < \frac{k_0}{n_0} \leq 2 \implies 2 \leq k_1 \leq n_1, \end{aligned}$$

и так далее, то есть  $t = \lceil \frac{k_0}{n_0} \rceil$ . Отсюда получаем  $P\{\xi_1 = k_0\}$ .  $\square$

### 1.5. Оценка параметров $\beta(\cdot|p_1, p_0; n_1, n_1)$

Пусть имеется вариационный ряд  $N_0, N_1, \dots, N_k$  с правым цензурированием. Поскольку для интерпретации параметров важна целочисленность параметров  $n_0$  и  $n_1$ , то по этим параметрам осуществлялся перебор, а параметры  $p_0$  и  $p_1$  оценивались численным образом из условия максимума логарифма функции правдоподобия

$$\ln \mathcal{L}(\cdot|p_1, p_0; n_1, n_1) = \sum_{j=0}^{k-1} N_j \ln P\{\xi = j\} + N_k \ln P\{\xi \geq k\}.$$

Оценки реинтрантных компонент равны  $\hat{\mu}_i = \hat{n}_i \hat{p}_i$ ,  $i = 0, 1$ . На рис. 3 представлены гистограммы эмпирических распределений с полигонами частот реинтрантного биномиального распределения  $\beta(\cdot|p_1, p_0; n_1, n_1)$ : а) *in vitro* (40 Гр),  $m_1 = 1.68$ ,  $m_0 = 1.32$ , б) *in vivo* (40 Гр)  $m_1 = 0.51$ ,  $m_0 = 3.29$ . Адекватность модели подтверждается высокими доверительными уровнями вероятности. Здесь следует обратить внимание на более симметричную при увеличении  $m_1$  форму распределения, аналогичную плотности гамма распределения при увеличении параметра экстенсивности.

На рис. 4 представлена динамика средних реинтрантных компонент в зависимости от дозы облучения в разных экспериментах. При увеличении дозы облучения в эксперименте *in vivo* значительно увеличивается внутренняя реинтрантная компонента  $m_0$  (корреляция равна  $r = 0.76$ ,  $p = 0.006$ ), в эксперименте *in vitro* внешняя реинтрантная компонента  $m_1$  ( $r = 0.95$ ,  $p = 0.00003$ ). Корреляции между дозой облучения и соответственно компонентами  $m_1$  в эксперименте *in vitro* и  $m_2$  в эксперименте *in vivo* отличаются от нуля незначимо.

Интерпретация реинтрантных компонент оказывается достаточно простой: внешняя реинтрантная компонента  $m_1$  рассматривается как экстенсивность или среднее число аномалий, полученное потомками от своих родителей, а внутренняя компонента  $m_0$  как среднее число аномалий, появившихся за период наблюдения (аналог параметра масштаба). Тогда логично, что у облученных клеток непосредственно от радиации зависит  $m_0$ , а у их потомков  $m_1$ .

Если сравнивать между собой характеристики клеток и их потомков, то можно обнаружить, что  $\hat{m}_0 = 0.64$  и  $\hat{m}_1 = 1.55$  у потомков ниже, чем у клеток  $\hat{m}_0 = 1.07$  и  $\hat{m}_1 = 1.86$  (статистически значимо по  $m_0$ ), что объясняется тем, что потомки не были под облучением непосредственно, кроме того при повышении дозы облучения имплантируемых клеток процент успешных прививок падает, а формирование привитых трансплантатов замедляется.



На рис. 5 отображены характеристики прививаемости опухолевых клеток: А) время от прививки до начала роста опухолей; Б) время от прививки до заполнения растущей опухолью глазной полости, увеличивающиеся по мере увеличения дозы облучения. В скобках указаны проценты успешных прививок. При дозе 45 Гр все прививки оказались неудачными.

