

Н.П. АЛЕКСЕЕВА

АНАЛИЗ  
МЕДИКО-  
БИОЛОГИЧЕСКИХ  
СИСТЕМ

РЕЦИПРОКНОСТЬ  
ЭРГОДИЧНОСТЬ  
СИНОНИМИЯ

Издательство С.-Петербургского университета

Санкт-Петербургский государственный университет

Н.П. Алексеева

АНАЛИЗ  
МЕДИКО-БИОЛОГИЧЕСКИХ  
СИСТЕМ

Реципрокность, эргодичность, синонимия

Санкт-Петербург  
2012

УДК 519.248:[57+61]

ББК 56.6+22.172

Б24

Р е ц е н з е н т ы: проф., д-р физ.мат.наук, *С.М.Ермаков* (С.-Петербург. гос. ун-т), проф., д-р физ.мат.наук, *В.Б.Невзоров* (С.-Петербург. гос. ун-т), проф., д-р мед.наук, *Б.Б.Бондаренко* (ФГУ ФЦСКЭ им.В.А.Алмазова)

*Печатается по постановлению  
Редакционно-издательского совета  
Ученого совета математико-механического факультета  
С.-Петербургского государственного университета*

Алексеева Н.П.

Б24 **Анализ медико-биологических систем. Реципрокность, эргодичность, синонимия.** — СПб.: Изд-во С.-Петерб. ун-та, 2012. — 184 с.

В монографии идет речь о расширении спектра математических методов статистического анализа медико-биологических наблюдений, в частности, о применении конечных групп в информационном структурировании категориальных данных, о марковских свойствах эволюционных процессов, об обобщенных на основе частичного обращения распределениях в моделях немонотонного риска и т.д. с приложениями в кардиологии, иммунологии, когнитивной биологии и т.п. Ключевые направления агрегированы в три феномена, важные для поддержания концептуальной целостности изучаемой системы: основу сбалансированности реципрокность, индикатор ее восстановления эргодичность и естественный атрибут неопределенности синонимия.

Монография ориентирована на математиков, но может оказаться полезной для других специалистов, которые интересуются проблемами обработки данных и анализа деятельности живых систем.

**ББК 22.172**

© Н.П.Алексеева, 2012  
© С.-Петербургский  
государственный  
университет, 2012

## Введение

Прогресс в развитии вычислительной техники и программного обеспечения открыл доступ к биостатистике специалистам широкого профиля и разного уровня компетентности. В настоящее время „экспериментаторы“, которыми по традиции биометрического семинара <sup>1</sup> называются хозяева статистических данных: врачи, биологи, лингвисты, историки, психологи — при небольшой математической поддержке могут самостоятельно воспользоваться нужной программой и интерпретировать результаты в соответствии с собственными задачами.

От специалистов с профессиональной математической подготовкой требуется четко ориентироваться в том, где, когда и при каких условиях можно использовать тот или иной метод. Более сложные задачи возникают при появлении каких-то нестандартных условий и в ситуации, когда известные методы не позволяют адекватно ответить на вопрос. В целом, в биостатистике используются те же методы, что и в других статистических приложениях, но в анализе живых систем обращают на себя внимание некоторые феномены, которые с завидным постоянством обнаруживаются в разных задачах и несут важную смысловую нагрузку для их концептуальных решений. Речь идет о реципрокности, эргодичности и синонимии, поскольку для любой живой системы характерна структура взаимодействия ее компонент, при которой осуществляется перераспределение напряжений, возвращение в стационарное состояние и взаимозаменяемость частей. В связи с этим актуальна разработка математических методов, имеющих непосредственное отношение к этим явлениям.

Кроме этого, за последние годы претерпели изменения и сами медико-биологические данные. Упоная на информационные технологии, врач-исследователь, имеющий дело, как правило, с огромным списком жалоб на самочувствие и внушительным набором снимаемых показателей, выгружает все это в свою базу данных. Но если с набором количественных показателей биостатистики научены справляться благодаря линейным статистическим методам, то обработка набора категориальных данных, тем более с пропусками,

---

<sup>1</sup>Биометрический семинар был организован на математико-механическом факультете в 1963 году под руководством О.М.Калинина. Долгое время соорганизатором семинара и его идейным вдохновителем был А.Г.Барт (1941-2007).

обычно выполняется поверхностно, однобоко и фрагментарно из-за высокой требовательности к объему выборки многофакторных моделей. Для более подробного исследования данных такого рода можно использовать то, что факторы выражаются через проективные пространства, интерпретируемые с симптомно-сидромальной точки зрения. Разработке концепции реципрокности и синонимии через проективные методы в статистике посвящена глава 1. В ней на основе информационных и теоретико-групповых методов рассматриваются задачи систематизации, классификации, дискриминации, прогнозирования.

Во второй главе исследуются проблемы неоднородности и неполноты данных. Через расслоение популяций по наиболее информативным признакам решается задача улучшения классификации. На основе марковских свойств описывается состояние равновесия в законе Харди-Вайнберга. Вычисляется эргодическое смещение для планов с расщепленными блоками, применяемых в дисперсионном анализе повторных наблюдений с пропусками. Обычно пропуски стараются заполнить чем-то искусственным. В таком случае неоправданно увеличиваются степени свободы. При удалении неполных данных просто теряется информация. Существование при небольших допущениях эргодического распределения средних по времени позволяет от смещенной модели с независимыми ошибками перейти к несмещенной модели с коррелированными ошибками и тем самым свести задачу к стандартному решению.

В главе 3 рассматриваются модели с управляющими параметрами: частичности в обобщенных распределениях по Барту и сжатия-растяжения в степенных гамма распределениях. В случае синонимии, то есть когда одно и то же явление описывается разными моделями, основная проблема заключается в построении дополнительных критериев оптимальности и дифференциации. Возникают новые задачи: либо поиска самого „удобного“ распределения, либо, если оснований к этому нет, проверки гипотез для всей совокупности моделей одновременно. Это направление имеет важное прикладное значение для описания структуры немоного риска (время восстановления после холодной пробы диаметра лучевой артерии, число ядерных аномалий у опухолевых клеток при облучении, в иммунологических процессах и т.д.).

## Глава 1.

# Проективные методы в анализе данных

Идея симптомально-синдромального подхода в анализе большого количества категориальных признаков состоит в том, чтобы рассматривать симптомы — признаки с одинаковым числом градаций, равным степени простого числа — в виде элементов проективной геометрии. В этом случае задача поиска наиболее информативных факторов сводится к выделению проективных подпространств с определенными свойствами — синдромов. Математическое описание этого метода опирается на факты, известные в комбинаторике и теории конечных проективных геометрий.

### 1.1. Конечные геометрии

#### 1.1.1. Проективные преобразования

Объектом исследования проективной геометрии являются свойства взаимного расположения геометрических образов [14]. Первоначальным, уже вышедшим из употребления ее названием является „геометрия положения“, в которой исключаются из рассмотрения метрические свойства фигур и понятие параллельности. Ее возникновение связывают с задачей изображения трехмерного предмета на плоскости. На рис. 1.1 изображено центральное проектирование

фигуры  $F$  на плоскость  $\alpha'$  из центра  $S$ . Как результат этого процесса мы получаем фигуру  $F'$ . Проведение лучей из центра  $S$  ко всем точкам фигуры  $F$  называется *проектированием*  $F$  из точки  $S$ . Фигура  $F'$  называется *центральной проекцией* фигуры  $F$ .

Соответствие между  $F$  и  $F'$  называется *перспективным*. Очевидно, для любой точки  $M \in F$  существует точка пересечения прямой  $(SM)$  с плоскостью  $\alpha'$ ,  $(SM) \cap \alpha' = M' \in F'$ .

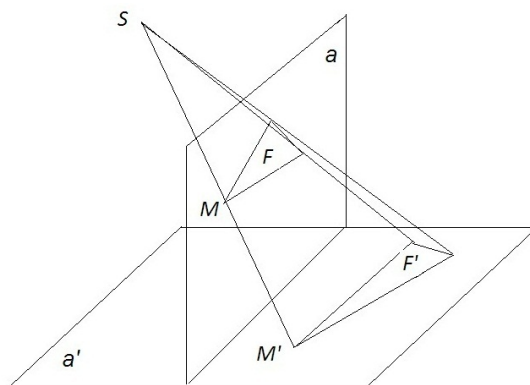


Рис. 1.1. Центральное проектирование фигуры  $F$ .

Результат ряда перспективных преобразований называют *проективным преобразованием*. Наблюдения убеждают нас в том, что в результате проективных преобразований окружность превращается в одно из конических сечений (эллипс, параболу или гиперболу), касательная к кривой всегда преобразуется в касательную. У многоугольников изменяются метрические свойства (стороны, углы, площади). Однако проекцией прямой всегда является прямая, а пересекающиеся прямые проектируются в пересекающиеся прямые. Параллельные прямые при центральном проектировании переходят в пересекающиеся.

Точку пересечения параллельных прямых называют *несобственной* или *бесконечно удаленной* точкой. Также принимают, что несобственную точку имеет каждая прямая, а все несобственные точки образуют несобственную (бесконечно удаленную) прямую.

Проективная плоскость представляет собой евклидову плоскость, дополненную несобственной прямой. Аналогично трехмерное евклидово пространство, дополненное бесконечно удаленной плоскостью, является трехмерным проективным пространством, а проективное пространство большей размерности получается добавлением к евклидову бесконечно удаленной гиперплоскости.

### 1.1.2. Теорема Дезарга

Основоположником проективной геометрии считается Жирар Дезарг (1591-1661), который впервые ввел понятия о бесконечно удаленных точках и прямых. Новые геометрические идеи, которые Дезарг разрабатывал в связи с исследованием вопросов перспективы, архитектуры, изобразительного искусства, были встречены грубыми нападками со стороны ученых своего времени. Этот выдающийся геометр находил время для просветительской деятельности, обучая мастеров и рабочих элементарным положениям математической науки, а также новым приемам решения практических задач изобразительного искусства.

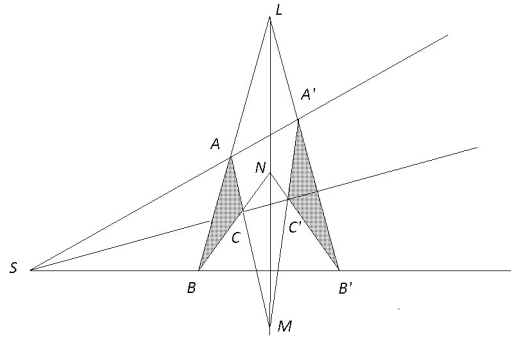


Рис. 1.2. Иллюстрация к теореме Дезарга.

**Теорема 1.** (Дезарга) Пусть  $S$  центр перспективы, треугольники  $\triangle ABC$  и  $\triangle A'B'C'$  перспективные,  $A' \in (SA)$ ,  $B' \in (SB)$ ,  $C' \in (SC)$ . Кроме того  $(AB) \cap (A'B') = L$ ,  $(AC) \cap (A'C') = M$ , тогда точка  $N = (BC) \cap (B'C')$  лежит на прямой  $(LM)$ .



Доказать теорему легко, если рассматривать положение перспективных треугольников в разных плоскостях. В этом случае ось перспективы, проходящая через точки  $L, M, N$ , является прямой пересечения этих двух плоскостей (рис. 1.2). В случае положения треугольников в одной плоскости доказательство удобнее разбить на два этапа: ввести еще один перспективный треугольник в новой плоскости, а потом отразить его во второй треугольник.

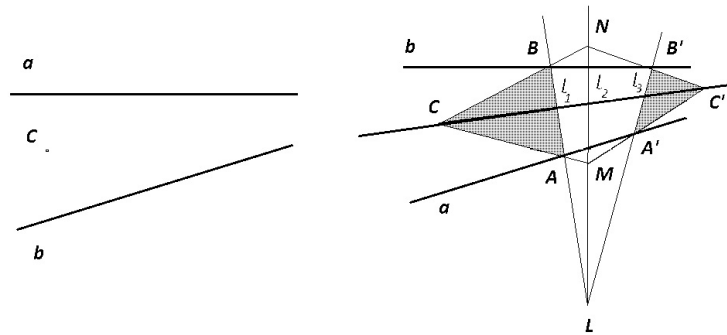


Рис. 1.3. Задача на применение теоремы Дезарга.

Одна из задач, которая, например, легко решается при помощи теоремы Дезарга, заключается в проведении через точку  $C$  прямой, проходящей через точку пересечения прямых  $a$  и  $b$  (рис. 1.3). Для решения из произвольной точки  $L$  проводятся три луча  $l_1, l_2, l_3$ . Луч  $l_1$  пересекает прямую  $a$  в точке  $A$ , прямую  $b$  в точке  $B$ . Луч  $l_3$  пересекает прямую  $a$  в точке  $A'$ , прямую  $b$  в точке  $B'$ . Из пересечения прямой  $l_2$  с прямыми  $(AC)$  и  $(BC)$  получаем точки  $M = (AC) \cap l_2$  и  $N = (BC) \cap l_2$ . По теореме Дезарга точка  $C' = (NB') \cap (MA')$  лежит на прямой, проходящей через центр перспективы  $S = a \cap b$ .

Следующий этап развития проективной геометрии связан с именем Блеза Паскаля (1623-1662) и его теоремой, заключающейся в том, что противоположные стороны любого шестиугольника, вписанного в коническое сечение, пересекаются в трех точках, лежащих на одной прямой. В частном случае конического сечения в виде пары прямых этот факт представляет собой теорему, известную древнегреческому геометру Паппу Александрийскому (рис. 1.4).

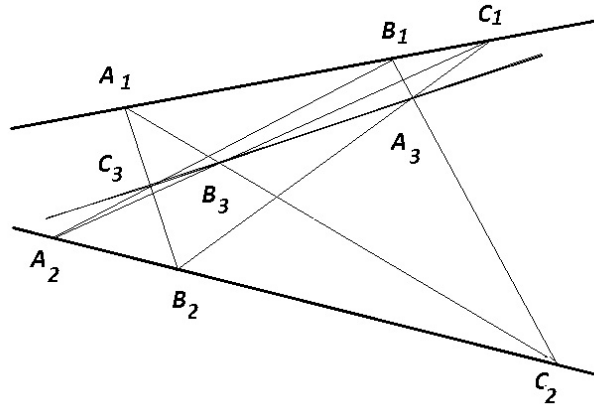


Рис. 1.4. Иллюстрация к теореме Паппа.

**Теорема 2.** (Паппа) Пусть точки  $A_1, B_1, C_1$  лежат на одной прямой,  $A_2, B_2, C_2$  на другой, кроме того  $C_3 = (A_1B_2) \cap (A_2B_1)$ ,  $B_3 = (A_1C_2) \cap (A_2C_1)$ , тогда  $(B_1C_2) \cap (B_2C_1) = A_3 \in (B_3C_3)$ .

Новое рождение проективной геометрии связано с именем Жана Понселе (1788-1865). В „Трактате о проективных свойствах фигур“ (1822), расцененном автором как „приложение к начертательной геометрии“, Понселе назвал проективными те свойства фигур, которые сохраняются при замене плоской фигуры ее проекцией, и определил проективную геометрию, как науку, изучающую свойства фигур, не изменяющиеся при проективных преобразованиях.

То, что пользуясь лишь одними аксиомами „плоской“ проективной геометрии и не используя метрических соотношений, теорему Дезарга доказать невозможно, было показано Гильбертом в конце XIX века. Таким образом он обнаружил, что справедливость теоремы Дезарга на плоскости не следует из аксиом проективной плоскости. Если не использовать выход в трехмерное пространство, то нужно присоединять теорему Дезарга в качестве новой аксиомы. Затем Гильберт показал, что, исключив эту новую аксиому, можно построить новую так называемую недезаргову геометрию.

### 1.1.3. Аксиомы проективной геометрии

Согласно [25], геометрия есть частный вид системы инцидентности с элементами в виде точек и с их подмножествами в виде прямых. Отношение инцидентности означает, что точка лежит на прямой или прямая содержит точку. Различные прямые не могут содержать одни те же точки. Конечной называется геометрия, в которой число точек конечно.

Аксиомы проективной геометрии.

- Существует одна единственная прямая, проходящая через две точки.
- Пусть три точки  $A, B, C$  не лежат на одной прямой. Тогда прямая  $(AC)$  пересекается с прямой, пересекающей прямые  $(AB)$  и  $(BC)$  (правило четырехугольника).
- Каждая прямая содержит не менее трех точек.

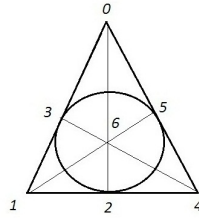


Рис. 1.5. Фигура Фано.

В качестве примера рассмотрим конечную проективную геометрию, состоящую из семи точек  $\{0, 1, 2, \dots, 6\}$  и семи прямых  $(1, 2, 4)$ ,  $(2, 3, 5)$ ,  $(3, 4, 6)$ ,  $(4, 5, 0)$ ,  $(5, 6, 1)$ ,  $(6, 0, 2)$ ,  $(0, 1, 3)$ , которые получены, начиная с орбиты  $(1, 2, 4)$ , прибавлением единицы по модулю 7 [25]. Эту геометрию изображают на плоскости в виде фигуры Фано (рис. 1.5). В справедливости аксиом можно убедиться непосредственно. Например, в правиле четырехугольника точки  $A = 0$ ,  $B = 1$  и  $C = 4$  не лежат на одной прямой, а прямые  $(AB) = (0, 1, 3)$  и  $(BC) = (1, 2, 4)$  пересекаются с прямой  $(5, 6, 1)$ , которая имеет общую точку 5 с прямой  $(AC) = (4, 5, 0)$ .

Через  $S_0$  обозначим точку, через  $S_1$  прямую, через  $S_2$  плоскость и так далее. Проективное пространство  $S_n$  получается из  $S_{n-1}$  добавлением точки  $P \notin S_{n-1}$  и всех прямых  $(PB)$ , где точка  $B \in S_{n-1}$ .

Основным условием введения координат в проективной плоскости является дезарговость, имеющая место в  $S_n$ ,  $n \geq 3$ . Итак, в дезарговой геометрии  $S_2$  или  $S_n$ ,  $n \geq 3$ , вводятся координаты точки  $P = u(x_0, \dots, x_n)$ , где  $u \in R$  (тело<sup>1</sup>),  $(x_0, \dots, x_n) \neq (0, 0, \dots, 0)$ .

Прямая описывается как множество точек вида

$$uQ = u_1P_1 + u_2P_2 = (u_1x_0 + u_2y_0, \dots, u_1x_n + u_2y_n), \quad (1.1)$$

где  $P_1 = u(x_0, \dots, x_n)$ ,  $P_2 = v(y_0, \dots, y_n)$ , и  $u_1$  и  $u_2$  не равны нулю одновременно. Также было доказано, что в проективном пространстве с  $R = F$  (поле) справедлива теорема Паппа.

Пусть  $F_q$  — произвольное поле характеристики  $q$ . Пространство векторов  $(a_0, a_1, \dots, a_n)$ , где  $a_i \in F_q$ , называется *проективной геометрией* размерности  $n$  над  $F_q$  и обозначается  $PG(n, F_q)$  или  $P_n^q$ . Нулевой вектор — это пустое пространство размерности  $-1$ . Точка  $P$  является пространством размерности  $0$  и представляет собой множество векторов  $bx = b(x_0, \dots, x_n)$ ,  $x = (x_1, \dots, x_n) \neq 0$ ,  $b \in F_q$ . Если  $y_0, \dots, y_k$  — это независимые векторы, то множество векторов вида  $b_0y_0 + \dots + b_ky_k$ , где  $b_i \in F_q$ , есть подпространство  $S_k$ . Подпространство  $S_{n-1}$  называется *гиперплоскостью*.

#### 1.1.4. Конечные поля

Поле  $F$  есть множество элементов с заданными операциями сложения и умножения, которые удовлетворяют аксиомам.

- |                                                    |                                                 |
|----------------------------------------------------|-------------------------------------------------|
| A0. $\forall a, b \in F \exists c = a + b \in F$ ; | M0. $\forall a, b \in F \exists c = ab \in F$ ; |
| A1. $(a + b) + c = a + (b + c)$ ;                  | M1. $(ab)c = a(bc)$ ;                           |
| A2. $a + b = b + a$ ;                              | M2. $ab = ba$ ;                                 |
| A3. $\exists 0 \in F, a + 0 = 0 + a = a$ ;         | M3. $\exists 1 \in F, a1 = 1a = a$ ;            |
| A4. $\exists -a \in F, a + (-a) = 0$ ;             | M3. $\exists a^{-1} \in F, aa^{-1} = 1$ ;       |
| D. $a(b + c) = ab + ac$ .                          |                                                 |

---

<sup>1</sup>Тело — это множество с двумя операциями (сложение и умножение), обладающее следующими свойствами: 1) абелевость группы относительно сложения; 2) все ненулевые элементы образуют группу относительно умножения; 3) умножение дистрибутивно относительно сложения. Если умножение коммутативно, тело называется полем.

В аксиомах А0 и М0 предполагается существование элемента  $c \in F$ , причем единственного. Основные примеры конечных полей — это классы вычетов по модулю простого числа:  $F_2, F_3, F_5, F_7$  и т.д. Из теории чисел известно, что характеристики  $q$  конечных полей исчерпываются степенями простых чисел. В случае  $q = p^r$ , где  $p$  — простое число, элементами поля являются векторы с  $r$  компонентами над полем  $F_p$  или полиномы степени не выше  $r - 1$  с коэффициентами над  $F_p$ .