## 1.6. Обобщенные положительно биномиальные распределения

Обобщенное негативно биномиальное распределение вводится как распределение суммы  $k$  независимых обобщенных геометрически распределенных случайных величин (теор. ??) и обозначается через  $\beta_{-}^{*A}(j|k, p, \alpha)$  или  $\beta_{-}^{*B}(j|k, p, \alpha)$  в зависимости от вида частично обратной  ${}_{\alpha}\xi^{-}(0)$  или  $\xi_{\alpha}^{-}(0)$ .

Обобщенное позитивно биномиальное распределение определяется через фидуциальное тождество:

$$\beta_{+}^{\leq}(k|n, p, \alpha) = 1 - \beta_{-}^{\leq}(n|k, p, \alpha). \quad (4)$$

В явном виде эти распределения громоздки, но для практических целей используются рекуррентные соотношения, непосредственно выводимые из обобщения случайного блуждания по целочисленной решетке. В некоторых приложениях применяется обобщенное биномиальное распределение  $\beta_{+}^{*C}(\cdot|n, p_1, p_2, p_3)$ , фидуциальным образом (4) построенное из свертки  $C$ -обобщения геометрического  $\beta_{-}^C(\cdot|p_1, p_2, p_3)$  с блужданием на решетке, изображенным на рис. 6.

## 1.7. Сложно-обобщенно-биномиальная модель

Для изучения структуры воздействия радиации используем расширение модели: биномиальные распределения внутренней (*in vivo*) или внешней (*in vitro*) компоненты сложного распределения заменяются обобщен-

ным распределением  $\beta_+^{*C}(\cdot|n, p_1, p_2, p_3)$  из раздела 1.6, которое характеризует работу нетренированной системы с более общими параметрами, не связанными частичностью  $\alpha$ . В качестве распределения противоположной компоненты используется биномиальное с оценками параметров, полученными из модели реинтрантного бинома ( $\hat{n}_0 = 3, \hat{p}_0 = 0.505$  *in vivo* и  $\hat{n}_0 = 5, \hat{p}_0 = 0.133$  *in vitro*).

Общий подход к интерпретации параметров обобщенного бинома заключается в непостоянстве вероятности успеха. В данном примере „успех“ означает образование аномалии, а “неудача” происходит за счет сопротивляемости клеток, одной из форм проявления которой является репарация. После первой неудачи (первичной репарации) вероятности образования аномалии, как показывают наблюдения, уменьшаются, но меняются циклически. В связи с этим параметр  $q_1$  интерпретируется как параметр первичной репарации,  $q_2$  и  $q_3$  как циклические параметры вторичной репарации, которая задействуется после того, как система „осознает“, что нужно предпринимать дополнительные меры. Более высокие  $q_2$  характерны для системы, которая сразу концентрируется для ответного удара, а  $q_3$  для той, которая „долго собирается“.

Оценки параметров  $p_1, p_2, p_3$  приведены в таблице 2. Графики зависимости вероятностей репарации  $q_1, q_2, q_3$  от дозы облучения представлены на рис. 7. Обращает на себя внимание тот факт, что до уровня 20-25 Гр параметры репарации *in vivo* и *in vitro* довольно похожи. Для эксперимента *in vivo* характерно ступенчатое снижение  $q_1$  на уровне 15 Гр до 0.6-0.7,  $q_2$  на уровне 30 Гр до 0.4, на  $q_3$  больше влияет случайность. У потомков, среди которых, разумеется, больше жизнеспособных, оказывается высокая вероятность вторичной репарации  $q_2$ .

Таблица 2. Оценки  $\hat{p}_1, \hat{p}_2, \hat{p}_3$  при  $n_1 = 3$  (in vivo),  $n_0 = 5$  (in vitro).