Сложение над  $F_q$ , где  $q = p^r$ , осуществляется по координатам над  $F_p$ . Для умножения используются неприводимые полиномы  $f(x)$ , не имеющие нетривиальных делителей, т. е. таких  $g_i(x) \neq 1$  и  $g_i(x) \neq f(x)$ , где  $i = 1, 2$ , что  $f(x) = g_1(x)g_2(x)$ .

Отличительной чертой полей четной характеристики  $q = 2^r$  является то, что в них как таковое отсутствует вычитание, и сумма одинаковых элементов равна нулю,  $x + x = 0$ . Для наглядности построим таблицы сложения и умножения (табл.1.1) над полем  $F_4$  с элементами в виде полиномов  $\{0, 1, x, x + 1\}$ . Коэффициенты этих полиномов задаются векторами  $(X_0, X_1)$  с компонентами над  $F_2$ :  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ ,  $(1, 1)$ , которым в двоичном коде  $M = X_0 + 2X_1$  соответствуют числа  $\{0, 1, 2, 3\}$ .

Таблица 1.1.

$M+$	0	1	2	3	$M*$	0	1	2	3
0	0	1	2	3	0	0	0	0	0
1	1	0	3	2	1	0	1	2	3
2	2	3	0	1	2	0	2	3	1
3	3	2	1	0	3	0	3	1	2

Например,  $2 + 3 = 1$ , так как  $x + (x + 1) = 1$ , а  $2 \cdot 3 = 1$ , так как  $x(x + 1) = (x^2 + x + 1) + 1 = 1$ , и остаток от деления на неприводимый полином  $x^2 + x + 1$  равен 1.

Приведем пример введения координат в конечной геометрии. Точками проективной геометрии  $P_2^2$  являются векторы  $(X_0, X_1, X_2)$  с компонентами над  $F_2$  кроме нулевого. Прямая, проходящая через точки  $P_1$  и  $P_2$ , представляет собой совокупность точек вида:  $u_1P_1 + u_2P_2$  над  $F_2$ , где  $u_1, u_2 \in F_2$ ,  $(u_1, u_2) \neq (0, 0)$ . Таким образом, прямую образуют точки вида  $(P_1, P_2, P_1 + P_2)$ .

Таблица 1.2.

$M$	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	3	2	5	4	7	6
2	2	3	0	1	6	7	4	5
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	5	4	7	6	1	0	3	2
6	6	7	4	5	2	3	0	1
7	7	6	5	4	3	2	1	0

Поскольку точкам геометрии  $P_2^2$  соответствуют ненулевые элементы поля  $F_8$ , сложение точек  $P_1 + P_2$  в  $P_2^2$  осуществляется над  $F_8$ . Результаты сложения над  $F_8$  представлены в табл. 1.2, где восьми бинарным векторам  $(X_0, X_1, X_2)$  в двоичном коде  $M = X_0 + 2X_1 + 4X_2$  соответствуют числа от 0 до 7.

Таблица 1.3.

Точки прямой	Уравнение прямой	Коэффициенты
	$c_0X_0 + c_1X_1 + c_2X_2 = 0$	$(c_0, c_1, c_2)$
(2 4 6)	$X_0 = 0$	(1,0,0)
(1 4 5)	$X_1 = 0$	(0,1,0)
(3 4 7)	$X_0 + X_1 = 0$	(1,1,0)
(1 2 3)	$X_2 = 0$	(0,0,1)
(2 5 7)	$X_0 + X_2 = 0$	(1,0,1)
(1 6 7)	$X_1 + X_2 = 0$	(0,1,1)
(3 5 6)	$X_0 + X_1 + X_2 = 0$	(1,1,1)

Например, используя результат  $1 + 5 = 4$  из табл. 1.2, по двум точкам  $P_1 = 1$  и  $P_2 = 5$  парами  $(u_1, u_2)$  определяем три точки  $(1, 0), (0, 1), (1, 1)$ , образующие прямую  $(1, 5, 4)$ . Имеем  $C_7^2 = 21$  всевозможных пар точек, но поскольку три пары точек задают одну и ту же прямую, число прямых равно  $C_7^2/3 = 7$  (табл.1.3).

## 1.2. Конечные геометрии и дизайны

### 1.2.1. Определение и свойства дизайнов

В комбинаторике такое размещение  $v$  элементов по  $b$  блокам размера  $k$ , что каждый элемент встречается  $r$  раз, а каждая пара  $\lambda$  раз, называется *дизайном*  $D(v, b, r, k, \lambda)$  или *неполной сбалансированной блок-схемой*. Если  $v = b$ ,  $r = k$ , то дизайн называется *симметричным* и обозначается  $D(v, k, \lambda)$ . Например, симметричный дизайн  $D(7, 3, 1)$  образуют точки и прямые в  $P_2^2$  (табл.1.3).

Для  $D(v, b, r, k, \lambda)$  справедливы два соотношения баланса:

$$vr = bk, \quad r(k - 1) = \lambda(v - 1).$$

Первое указывает на то, что общее количество элементов в размещении может быть вычислено двумя способами: либо число всевозможных элементов  $v$  умножается на  $r$  (сколько раз каждый элемент встречается), либо количество блоков  $b$  умножается на число элементов в каждом блоке  $k$ . Для доказательства второго выделим какой-то элемент, например,  $a$ , который встречается  $r$  раз. Число пар с участием  $a$  равно  $r(k - 1)$ . С другой стороны, число пар вида  $\{a, *\}$ , каждая из которых встречается  $\lambda$  раз, равно  $\lambda(v - 1)$ , так как для элемента  $*$  имеет место  $v - 1$  вариант.

Таблица 1.4.

0246	0145	0347	0123	0257	0167	0356
1357	2367	1256	4567	1346	2345	1247

Дизайн с элементами над полем  $F_q$  и блоками, в которых сумма элементов над  $F_q$  равна нулю, называется *каноническим* и обозначается  $D^0(v, b, r, k, \lambda|F_q)$ . Например,  $D(7, 3, 1)$  из таблицы 1.3 является каноническим  $D^0(7, 3, 1|F_8 \setminus \{0\})$ , поскольку сумма элементов в блоках  $(P_1, P_2, P_1 + P_2)$  над  $F_8$  равна нулю. Если каждый из семи блоков этого дизайна дополнить нулевым элементом и ввести ему противоположный блок, то получится канонический дизайн  $D^0(8, 14, 7, 4, 3|F_8)$  с 14 блоками (табл.1.4).

### 1.2.2. Теорема Зингера

Речь идет об утверждении [25], позволяющем рассматривать конечные геометрии как блок-схемы. Заметим, что теорема представлена в усеченном виде и не содержит положения о цикличности блок-схемы.

**Теорема 3.** *Гиперплоскости геометрии  $P_n^q$ ,  $q = p^r$ , взятые в качестве блоков, и точки, взятые в качестве элементов, образуют симметричную блок-схему с параметрами*

$$v = \frac{q^{n+1} - 1}{q - 1}, \quad k = \frac{q^n - 1}{q - 1}, \quad \lambda = \frac{q^{n-1} - 1}{q - 1}.$$

**Доказательство.** Число векторов  $(x_0, x_1, \dots, x_n)$  с компонентами над  $F_q$  равно  $q^{n+1}$ . Каждый из  $q^{n+1} - 1$  ненулевых векторов определяет точку, и поскольку векторы

$$(x_0, x_1, \dots, x_n) \quad \text{и} \quad b(x_0, x_1, \dots, x_n), \quad \text{где } b \neq 0,$$

определяют одну и ту же точку, всего имеется  $v = \frac{q^{n+1}-1}{q-1}$  точек.

Если  $(c_0, c_1, \dots, c_n) \neq (0, 0, \dots, 0)$ , то множество векторов с компонентами  $x_0, \dots, x_n$ , удовлетворяющих равенству

$$c_0x_0 + \dots + c_nx_n = 0,$$

есть гиперплоскость. С другой стороны, коэффициенты

$$(c_0, c_1, \dots, c_n) \quad \text{и} \quad (sc_0, sc_1, \dots, sc_n), \quad \text{где } s \neq 0,$$

определяют одну и ту же гиперплоскость. Следовательно, количество различных гиперплоскостей равно  $b = \frac{q^{n+1}-1}{q-1}$ .

Если  $y_0, \dots, y_t$  — это  $t + 1$  независимых векторов, то линейные комбинации  $b_0y_0 + \dots + b_t y_t$ ,  $b_i \in F_q$ , дают  $q^{t+1} - 1$  ненулевых векторов. Таким образом, подпространство  $S_t$  содержит  $\frac{q^{t+1}-1}{q-1}$  векторов, а гиперплоскость  $S_{n-1}$  имеет  $k = \frac{q^n-1}{q-1}$  различных точек. Две различные гиперплоскости пересекаются по подпространству  $S_{n-2}$ , которое имеет  $\lambda = \frac{q^{n-1}-1}{q-1}$  точек.  $\square$



### 1.2.3. Аффинная геометрия

Пространство точек  $(x_1, \dots, x_n)$ ,  $x_i \in F_q$ , называется *аффинной геометрией*<sup>2</sup> и обозначается  $E_n^q$ .

Аксиомы аффинной геометрии.

- Для двух различных точек существует только одна прямая, которая содержит обе точки.
- Для прямой  $P$  и точки  $p \notin P$ , существует одна и только одна прямая  $P'$ , такая, что  $p \in P'$ ,  $P \cap P' = \emptyset$ .
- Существует множество из четырех точек, никакие три из которых не лежат на одной прямой.

**Определение 1.** *Остаточным* называется дизайн, который получается из блок-схемы при вычеркивании одного блока и всех его элементов из остальных блоков. Дизайн с блоками из вычеркнутых элементов называется *производным*.

Для  $P_n^q$  производный дизайн соответствует  $P_{n-1}^q$ , а остаточный  $E_n^q$ . Покажем это на примере дизайна  $D(7, 3, 1)$  из табл.1.3. Удаляя из него элементы 1, 2, 3, получим остаточный дизайн  $D(4, 6, 3, 2, 1)$ , который состоит из  $v = 4$  элементов  $\{4, 5, 6, 7\}$  и  $b = 6$  блоков.

$$\begin{array}{l}
 (2, 4, 6) \\
 (1, 4, 5) \\
 (3, 4, 7) \\
 (1, 2, 3) \\
 (2, 5, 7) \\
 (1, 6, 7) \\
 (3, 5, 6)
 \end{array}
 \rightarrow
 \begin{array}{l}
 (4, 5) (6, 7) \\
 (4, 6) (5, 7) \\
 (4, 7) (5, 6)
 \end{array}$$

---

<sup>2</sup>Аффинная геометрия (лат. *affinis* — родственный) — раздел геометрии, в котором изучаются свойства фигур, инвариантные относительно линейных преобразований. Длины отрезков и величины углов в аффинной геометрии в отличие от евклидовой не имеют существенного значения, поэтому на аффинной плоскости таких понятий, как равносторонний или прямоугольный треугольник нет. Там есть просто треугольник. Однако такие свойства, как прямолинейное расположение трех точек или параллельность двух прямых, инвариантны не только в евклидовой, но и аффинной геометрии.

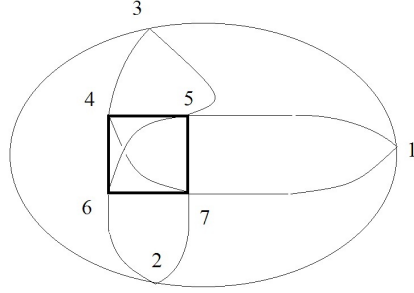


Рис. 1.6. Интерпретация  $E_2^2$  и  $P_2^2$  на плоскости.

Элементами  $D(7, 3, 1|F_8 \setminus \{0\})$  являются точки  $P_2^2$  с координатами  $X = (X_0, X_1, X_2)$ ,  $X_i \in F_2$ ,  $X \neq (0, 0, 0)$ . При удалении элементов  $\{1, 2, 3\}$  в коде  $M = X_0 + 2X_1 + 4X_2$  с  $X_2 = 0$  остаются элементы  $4 = (0, 0, 1)$ ,  $5 = (1, 0, 1)$ ,  $6 = (0, 1, 1)$ ,  $7 = (1, 1, 1)$  с  $X_2 = 1$ , которым соответствуют точки  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$ ,  $(1, 1)$  геометрии  $E_2^2$ .

Геометрически дизайн  $D(4, 6, 3, 2, 1)$  и соответственно аффинную геометрию  $E_2^2$  удобно представлять в виде четырехугольника (рис. 1.6), в котором имеются три пары „параллельных“ прямых: две пары противоположных сторон (45) и (67), (46) и (57) и пара диагоналей (47) и (56).

В общем виде параметры дизайна  $D(v, b, r, k, \lambda)$ , соответствующего аффинной геометрии  $E_n^q$ , имеют вид:

$$v = \frac{q^{n+1} - 1}{q - 1} - \frac{q^n - 1}{q - 1} = q^n, \quad b = \frac{q^{n+1} - 1}{q - 1} - 1 = \frac{q(q^n - 1)}{q - 1},$$

$$k = \frac{q^n - 1}{q - 1} - \frac{q^{n-1} - 1}{q - 1} = q^{n-1}, \quad \lambda = \frac{q^{n-1} - 1}{q - 1},$$

$$r = \frac{bk}{v} = \frac{q(q^n - 1)q^{n-1}}{q^n(q - 1)} = \frac{q^n - 1}{q - 1}.$$

Для того, чтобы из аффинной плоскости  $E_2^2$  построить проективную плоскость  $P_2^2$ , нужно ввести три бесконечно удаленные точки пересечения параллельных прямых. Дополним прямые (4,5) и (6,7) точкой 1, прямые (4,6) и (5,7) точкой 2, прямые (4,7) и (5,6)

Таблица 1.5.

Порядок блока	Нормаль к плоскости	Блоки	
		$D(8, 14, 7, 4, 3)$	
$i$	$(c_1, c_2, c_3)$	$B_i$	$\overline{B}_i$
1	(1,0,0)	0246	1357
2	(0,1,0)	0145	2367
3	(1,1,0)	0347	1256
4	(0,0,1)	0123	4567
5	(1,0,1)	0257	1347
6	(0,1,1)	0167	2345
7	(1,1,1)	0356	1247

точкой 3. Эти шесть прямых вместе с бесконечно удаленной прямой (1,2,3) образуют дизайн  $D(7, 3, 1)$  или  $P_2^2$ .

Точки аффинной геометрии  $E_3^2$ , взятые в качестве элементов, плоскости  $c_1X_1 + c_2X_2 + c_3X_3 = 0$  и  $c_1X_1 + c_2X_2 + c_3X_3 = 1$ , взятые в качестве блоков  $B_i$  и  $\overline{B}_i$ , где  $i = c_1 + 2c_2 + 4c_3$ , образуют канонический дизайн  $D(8, 14, 7, 4, 3|F_8)$ , который представлен в табл.1.5.

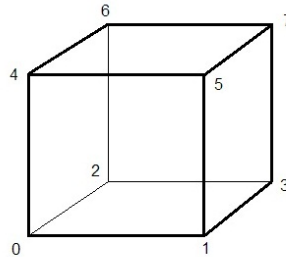


Рис. 1.7. Интерпретация  $D(8, 14, 7, 4, 3|F_8)$  на кубе.

Геометрически аффинную геометрию  $E_3^2$  и соответствующий ей дизайн  $D(8, 14, 7, 4, 3)$  можно изобразить в виде куба (рис. 1.7). Блоки образуют семь пар противоположных граней: три пары внешних граней, три пары диагональных и два внутренних тетраэдра.

### 1.2.4. Изоморфные и неизоморфные подстановки дизайнов

Изоморфные подстановки дизайнов возникают при установлении взаимно-однозначного соответствия между элементами. Рассмотрим подстановки дизайна  $D(15, 7, 3)$  с множеством элементов над  $F_{16} \setminus \{0\}$ , для которых используем обозначения  $K = X_0 + 2X_1 + 4X_2 + 8X_3$  при  $X_3 = 0$  и  $(K - 8)'$  при  $X_3 = 1$ . Если в  $D^0(8, 14, 7, 4, 3)$  из табл. 1.5 дополнить блоки  $(B_i, \bar{B}_i)$ , где  $B_i = (0abc)$ , блоком вида  $(a'b'c')$ , то получится  $D(15, 7, 3)$ , соответствующий  $P_3^2$  (табл. 1.6, А). Можно непосредственно убедиться, что ему изоморфен дизайн с орбитой  $0, 1, 2, 4, 5, 8, 10 \pmod{15}$  (табл. 1.6, В).

Таблица 1.6.

1'2'3'4'5'6'7'	
0246	2'4'6'
1357	2'4'6'
0145	1'4'5'
2367	1'4'5'
0347	3'4'7'
1256	3'4'7'
0123	1'2'3'
4567	1'2'3'
0257	2'5'7'
1346	2'5'7'
0167	1'6'7'
2345	1'6'7'
0356	3'5'6'
1247	3'5'6'

А

1'2'3'4'5'6'7'	
0246	2'4'6'
3246	1'5'7'
0145	1'4'5'
6145	2'3'7'
0347	3'4'7'
1347	2'5'6'
0123	1'2'3'
5123	4'6'7'
0257	2'5'7'
4257	1'3'6'
0167	1'6'7'
2167	3'4'5'
0356	3'5'6'
7356	1'2'4'

Б

0	1	2	4	5	8	10
1	2	3	5	6	9	11
2	3	4	6	7	10	12
3	4	5	7	8	11	13
4	5	6	8	9	12	14
5	6	7	9	10	13	0
6	7	8	10	11	14	1
7	8	9	11	12	0	2
8	9	10	12	13	1	3
9	10	11	13	14	2	4
10	11	12	14	0	3	5
11	12	13	0	1	4	6
12	13	14	1	2	5	7
13	14	0	2	3	6	8
14	0	1	3	4	7	9

В

Неизоморфный дизайн  $D(15, 7, 3)$  (табл. 1.6, Б) получится, если в остаточной схеме  $D(8, 14, 7, 4, 3)$  противоположный  $B_i = (0abc)$  блок  $\bar{B}_i$  заменить блоком вида  $(iabc)$ , где  $\tilde{i} = \tilde{a} + \tilde{b} + \tilde{c}$  над  $F_8$  в соответствии с (1.14), а один из производных блоков вида  $(a'b'c')$  заменить на  $(\bar{B}_i \setminus i)'$ . В случае (А) непротивоположные блоки имеют по два общих элемента („правый“ куб  $E_3^2$ ), а в случае (Б) семь пар блоков имеют по три общих элемента („левый куб“).

### 1.2.5. Первый исключительный изоморфизм

На основе автоморфизмов производной и остаточной блок-схем дизайна  $D(11, 5, 2)$  можно построить изоморфизм между простыми группами  $PSL_2^{F_5}$  и  $PSL_2^{F_4}$  порядка 60. <sup>3</sup> Через  $PSL_2^{F_5}$  и  $PSL_2^{F_4}$  обозначены проективные <sup>4</sup> специальные <sup>5</sup> группы матриц порядка два над  $F_5$  и  $F_4$  соответственно. В табл. 1.7 представлены блоки дизайна  $D(11, 5, 2)$ , полученные из орбиты 1, 3, 4, 5, 9 над полем  $F_{11}$ .

Таблица 1.7.

$i$	Блок $B_i$				
1	<b>1</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>9</b>
2	2	<b>4</b>	<b>5</b>	6	10
3	<b>3</b>	<b>5</b>	6	7	0
4	<b>4</b>	6	7	8	<b>1</b>
5	<b>5</b>	7	8	<b>9</b>	2
6	6	8	<b>9</b>	10	<b>3</b>
7	7	<b>9</b>	10	0	<b>4</b>
8	8	10	0	<b>1</b>	<b>5</b>
9	<b>9</b>	0	<b>1</b>	2	6
10	10	<b>1</b>	2	<b>3</b>	7
11	0	2	<b>3</b>	<b>4</b>	8

Точки проективной геометрии  $P_1^5$  имеют вид  $(x_0, x_1) \neq 0$  над  $F_5$ , причем для  $c \neq 0$  векторы  $c(x_0, x_1)$  определяют одну и ту же точку. Каждой точке можно сопоставить отношение

$$t = \frac{x_1}{x_0} \pmod{5}, \quad \text{где } t \in \{0, 1, 2, 3, 4, 5\},$$

где  $5 = \infty$  при  $x_0 = 0$ . Аналогично поставим в соответствие пяти точкам  $P_1^4$  вида  $(x_0, x_1)$  над  $F_4$  значения

$$t' = \frac{x_1}{x_0} \in \{0, 1, 2, 3, 4 = \frac{x_1}{0} = \infty\} \quad \text{над } F_4.$$

<sup>3</sup>Подстановки с одинаковыми множествами элементов называются *автоморфными*, если множества блоков совпадают, иначе неавтоморфными.

<sup>4</sup>Определяемые с точностью до умножения на константу.

<sup>5</sup>Определитель равен 1.