доза (in vivo )	$n_0$	$p_1$	$p_2$	$p_3$	$P_{\chi^2}$	$n_0$	$p_1$	$p_2$	$p_3$	$P_{\chi^2}$
0	2	0.07	0.20	0.68	0.80	4	0.04	0	0.21	0.78
5	2	0.167	0.30	0.63	0.99	4	0	0	0.53	0.54
10	2	0.20	0.39	0.54	0.96	4	0.02	0	0.55	0.79
15	3	0.21	0.00	0.63	0.44	5	0.4	0	0.2	0.39
20	2	0.60	0.43	0.44	0.97	4	0.26	0	0.7	0.94
25	3	0.28	0.02	0.44	0.98	3	0.28	0.02	0.44	0.98
30	3	0.50	0.00	0.62	1.00	4	0.25	0.38	0.2	0.93
35	3	0.30	0.76	0.00	0.73	4	0.13	0.58	0	0.73
40	4	0.21	0.63	0.13	0.99	4	0.21	0.63	0.13	0.99
45	3	0.52	0.60	0.86	0.96	4	0.3	0.44	0.89	0.93
доза (in vitro)	$n_1$	$p_1$	$p_2$	$p_3$	$P_{\chi^2}$	$n_1$	$p_1$	$p_2$	$p_3$	$P_{\chi^2}$
0	1	0.63	0.45	0.42	0.97	4	0	0	0.68	0.96
5	2	0.20	0.37	0.65	0.99	4	0.00	0	0.53	0.96
10	2	0.34	0.62	0.47	0.53	4	0.06	0.14	0.65	0.51
15	5	0.38	0	0.42	0.35	5	0.38	0	0.42	0.37
20	3	0.33	0	0.62	0.63	3	0.33	0	0.62	0.63
25	4	0.48	0	0.80	0.95	4	0.48	0	0.80	0.95
30	3	0.64	0	1	0.63	4	0.49	0	0.82	0.58
35	5	0.51	0	0.82	0.94	4	0.63	0	0.90	0.92
40	4	0.72	0.03	1	0.98	4	0.726	0.02	1.00	0.98