Разделим  $D(11, 5, 2)$  на производный  $D(5, 10, 4, 2, 1|F_{11})$  и остаточный  $D(6, 10, 5, 3, 2|F_{11})$ , а затем переведем элементы  $\{1, 3, 4, 5, 9\}$  производного дизайна в точки  $P_1^4$ , а элементы  $\{0, 2, 6, 7, 8, 10\}$  остаточного в точки  $P_1^5$ , например, таким образом

$$\left( \begin{array}{c|ccccc} F_{11} & 0 & 7 & 2 & 6 & 8 & 10 \\ \hline P_1^5 & 0 & 1 & 2 & 3 & 4 & 5 \end{array} \right) \text{ и } \left( \begin{array}{c|ccccc} F_{11} & 4 & 3 & 5 & 9 & 1 \\ \hline P_1^4 & 4' & 0' & 1' & 2' & 3' \end{array} \right).$$

В результате этих подстановок получаем соответствие между дизайнами  $D(5) = D(5, 10, 4, 2, 1|P_1^4)$ ,  $D(6) = D(6, 10, 5, 3, 2|P_1^5)$ , которые вместе с блоком  $(0', 1', 2', 3', 4')$  образуют представленный в табл. 1.8 дизайн  $D(11) = D(11, 5, 2|P_1^4 \cup P_1^5)$ . Это соответствие можно интерпретировать геометрически (рис. 1.8). Дизайн  $D(5)$ , состоящий из всех сочетаний двух элементов из пяти, очевидно инвариантен относительно группы  $PSL_2^{F_4}$ , то есть при любой подстановке на элементах  $P_1^4$ , в том числе при помощи группы  $PSL_2^{F_4}$ , множество блоков останется тем же самым, хотя расположение блоков  $D(5)$  в  $D(11)$  изменится. Но при этом изменится и расположение блоков  $D(6)$ , жестко связанных с блоками  $D(5)$  по соответствию из табл. 1.8. Тогда можно будет найти соответствующую подстановку на элементах  $P_1^5$ . Таким образом, остается выяснить, является ли инвариантным относительно группы  $PSL_2^{F_5}$  дизайн  $D(6)$ .

Таблица 1.8.

$D(5, 10, 4, 2, 1 P_1^4)$	$D(6, 10, 5, 3, 2 P_1^5)$	$D(5, 10, 4, 2, 1 P_1^4)$
$(1', 4')$	$(2, 3, 5)$	$(2, 3, 0)$
$(0', 2')$	$(3, 4, 5)$	$(3, 4, 1)$
$(1', 3')$	$(4, 0, 5)$	$(4, 0, 2)$
$(2', 4')$	$(0, 1, 5)$	$(0, 1, 3)$
$(0', 3')$	$(1, 2, 5)$	$(1, 2, 4)$
		$(2', 3')$
		$(3', 4')$
		$(0', 4')$
		$(0', 1')$
		$(1', 2')$

**Предложение 1.** *Дизайн  $D(6, 10, 5, 3, 2|P_1^5)$  из табл. 1.8 инвариантен относительно группы  $PSL_2^{F_5}$ .*

**Доказательство.** Покажем, что линейные подстановки матрицами  $A \in PSL_2^{F_5}$  элементов блока  $(015) = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$  со-

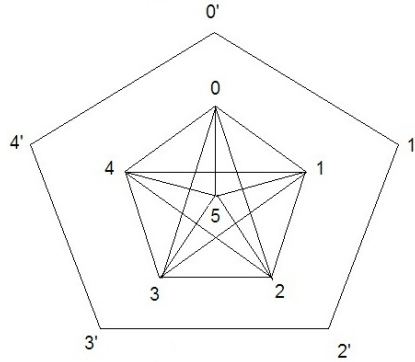


Рис. 1.8. Соответствие между  $P_1^4$  и  $P_1^5$ .

ответствуют только блокам дизайна  $D(6, 10, 5, 3, 2|P_1^5)$ . При помощи матрицы

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in PSL_2^{F_5}$$

задается преобразование на элементах геометрии  $P_1^5$  вида

$$(k, \tau_k), \quad \text{где} \quad \tau_k = \frac{c + kd}{a + kb}.$$

При  $k = \infty = 5$  очевидно  $\tau_5 = \frac{d}{b}$ . Если  $a = 0$ , тогда

$$\tau_0 = \frac{c}{a} = \infty = 5, \quad \tau_1 = \frac{c + d}{a + b} = \frac{c + d}{b} = \frac{c}{b} + \tau_5.$$

Из условия равенства единице определителя получаем  $bc = 4$ , откуда следует

$$\frac{c}{b} = \frac{4}{b^2} = \begin{cases} -1 & \text{при } b^2 = 1, \\ 1 & \text{при } b^2 = 4. \end{cases}$$

Таким образом, блоки, содержащие 5, имеют вид  $(5, t, t \pm 1)$ .

Если  $a \neq 0$ , тогда

$$\tau_1 = \frac{c + d}{a + b} = \frac{a\tau_0 + b\tau_5}{a + b} = \frac{a\tau_0 + \frac{\tau_5}{a(\tau_5 - \tau_0)}}{a + \frac{1}{a(\tau_5 - \tau_0)}} = \frac{a^2\tau_0(\tau_5 - \tau_0) + \tau_5}{a^2(\tau_5 - \tau_0) + 1}.$$

Пусть  $\tau_0 = 0$ , и в блоке нет 5, тогда  $\tau_1^{-1} = \tau_5^{-1} + a^2 = \tau_5^{-1} \pm 1$ , следовательно, пары  $(\tau_1, \tau_5)$  имеют вид  $(1, 3), (2, 3), (2, 4)$ , так как  $\frac{1}{1} - \frac{1}{3} = 1 - 2 = -1$ ,  $\frac{1}{2} - \frac{1}{3} = 3 - 2 = 1$ ,  $\frac{1}{2} - \frac{1}{4} = 3 - 4 = -1$ . Остальные блоки строятся аналогично.  $\square$

ПРИМЕР. При помощи матрицы из группы  $PSL_2^{F_5}$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

задается подстановка на  $P_1^5$

$$\begin{pmatrix} 5 & 0 & 1 & 2 & 3 & 4 \\ 1 & 3 & 0 & 5 & 2 & 4 \end{pmatrix}.$$

Это можно показать непосредственным умножением матриц.

$$\begin{aligned} & \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 \end{bmatrix} = \\ & = \begin{bmatrix} 2 & 1 & 3 & 0 & 2 & 4 \\ 2 & 3 & 0 & 2 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 3 & 0 & 1 & 2 & 4 \end{bmatrix}. \end{aligned}$$

Для вычисления подстановки на  $P_1^4$  воспользуемся соотношением между блоками дизайнов  $D(5)$  и  $D(6)$  в табл. 1.8.

$$\begin{aligned} \left. \begin{aligned} (01) \sim (013) &\longrightarrow (023) \sim (23) \\ (02) \sim (345) &\longrightarrow (124) \sim (12) \end{aligned} \right\} &\implies 0 \longrightarrow 2, 1 \longrightarrow 3, 2 \longrightarrow 1, \\ (03) \sim (025) &\longrightarrow (015) \sim (24) &\implies 3 \longrightarrow 4. \end{aligned}$$

Для оставшегося элемента справедливо  $4 \longrightarrow 0$ . Таким образом получена подстановка на  $P_1^4$  вида

$$\begin{pmatrix} 4 & 0 & 1 & 2 & 3 \\ 0 & 2 & 3 & 1 & 4 \end{pmatrix}.$$

Ей соответствует матрица

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} \in PSL_2^{F_4},$$

так как  $\tau_4 = 0, \Rightarrow d = 0$ ;  $\tau_4 = 0, \Rightarrow c = 2a$ ;  $\tau_3 = 4, \Rightarrow a + 3b = 0, \Rightarrow b = 2a$ ; над  $F_4$  определитель равен 1,  $ad - bc = bc = 2a \cdot 2a = 3a^2 = 1$ , откуда  $a^2 = 2, a = 3, b = c = 2 \cdot 3 = 1$ .

**Замечание.** Указанное соответствие между матрицами определяется с точностью до автоморфизма. Вместо матрицы  $B \in PSL_2^{F_4}$  можно рассматривать матрицу  $J^{-1}BJ$  того же спектра, где  $J \in PSL_2^{F_4}$ . Действительно, если  $k$  – спектр матрицы  $B$ , то есть  $B^k = I$ , тогда

$$(J^{-1}BJ)^k = (J^{-1})^k B^k J^k = (J^{-1})^k J^k = (J^{-1}J)^k = I^k = I.$$

Кроме того, для  $B_1 \cdot B_2 = B_3$ ,  $B_i \in PSL_2^{F_4}$  справедливо

$$(J^{-1}B_1J) \cdot (J^{-1}B_2J) = J^{-1}B_1B_2J = J^{-1}B_3J.$$



### 1.3. Бинарные признаки и $P_n^2$

Статистический аспект применения конечных геометрий связан с возможностью явного привлечения перекрестных факторов. Например, пусть имеется фактор роста: низкий ( $X_1 = 0$ ) и высокий ( $X_1 = 1$ ), а также фактор массы:  $X_2 = 0$ , если индивид худой,  $X_2 = 1$ , если индивид толстый. В двухфакторном дисперсионном анализе эффект взаимодействия этих факторов<sup>6</sup> можно выразить через новый признак  $X_{12} = |X_1 - X_2| = 0$ , если рост и вес ассоциированы,  $X_1 = X_2$ , то есть индивид либо худенький и маленький, либо большой и толстый;  $X_{12} = 1$  при  $X_1 \neq X_2$ , если вес не соответствует росту, то есть либо индивид маленький и толстый, либо высокий и худой.

С информационной точки зрения эти признаки равноправны, так как большие или маленькие рост и вес можно узнать по любой паре этих признаков:  $(X_1, X_2)$ ,  $(X_1, X_{12})$  или  $(X_2, X_{12})$ . Если нам известно, что индивид высокий, и его рост и вес ассоциированы, то, очевидно, что он не худой.

Для возможности распространения операции  $X_{12} = |X_1 - X_2|$  на большее число факторов заменим ее эквивалентной операцией сложения над полем  $F_2$

$$X_{12} = X_1 + X_2 \pmod{2},$$

которая позволяет рассматривать дихотомические признаки и их линейные комбинации как точки проективной геометрии.

#### 1.3.1. Определение симптома

Для обозначения  $m$  дихотомических признаков используем случайный вектор  $X = (X_1, \dots, X_m)^T$  с компонентами, принимающими два значения 0 или 1. Множество  $\Omega_m$  размерности  $2^m$  возможных значений этого вектора  $x = (x_1, \dots, x_m)^T$ ,  $x_i \in F_2$ , имеет структуру конечной аффинной геометрии  $E_m^2$  [25]. Каждому элементу из  $\Omega_m$  можно сопоставить число

$$j = x_1 + 2x_2 + \dots + 2^{m-1}x_m.$$

---

<sup>6</sup>Эффект взаимодействия двух указанных бинарных факторов проявляется, например, при исследовании астении, так как она чаще проявляется либо у людей истощенных (высокий, худой), либо при нарушении обмена веществ (маленький, толстый).

**Определение 2.** Обозначим  $k$ -подмножество из  $m$  натуральных чисел через

$$\tau = (t_1, \dots, t_k) \subseteq (1, 2, \dots, m)$$

и зададим вектор-строку  $A_\tau = (a_1, \dots, a_m)$  с компонентами

$$a_j = \begin{cases} 1, & \text{если } j \in \tau, \\ 0 & \text{иначе.} \end{cases}$$

Линейная комбинация вида  $X_\tau = A_\tau X \pmod{2}$  называется симптомом ранга  $k$ .

Компоненты вектора  $X$  являются тривиальными симптомами единичного ранга  $X_i$ ,  $i = 1, \dots, m$ . Симптом нулевого ранга, то есть со всеми нулевыми коэффициентами, является вырожденным  $X_\emptyset$  и принимает значение 0 с вероятностью 1.

### 1.3.2. Проективные свойства синдрома

Если точки проективной геометрии  $P_m^2$  со статистической точки зрения можно рассматривать как симптомы, то подпространства  $S_k$ , где  $k \leq m$ , представляют собой различные сочетания этих симптомов или синдромы.

**Определение 3.** Пусть имеется  $k + 1 > 0$  симптомов  $X_0, \dots, X_k$ . Совокупность  $2^{k+1} - 1$  симптомов вида

$$\beta_1 X_0 + \dots + \beta_k X_k \pmod{2}, \quad (1.2)$$

где коэффициенты  $\beta_i \in F_2$  не равны нулю одновременно, называется синдромом  $k$ -го порядка  $S_k$ .

Единичный симптом можно рассматривать как синдром нулевого порядка  $S_0 = S_0(X_\tau)$ . Синдром первого порядка  $S_1$  состоит из симптомов  $X_\tau, X_\mu, X_{\tau\mu}$ , которые устроены (1.1) как точки, образующие проективную прямую. Действительно,

$$\begin{aligned} 1 \cdot X_\tau + 0 \cdot X_\mu \pmod{2} &= X_\tau, \\ 0 \cdot X_\tau + 1 \cdot X_\mu \pmod{2} &= X_\mu, \\ 1 \cdot X_\tau + 1 \cdot X_\mu \pmod{2} &= X_{\tau\mu}. \end{aligned}$$

Из линейной зависимости  $X_\tau + X_\mu \pmod{2} = X_{\tau\mu}$  получаем, что синдром  $S_1$  задается любой парой симптомов:  $(X_\tau, X_\mu)$ ,  $(X_\tau, X_{\tau\mu})$  или  $(X_{\tau\mu}, X_\mu)$ . Сумма трех симптомов одного и того же синдрома равна вырожденному симптому  $X_\emptyset$ . Синдром 2-го порядка

$$S_2 = (X_\tau, X_\mu, X_{\tau\mu}, X_\kappa, X_{\kappa\tau}, X_{\kappa\mu}, X_{\kappa\tau\mu}) \quad (1.3)$$

устроен как проективная геометрия  $P_2^2$  и получается добавлением к  $S_1$  нового признака  $X_\kappa \notin S_1(X_\tau, X_\mu)$  и симптомов, которые образуются из симптомов синдрома  $S_1$  и признака  $X_\kappa$  линейным образом:

$$(X_{\kappa\tau}, X_{\kappa\mu}, X_{\kappa\tau\mu}) = S_1(X_\tau, X_\mu) + X_\kappa \pmod{2}.$$

Рассматривая в качестве множества элементов синдром  $S_2$ , а в качестве блоков все синдромы  $S_1$ , получаем дизайн  $D(7, 3, 1|S_2)$  вида

$$\begin{aligned} & (X_\tau \ X_\mu \ X_{\tau\mu} ), \\ & (X_\tau \ X_\kappa \ X_{\tau\kappa} ), \\ & (X_\tau \ X_{\kappa\mu} \ X_{\kappa\tau\mu} ), \\ & (X_\mu \ X_\kappa \ X_{\kappa\mu} ), \\ & (X_\mu \ X_{\tau\kappa} \ X_{\kappa\tau\mu} ), \\ & (X_\kappa \ X_{\tau\mu} \ X_{\kappa\tau\mu} ), \\ & (X_{\tau\mu} \ X_{\tau\kappa} \ X_{\kappa\mu} ). \end{aligned} \quad (1.4)$$

Аналогично индуктивному построению проективных геометрий через введение новой точки, не принадлежащей конечной геометрии, и через построение всевозможных прямых, проходящих через эту точку, может быть определен синдром  $k + 1$ -го порядка

$$S_{k+1} = (S_k, X_\nu, S_k + X_\nu \pmod{2}), \quad X_\nu \notin S_k. \quad (1.5)$$

Порядок симптомов в (1.5) будем называть *импульсным*<sup>7</sup>. Симптомы, находящиеся импульсно упорядоченном синдроме  $S_k$  на  $2^m$ -м месте,  $m = 0, 1, \dots, k$ , называются *базовыми*. В качестве базовых могут рассматриваться любые линейно независимые симптомы.

<sup>7</sup>Импульсными последовательностями в алгебре называют последовательности чисел, построенные в конечном поле в результате рекуррентных соотношений типа Фибоначчи.

### 1.3.3. Автоморфизмы синдрома

Всевозможные варианты выбора базовых симптомов в синдроме  $S_n$  исчерпываются группой, изоморфной группе автоморфизмов проективной геометрии  $P_n^q$ .

Согласно [25], группой автоморфизмов симметричного дизайна  $D(2^n - 1, 2^{n-1} - 1, 2^{n-2} - 1)$ , соответствующего по теореме Зингера  $P_{n-1}^2$ , является группа  $PSL_n^{F_2} = SL_n^{F_2}$  невырожденных матриц размерности  $n$  на  $n$  над полем  $F_2$  порядка

$$(2^n - 1)(2^n - 2) \dots (2^n - 2^{n-1}).$$

В случае характеристики 2 „проективность“ ( $P$ ) тривиальна, а невырожденность ( $L$ ) и специальность<sup>8</sup> ( $S$ ) означает одно и то же. Матрица группы  $SL_n^{F_2}$

$$A = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

состоит из  $n$  ненулевых  $n$ -столбцов над  $F_2$ , соответствующих  $n$  элементам поля  $F_q$ , где  $q = 2^n$ . Так как речь идет о матрицах полного ранга, ни один из векторов не является линейной комбинацией других. В качестве  $\alpha_1$  может быть использован любой из  $2^n - 1$  элементов  $F_q \setminus \{0\}$ , в качестве  $\alpha_2$  — любой из оставшихся  $2^n - 2$  векторов, так как одинаковыми в невырожденной матрице столбцы быть не могут. Третьим элементом  $\alpha_3$  не могут быть  $\alpha_1, \alpha_2$  и  $\alpha_1 + \alpha_2$  (сложение в  $F_q$  покоординатное), поэтому остается  $2^n - 1 - (2^2 - 1) = 2^n - 2^2$  возможных варианта. Вычитаемое  $2^2 - 1$  равно мощности подпространства  $S_1$ , образованного  $\alpha_1$  и  $\alpha_2$ , и так далее. Столбец  $\alpha_k$  выбирается

$$2^n - 1 - |S_{k-2}| = 2^n - 1 - (2^{k-1} - 1) = 2^n - 2^{k-1}$$

способами, и для последнего  $\alpha_n$  число вариантов равно  $2^n - 2^{n-1}$ .

В частности, порядок группы  $SL_3^{F_2}$  автоморфизмов  $D(7, 3, 1)$  равен  $168 = 7 \cdot 6 \cdot 4$ . Умножение матрицы  $A = (\alpha_1, \alpha_2, \alpha_3) \in SL_3^{F_2}$ ,  $\alpha_i \in F_8 \setminus \{0\}$ , на набор  $B$  ненулевых векторов из  $F_8 \setminus \{0\}$  вида

$$A \cdot B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

<sup>8</sup>Определитель матрицы равен 1.

приводит к состоящей из всевозможных сумм столбцов  $\alpha_i$  матрице

$$\begin{bmatrix} a_{13} & a_{12} & a_{12} + a_{13} & a_{11} & a_{11} + a_{13} & a_{11} + a_{12} & a_{11} + a_{12} + a_{13} \\ a_{23} & a_{22} & a_{22} + a_{23} & a_{21} & a_{21} + a_{23} & a_{21} + a_{22} & a_{21} + a_{22} + a_{23} \\ a_{33} & a_{32} & a_{32} + a_{33} & a_{31} & a_{31} + a_{33} & a_{31} + a_{32} & a_{31} + a_{32} + a_{33} \end{bmatrix}$$

и к подстановке ненулевых элементов поля  $F_8$

$$\left( \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \alpha_3 & \alpha_2 & \alpha_2 + \alpha_3 & \alpha_1 & \alpha_1 + \alpha_3 & \alpha_1 + \alpha_2 & \alpha_1 + \alpha_2 + \alpha_3 \end{array} \right). \quad (1.6)$$

Нулевой элемент инвариантен относительно этого преобразования. Например, для матрицы

$$A = (234) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

подстановка на всех элементах поля  $F_8$  имеет вид

$$(234) : \left( \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 4 & 3 & 7 & 2 & 6 & 1 & 5 \end{array} \right).$$

В результате этой подстановки блоки дизайна  $D(7, 3, 1)$  из табл.1.3 сохранят свою структуру, но порядок блоков в дизайне изменится.

$$\begin{bmatrix} 2, 4, 6 \\ 1, 4, 5 \\ 3, 4, 7 \\ 1, 2, 3 \\ 2, 5, 7 \\ 1, 6, 7 \\ 3, 5, 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1, 2, 3 \\ 2, 4, 6 \\ 2, 5, 7 \\ 3, 4, 7 \\ 3, 5, 6 \\ 1, 4, 5 \\ 1, 6, 7 \end{bmatrix}.$$

Также существенным для биометрических приложений<sup>9</sup> оказывается частный случай простой группы  $SL_4^{F_2}$  порядка  $(2^4 - 1)(2^4 - 2)(2^4 - 2^2)(2^4 - 2^3) = 15 \cdot 14 \cdot 12 \cdot 8 = 20160$  автоморфизмов дизайна  $D(15, 7, 3)$ , соответствующего  $P_3^2$ .

<sup>9</sup>В дальнейшем будет показано его использование для параметризации всевозможных искажений симметрий в структуре трех факторов.

### 1.3.4. Двойственность симптомов и синдромов

Пусть имеется случайный бинарный вектор  $X = (X_1, \dots, X_m)$  с множеством всех реализаций  $\Omega_m$ . Множество точек из  $\Omega_m$ , при которых симптом  $X_\tau$  принимает значения 0 или 1, будем обозначать через  $B_\tau$  и  $\overline{B}_\tau$  соответственно, то есть  $X_\tau$  задает разбиение множества  $\Omega_m = B_\tau \cup \overline{B}_\tau$  на непересекающиеся множества  $B_\tau \cap \overline{B}_\tau = \emptyset$  одинаковой размерности  $|B_\tau| = |\overline{B}_\tau| = 2^{m-1}$ . Блок  $B_\tau$  с точками, удовлетворяющими уравнению  $X_\tau = A_\tau X \pmod{2} = 0$ , будем называть *основным*, а блок  $\overline{B}_\tau$  с точками, удовлетворяющими уравнению  $X_\tau = A_\tau X \pmod{2} = 1$ , *дополнительным*. Вырожденному симптому  $X_\emptyset$  соответствует разбиение  $(\Omega_m, \emptyset)$ .

**Определение 4.** Блок  $B_\tau \oplus B_\mu$  вида

$$(B_\tau \cup B_\mu) \setminus (B_\tau \cap B_\mu) = (B_\tau \setminus B_\mu) \cup (B_\mu \setminus B_\tau)$$

будем называть суммой<sup>10</sup> блоков  $B_\tau$  и  $B_\mu$ .

**Теорема 4.** 1. Симптом  $X_{\tau\mu}$  задает разбиение множества  $\Omega_m$  на  $(B_{\tau\mu}, \overline{B}_{\tau\mu})$ , для которого справедливы соотношения:

$$\begin{aligned} B_{\tau\mu} &= B_\tau B_\mu + \overline{B}_\tau \overline{B}_\mu = B_\tau \oplus \overline{B}_\mu = \overline{B}_\tau \oplus B_\mu, \\ \overline{B}_{\tau\mu} &= \overline{B}_\tau B_\mu + B_\tau \overline{B}_\mu = B_\tau \oplus B_\mu = \overline{B}_\tau \oplus \overline{B}_\mu. \end{aligned} \quad (1.7)$$

2. Если над полем четной характеристики сумма элементов в блоках  $B_\tau$  и  $B_\mu$  равна соответственно  $|B_\tau| = k_1 = |\overline{B}_\tau|$  и  $|B_\mu| = k_2 = |\overline{B}_\mu|$ , то  $|B_{\tau\mu}| = k_1 + k_2$ .

**Доказательство.** 1. Блок  $B_{\tau\mu}$  состоит из элементов, удовлетворяющих условию  $X_{\tau\mu} = 0$ . Это означает, что или  $X_\tau = 0, X_\mu = 0$ , или  $X_\tau = 1, X_\mu = 1$ . Условиям  $X_\tau = 0$  и  $X_\mu = 0$  удовлетворяют наборы элементов  $B_\tau$  и  $B_\mu$ , а условиям  $X_\tau = 1$  и  $X_\mu = 1$  удовлетворяют  $\overline{B}_\tau$  и  $\overline{B}_\mu$ . Отсюда получаем необходимое выражение для  $B_{\tau\mu}$ . Кроме того, по определению 4 имеем

$$\begin{aligned} B_\tau \oplus \overline{B}_\mu &= (B_\tau \cup \overline{B}_\mu) \setminus (B_\tau \cap \overline{B}_\mu) = \\ &= (B_\tau \setminus (B_\tau \cap \overline{B}_\mu)) \cup (\overline{B}_\mu \setminus (B_\tau \cap \overline{B}_\mu)) = \\ &= (B_\tau \setminus \overline{B}_\mu) \cup (\overline{B}_\mu \setminus B_\tau) = B_\tau B_\mu \cup \overline{B}_\mu \overline{B}_\tau = B_{\tau\mu}. \end{aligned}$$

<sup>10</sup>С точки зрения теории множеств, операции  $B_\tau \oplus B_\mu$  соответствует симметрическая разность.

Аналогично могут быть получены выражения для  $\overline{B_{\tau\mu}}$ .

2. В силу того, сумма одинаковых элементов в поле четной характеристики равна нулю, получаем

$$\begin{aligned} k_1 &= |B_\tau| = |B_\tau \cap \overline{B_\mu}| + |B_\tau \setminus \overline{B_\mu}|, \\ k_2 &= |\overline{B_\mu}| = |B_\tau \cap \overline{B_\mu}| + |\overline{B_\mu} \setminus B_\tau|, \\ |B_{\tau\mu}| &= |B_\tau \oplus \overline{B_\mu}| = |B_\tau \setminus \overline{B_\mu}| + |\overline{B_\mu} \setminus B_\tau| = \\ &= k_1 + |B_\tau \cap \overline{B_\mu}| + k_2 + |B_\tau \cap \overline{B_\mu}| = k_1 + k_2. \quad \square \end{aligned}$$

Так как для любого  $\tau$  справедливо  $B_\tau \oplus B_\tau = \emptyset$  и  $B_\tau \oplus \overline{B_\tau} = \Omega_m$ , нетрудно убедиться в том, что

$$\begin{aligned} (B_\tau, \overline{B_\tau}) \oplus (B_\tau, \overline{B_\tau}) &= (B_\tau \oplus \overline{B_\tau}, B_\tau \oplus B_\tau) = (\emptyset, \Omega_m), \\ (B_\tau, \overline{B_\tau}) \oplus (B_\mu, \overline{B_\mu}) &= (B_{\tau\mu}, \overline{B_{\tau\mu}}) \iff \\ \iff (B_\tau, \overline{B_\tau}) \oplus (B_{\tau\mu}, \overline{B_{\tau\mu}}) &= (B_\mu, \overline{B_\mu}). \\ (B_\tau, \overline{B_\tau}) \oplus (B_{\tau\mu}, \overline{B_{\tau\mu}}) &= (B_\tau \oplus \overline{B_{\tau\mu}}, B_\tau \oplus B_{\tau\mu}) = \\ &= B_\tau \oplus B_\tau \oplus B_\mu, B_\tau \oplus B_\tau \oplus \overline{B_\mu} = (B_\mu, \overline{B_\mu}). \end{aligned}$$

Это означает, что через индуктивное построение пар противоположных блоков можно задавать ту же последовательность симптомов в синдроме  $S_{m-1}$ , что и в (1.5), то есть синдром  $S_{m-1}$  можно описывать с двух сторон: как проективную геометрию, в которой симптомы являются точками, или как структуру симметрий  $m$ -мерной аффинной геометрии в виде канонического дизайна над  $F_{2^m}$  с параметрами

$$v = 2^m, b = 2(2^m - 1), r = 2^m - 1, k = 2^{m-1}, \lambda = 2^{m-1} - 1. \quad (1.8)$$

В каноническом дизайне (табл.1.5), геометрической интерпретацией которого является трехмерный куб (рис.1.7), пары противоположных блоков упорядочены в соответствии с импульсным порядком симптомов в синдроме  $(X_0, X_1, X_{01}, X_2, X_{02}, X_{12}, X_{012})$ . Три пары внешних граней соответствуют разбиениям множества значений трехмерного дихотомического вектора (вершин куба) тремя базовыми симптомами, в качестве которых рассматриваются исходные признаки (симптомы единичного ранга). Три пары диагональных граней соответствуют разбиениям симптомами ранга два, тетраэдры — разбиениям симптомами ранга три.

## 1.4. Второй исключительный изоморфизм $SL_3^{F_2} \cong PSL_2^{F_7}$

Соответствие, установленное между парами параллельных гиперплоскостей аффинной геометрии  $E_n^2$  и гиперплоскостями проективной геометрии  $P_{n-1}^2$  (раздел 1.3.4), свидетельствует о том, что изменение в упорядочивании симптомов в синдроме  $S_{n-1}$  при помощи группы  $SL_n^{F_2}$  эквивалентно изменению порядка блоков соответствующего  $E_n^2$  дизайна  $D(v, b, r, k, \lambda)$  с параметрами из (1.8), которому соответствует некоторая подстановка на точках  $E_n^2$ . При  $n = 3$  эта подстановка осуществляется при помощи элементов группы  $PSL_2^{F_7}$ , что позволяет построить второй исключительный изоморфизм простых классических конечных групп порядка 168.

### 1.4.1. Группа $PSL_2^{F_7}$ и подстановки на $P_1^7$

Дизайн  $D(8, 14, 7, 4, 3)$ , геометрической интерпретацией которого является трехмерный куб, образуют элементы поля  $F_8$  (табл. 1.4). Между элементами  $F_8$  и  $P_1^7$  можно установить взаимно-однозначное соответствие. При введении системы координат точки проективной прямой  $P_1^7$  выражаются при помощи двумерных векторов:

$$\begin{bmatrix} 0 \\ c \end{bmatrix}, \begin{bmatrix} c \\ 0 \end{bmatrix}, \begin{bmatrix} c \\ c \end{bmatrix}, \begin{bmatrix} c \\ 2c \end{bmatrix}, \begin{bmatrix} c \\ 3c \end{bmatrix}, \begin{bmatrix} c \\ 4c \end{bmatrix}, \begin{bmatrix} c \\ 5c \end{bmatrix}, \begin{bmatrix} c \\ 6c \end{bmatrix},$$

где  $c \in F_7$ ,  $c \neq 0$ . Для линейных преобразований удобнее использовать нормировку, а нормированные вектора собрать в матрице

$$T_0 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}. \quad (1.9)$$

В кодировке точек  $P_1^7$  используем отношения  $k = \frac{x_0}{x_1}$ , которые над  $F_7$  принимают значения  $0, 1, \dots, 6, 7$ . Через 7 обозначено отношение  $x_1/0$ , имеющее смысл бесконечности. Итак, любой матрице над  $F_7$

$$T = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ t_{21} & \dots & t_{2n} \end{bmatrix} \quad (1.10)$$

ставится в соответствие вектор  $\pi(T) = \left[ \frac{t_{21}}{t_{11}} \quad \dots \quad \frac{t_{2n}}{t_{1n}} \right]$  над  $P_1^7$ . Нам будут интересовать матрицы с  $n = 8$  столбцами, для которых



Таблица 1.9.

$D^1(8, 14, 7, 4, 3 P_1^7)$			$D^0(8, 14, 7, 4, 3 P_1^7)$		
$i$	$B_i$	$\overline{B}_i$	$j$	$S_i$	$\overline{S}_i$
1	0256	1347	1	0125	3467
2	0345	1267	2	0234	1567
3	0157	2346	3	0267	1345
4	0136	2457	4	0146	2357
5	0467	1235	5	0137	2456
6	0237	1456	6	0457	1236
7	0124	3567	7	0356	1247

вектор  $\pi(T)$  содержит все элементы  $P_1^7$ . Матрица  $T_0$  из (1.9) задает вектор  $\pi(T_0) = [70123456]$ . Преобразование точек проективной прямой  $P_1^7$  при помощи матрицы  $A \in PSL_2^{F_7}$  приводит к подстановке

$$\begin{aligned} & \begin{pmatrix} \pi(T_0) \\ \pi(AT_0) \end{pmatrix} = \begin{pmatrix} k \\ \tau_k \end{pmatrix}_{k, \tau_k \in P_1^7} = \\ & = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \tau_7 & \tau_0 & \tau_1 & \tau_2 & \tau_3 & \tau_4 & \tau_5 & \tau_6 \end{pmatrix}. \end{aligned}$$

#### 1.4.2. Подстановки инвариантных относительно $PSL_2^{F_7}$ дизайнов $D(8, 14, 7, 4, 3|P_1^7)$

По аналогии с предл. 1 (раздел 1.2.5) могут быть получены две неавтоморфные подстановки дизайна  $D(8, 14, 7, 4, 3|P_1^7)$ , инвариантные относительно группы  $PSL_2^{F_7}$  (табл. 1.9). Обозначим через  $(B_i, \overline{B}_i)_A, (S_j, \overline{S}_j)_A$  подстановки блоков  $(B_i, \overline{B}_i)$  и  $(S_j, \overline{S}_j)$  дизайнов  $D^1(8|P_1^7) = D^1(8, 14, 7, 4, 3|P_1^7)$  и  $D^0(8|P_1^7) = D^0(8, 14, 7, 4, 3|P_1^7)$  после преобразования матрицей  $A \in PSL_2^{F_7}$ . Для подстановок на блоках используем следующие обозначения:

$$\begin{pmatrix} i \\ \alpha_i \end{pmatrix} \sim \begin{pmatrix} (B_i, \overline{B}_i) \\ (B_i, \overline{B}_i)_A \end{pmatrix}, \quad \begin{pmatrix} j \\ \beta_j \end{pmatrix} \sim \begin{pmatrix} (S_j, \overline{S}_j) \\ (S_j, \overline{S}_j)_A \end{pmatrix}, \quad (1.11)$$

где  $i, j \in \{1, \dots, 7\} = F_8 \setminus \{0\}$ . Используем сложение (1.7) на полных блоках  $(B_i, \overline{B}_i) \oplus (B_j, \overline{B}_j) = (B_k, \overline{B}_k)$ , где  $B_k = B_i B_j + \overline{B}_i \overline{B}_j$ ,  $\overline{B}_k =$

$B_i\overline{B_j} + \overline{B_i}B_j$ . Тогда через первый, второй и четвертый блоки можно выразить остальные. Для блоков  $S_i$  все аналогично. Отсюда могут быть получены соответствия для подстановок  $\alpha$  и  $\beta$  над  $F_8$ .

$$\begin{aligned} \alpha_3 &= \alpha_1 + \alpha_2, & \beta_3 &= \beta_1 + \beta_2, \\ \alpha_5 &= \alpha_1 + \alpha_4, & \beta_5 &= \beta_1 + \beta_4, \\ \alpha_6 &= \alpha_2 + \alpha_4, & \beta_6 &= \beta_2 + \beta_4, \\ \alpha_7 &= \alpha_1 + \alpha_2 + \alpha_4, & \beta_7 &= \beta_1 + \beta_2 + \beta_4. \end{aligned} \quad (1.12)$$

По этой же причине номера полных блоков дизайнов  $D^0(8|P_1^7)$  и  $D^1(8|P_1^7)$ , в сумме над полем  $F_8$  равные нулю, образуют канонический дизайн с соответствующими прямым в  $P_2^2$  (табл. 1.3) блоками дизайна  $D(7, 3, 1|F \setminus \{0\})$ :

$$246, \quad 145, \quad 347, \quad 123, \quad 257, \quad 167, \quad 356. \quad (1.13)$$

Очевидно при подстановке матрицей  $A \in PSL_2^{F_7}$  блоки дизайна (1.13) изменят свой порядок. Для автоматического вычисления номера блока  $(abc)$  введем правило  $m = \tilde{a} + \tilde{b} + \tilde{c}$ , где через  $\tilde{x}$  обозначено соответствие жестко-канонической подстановки вида

$$\left( \begin{array}{c|cccccccc} x & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline \tilde{x} & 0 & 3 & 6 & 1 & 5 & 4 & 2 & 7 \end{array} \right). \quad (1.14)$$

Из (1.14) получаем, что блок (246) имеет порядковый номер  $1 = \tilde{2} + \tilde{4} + \tilde{6} = 6 + 5 + 2$ , блок (145) имеет порядковый номер  $2 = \tilde{1} + \tilde{4} + \tilde{5} = 3 + 5 + 4$  и так далее.

**Замечание.** „Левый куб“ из табл. 1.6 образуют блоки:

$$(B_1, \overline{S}_3), (B_2, \overline{S}_6), (B_3, \overline{S}_1), (B_4, \overline{S}_5), (B_5, \overline{S}_4), (B_6, \overline{S}_2), (B_7, \overline{S}_7).$$

ПРИМЕР 1.  $A = \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \in PSL_2^{F_7}$ ,  $\text{Tr}A = 6$  над  $F_7$ .

$$\begin{aligned} AT_0 &= \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 4 & 1 & 5 & 2 & 6 & 3 & 0 & 4 \end{bmatrix}. \end{aligned}$$

С точностью до умножения на константу эта матрица равна матрице

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 6 & 1 & 3 & 5 & 0 & 2 \end{bmatrix}.$$

Отсюда получаем подстановку  $(k, \tau_k)$  на элементах  $P_1^7$  вида

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix}_{k, \tau_k \in P_1^7} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 4 & 6 & 1 & 3 & 5 & 0 & 2 \end{pmatrix}. \quad (1.15)$$

ПРИМЕР 2.  $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \in PSL_2^{F_7}$ ,  $\text{Tr} A = 2 + 1 = 3$  над  $F_7$ .

$$\begin{aligned} AT &= \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Получаем подстановку на элементах  $P_1^7$

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix}_{k, \tau_k \in P_1^7} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 2 & 3 & 4 & 5 & 6 & 0 & 1 \end{pmatrix} \quad (1.16)$$

и подстановки на полных блоках дизайнов (табл. 1.9):

$$\begin{aligned} (B_1, \overline{B_1})_A &= (0256, 1347)_A = (2401, 3567) = (B_7, \overline{B_7}), \\ (B_2, \overline{B_2})_A &= (0345, 1267)_A = (2560, 1347) = (B_1, \overline{B_1}), \\ (B_4, \overline{B_4})_A &= (0136, 2457)_A = (0467, 1235) = (B_5, \overline{B_5}), \\ (S_1, \overline{S_1})_A &= (0125, 3467)_A = (2340, 1567) = (S_2, \overline{S_2}), \\ (S_2, \overline{S_2})_A &= (0234, 1567)_A = (2456, 0137) = (S_5, \overline{S_5}), \\ (S_4, \overline{S_4})_A &= (0267, 2357)_A = (2361, 0457) = (S_6, \overline{S_6}), \end{aligned}$$

соответствующие подстановкам  $(i, \alpha_i)$  и  $(j, \beta_j)$ :

$$\begin{aligned} \begin{pmatrix} i \\ \alpha_i \end{pmatrix} &\sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 1 & 6 & 5 & 2 & 4 & 3 \end{pmatrix}, \\ \begin{pmatrix} j \\ \beta_j \end{pmatrix} &\sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 5 & 7 & 6 & 4 & 3 & 1 \end{pmatrix}. \end{aligned} \quad (1.17)$$

Заметим, что подстановки  $\alpha$  и  $\beta$  связаны между собой следующим образом:

$$\begin{aligned} \alpha_1 = 7 &= \widetilde{\beta}_2 + \widetilde{\beta}_4 + \widetilde{\beta}_6 = \widetilde{5} + \widetilde{6} + \widetilde{3} = 4 + 2 + 1 = 7, \\ \alpha_2 = 1 &= \widetilde{\beta}_1 + \widetilde{\beta}_4 + \widetilde{\beta}_5 = \widetilde{2} + \widetilde{6} + \widetilde{4} = 6 + 2 + 5 = 1, \\ \alpha_4 = 5 &= \widetilde{\beta}_1 + \widetilde{\beta}_2 + \widetilde{\beta}_3 = \widetilde{2} + \widetilde{5} + \widetilde{7} = 6 + 4 + 7 = 5. \end{aligned}$$

Аналогичные соотношения имеют место при переходе к двойственным обозначениям за счет того, что все  $\alpha$  и  $\beta$  меняются местами.

$$\begin{aligned} \beta_1 = 2 &= \widetilde{\alpha}_2 + \widetilde{\alpha}_4 + \widetilde{\alpha}_6 = \widetilde{1} + \widetilde{5} + \widetilde{4} = 3 + 4 + 5 = 2, \\ \beta_2 = 5 &= \widetilde{\alpha}_1 + \widetilde{\alpha}_4 + \widetilde{\alpha}_5 = \widetilde{7} + \widetilde{2} + \widetilde{5} = 7 + 6 + 4 = 5, \\ \beta_4 = 6 &= \widetilde{\alpha}_1 + \widetilde{\alpha}_2 + \widetilde{\alpha}_3 = \widetilde{7} + \widetilde{1} + \widetilde{6} = 7 + 3 + 2 = 6. \end{aligned}$$

Таблица 1.10.

*	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	4	6	5	7	1	3
3	3	6	5	1	2	7	4
4	4	5	1	7	3	2	6
5	5	7	2	3	6	4	1
6	6	1	7	2	4	3	5
7	7	3	4	6	1	5	2

### 1.4.3. Линейно-логарифмическое соответствие между $P_1^7$ и $F_8$

Для наведения контрастности за счет явного проявления осевых симметрий переведем дизайны с элементами над  $P_1^7$  в дизайны с элементами над  $F_8$ . В качестве элементов  $F_8$  будем рассматривать полиномы  $c_0 + c_1x + c_2x^2$ , где  $c_i \in F_2$ , а именно,  $\{0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1\} = \{0, 1, \dots, 7\}$ . Известно [25], что степени примитивного корня  $x \in F_q$  образуют циклическую группу порядка  $q-1$ ,  $x^{q-1} = 1$ . Построим над  $F_8$  таблицу умножения (табл. 1.10) и степеней примитивного корня 2 (табл. 1.11), принимая  $f(x) = x^3 + x^2 + 1$  в качестве неприводимого полинома и  $2^\infty = 0$ , где  $\infty = 7$ .

Таблица 1.11.

$y \in P_1^7$	7	0	1	2	3	4	5	6
$x = 2^y \in F_8$	0	1	2	4	5	7	3	6

Показатели степеней примитивного корня  $\{0, 1, 2, \dots, 6\}$  являются элементами  $F_7$  и вместе с  $7 = \infty$  могут рассматриваться как точки проективной прямой  $P_1^7$ .

Построим линейно-степенное преобразование  $F(y) = A(2^y)$ , где  $y \in P_1^7$ ,  $2^y \in F_8$ , матрица  $A = [147] \in SL_3^{F_2}$  задает подстановку на элементах поля  $F_8$  вида

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 7 & 4 & 3 & 1 & 6 & 5 & 2 \end{pmatrix}.$$

Таблица 1.12.