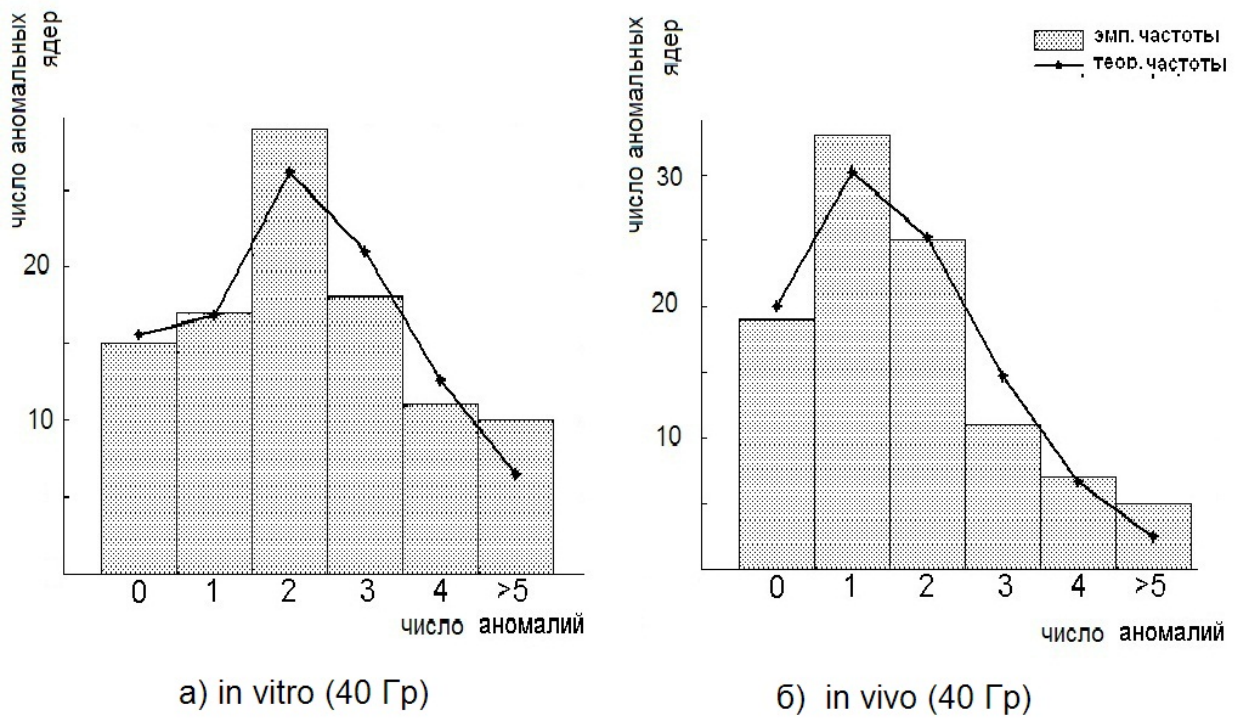


Рис. 3. Оценка адекватности модели  $\beta(\cdot|p_1, p_0; n_1, n_1)$ .

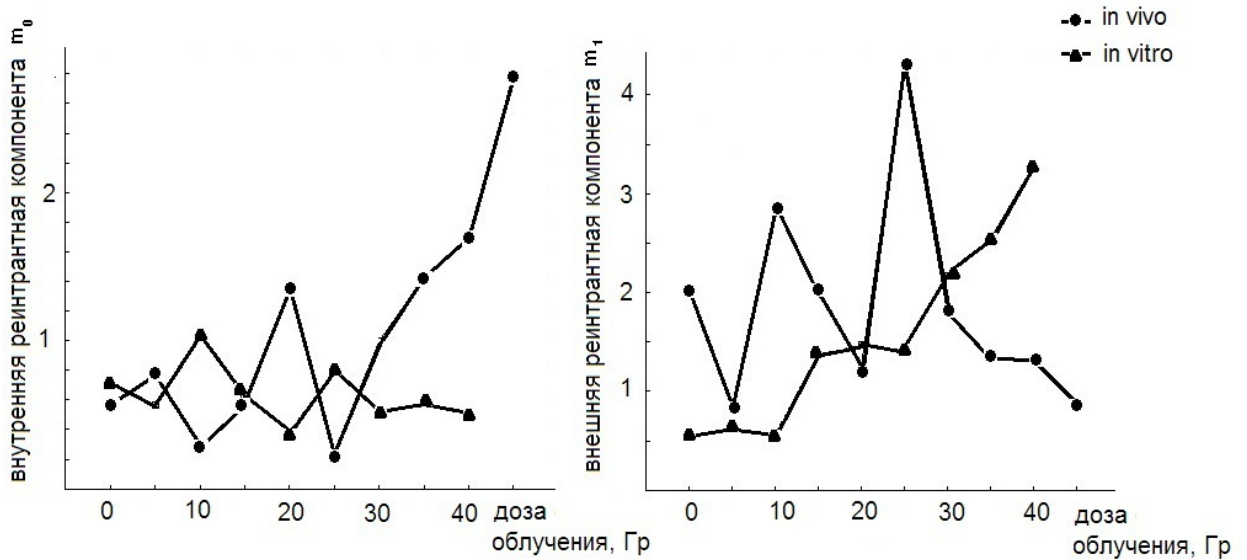


Рис. 4. Динамика средних реинтрантных компонент.

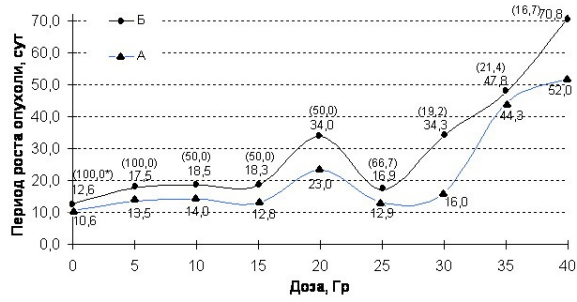


Рис. 5. Характеристики прививаемости.