$y \in P_1^7$	7	0	1	2	3	4	5	6
$F(y) \in F_8$	0	7	4	1	6	2	3	5

$x \in F_8$	0	1	2	3	4	5	6	7
$G(x) \in P_1^7$	7	2	4	5	1	6	3	0

Результаты обратного логарифмически-линейное преобразования  $G(x) = \log_2(A^{-1}x) \in P_1^7$ , где  $x \in F_8$ , представлены в табл. 1.12. Подстановка, задаваемая матрицей  $A^{-1} = [274] \in SL_3^{F_2}$ , имеет вид

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 4 & 7 & 3 & 2 & 6 & 5 & 7 \end{pmatrix}.$$

При помощи этого линейно-степенного преобразования  $F(y)$  дизайны  $D^1(8, 14, 7, 4, 3|P_1^7)$  и  $D^0(8, 14, 7, 4, 3|P_1^7)$  (табл. 1.9) перейдут в канонический  $D^0(8, 14, 7, 4, 3|F_8)$  и порядковый<sup>11</sup>  $D^1(8, 14, 7, 4, 3|F_8)$  дизайны с точным чередованием блоков (табл. 1.13), соответствующим импульсному порядку симптомов в синдроме.

Таблица 1.13.

$D^0(8, 14, 7, 4, 3 F_8)$			$D^1(8, 14, 7, 4, 3 F_8)$		
$i$	$B_i$	$\overline{B}_i$	$j$	$S_i$	$\overline{S}_i$
1	0246	1357	1	0256	1347
2	0145	2367	2	0345	1267
3	0347	1256	3	0157	2346
4	0123	4567	4	0136	2457
5	0257	1346	5	0467	1235
6	0167	2345	6	0237	1456
7	0356	1247	7	0124	3567

<sup>11</sup>Под порядковым мы понимаем дизайн  $D^1(8, 14, 7, 4, 3|F_8)$ , для которого суммы элементов в парах противоположных блоков над  $F_8$  исчерпывают множество  $F_8 \setminus \{0\}$ .

#### 1.4.4. Двойственность $\alpha$ - и $\beta$ -подстановок

Рассмотрим комплекс следующих условий для подстановок дизайнов  $D^k(8|P_1^7) = D^k(8, 14, 7, 4, 3|P_1^7)$  и  $D^k(8|F_8)$ ,  $k = 0; 1$ .

- В результате преобразования на  $P_1^7$  при помощи  $A \in PSL_2^{F_7}$  имеются подстановки дизайнов  $D^k(8|P_1^7)$  (табл. 1.9).
- Через соответствие в табл. 1.12 подстановки  $D^k(8|P_1^7)$  переведены в  $D^k(8|F_8)$  (табл. 1.13).
- Подстановки  $(i, \alpha_i)$  и  $(j, \beta_j)$  над  $F_8$  из (1.11) соответствуют подстановкам на полных блоках  $D^k(8|F_8)$ .
- Сумма (1.7) трех полных блоков в  $D^k(8|P_1^7)$  с номерами  $a, b, c$  является вырожденным блоком  $(\emptyset, F_8)$ .

**Теорема 5.** Для подстановок  $(i, \alpha_i)$  и  $(j, \beta_j)$  из (1.11) выполнены соотношения двойственности:

$$\alpha_i = \widetilde{\beta}_a + \widetilde{\beta}_b + \widetilde{\beta}_c, \quad \text{где } i = \widetilde{a} + \widetilde{b} + \widetilde{c}; \quad (1.18)$$

$$\beta_j = \widetilde{\alpha}_a + \widetilde{\alpha}_b + \widetilde{\alpha}_c, \quad \text{где } j = \widetilde{a} + \widetilde{b} + \widetilde{c}. \quad (1.19)$$

**Доказательство.** Определим четный гипосиндром  $\Delta_j = \Delta_j(abc)$ , где  $j = \widetilde{a} + \widetilde{b} + \widetilde{c}$ , как образованное блоками канонического дизайна  $D^0(8|F_8)$  множество пар элементов поля  $F_8$

$$\Delta_j = \Delta_j(abc) = \{B_a B_b, B_a \overline{B_b}, \overline{B_a} B_b, \overline{B_a} \overline{B_b}\}. \quad (1.20)$$

Нечетный гипосиндром  $\nabla_i = \nabla_i(abc)$ , где  $\widetilde{i} = \widetilde{a} + \widetilde{b} + \widetilde{c}$ , — это образованное блоками порядкового дизайна  $D^1(8|F_8)$  множество пар элементов поля  $F_8$

$$\nabla_i = \nabla_i(abc) = \{S_a S_b, S_a \overline{S_b}, \overline{S_a} S_b, \overline{S_a} \overline{S_b}\}. \quad (1.21)$$

Нумерация гипосиндромов устроена таким образом, что для любого  $j = 1, \dots, 7$  выполняются соотношения инцидентности вида  $(0j) \in \Delta_j$  или  $(0i) \in \nabla_i$ . Например,  $\Delta_1 = \Delta_1(246) = \{01, 23, 45, 67\}$ ,  $\nabla_1 = \nabla_3 = \{01, 24, 36, 57\}$ .

Геометрически гипосиндром представляет собой пучок из четырех параллельных прямых. Для каждого такого пучка существует

единственная пара противоположных плоскостей, на которую он опирается. Например, для нечетного гипосиндрома

$$\nabla_1 = \nabla_1(347) = \{01, 24, 36, 57\}$$

среди полных блоков  $(B_i, \overline{B}_i)$  выберем блоки, не инцидентные гипосиндрому  $\nabla_1 = \nabla_1(347)$ , т.е. те, которые не содержат ни одну пару элементов из  $\nabla_1$ . Непосредственно убеждаемся, что существует всего один такой блок  $(B_3, \overline{B}_3) = (0347, 1256)$ . Остальные блоки не подходят, так как первый блок  $B_1 = (0246, 1357)$  содержит пару (24), второй  $B_2 = (0145, 2367)$ , четвертый  $B_4 = (0123, 4567)$  и шестой  $B_6 = (0167, 2345)$  пару (01), пятый  $B_5 = (0257, 1346)$  пару (57), седьмой  $B_7 = (0356, 1247)$  пару (36).

Подобное взаимно-однозначное соответствие может быть установлено между всеми остальными нечетными гипосиндромами и блоками канонического дизайна, а также между четными гипосиндромами и блоками порядкового дизайна (табл. 1.14). Нетрудно заметить, что подстановке  $(i, \alpha_i)$ , эквивалентной подстановке на блоках  $((B_i, \overline{B}_i), (B_i, \overline{B}_i)_A)$ , ставится в соответствие подстановка на гипосиндромах  $(\nabla_{\tilde{i}}, \nabla_{\tilde{\alpha}_i})$ , а подстановке  $(j, \beta_j)$ , эквивалентной подстановке на  $((S_j, \overline{S}_j), (S_j, \overline{S}_j)_A)$ , ставится в соответствие подстановка на гипосиндромах  $(\Delta_j, \Delta_{\beta_j})$ . Например, из

$$(\Delta_1(246))_A = \Delta_{\beta_1}(\alpha_2\alpha_4\alpha_6)$$

по определению гипосиндрома  $\Delta_j(abc)$  получаем  $\beta_1 = \widetilde{\alpha}_2 + \widetilde{\alpha}_4 + \widetilde{\alpha}_6$ . В общем виде  $\beta_j = \widetilde{\alpha}_a + \widetilde{\alpha}_b + \widetilde{\alpha}_c$  следует из

$$(\Delta_j(abc))_A = \Delta_{\beta_j}(\alpha_a\alpha_b\alpha_c).$$

Аналогично, из

$$(\nabla_{\tilde{1}}(246))_A = \nabla_{\tilde{\alpha}_1}(\beta_2\beta_4\beta_6)$$

по определению гипосиндрома  $\nabla_j(abc)$  получаем  $\alpha_1 = \widetilde{\beta}_2 + \widetilde{\beta}_4 + \widetilde{\beta}_6$ . И в общем виде из

$$(\nabla_{\tilde{i}}(abc))_A = \nabla_{\tilde{\alpha}_i}(\beta_a\beta_b\beta_c) \quad (1.22)$$

по определению гипосиндрома  $\nabla_j(abc)$  получаем  $\alpha_i = \widetilde{\beta}_a + \widetilde{\beta}_b + \widetilde{\beta}_c$ . Равенства (1.18) и (1.18) доказаны.

Таблица 1.14.

$abc$	$B_1$	$S_a S_b$	$S_a \bar{S}_b$	$\bar{S}_a S_b$	$\bar{S}_a \bar{S}_b$	$S_a S_b$	$S_a \bar{S}_b$	$\bar{S}_a S_b$	$\bar{S}_a \bar{S}_b$	$\nabla_3$	$S_1$	$B_a B_b$	$B_a \bar{B}_b$	$\bar{B}_a B_b$	$\bar{B}_a \bar{B}_b$	$\Delta_1$
246	$\frac{B_1}{B_1}$	0	4	6	2	2	7	7	7	$\bar{1}=3$	$\frac{S_1}{S_1}$	0	5	2	6	$\Delta_1$
145	$\frac{B_2}{B_2}$	0	5	1	4	4	7	7	7	$\nabla_6$	$\frac{S_2}{S_2}$	0	4	3	5	$\Delta_2$
347	$\frac{B_3}{B_3}$	0	7	3	4	4	7	7	7	$\nabla_1$	$\frac{S_3}{S_3}$	0	7	1	5	$\Delta_3$
123	$\frac{B_4}{B_4}$	0	2	3	1	1	7	7	7	$\bar{3}=1$	$\frac{S_4}{S_4}$	0	6	1	3	$\Delta_4$
257	$\frac{B_5}{B_5}$	0	5	7	2	2	7	7	7	$\nabla_4$	$\frac{S_5}{S_5}$	0	4	7	6	$\Delta_5$
167	$\frac{B_6}{B_6}$	0	6	7	1	1	7	7	7	$\bar{5}=4$	$\frac{S_6}{S_6}$	0	2	7	3	$\Delta_6$
356	$\frac{B_7}{B_7}$	0	5	6	3	3	4	4	4	$\bar{6}=2$	$\frac{S_7}{S_7}$	0	4	2	1	$\Delta_7$



### 1.4.5. Сдвиговой параметр

Матрица из группы  $PSL_2^{F_7}$  задает подстановку  $(m, \tau_m)$  над  $P_1^7$ , из которой при помощи логарифмически-линейного преобразования из табл. 1.12 получается подстановка  $(k, t_k)$  над  $F_8$  элементов канонического и порядкового дизайнов  $D(8) = D(8, 14, 7, 4, 3|F_8)$  из табл. 1.13. Аналитически  $\alpha$ - и  $\beta$ -подстановки на полных блоках этих дизайнов выражаются над  $F_8$  следующим образом:

$$\begin{aligned} \alpha_1 &= \tilde{t}_0 + \tilde{t}_2 + \tilde{t}_4 + \tilde{t}_6, & \beta_1 &= t_0 + t_2 + t_5 + t_6, \\ \alpha_2 &= \tilde{t}_0 + \tilde{t}_1 + \tilde{t}_4 + \tilde{t}_5, & \beta_2 &= t_0 + t_3 + t_4 + t_5, \\ \alpha_4 &= \tilde{t}_0 + \tilde{t}_1 + \tilde{t}_2 + \tilde{t}_3, & \beta_4 &= t_0 + t_1 + t_3 + t_6. \end{aligned} \quad (1.23)$$

Из равенства нулю четной суммы одинаковых слагаемых над полем четной характеристики получаем, что  $\beta$ -подстановка и, следовательно, двойственная ей  $\alpha$ -подстановка инвариантны относительно сдвига, то есть подстановки  $(k, t_k + s)$ ,  $s \in F_8$ .

**Предложение 2.** *Блоки канонического дизайна инвариантны относительно сдвига  $(k, t_k + s)$ ,  $s \in F_8$ .*

Действительно, пусть основной и дополнительный блоки имеют соответственно вид  $B_i = (0, a, b, a + b)$  и  $\overline{B}_i = (c, a + c, b + c, a + b + c)$ . Если  $s \in B_i$ , то при сдвиге основной блок  $(s, a + s, b + s, a + b + s)$  не изменится. Например, при  $s = a$  и  $s = a + b$  имеем соответственно  $B_i = (a, 0, b + a, b)$  и  $B_i = (a + b, b, a, 0)$ , далее аналогично. Если  $s \in \overline{B}_i$ , то  $(s, a + s, b + s, a + b + s) = \overline{B}_i$ , например, при  $s = c$  и  $s = a + c$  имеем соответственно  $(c, a + c, b + c, a + b + c) = \overline{B}_i$  и  $(a + c, c, a + b + c, b + c) = \overline{B}_i$ .  $\square$

Порядковый дизайн  $D_0^1(8)$  при  $(k, t_k + s)$ , где  $s \neq 0$ , преобразуется в порядковый дизайн  $D_s^1(8)$  (табл.1.21). Так как над  $F_8$  выполняется  $(a_1 + s) + (a_2 + s) + (a_3 + s) + (a_4 + s) = a_1 + a_2 + a_3 + a_4$ , дизайн останется порядковым, т.е.  $\beta$ -подстановка инвариантна относительно сдвига. Если  $(a_1, a_2, a_3, a_4) = (0abc)$ ,  $s \in (0abc)$ , тогда, например, при  $s = a$  имеем также  $(a + b + c)$ -блок, но  $(a, 0, a + b, a + c) \neq (0abc)$ , а при  $s = a + b + c$  получим противоположный блок  $(a + b + c, b + c, a + c, a + b)$ , и  $s$ -блок дизайна  $D_0^1$  содержится в  $D_s^1$ .

Подстановка  $(k, t_k)$  может быть выражена через  $\beta$ -подстановку в виде  $t_k = \beta_k + t_0$ . Например, первые базовые элементы равны

$$\begin{aligned} \beta_1 &= t_0 + t_2 + t_5 + t_6 = t_0 + (t_2 + t_5 + t_6 + t_1) + t_1 = t_0 + t_1, \\ \beta_2 &= t_0 + t_3 + t_4 + t_5 = t_0 + (t_3 + t_4 + t_5 + t_2) + t_2 = t_0 + t_2, \end{aligned}$$

откуда  $\beta_3 = \beta_1 + \beta_2 = t_0 + (t_0 + t_1 + t_2 + t_3) + t_3 = t_0 + t_3$ . Таким образом, если нужно построить переход от  $\alpha$ - и  $\beta$ -подстановок над  $F_8 \setminus \{0\}$  к  $(k, t_k)$ , то достаточно найти только  $t_0$ .

**Замечание.** Если  $D_0^1(8)$  инвариантен относительно  $(k, \beta_k)$ , то  $t_0 = 0$ . В случае, когда при подстановке  $(k, \beta_k)$  дизайн  $D_0^1(8)$  переходит в  $D_s^1(8)$ , а это видно по наличию соответствующего  $s$ -блока, то  $t_0 = s$ , так как для того чтобы получить вместо  $D_s^1(8)$  дизайн  $D_0^1(8)$ , нужно использовать подстановку  $(k, \beta_k + s)$ . Помимо нахождения сдвига через порядковые дизайны, существует еще один способ вычисления  $t_0$  через условие совпадения индексов или когерентность.

#### 1.4.6. Теорема когерентности

**Теорема 6.** Пусть имеются  $\alpha$ -,  $\beta$ -подстановки и  $(k, t_k)$  на  $F_8$  из (1.23). Тогда или существует единственное  $k$ , при котором  $(k, \beta_k) = (k, \widetilde{\alpha}_k)$  и  $t_0 = \beta_k$ , или  $\forall k \in F_8$   $(k, \beta_k) = (k, \widetilde{\alpha}_k)$  и  $t_0 = 0$ .

**Доказательство.** Для нечетных гипосиндромов  $\nabla$ , согласованных с подстановкой  $(i, \alpha_i)$ , имеет место преобразование

$$(\nabla_k)_A = (\nabla_{\widetilde{k}})_A = \nabla_{\widetilde{\alpha}_k}.$$

Следовательно, условие  $\beta_k = \widetilde{\alpha}_k = m$  означает, что разного рода гипосиндромы  $\Delta_k$  и  $\nabla_k$ , содержащие один и тот же блок  $(0k)$ , отображаются в соответствующие гипосиндромы, содержащие один и тот же блок  $(0m)$ .

$$\begin{cases} (\Delta_k)_A = \Delta_m \\ (\nabla_k)_A = \nabla_m \end{cases} \implies (0k)_A = (0m),$$

$$\begin{cases} (\Delta_j)_A = \Delta_n \\ (\nabla_j)_A = \nabla_l \end{cases} \implies (0j)_A = (\Delta_n \cap \nabla_l).$$

Если  $n = l$ , тогда  $(0j)_A = (\Delta_j \cap \nabla_j)_A = \Delta_n \cap \nabla_n = (0n)$ , и блок  $(\Delta_n \cap \nabla_l)$  содержит ноль. Следовательно,  $(0)_A = ((0k) \cap (0j))_A = (0m) \cap (0n) = 0$ , то есть  $t_0 = 0$ . При этом если  $(\Delta_i)_A = \Delta_j$ , то  $(\nabla_i)_A = \nabla_j$ .

В случае, когда последний блок  $(\Delta_n \cap \nabla_l)$  не содержит ноль ( $n \neq l$ ), тогда  $t_0 = m$ , так как

$$(0)_A = ((0k) \cap (0j))_A = (0m) \cap (\Delta_n \cap \nabla_l) = m. \quad \square$$

ПРИМЕР 2 (продолжение).

Для  $(i, \alpha_i)$  и  $(j, \beta_j)$  из (1.17) проверим условие когерентности. Подстановки

$$\begin{pmatrix} \tilde{i} \\ \tilde{\alpha}_i \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 5 & 7 & 6 & 4 & 3 & 1 \end{pmatrix},$$

$$\begin{pmatrix} j \\ \beta_j \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 5 & 7 & 6 & 4 & 3 & 1 \end{pmatrix}$$

совпадают, следовательно,  $t_0 = 0$ ,

$$\begin{pmatrix} l \\ t_l \end{pmatrix}_{l, t_l \in F_8} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 2 & 5 & 7 & 6 & 4 & 3 & 1 \end{pmatrix}.$$

При помощи табл. 1.12 от  $(l, t_l)$  из  $F_8$  переходим к обозначениям в  $P_1^7$ .

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 2 & 3 & 4 & 5 & 6 & 0 & 1 \end{pmatrix}.$$

Для вычисления матрицы

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in PSL_2^{F_7}$$

в поле  $F_7$ , подразумевая, что 7 имеет смысл бесконечности, находим

$$a^2 = 1/(\tau_1 - \tau_0) - 1/(\tau_7 - \tau_0) = 1/(3 - 2) - 1/(7 - 2) = 1 - 0 = 1,$$

откуда  $a = 1$  или  $a = 6$ . Используем

$$c = a\tau_0 = 1 \cdot 2 = 2; d = b\tau_7 = b \cdot 7, \implies b = 0.$$

Определитель матрицы  $A = \begin{bmatrix} 1 & 0 \\ 2 & d \end{bmatrix}$  равен  $d = 1$ . Получаем две матрицы

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \text{ и } A = \begin{bmatrix} 6 & 0 \\ 5 & 6 \end{bmatrix},$$

определяемые с точностью до умножения на  $-1 = 6$  в  $F_7$ , и проверяем вычисление.

$$\begin{aligned} AT &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}, \end{aligned}$$

где матрица  $T$  имеет вид (1.10). Получаем искомую подстановку

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix} = \begin{pmatrix} \pi(T) \\ \pi(AT) \end{pmatrix} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 2 & 3 & 4 & 5 & 6 & 0 & 1 \end{pmatrix}.$$

**Замечание.** Если  $a = 0$ , то для вычисления остальных коэффициентов можно воспользоваться выражением  $b^2 = \frac{1}{\tau_7 - \tau_1}$ .

ПРИМЕР 3.

$$\begin{pmatrix} i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 4 & 1 & 2 & 7 & 6 & 3 \end{pmatrix}.$$

Из соотношений двойственности (1.18) и (1.19)

$$\begin{aligned} \widetilde{\alpha}_2 + \widetilde{\alpha}_4 + \widetilde{\alpha}_6 &= \widetilde{4} + \widetilde{2} + \widetilde{6} = 5 + 6 + 2 = 1 = \beta_1, \\ \widetilde{\alpha}_1 + \widetilde{\alpha}_4 + \widetilde{\alpha}_5 &= \widetilde{5} + \widetilde{2} + \widetilde{7} = 4 + 6 + 7 = 5 = \beta_2, \\ \widetilde{\alpha}_1 + \widetilde{\alpha}_2 + \widetilde{\alpha}_3 &= \widetilde{5} + \widetilde{4} + \widetilde{1} = 4 + 5 + 3 = 2 = \beta_4 \end{aligned}$$

получаем подстановки:

$$\begin{pmatrix} j \\ \beta_j \end{pmatrix} = \begin{pmatrix} 1 & 2 & \mathbf{3} & 4 & 5 & 6 & 7 \\ 1 & 5 & \mathbf{4} & 2 & 3 & 7 & 6 \end{pmatrix},$$

$$\begin{pmatrix} \widetilde{i} \\ \widetilde{\alpha}_i \end{pmatrix} = \begin{pmatrix} 1 & 2 & \mathbf{3} & 4 & 5 & 6 & 7 \\ 3 & 2 & \mathbf{4} & 7 & 6 & 5 & 1 \end{pmatrix}.$$

Следовательно,  $t_0 = 4$ ,

$$\begin{pmatrix} l \\ t_l \end{pmatrix}_{l, t_l \in F_8} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 1 & 0 & 6 & 7 & 3 & 2 \end{pmatrix}.$$

При помощи табл. 1.12 от  $(l, t_l)$  из  $F_8$  переходим к обозначениям в  $P_1^7$ .

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 3 & 6 & 5 & 2 & 7 & 0 \end{pmatrix}.$$

Для вычисления матрицы  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in PSL_2^{F_7}$  в поле  $F_7$  находим

$$\begin{aligned} a^2 &= 1/(\tau_1 - \tau_0) - 1/(\tau_7 - \tau_0) = \\ &= 1/(3 - 4) - 1/(1 - 4) = 1/6 - 1/4 = 6 - 2 = 4, \end{aligned}$$

откуда  $a = 2$  или  $a = 5$ . Используем

$$c = a\tau_0 = 2 \cdot 4 = 1; d = b\tau_7 = b.$$

Определитель матрицы  $A = \begin{bmatrix} 2 & b \\ 1 & b \end{bmatrix}$  равен  $2b - b = b = 1$ . Получаем две матрицы

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \text{ и } A = \begin{bmatrix} 5 & 6 \\ 6 & 6 \end{bmatrix},$$

определяемые с точностью до умножения на  $-1 = 6$  в  $F_7$ . Проверяем правильность вычислений.

$$\begin{aligned} AT &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 0 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 & 6 & 0 \end{bmatrix}, \end{aligned}$$

где матрица  $T$  имеет вид (1.10). Получаем искомую подстановку

$$\begin{pmatrix} k \\ \tau_k \end{pmatrix} = \begin{pmatrix} \pi(T) \\ \pi(AT) \end{pmatrix} = \begin{pmatrix} 7 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 3 & 6 & 5 & 2 & 7 & 0 \end{pmatrix}.$$

## 1.5. Информативность симптомов

### 1.5.1. Распределение синдрома

Для категориальных признаков теряют всякий смысл понятия среднего и дисперсии, и основными объектами исследования являются законы распределения или информационные характеристики.

Пронумеруем элементы множества  $\Omega_n$  значений  $(x_1, \dots, x_n)$  случайного дихотомического вектора  $X = (X_1, \dots, X_n)^T$  по правилу  $j = x_1 + 2x_2 + \dots + 2^{n-1}x_n$ ,  $j = 0, 1, \dots, N = 2^n - 1$ , и обозначим совместное распределение через  $(p_0, \dots, p_N)$ , где

$$p_j = P(X_1 = x_1, \dots, X_n = x_n), \quad \sum_{j=0}^N p_j = 1.$$

Вероятность того, что  $X$  принимает значения с номерами из множества  $B \subseteq \{0, 1, 2, \dots, N\}$ , равна

$$P(B) = \sum_{j \in B} p_j. \quad (1.24)$$

Если известно совместное распределение симптомов  $X_\tau$  и  $X_\mu$

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} P(B_\tau B_\mu) & P(B_\tau \bar{B}_\mu) \\ P(\bar{B}_\tau B_\mu) & P(\bar{B}_\tau \bar{B}_\mu) \end{bmatrix},$$

где через  $B_\tau \subseteq \{0, 1, 2, \dots, N\}$  обозначено множество значений  $X$ , удовлетворяющих условию  $X_\tau = 0$  (раздел 1.3.4),  $|B_\tau| = 2^{n-1}$ , то распределение симптомов  $X_\tau$ ,  $X_\mu$ , и  $X_{\tau\mu} = X_\tau + X_\mu \pmod{2}$  может быть получено соответственно в виде

$$(p_{00} + p_{01}, p_{10} + p_{11}), (p_{00} + p_{10}, p_{01} + p_{11}), (p_{00} + p_{11}, p_{01} + p_{10}).$$

**Определение 5.** Распределение синдрома  $k$ -го порядка есть совместное распределение базовых симптомов  $X_{\tau_0}, X_{\tau_1}, \dots, X_{\tau_k}$ .

Для вычисления распределения синдромов нулевого порядка или симптомов  $X_{\tau_i}$ , где  $\tau_i \subseteq \{1, 2, \dots, n\}$ , воспользуемся тем, что пара противоположных блоков симптома  $X_{\tau_i}$  задается с точностью

до перестановки основного и дополнительного блоков в виде прямой суммы

$$\{B_{\tau_i}, \overline{B}_{\tau_i}\} = \left\{ \sum_{t \in \tau_i} B_t, \overline{\sum_{t \in \tau_i} B_t} \right\},$$

а вероятности  $P(B_{\tau_i})$  и  $P(\overline{B}_{\tau_i})$  вычисляются при помощи (1.24). Таким образом, в  $S_0$  распределение задается вероятностями на двух множествах мощности  $2^{n-1}$ .

Если известны частные распределения симптомов  $X_\tau$  и  $X_\mu$ , то их совместное распределение задается вероятностями, определяемыми на  $2^2$  множествах мощности  $2^{n-2}$ , полученных через тензорное произведение

$$(B_\tau, \overline{B}_\tau) \otimes (B_\mu, \overline{B}_\mu) = (B_\tau B_\mu, B_\tau \overline{B}_\mu, \overline{B}_\tau B_\mu, \overline{B}_\tau \overline{B}_\mu).$$

Для трех и более независимых симптомов  $X_{\tau_1}, \dots, X_{\tau_k}$ ,  $k \leq n$ , совместное распределение определяется вероятностями на  $2^k$  множествах мощности  $2^{n-k}$  вида

$$(B_{\tau_1}, \overline{B}_{\tau_1}) \otimes \dots \otimes (B_{\tau_n}, \overline{B}_{\tau_n}). \quad (1.25)$$

При вычислении совместного распределения  $n$  независимых симптомов в соответствии с (1.25) через тензорные произведения получаем  $2^n$  множеств единичной мощности, которые исчерпывают все компоненты вектора распределения  $(p_0, p_1, \dots, p_N)$ , где  $N = 2^n - 1$ . Таким образом, доказано следующее утверждение.

**Теорема 7.** *Распределение синдрома  $(n-1)$ -го порядка с базовыми симптомами  $X_{\tau_0}, X_{\tau_1}, \dots, X_{\tau_{n-1}}$  совпадает с совместным распределением компонент исходного дихотомического случайного вектора  $(X_1, \dots, X_n)^T$  с точностью до перестановки.*

### 1.5.2. Информационные главные компоненты

Энтропия случайного бинарного вектора  $X = (X_1, \dots, X_n)^T$  с совместным распределением  $p_0, \dots, p_N$ ,  $N = 2^n - 1$ , вычисляется по формуле

$$H(X_1, \dots, X_n) = - \sum_{i=0}^N p_i \log_2 p_i. \quad (1.26)$$

Таблица 1.15.

$i$	0	1	2	3	4	5	6	7
$x_3$	0	0	0	0	1	1	1	1
$x_2$	0	0	1	1	0	0	1	1
$x_1$	0	1	0	1	0	1	0	1
$n_i$	117	26	74	21	39	9	37	8
$\frac{n_i}{n}$	0.353	0.079	0.224	0.063	0.118	0.027	0.112	0.024

С помощью энтропии в битах измеряется степень неопределенности вектора  $X$ . За один бит принимается энтропия дихотомической случайной величины с равновероятными значениями. Используя условную энтропию

$$H(Y|X) = \sum_{i=0}^N H(Y|X = x_i)P\{X = x_i\},$$

можно показать, насколько уменьшается неопределенность одного признака, если известно значение другого. Нетрудно убедиться в том, что

$$H(Y|X) = H(X, Y) - H(X). \quad (1.27)$$

ПРИМЕР. В табл. 1.15 представлены данные о числе наркоманов, которые характеризуются тремя бинарными признаками. Если  $X_1 = 1$ , то длительность употребления героина у мужчин больше пяти лет, у женщин больше трех лет, иначе  $X_1 = 0$ . При  $X_2 = 1$  назначался антидепрессант, при  $X_2 = 0$  плацебо. Признак  $X_3$  означает выполнение программы отказа от употребления наркотиков. Возможны восемь вариантов значений вектора  $(X_1, X_2, X_3)$ , которые через  $i = X_1 + 2X_2 + 4X_3$  кодируются числами от 0 до 7.

Используя относительные частоты  $\hat{p}_i = \frac{n_i}{n}$ ,  $i = 0, 1, \dots, 7$ , из табл. 1.15 и разбиение  $\Omega_3$  при помощи симптомов на блоки дизайна  $D(8, 14, 7, 4, 3)$  из табл. 1.4, вычисляем вероятности  $P\{X_\tau = 0\} = P(B_\tau)$  и  $P\{X_\tau = 1\} = P(\bar{B}_\tau)$  в распределении симптомов по (1.24) и по (1.26) их энтропии (табл. 1.16).

Поскольку, как было показано в теор.7, распределение синдрома при изменении базовых симптомов изменяется с точностью до перестановки, энтропия синдрома оказывается инвариантной относительно выбора базовых симптомов. На основе этого доказывается следующее утверждение.

Таблица 1.16.

$\tau$	$(B_\tau, \bar{B}_\tau)$	$P(B_\tau)$	$P(\bar{B}_\tau)$	$H(X_\tau)$
1	(0246, 1357)	0.807	0.193	0.708
2	(0145, 2367)	0.577	0.423	0.983
12	(0347, 1256)	0.559	0.441	0.990
3	(0123, 4567)	0.719	0.281	0.857
13	(0257, 1346)	0.628	0.372	0.952
23	(0167, 2345)	0.586	0.414	0.987
123	(0356, 1247)	0.556	0.444	0.991

**Предложение 3.** Пусть  $X_1, \dots, X_n$  являются базовыми симптомами в синдроме  $S_{n-1}$ . Тогда для симптомов  $X_{\tau_j} \in S_{j-1} \cup X_\emptyset$  имеет место разложение энтропии синдрома в виде

$$H(S_{n-1}) = H(X_1) + \sum_{j=1}^{n-1} H(X_{j+1} + X_{\tau_j} | S_{j-1}).$$

Например, энтропию  $H(S_2)$  синдрома  $S_2$ , который задается базовыми симптомами  $(X_1, X_2, X_3)$ , можно разложить восемью способами.

$$H(S_2) = H(X_1) + \left\{ \begin{array}{l} H(X_2|X_1) \\ H(X_{12}|X_1) \end{array} \right\} + \left\{ \begin{array}{l} H(X_3|X_1, X_2) \\ H(X_{13}|X_1, X_2) \\ H(X_{23}|X_1, X_2) \\ H(X_{123}|X_1, X_2) \end{array} \right\}.$$

В многомерном статистическом анализе метод главных компонент предполагает переход от коррелированных признаков к наиболее информативным некоррелированным главным компонентам. Применение этого метода к дихотомическим данным не позволяет сохранить дихотомическую структуру главных компонент.

Возможен подход, при котором главные компоненты остаются дихотомическими, а некоррелированность заменяется меньшей связностью. Первой информационной главной компонентой считается симптом с наибольшей энтропией. Симптомы с наибольшей условной энтропией относительно первой главной компоненты относятся ко второй информационной главной компоненте, с наибольшей условной энтропией относительно первых двух главных компонент относятся к третьей и так далее. Количество информационных главных компонент совпадает с количеством признаков.



Можно упорядочивать главные компоненты при последовательном удалении наименее информативных симптомов.

Согласно (1.26), энтропия синдрома  $S_2$  равна  $H(X_1, X_2, X_3) = 2.542$ . Самая большая энтропия  $H(X_{123}) = 0.991$  у симптома  $X_{123}$ . Его вклад в общую энтропию равен 39%. Для нахождения следующей информационной главной компоненты нужны условные энтропии, но на основании (1.27) достаточно значений энтропии синдромов  $S_1$  (табл.1.17). Среди синдромов  $S_1$ , содержащих симптом  $X_{123}$ , наибольшую энтропию имеет синдром, состоящий из симптомов  $X_2, X_{13}$  и  $X_{123}$ , с 76%-м вкладом в общую энтропию,  $H(X_2, X_{13}) = 1.931$ .

Можно поступить иначе: удалить менее информативные симптомы, выбирая синдром 1-го порядка с наибольшей энтропией. Энтропия синдрома  $S_1$ , содержащего симптомы  $X_{12}, X_{13}, X_{23}$ , максимальна и равна  $H(X_{12}, X_{13}) = 1.933$ . Из  $X_{12}, X_{13}, X_{23}$  отбрасываем менее информативные симптомы, оставляя  $X_{12}$  с энтропией  $H(X_{12}) = 0.990$ . По информативности разложение энтропии синдрома  $S_2$  оказывается таким же, как и в первом случае: 39% и 76% информации объясняется одним и двумя симптомами соответственно.

Таблица 1.17.

$\tau, \mu, (\tau \oplus \mu)$	$B_\tau B_\mu$	$\overline{B}_\tau B_\mu$	$B_\tau \overline{B}_\mu$	$\overline{B}_\tau \overline{B}_\mu$	$H(X_\tau, X_\mu)$
2,3,23	10	23	45	67	1.835
1,3,13	20	13	46	57	1.565
12,3,123	30	12	47	56	1.843
1,2,3	40	15	26	37	1.691
2,13,123	50	27	14	36	1.931
1,23,123	60	17	24	35	1.694
12,13,23	70	25	34	16	1.933

Наиболее связными являются симптомы  $X_2$  и  $X_{12}$  с коэффициентом неопределенности  $J = 28\%$ ,  $X_3$  и  $X_{13}$  ( $J = 26\%$ ),  $X_{23}$  и  $X_{123}$  ( $J = 28.5\%$ ).

Особый интерес представляет интерпретация расслоения выборки по наиболее информативному симптому  $X_{123}$ . При  $X_{123} = 0$  среди „начинающих“ наркоманов положительный итог ассоциировался с назначением антидепрессанта, а среди „бывалых“ нет. При  $X_{123} = 1$  адекватное лечение антидепрессантом получали наркоманы со стажем, а новички нет. Скорее всего, градация  $X_{123} = 0$  отвечает за эндогенный, а  $X_{123} = 1$  за экзогенный характер депрессии, вызываемый длительностью употребления наркотиков [32].

### 1.5.3. Генотип АПФ и прогноз артериальной гипертензии

В табл. 1.18 представлены данные о численности больных с тремя факторами, влияющими на артериальную гипертензию. Признак  $X_1 = 1$  означает концентрическую гипертрофию левого желудочка,  $X_1 = 0$  иначе; признак  $X_2 = 1$  соответствует плохому прогнозу (наличие неблагоприятного события в последующие 12 лет),  $X_2 = 0$  хорошему; признак  $X_3 = 1$  означает наличие в генотипе ( $ID, DD$ ) ангиотензинпревращающего фермента аллеля  $D$ , значение  $X_3 = 0$  соответствует генотипу  $II$ .

При исследовании влияния факторов генетики и формы гипертрофии на прогноз было получено, что фактор генетики  $X_3$  влияет незначимо ( $p = 0.35$  по критерию независимости хи-квадрат), а фактор формы гипертрофии  $X_1$  значимо ( $p = 0.01$ ). Но обращает на себя внимание значимая корреляция ( $p < 10^{-5}$ ) между фактором ассоциированности прогноза с формой гипертрофии  $X_{12}$  (в случае концентрической гипертрофии прогноз менее благополучен) и фактором ассоциированности аллеля  $D$  с прогнозом  $X_{13}$  (предполагается наличие аллеля  $D$  менее благополучным). Поясним это на расслоенной по симптому  $X_{13}$  выборке.

Таблица 1.18. Данные об артериальной гипертензии.

$i$	0	1	2	3	4	5	6	7
$x_3$	0	0	0	0	1	1	1	1
$x_2$	0	0	1	1	0	0	1	1
$x_1$	0	1	0	1	0	1	0	1
$n_i$	32	7	4	2	97	49	14	17
$\frac{n_i}{n}$	0.14	0.03	0.02	0.01	0.44	0.22	0.06	0.08

При концентрической гипертрофии аллель  $D$  встречается у 90% больных по сравнению с 75% у остальных больных ( $p = 0.02$ ). Если  $X_{13} = 0$  (форма гипертрофии адекватна генотипу), то 26% больных с наличием аллеля  $D$  и 11% носителей  $II$  имеют неблагоприятный прогноз. При  $X_{13} = 1$  (форма гипертрофии не адекватна генотипу), наоборот, 12.6% больных с наличием аллеля  $D$  и 22% носителей  $II$  имеют неблагоприятный прогноз.

#### 1.5.4. Симптомы и эффекты взаимодействия в дисперсионном анализе

Двухфакторная модель дисперсионного анализа с дихотомическими факторами  $X_1$  и  $X_2$  имеет вид

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \quad (1.28)$$

где  $\mu$  — генеральное среднее,  $\alpha_i$  и  $\beta_j$  — дифференциальные эффекты факторов  $X_1$  и  $X_2$ ,  $(\alpha\beta)_{ij}$  — эффекты взаимодействия факторов,  $i, j = 0, 1, k = 1, \dots, n_{ij}$ ,  $n_{11} + n_{12} + n_{21} + n_{22} = n$ ,  $e_{ijk}$  независимые нормально распределенные величины с нулевым средним и дисперсией  $\sigma^2$ . Для единственности решения по методу наименьших квадратов (МНК) на параметры накладываются ограничения.

$$\begin{aligned} u_1\alpha_1 + u_2\alpha_2 &= 0, & v_1\beta_1 + v_2\beta_2 &= 0, \\ w_{11}(\alpha\beta)_{11} + w_{12}(\alpha\beta)_{12} &= 0, & w_{11}(\alpha\beta)_{11} + w_{21}(\alpha\beta)_{21} &= 0, \\ w_{21}(\alpha\beta)_{21} + w_{22}(\alpha\beta)_{22} &= 0, & w_{12}(\alpha\beta)_{12} + w_{22}(\alpha\beta)_{22} &= 0, \end{aligned} \quad (1.29)$$

где  $u_1 + u_2 = 1$ ,  $v_1 + v_2 = 1$ ,  $w_{11} + w_{12} + w_{21} + w_{22} = 1$ . Выражения

$$u_i = \frac{n_{i.}}{n}, \quad v_j = \frac{n_{.j}}{n}, \quad w_{ij} = \frac{n_{ij}}{n}$$

соответствуют так называемому частичному плану.

Поскольку параметры модели (1.28) связаны через (1.29), искомым является вектор параметров

$$\Theta = (\mu, \alpha_1, \beta_1, (\alpha\beta)_{11})^T.$$

Остальные параметры имеют вид:

$$\begin{aligned} \alpha_2 &= -\frac{n_{1.}}{n_{2.}}\alpha_1, \\ \beta_2 &= -\frac{n_{.1}}{n_{.2}}\beta_1, \\ (\alpha\beta)_{12} &= -\frac{n_{11}}{n_{12}}(\alpha\beta)_{11}, \\ (\alpha\beta)_{21} &= -\frac{n_{11}}{n_{21}}(\alpha\beta)_{11}, \\ (\alpha\beta)_{22} &= \frac{n_{11}}{n_{22}}(\alpha\beta)_{11}. \end{aligned}$$

Таблица 1.19.

	$\mu$	$\alpha_1$	$\beta_1$	$(\alpha\beta)_{11}$
$y_{111}$	1	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{11n_{11}}$	1	1	1	1
$y_{121}$	1	1	$-\frac{n_{.1}}{n_{.2}}$	$-\frac{n_{11}}{n_{12}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{12n_{11}}$	1	1	$-\frac{n_{.1}}{n_{.2}}$	$-\frac{n_{11}}{n_{12}}$
$y_{211}$	1	$-\frac{n_{1.}}{n_{2.}}$	1	$-\frac{n_{11}}{n_{21}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{21n_{21}}$	1	$-\frac{n_{1.}}{n_{2.}}$	1	$-\frac{n_{11}}{n_{21}}$
$y_{221}$	1	$-\frac{n_{1.}}{n_{2.}}$	$-\frac{n_{.1}}{n_{.2}}$	$\frac{n_{11}}{n_{22}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_{22n_{22}}$	1	$-\frac{n_{1.}}{n_{2.}}$	$-\frac{n_{.1}}{n_{.2}}$	$\frac{n_{11}}{n_{22}}$

Минимизация квадратичной формы  $(Y - X\Theta)^T(Y - X\Theta)$  приводит к оценке вектора параметров

$$\hat{\Theta} = (X^T X)^{-1} X^T Y. \quad (1.30)$$

Для проверки гипотезы  $H_0 : \alpha_1 = \alpha_2 = 0$  достаточно ввести в рассмотрение усеченную модель с матрицей плана  $X_\alpha$ , представляющей собой матрицу плана  $X$  без второго столбца, вычислить соответственно ошибки основной и усеченной моделей

$$\begin{aligned} R_0^2 &= (Y - X\hat{\Theta})^T(Y - X\hat{\Theta}), \\ R_1^2 &= (Y - X_\alpha\hat{\Theta}_\alpha)^T(Y - X_\alpha\hat{\Theta}_\alpha) \end{aligned} \quad (1.31)$$

и построить статистику

$$F = \frac{R_1^2 - R_0^2}{R_0^2} \cdot \frac{n - r}{r_1 - r}, \quad (1.32)$$

которая при справедливости нулевой гипотезы имеет распределение Фишера с  $r_1 - r$  и  $n - r$  степенями свободы. Через  $r = 4$  и  $r_1 = 3$  обозначены ранги основной и усеченной матриц плана.