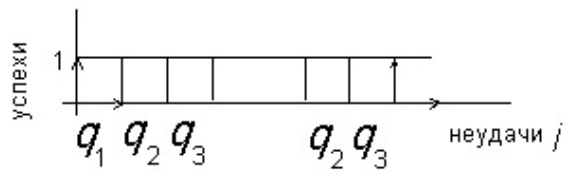


Рис. 6. Схема  $\beta_-^C(j|p, \alpha)$ -блуждания по решетке.

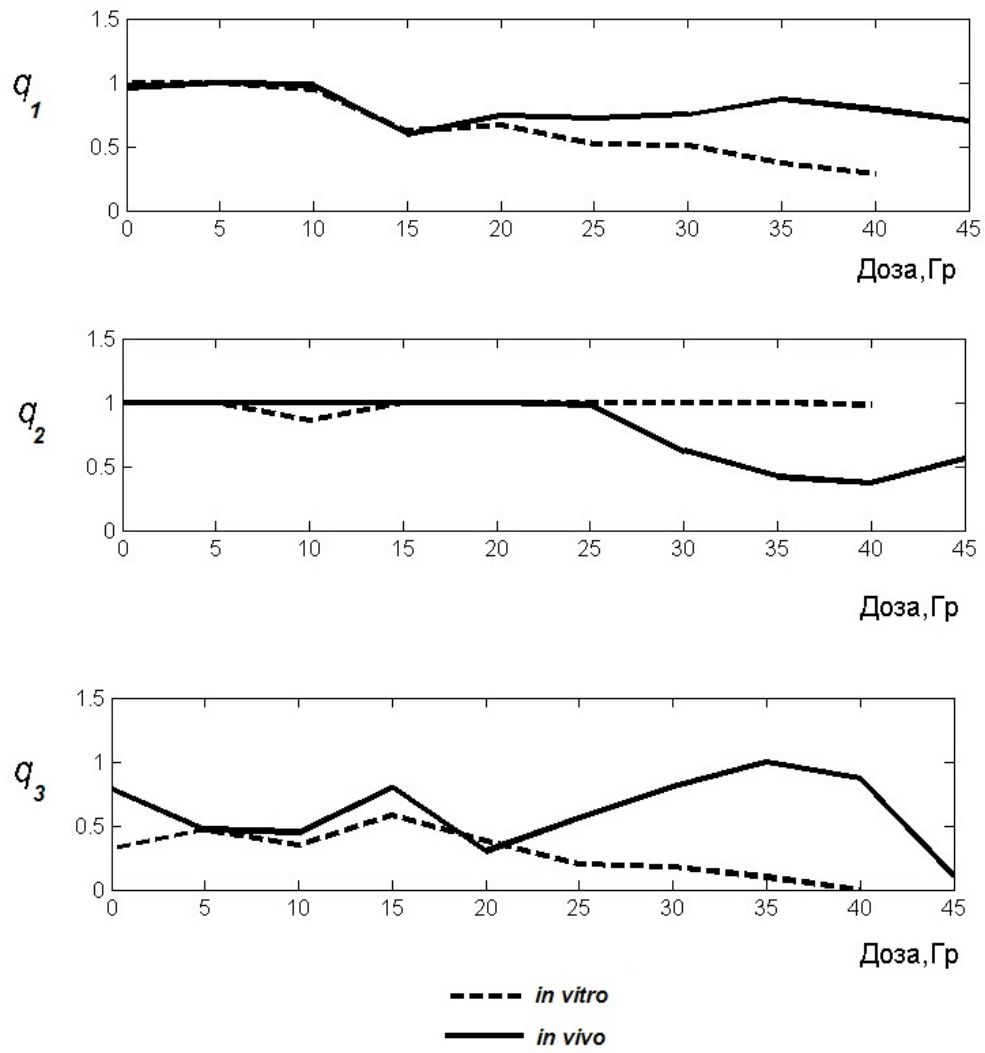


Рис. 7. Динамика вероятностей репарации.