Обозначим через  $\gamma_i$ , где  $i = 1, 2$ , эффект симптома

$$X_{12} = |X_1 - X_2| = X_1 + X_2 \pmod{2}$$

и построим модель

$$y_{ijk} = \mu + \gamma_i + \beta_j + (\gamma\beta)_{ij} + e_{ijk}. \quad (1.33)$$

**Предложение 4.** При одинаковых весах в ограничениях на параметры (1.29) проверка гипотезы  $H_0 : \gamma_1 = \gamma_2 = 0$  в модели (1.33) равносильна проверке гипотезы  $H_0 : (\alpha\beta)_1 = (\alpha\beta)_2 = 0$  в (1.28).

Обозначим через  $e_1, \dots, e_4$  и  $I_1, \dots, I_4$  векторы из  $n_{11}$ ,  $n_{12}$ ,  $n_{21}$ ,  $n_{22}$  единиц и единичные матрицы соответствующей размерности. Матрицы плана  $X$  и  $\tilde{X}$  моделей (1.28) и (1.33) схематично можно изобразить следующим образом.

$\mu$	$\alpha_1$	$\beta_1$	$(\alpha\beta)_{11}$
$e_1$	$e_1$	$e_1$	$e_1$
$e_2$	$e_2$	$-e_2$	$-e_2$
$e_3$	$-e_3$	$e_3$	$-e_3$
$e_4$	$-e_4$	$-e_4$	$e_4$

$\mu$	$\gamma_1$	$\beta_1$	$(\gamma\beta)_{11}$
$e_1$	$e_1$	$e_1$	$e_1$
$e_2$	$-e_2$	$-e_2$	$e_2$
$e_3$	$-e_3$	$e_3$	$-e_3$
$e_4$	$e_4$	$-e_4$	$-e_4$

Между собой  $X$  и  $\tilde{X}$  отличаются перестановкой столбцов, которая выражается через матричное произведение  $\tilde{X} = XJ$ , где

$$J = \begin{bmatrix} I_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_2 \\ 0 & 0 & I_3 & 0 \\ 0 & I_4 & 0 & 0 \end{bmatrix}.$$

Покажем, что для моделей (1.28) и (1.33) ошибки (1.31) совпадают, то есть  $R_0^2 = \tilde{R}_0^2$ . Подставляя в  $R_0^2$  решение (1.30), получаем, что ошибка  $R_0^2$  имеет вид

$$(Y - X(X^T X)^{-1} X^T Y)^T (Y - X(X^T X)^{-1} X^T Y) = Y^T A^T A Y,$$

где  $A = I - X(X^T X)^{-1} X^T$  инвариантна относительно перестановки столбцов в  $X$ . Действительно,

$$\begin{aligned} \tilde{A} &= I - \tilde{X}(\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T = \\ &= I - X J J^{-1} X^{-1} (X^T)^{-1} (J^T)^{-1} J^T X^T = \\ &= I - X X^{-1} (X^T)^{-1} X^T = I - X(X^T X)^{-1} X^T = A. \end{aligned}$$

Очевидно, одинаковыми оказываются усеченные матрицы плана  $X_\alpha$  и  $X_\gamma$ , как и соответствующие оценки параметров (1.30) и значения статистики (1.32).

### 1.5.5. Энтропия дизайна

Исследуем подстановки<sup>12</sup> дизайна  $\mathcal{D} = D(v, b, r, k, \lambda | \Omega_m)$  с параметрами из (1.8), в которых пары противоположных блоков импульсно упорядочены согласно (1.5), а элементы в блоках удовлетворяют уравнениям  $X_\tau = 0$  и  $X_\tau = 1$ , где  $X_\tau$  — симптомы синдрома  $S_{m-1}$ ,  $\tau \subseteq \{1, 2, \dots, m\}$ . Через  $I_{\tau, \mu}$ , где  $\mu \subseteq \{1, 2, \dots, m\}$ , будем обозначать совместную информацию симптомов  $X_\tau$  и  $X_\mu$ ,

$$\begin{aligned} I(X_\tau, X_\mu) &= H(X_\tau) - H(X_\tau | X_\mu) = H(X_\mu) - H(X_\mu | X_\tau) = \\ &= H(X_\tau) + H(X_\mu) - H(X_\tau, X_\mu). \end{aligned}$$

**Определение 6.** Под *энтропией дизайна*  $\mathcal{D}$  будем понимать энтропию распределения соответствующего синдрома  $S_{m-1}$ , а под *суммарной энтропией дизайна*  $\mathcal{D}$  сумму энтропий симптомов, образующих синдром  $S_{m-1}$ , соответствующий дизайну  $\mathcal{D}$ .

Согласно теореме 7, энтропия дизайна  $\mathcal{D}$  инвариантна относительно его подстановки. Что же касается суммарной энтропии, то за некоторым исключением, как например при равномерном распределении, она, наоборот, зависит от его подстановки. Рассматривая существенный частный случай  $m = 3$ , покажем, что при минимуме суммарной энтропии симптомы связаны наименьшим образом.

**Теорема 8.** *Пары противоположных блоков, соответствующих симптомам синдрома  $S_2(X_\tau, X_\mu, X_\kappa)$ ,  $\tau, \mu, \kappa \subseteq \{1, 2, \dots, m\}$ , образуют дизайн  $D(8, 14, 7, 4, 3 | F_8)$  (табл. 1.20). Тогда суммарная энтропия прямо пропорциональна суммарной совместной информации между симптомами и равна*

$$\sum_{\lambda \subseteq G(\tau, \mu, \kappa)} H(X_\lambda) = \frac{1}{6} \sum_{\lambda_1, \lambda_2 \subseteq G(\tau, \mu, \kappa)} I_{\lambda_1 \lambda_2} + c, \quad (1.34)$$

$$\text{где } c = \frac{1}{2} \sum_{i=1}^7 H(S_1^i) = \text{const},$$

$$G(\tau, \mu, \kappa) = \{\tau, \mu, \tau \oplus \mu, \kappa, \kappa \oplus \tau, \kappa \oplus \mu, \kappa \oplus \tau \oplus \mu\}.$$

**Доказательство.** Симптомы синдрома  $S_2(X_\tau, X_\mu, X_\kappa)$  образуют (табл. 1.20) дизайн  $D(7, 3, 1)$ , в котором каждому блоку соответствует синдром первого порядка  $S_1^i$ ,  $i = 1, \dots, 7$ . Для любой пары из

<sup>12</sup>Поскольку далее речь идет только об изоморфных подстановках, термин „изоморфные“ опускается.

симптомов  $X_\tau, X_\mu, X_{\tau\mu}$  синдрома  $S_1^1$  справедливы следующие равенства:

$$\begin{aligned} I_{\tau,\mu} &= H(X_\tau) + H(X_\mu) - H(X_\tau, X_\mu), \\ I_{\tau,\tau\mu} &= H(X_\tau) + H(X_{\tau\mu}) - H(X_\tau, X_{\tau\mu}), \\ I_{\tau\mu,\mu} &= H(X_{\tau\mu}) + H(X_\mu) - H(X_{\tau\mu}, X_\mu), \end{aligned}$$

в которых последние слагаемые равны одному и тому же числу  $H(S_1^1)$  (теор. 7). Отсюда

$$I_{\tau,\mu} + I_{\tau,\tau\mu} + I_{\tau\mu,\mu} = 2(H(X_\tau) + H(X_\mu) + H(X_{\tau\mu})) - 3H(S_1^1).$$

Таблица 1.20.

$S_1^1$	$S_1^2$	$S_1^3$	$S_1^4$	$S_1^5$	$S_1^6$	$S_1^7$
$X_\mu$	$X_\mu$	$X_\mu$	$X_\tau$	$X_\tau$	$X_\kappa$	$X_{\mu\tau}$
$X_\tau$	$X_\kappa$	$X_{\tau\kappa}$	$X_\kappa$	$X_{\mu\kappa}$	$X_{\mu\tau}$	$X_{\mu\kappa}$
$X_{\mu\tau}$	$X_{\mu\kappa}$	$X_{\mu\tau\kappa}$	$X_{\tau\kappa}$	$X_{\mu\tau\kappa}$	$X_\kappa$	$X_{\tau\kappa}$

Таких уравнений всего  $7 = C_7^2/3$  и равно количеству синдромов  $S_1$  в синдроме  $S_2$  (табл. 1.20). Слагаемое  $H(X_\tau)$  встречается в трех уравнениях из семи, выделяя которые, получаем выражение

$$\sum_{\lambda_1, \lambda_2 \subseteq G(\tau, \mu, \kappa)} I_{\lambda_1, \lambda_2} = 6 \sum_{\lambda \subseteq G(\tau, \mu, \kappa)} H(X_\lambda) - 3 \sum_{i=1}^7 H(S_1^i),$$

эквивалентное (1.34). Распределение синдрома  $S_1^1$  задается суммами вероятностей на парах точек  $E_2^3$ , которые задаются четырьмя множествами вида

$$\begin{aligned} &(B_\tau, \overline{B}_\tau) \otimes (B_\mu, \overline{B}_\mu) = \\ &= (B_\tau B_\mu, B_\tau \overline{B}_\mu, \overline{B}_\tau B_\mu, \overline{B}_\tau \overline{B}_\mu). \end{aligned}$$

Семью синдромами  $S_1^i$  исчерпывается все множество из  $28 = C_8^2$  пар точек  $F_8$ , поэтому для любой подстановки  $D(8, 14, 7, 4, 3|F_8)$  сумма энтропий  $H(S_1^i)$  остается одной и той же.  $\square$

### 1.5.6. Разнообразие блоков

**Определение 7.** *Разнообразием* совокупности  $n$  элементов по признаку  $X$  с  $N$  градациями называют характеристику, которая вычисляется по формуле

$$J_X = n \ln n - \sum_{i=0}^N n_i \ln n_i,$$

где  $n_i$  — количество элементов  $i$ -й градации,  $n_1 + \dots + n_N = n$ .

Очевидно, разнообразие совокупности, состоящей из одинаковых элементов, равно нулю. Если совокупность характеризуется разными признаками, то разнообразия по отдельным признакам суммируются [27].

Как и ранее, считаем  $\Omega_m = \{0, 1, \dots, N\}$  множеством реализаций  $m$ -мерного случайного дихотомического вектора, элементы которого кодируются в двоичной системе от 0 до  $N = 2^m - 1$ . Числу  $n_i$  соответствует количество наблюдений со значением  $i \in \Omega_m$ .

Пусть  $Y$  — дихотомический признак. Обозначим множества точек, удовлетворяющие условиям  $Y = 0$  и  $Y = 1$  через  $B$  и  $\bar{B} \subseteq \Omega_m$  с количеством наблюдений соответственно

$$n_B = \sum_{i \in B} n_i \quad \text{и} \quad n_{\bar{B}} = \sum_{i \in \bar{B}} n_i.$$

**Предложение 5.** *Выборочная оценка энтропии признака  $Y$  имеет вид  $\hat{H}(Y) = \frac{1}{n} (g(\Omega_m) - g(B) - g(\bar{B}))$ , где  $g(B) = n_B \ln n_B$ .*

Это нетрудно показать, используя формулу Шеннона [28].

$$\begin{aligned} \hat{H}(Y) &= \frac{n_B}{n} \ln \frac{n}{n_B} + \frac{n_{\bar{B}}}{n} \ln \frac{n}{n_{\bar{B}}} = \\ &= \frac{1}{n} (n_B \ln n + n_{\bar{B}} \ln n - n_B \ln n_B - n_{\bar{B}} \ln n_{\bar{B}}) = \\ &= \frac{1}{n} (g(\Omega_m) - g(B) - g(\bar{B})). \end{aligned}$$

Например,  $m = 3$ , признак  $Y$  задает разбиение по блокам

$$B = (0123) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \bar{B} = (4567) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$



По координате  $X_1$  разобразие равно

$$J_{X_1}(B) = (n_0 + n_1 + n_2 + n_3) \ln(n_0 + n_1 + n_2 + n_3) - (n_0 + n_2) \ln(n_0 + n_2) - (n_1 + n_3) \ln(n_1 + n_3).$$

Его можно рассматривать как

$$J_{X_1}(B) = g(B) - g(BB_1) - g(B\bar{B}_1),$$

где блоки  $B_1 = (0246)$ ,  $\bar{B}_1 = (1357)$  соответствуют разбиению множества значений при помощи  $X_1$ . Обобщая, получаем, что разобразие в блоках  $B$  и  $\bar{B}$  по  $i$ -й координате,  $i = 1, \dots, m$ , можно представить в виде

$$\begin{aligned} J_{X_i}(B) &= g(B) - g(BB_i) - g(B\bar{B}_i), \\ J_{X_i}(\bar{B}) &= g(\bar{B}) - g(\bar{B}B_i) - g(\bar{B}\bar{B}_i). \end{aligned}$$

Если в каждом слагаемом суммы покомпонентных разобращий

$$J(B, \bar{B}) = \sum_{i=1}^m J_{X_i}(B) + J_{X_i}(\bar{B})$$

прибавить и отнять  $g(\Omega_m)$ , то разобразие  $J(B, \bar{B})$  можно выразить через условную энтропию следующим образом:

$$\begin{aligned} J(B, \bar{B}) &= \sum_{i=1}^m (g(B) - g(BB_i) - g(B\bar{B}_i) + \\ &+ g(\bar{B}) - g(\bar{B}B_i) - g(\bar{B}\bar{B}_i) + g(\Omega_m) - g(\Omega_m)) = \\ &= \sum_{i=1}^m (n\hat{H}(YX_i) - n\hat{H}(Y)) = \sum_{i=1}^m n\hat{H}(X_i|Y). \end{aligned}$$

Это означает, что разобразие, порожаемое симптомом  $Y$ , представляют собой не что иное, как остаточную по  $Y$  неопределенность о компонентах дихотомического вектора. *Чем меньше разобразие, тем более информативен для  $X_1, \dots, X_m$  оказывается симптом  $Y$ .* Можно показать, что в случае равновероятных исходов минимальным разобразием отличается канонический дизайн.

## 1.6. Обобщенный синдром

В разделе 1.3.4 было показано, что при использовании в качестве базовых симптомов в синдроме  $S_{m-1}$  компонент дихотомического вектора  $X = (X_1, \dots, X_m)^T$  пары блоков, соответствующие симптомам, образуют канонический дизайн  $D^0(v, b, r, k, \lambda | F_{2^m})$  с параметрами

$$v = 2^m, b = 2(2^m - 1), r = 2^m - 1, k = 2^{m-1}, \lambda = 2^{m-1} - 1. \quad (1.35)$$

При выборе в качестве базовых других симптомов мы получаем автоморфные<sup>13</sup> подстановки этого же дизайна, и остается неизменной структура симметрий  $\Omega_m$ , задаваемая разбиениями  $\Omega_m = B_\tau \cup \bar{B}_\tau$  при помощи симптомов  $X_\tau$  на два множества точек одинаковой мощности  $B_\tau$  и  $\bar{B}_\tau$ , где  $\tau \in \{1, 2, \dots, m\}$ .

Предположим, что в результате некоторой подстановки на элементах  $F_{2^m}$  получена новая подстановка дизайна  $D(v, b, r, k, \lambda | F_{2^m})$ , неавтоморфная каноническому дизайну. Структура блоков сохранит импульсную упорядоченность обобщенных „симптомов“, определяющих разбиения

$$\Omega_m = B_\tau \cup \bar{B}_\tau, \quad |B_\tau| = |\bar{B}_\tau|,$$

но  $\tau$  уже не может нести на себе смысловой нагрузки подмножества из  $m$  натуральных чисел. Хотя эти „симптомы“ нельзя выразить через линейную комбинацию дихотомических признаков, появляется возможность описания через  $B_\tau$  и  $\bar{B}_\tau$  искажений симметрий, внесенных подстановкой на  $F_q$ ,  $q = 2^m$ .

Распределение обобщенного синдрома строится также, как и в случае обычных симптомов (1.25). Энтропия обобщенного синдрома по  $m$  симптомам эквивалентна энтропии обычного синдрома  $S_{m-1}$ , поскольку согласно (теор. 7) дробление пространства  $\Omega_m$  по  $m$  симптомам в конечном итоге приводит к элементам  $\Omega_m$ , в том числе в обобщенном случае.

<sup>13</sup>Автоморфные подстановки дизайна означают, что при подстановке на элементах множество блоков остается тем же самым. Если множество блоков изменяется при заданной подстановке на элементах, то подстановки неавтоморфные (раздел 1.2.4).

### 1.6.1. Порядковые дизайны $D^1(7, 3, 1|F_8 \setminus \{0\})$

Возможности параметризации изоморфных подстановок дизайнов, соответствующих аффинной геометрии  $E_n^2$ , рассмотрим на примере существенного частного случая  $n = 3$ , в котором для этих целей может быть использована простая группа  $SL_4^{F_2}$  автоморфизмов дизайна  $D(15, 7, 3)$ .

Одной из изоморфных, но не автоморфных каноническому дизайну  $D(8, 14, 7, 4, 3)$  подстановок является подстановка, в которой суммы элементов в парах противоположных блоков образуют множество ненулевых элементов поля  $F_8$  (раздел 1.4.3). В [9] такой дизайн был назван *порядковым*.

**Определение 8.** Пусть число блоков в дизайне  $D(v, b, r, k, \lambda|F_q)$  с множеством блоков  $\mathcal{B} = \{B_j\}_{j=1}^b$  равно  $b = q-1$  (или  $b = 2q$  в случае наличия пар противоположных блоков  $\mathcal{B} = \{(B_j, \bar{B}_j)\}_{j=1}^b$ ). Если существует взаимно-однозначное отображение  $S : \mathcal{B} \rightarrow F_q \setminus \{0\}$ , то дизайн называется *порядковым* и обозначается  $D^1(v, b, r, k, \lambda|F_q)$ .

Для определенности в качестве отображения  $S$  будем рассматривать суммирование элементов в блоках над  $F_q$ . Сцепленность дизайнов  $D(8, 14, 7, 4, 3|F_8)$  и  $D(7, 3, 1|F_8 \setminus \{0\})$  проявляется в том, что ненулевые элементы основных блоков дизайна  $D(8, 14, 7, 4, 3|F_8)$  образуют  $D(7, 3, 1|F_8 \setminus \{0\})$ , а расширение дизайна  $D(7, 3, 1|F_8 \setminus \{0\})$  на нулевой элемент с рассмотрением всех дополнительных блоков приводит к  $D(8, 14, 7, 4, 3|F_8)$ . То есть в дизайне  $D^1(8, 14, 7, 4, 3|F_8)$  мы можем рассматривать только его существенную часть в виде дизайна  $D^1(7, 3, 1|F_8 \setminus \{0\})$  с блоками

$$(256), (345), (157), (136), (467), (237), (124). \quad (1.36)$$

В отличие от канонического порядковый дизайн не является единственным.

**Лемма 1.** *Существует четыре блока неповторяющихся ненулевых элементов поля  $F_8$ , которые в сумме дают число  $k \in F_8$ .*

**Доказательство.** Очевидно,  $k$  не входит в этот блок, так как иначе оставшаяся пара элементов должна в сумме давать ноль, что возможно только при двух одинаковых элементах. Таким образом, можно рассматривать  $C_6^2 = 15$  пар из элементов поля  $F_8 \setminus \{0, k\}$ ,

третий элемент равен сумме двух и  $k$ . Далее из этих 15 пар следует исключить три пары, которые в сумме равны  $k$ , так как для того, чтобы получить итоговую сумму  $k$ , нужно будет прибавлять ноль, что невозможно. Пар  $\{a, a + k\}$ , где  $a$  выбирается из шести разнообразных элементов, с точностью до перестановки может быть только три. Из оставшихся 12 пар с точностью до  $C_3^2 = 3$  перестановок получаем четыре разнообразные тройки, так как одну и ту же тройку  $(abc)$  образуют три возможные пары  $(ab)$ ,  $(ac)$  и  $(bc)$ .  $\square$

**Теорема 9.** *Каждая пара из восьми порядковых  $D^1(7, 3, 1|F_8 \setminus \{0\})$  дизайнов имеет по одному общему блоку.*

**Доказательство.** В качестве блока номер 1, согласно по лемме 3, может быть выбрана любая из четырех троек, в сумме дающих 1. Это могут быть (247), (256), (346) или (357). Выберем для примера блок (256).

Из четырех троек (147), (156), (345), (367), в сумме дающих 2, выбираем блоки, имеющие по одному общему элементу с (256). Это могут быть (345) или (367). Блок (147) не подходит, так как у него нет общих элементов с блоком (256), а блок (156) не подходит из-за того, что имеет с боком (256) два общих элемента.

В случае, когда первое и второе место занимают блоки (256) и (345), на четвертое место из множества

$$(127), (136), (235), (567)$$

блоков с суммой 4 можно поставить только один блок

$$(136) = \overset{(256)}{\{(127), (136)\}} \cap \overset{(345)}{\{(136), (567)\}},$$

а в случае, когда первые два места занимают блоки (256) и (367), на четвертое место ставится блок

$$(127) = \overset{(256)}{\{(127), (136)\}} \cap \overset{(367)}{\{(127), (235)\}}.$$

Таким образом количество порядковых дизайнов равно восьми.

Для построения порядковых дизайнов используются  $C_7^3 - 7 = 28$  троек ненулевых элементов поля  $F_8$ , так как исключается семь канонических блоков. Предполагая, что каждая пара порядковых дизайнов имеет по одному общему элементу, получаем исчерпывающую это множество троек сумму  $7 + 6 + 5 + 4 + 3 + 2 + 1 = 28$ .  $\square$

### 1.6.2. Субпорядковые дизайны $D^1(7, 3, 1|F_8 \setminus \{0\})$

Запишем ненулевые элементы поля  $F_8$  в виде импульсной последовательности

$$(a, b, a + b, k, a + k, b + k, a + b + k)$$

и исключим из этого набора элемент, равный  $k$ . Из шести оставшихся элементов выбираем четыре (по лемме 1) тройки с суммой, равной  $k$ .

$$\begin{aligned} & ( a \quad b \quad a + b + k ), \\ & ( a \quad a + b \quad b + k ), \\ & ( b \quad a + b \quad a + k ), \\ & ( a + k \quad b + k \quad a + b + k ). \end{aligned}$$

В результате получаем блок-схему  $D(6, 4, 2, 3, \lambda_*) = D^*(4, 6, 3, 2, 1)$ , двойственную  $D(4, 6, 3, 2, 1)$ . Это частично-сбалансированная блок-схема, для которой выполнено основное соотношение баланса  $vb = rk$ , а число встречающихся пар равно дробному числу

$$\lambda_* = \frac{r(k-1)}{v-1} = \frac{2(3-1)}{6-1} = \frac{4}{5}.$$

Рациональная конструкция параметра  $\lambda_*$  указывает на то, что из потенциальных  $15 = C_6^2$  пар в блок-схеме присутствуют только  $12 = \frac{4}{5} \cdot 15$ , а три пары  $(a, a+k)$ ,  $(b, b+k)$ ,  $(a+b, a+b+k)$  отсутствуют. Если их восполнить, добавляя к четырем блокам элементов, сумма которых равна  $k$ , три канонических  $k$ -блока

$$\begin{aligned} & ( a \quad a + k \quad k ) \\ & ( b \quad b + k \quad k ) \\ & ( a + b \quad a + b + k \quad k ), \end{aligned}$$

то получим дизайн  $D(7, 3, 1|F_8 \setminus \{0\})$ , который назовем *субпорядковым*, поскольку в нем оказываются три блока с нулевой суммой элементов, содержащие элемент  $k$ , и четыре блока порядковых дизайнов с суммой элементов, равной  $k = 1, 2, \dots, 7$ . Очевидно, имеется семь субпорядковых дизайнов.

Таблица 1.21.

Четные дизайны $D(7, 3, 1 F_8 \setminus \{0\})$														
$0'$	$1'$	$1$	$2'$	$2$	$3'$	$3$	$4'$	$4$	$5'$	$5$	$6'$	$6$	$7'$	$7$
246	246	246	567	136	567	235	147	367	147	345	125	237	125	237
145	127	567	145	145	127	136	357	256	234	467	357	346	234	126
347	235	136	235	127	347	347	267	146	456	124	456	135	267	245
123	367	345	247	357	146	157	123	123	367	156	247	256	146	157
257	156	147	137	234	236	456	156	345	257	257	236	124	137	467
167	457	125	346	256	135	124	346	247	135	236	167	167	547	134
365	134	237	126	467	245	267	245	157	126	137	134	547	365	365
Нечетные дизайны $D(7, 3, 1 F_8 \setminus \{0\})$														
$0'$	$1'$	$1$	$2'$	$2$	$3'$	$3$	$4'$	$4$	$5'$	$5$	$6'$	$6$	$7'$	$7$
256	256	256	247	123	247	146	357	127	357	145	346	167	346	167
345	367	247	345	345	367	123	147	246	156	257	147	125	156	236
157	146	123	146	367	157	157	267	235	245	356	245	134	267	456
136	127	145	567	147	235	347	136	136	127	234	567	246	235	347
467	234	357	137	156	126	245	234	145	467	467	126	356	137	257
237	457	346	125	246	134	356	125	567	134	126	237	237	457	135
124	135	167	236	257	456	267	456	347	236	137	135	457	124	124

### 1.6.3. О сумме порядковых дизайнов

$$D^1(7, 3, 1|F_8 \setminus \{0\})$$

Соберем в табл.1.21 порядковые дизайны<sup>14</sup> и присвоим дизайну из (1.36) номер  $0' = 8+0$ . Номера  $j' = 8+j$ , где  $j \in F_8 \setminus \{0\}$ , присвоим остальным порядковым дизайнам  $D_{j'}^1 = D_j^1(7, 3, 1|F_8 \setminus \{0\})$  таким образом, чтобы общий блок дизайнов  $D_{0'}^1$  и  $D_{j'}^1$  был  $j$ -блоком.

Рассмотрим два порядковых дизайна  $D_{i'}^1$  и  $D_{j'}^1$  с общим  $k$ -блоком  $(a, b, a+b+k)$ , и найдем множество  $L(a, b, a+b+k)$  всех троек ненулевых элементов поля  $F_8$ , которые имеют один и только один общий элемент с блоком  $(a, b, a+b+k)$ .

**Лемма 2.** Пусть  $a_i \in F_8 \setminus \{0\}$ ,  $i = 1, 2, 3$ . Тогда  $|L(a_1, a_2, a_3)| = 18$ .

**Доказательство.** Пусть  $B \cap (a_1, a_2, a_3) = a_i$ . Тогда  $B = (a_i, c_k, c_j)$ , где  $c_i, c_k \in \{c_1, c_2, c_3, c_4\} = F_8 \setminus \{0, a_1, a_2, a_3\}$ . Количество пар вида  $(c_k, c_j)$  равно  $C_4^2 = 6$ , следовательно,  $|L(a_1, a_2, a_3)| = 18 = 6 \cdot 3$ .  $\square$   
Введем операцию сложения на дизайнах в виде

$$D_k \oplus D_m = D_k \cap D_m \cup L(D_k \cap D_m) \setminus (D_k \cup D_m). \quad (1.37)$$

**Предложение 6.** В табл.1.21 суммой двух порядковых дизайнов является субпорядковый дизайн  $D_{i'+j'} = D_{i'} \oplus D_{j'}$ .

**Доказательство.** Из 18 троек множества  $L(a, b, a+b+k)$  три блока оказываются каноническими:  $(k, a, a+k)$ ,  $(k, b, b+k)$  и  $(k, a+b, a+b+k)$ ; три блока являются  $k$ -блоками:  $(a, a+b, b+k)$ ,  $(b, a+b, a+k)$  и  $(a+k, b+k, a+b+k)$ . Остальные 12 блоков попарно оказываются  $l$ -блоками,  $l \neq k$ .

$$\begin{array}{ll} a = b + (a+k) + (b+k); & a = (a+b+k) + (a+b) + (a+k); \\ b = a + (a+k) + (b+k); & b = (a+b+k) + (a+b) + (b+k); \\ a+b+k = a + (a+b) + (a+k); & a+b+k = b + (a+b) + (b+k); \\ a+b = a+k + (b+k); & a+b = b+k + (a+k); \\ a+k = b + (a+b) + k; & a+k = (a+b+k) + (b+k) + k; \\ b+k = a + (a+b) + k; & b+k = (a+b+k) + (a+k) + k. \end{array}$$

Поскольку порядковые дизайны исчерпывают множество  $l$ -блоков,  $l \neq k$ , для  $D_{i'+j'}$  остаются наряду с блоком  $(a, b, a+b+k)$  три канонических блока, содержащие элемент  $k$ , и три  $k$ -блока, что и соответствует структуре субпорядкового дизайна.  $\square$

<sup>14</sup>Порядковые дизайны относятся к нечетным.

#### 1.6.4. Неавтоморфные подстановки дизайна

$$D(7, 3, 1)$$

Порядковый дизайн  $D^1(7, 3, 1|F_8 \setminus \{0\})$  связан с каноническим  $D^0(7, 3, 1|F_8 \setminus \{0\})$  жестко-канонической подстановкой (1.14), соответствующей трем транспозициям:  $1 \leftrightarrow 3$ ,  $2 \leftrightarrow 6$  и  $4 \leftrightarrow 5$ . Дизайны, полученные из канонического в результате нечетного числа транспозиций называются *нечетными* [9], иначе *четными*. Порядковые и субпорядковые дизайны оказываются нечетными.

Общее число подстановок дизайнов  $D(7, 3, 1|F_8 \setminus \{0\})$  равно

$$30 = \frac{C_7^3 \cdot (3 \cdot C_4^2) \cdot (2 \cdot (C_4^2 - 2))}{(2^3 - 2^0)(2^3 - 2^1)(2^3 - 2^2)} = \frac{35 \cdot 18 \cdot 8}{24 \cdot 7} = \frac{5040}{168},$$

так как в качестве первого блока  $B_1$  можно выбрать любую из  $35 = C_7^3$  троек ненулевых элементов поля  $F_8$ , например,  $(a_1, a_2, a_3)$ . Согласно лемме 2, второй блок можно выбрать 18-ю способами. Предположим, что  $B_1 \cap B_2 = a_1$ . Блок  $B_3$  вычисляется из  $B_1$  и  $B_2$  по правилу

$$B_3 = B_1 \cap B_2 \cup (F_8 \setminus \{0, B_1 \cup B_2\}). \quad (1.38)$$

Выбирая  $B_4$  из 18 блоков, имеющих один общий элемент с  $B_1$ , имеем в виду, что  $B_4$  не содержит  $a_1$  (осталось 12), а также в него не входят пары  $B_2 \setminus a_1$  и  $B_3 \setminus a_1$ , которые в  $L(a_1, a_2, a_3)$  встречаются по три раза: с  $a_1$ ,  $a_2$  и  $a_3$ . Таким образом, остается 8 вариантов построения блока  $B_4$ . Например,  $B_1 = (1, 2, 3)$ , тогда 18 блоков вида

$$\begin{array}{cccccc} \mathbf{145} & \mathbf{146} & \mathbf{147} & \mathbf{156} & \mathbf{157} & \mathbf{167} \\ \mathbf{245} & 246 & 247 & 256 & 257 & \mathbf{267} \\ \mathbf{345} & 346 & 347 & 356 & 357 & \mathbf{367} \end{array},$$

образуют  $L(1, 2, 3)$ . Пусть  $B_2 = (1, 4, 5)$ , тогда  $B_3 = (1, 6, 7)$ . При выборе  $B_4$  блоки, содержащие (45) и (67), то есть (245), (345), (267), (367) из  $L(1, 2, 3)$  исключаются. Блоки  $B_5, B_6, B_7$  выражаются как и  $B_3$  из  $B_1, B_2, B_4$  согласно (1.38).

Полученное число подстановок  $5040 = 35 \cdot 18 \cdot 8$  нужно поделить на 168, равное числу перестановок блоков внутри  $D(7, 3, 1)$ . Поскольку в результате одной транспозиции 15 нечетных дизайнов отображаются в 15 четных, доказано следующее утверждение.

**Теорема 10.** *1. Имеют место 30 подстановок  $D(7, 3, 1|F_8 \setminus \{0\})$ . Из них половина из них четные, половина нечетные.*



### 1.6.5. Комбинаторная структура нечетных дизайнов $D^1(7, 3, 1|F_8 \setminus \{0\})$

В (1.37) для двух подстановок  $D_i$  и  $D_j$  дизайна  $D(7, 3, 1|F_8 \setminus \{0\})$ , для которых пересечение  $D_i \cap D_j$  состоит из одного блока, определяется сумма в виде дизайна

$$D_{i+j} = D_i \oplus D_j = (D_i \cap D_j) \cup L(D_i \cap D_j) \setminus (D_i \cup D_j),$$

где через  $L(D_i \cap D_j)$  обозначена совокупность блоков, имеющая с блоком  $D_i \cap D_j$  по одному общему элементу. Полный набор сочетаний из трех элементов в виде дизайна  $D(7, 35, 15, 3, 1|F_8 \setminus \{0\})$  будем использовать в качестве вырожденного дизайна  $D_\emptyset$  и будем считать  $L(D_i) = D_\emptyset$ .

**Лемма 3.** 1. Суммой двух субпорядковых дизайнов является субпорядковый дизайн, то есть,  $D_i^1 \oplus D_j^1 = D_{i+j}^1$ . 2. Для нечетных (четных) дизайнов  $D_i$  справедливо  $D_i \oplus D_\emptyset = D_i$  и  $D_i + D_i = D_\emptyset$ .

**Доказательство.** 1. Субпорядковые дизайны  $D_i^1$  и  $D_j^1$  содержат по четыре  $i$ - и  $j$ -блока соответственно, следовательно, по этим блокам пересечений нет. В  $D_i^1$  три канонических блока содержат  $i$ , в  $D_j^1$  три канонических блока содержат  $j$ . Общий блок должен содержать  $i$  и  $j$ , то есть имеет вид  $(i, i, i + j)$  над полем  $F_8$ . Но этот блок содержится в субпорядковом дизайне  $D_{i+j}^1$ , откуда  $D_i^1 \oplus D_j^1 = D_{i+j}^1$ .

2. В первом случае  $D_i \oplus D_\emptyset = (D_i \cap D_\emptyset) \cup L(D_i \cap D_\emptyset) \setminus (D_i \cup D_\emptyset) = D_i \cup L(D_i) \setminus D_\emptyset = D_i \cup D_\emptyset \setminus D_\emptyset = D_i \cup \emptyset = D_i$ , а во втором  $D_i + D_i = (D_i \cap D_i) \cup L(D_i \cap D_i) \setminus (D_i \cup D_i) = D_i \cup L(D_i) \setminus D_i = D_i \cup D_\emptyset \setminus D_i = D_\emptyset$ .  $\square$

**Теорема 11.** Нечетные дизайны образуют дизайн  $D(15, 7, 3)$ .

**Доказательство.** На основании леммы 3 и нумерации порядковых дизайнов  $D_{0'}^1 + D_{i'}^1 = D_i^1$ , получаем  $D_{j'}^1 + D_{i'}^1 = D_{0'}^1 + D_j^1 + D_{0'}^1 + D_i^1 = D_j^1 + D_i^1 = D_{i+j}^1$ , то есть сложение дизайнов осуществляется в соответствии со сложением в поле  $F_{16}$ , ненулевые элементы которого образуют дизайн  $D(15, 7, 3|F_{16} \setminus \{0\})$ .  $\square$

Одной транспозиции достаточно для того, чтобы преобразовать дизайн  $D(15, 7, 3|\{D_i^1\})$ , состоящий из нечетных дизайнов, в блок-схему  $D(15, 7, 3|D_i^0)$  с четными дизайнами в качестве элементов, например, используется жестко-каноническая подстановка (1.14).

**Замечание.** Тот факт, что нечетные (четные) дизайны  $D(7, 3, 1)$ , а следовательно, и сцепленные с ними  $D(8, 14, 7, 4, 3)$  (раздел 1.6.1),

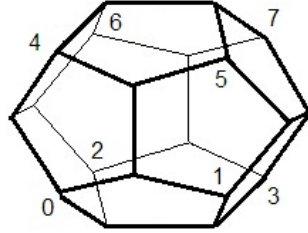


Рис. 1.9. Расположение куба в додекаэдре.

взяты в качестве элементов, образуют дизайн  $D(15, 7, 3)$ , лежит в основе четвертого исключительного изоморфизма простых классических конечных групп порядка  $20160 = \frac{8!}{2}$  между  $SL_4^{F_2}$  и группой четных подстановок  $A_8$  на восьми элементах. Элементом группы  $A_8$  задается четная подстановка на элементах  $F_8$ , приводящая к подстановкам четных (или нечетных) дизайнов  $D(8, 14, 7, 4, 3)$ . Дизайны  $D(8, 14, 7, 4, 3)$  образуют  $D(15, 7, 3|\{D_i^0\})$ , подстановки которого и определяются элементами группы  $SL_4^{F_2}$ .

### 1.6.6. Интерпретация дизайна $D(15, 7, 3)$ на додекаэдре

На рис. 1.9 изображен додекаэдр с длиной стороны, равной числу Фидия  $(\sqrt{5} - 1)/2 = 0.618\dots$ , внутри которого располагается единичный куб. Стороны этого куба являются диагоналями пятиугольных граней додекаэдра. Наличие пяти диагоналей в пятиугольнике указывает на то, что таких кубов внутри додекаэдра ровно пять  $c_1, c_2, \dots, c_5$ , и каждый куб в додекаэдре определяется четверкой пространственных диагоналей, то есть четырьмя из десяти отрезков  $d_i$ ,  $i = 0, 1, \dots, 9$ , соединяющих две противоположные вершины. Вершин в додекаэдре 20, поэтому диагоналей 10. Каждая пара внутренних кубов в додекаэдре имеет по одной общей диагонали, и наоборот, каждая диагональ является пересечением пары кубов. Построим соответствие между ними, используя лексико-графическое упорядочивание.

Таблица 1.22.

	куб	общие диагонали					
$F_{16}$	$(i_1 i_2 i_3 i_4)$	$i_1 i_2$	$i_1 i_3$	$i_1 i_4$	$i_2 i_3$	$i_2 i_4$	$i_3 i_4$
1	0123	0149	0258	0367	1267	1358	2349
2	0456	0149	0258	0367	3457	2468	1569
3	0789	0367	0258	0149	1478	2579	3689
4	1478	0149	1267	1358	3457	2468	0789
5	1569	1358	1267	0149	0456	2579	3689
6	2349	0123	2468	2579	3457	3689	0149
7	0149	0123	0456	0789	1478	1569	2349
0'	2579	0258	1267	2349	3457	1569	0789
1'	2468	2349	1267	0258	0456	1478	3689
2'	1358	0123	1569	1478	3457	3689	0258
3'	0258	0123	0456	0789	2579	2468	1358
4'	0367	0123	0456	0789	3689	3457	1267
5'	1267	0123	1569	1478	2468	2579	0367
6'	3457	2349	1358	0367	0456	1478	2579
7'	3689	0367	1358	2349	2468	1569	0789

$$\begin{pmatrix} d_0 & d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 & d_9 \\ c_1 c_2 & c_1 c_3 & c_1 c_4 & c_1 c_5 & c_2 c_3 & c_2 c_4 & c_2 c_5 & c_3 c_4 & c_3 c_5 & c_4 c_5 \end{pmatrix}.$$

Кубы  $c_1, \dots, c_5$  можно соответственно выразить через совокупности из четырех пространственных диагоналей  $d_i$ .

$$\begin{pmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ d_0 d_1 d_2 d_3 & d_0 d_4 d_5 d_6 & d_1 d_4 d_7 d_8 & d_2 d_5 d_7 d_9 & d_3 d_6 d_8 d_9 \end{pmatrix}.$$

Введем в рассмотрение  $10 = C_5^2$  так называемых диагональных кубов, а именно, для каждой пары из пяти кубов  $c_i$  и  $c_j$  как совокупностей из четырех диагоналей  $d_i$  определим *диагональный куб* как совокупность также из четырех диагоналей, состоящую из одной общей диагонали, равной  $c_i \cap c_j$ , и трех дополнительных диагоналей, представляющих собой  $c_i \cup c_j$ , т.е.

$$c_k = c_i \oplus c_j = (c_i \cap c_j) \cup \overline{c_i \cup c_j}.$$

Для 15 кубов (5 основных и 10 диагональных) введем операцию

„сложения“, которая зависит от мощности их пересечения и выражается через симметрическую разность и ее дополнение в виде

$$c_i \oplus c_j = \begin{cases} \overline{c_i \Delta c_j}, & \text{card}(c_i \cap c_j) = 1, \\ c_i \Delta c_j, & \text{card}(c_i \cap c_j) = 2, \end{cases} \quad (1.39)$$

где  $c_i \Delta c_j = c_i \setminus c_j \cup c_j \setminus c_i$  и  $\overline{c_i \Delta c_j} = (c_i \cap c_j) \cup \overline{c_i \cup c_j}$ . Очевидно  $c_k \oplus c_i = c_j$  и  $c_k \oplus c_j = c_i$ . Таким образом, этим комбинаторным кубам (табл.1.22) можно сопоставить ненулевые элементы поля  $F_{16}$  и соответственно элементы дизайна  $D(15, 7, 3)$ . Поскольку автоморфизмы  $D(15, 7, 3)$  задаются четырьмя элементами, покажем, что пятый куб можно получить из остальных четырех. Действительно,

$$\begin{aligned} c_1 &= \{\mathbf{c}_1 \mathbf{c}_2, c_1 c_3, c_1 c_4, c_1 c_5\}, \\ c_2 &= \{\mathbf{c}_1 \mathbf{c}_2, c_2 c_3, c_2 c_4, c_2 c_5\}, \\ c_1 \oplus c_2 &= \{\mathbf{c}_1 \mathbf{c}_2, c_3 c_4, c_3 c_5, c_4 c_5\}, \\ c_3 &= \{c_3 c_1, c_3 c_2, \mathbf{c}_3 \mathbf{c}_4, c_3 c_5\}, \\ c_4 &= \{c_4 c_1, c_4 c_2, \mathbf{c}_4 \mathbf{c}_3, c_4 c_5\}, \\ c_3 \oplus c_4 &= \{\mathbf{c}_3 \mathbf{c}_4, c_1 c_2, \underline{c_1 c_5}, \underline{c_2 c_5}\}, \\ (c_1 \oplus c_2) \oplus (c_3 \oplus c_4) &= \{c_1 c_5, c_2 c_5, c_3 c_5, c_4 c_5\} = c_5. \end{aligned}$$

### 1.6.7. Симметрии в стоматологических данных

В методе главных компонент [29] матрица факторных нагрузок состоит из коэффициентов корреляции между  $p$  признаками и  $p$  главными компонентами. В [9] приведены результаты анализа знаковой структуры факторных весов стоматологических данных на основе различных подстановок дизайнов. В качестве признаков рассматривались характеристики гигиены полости рта, а в качестве факторов: внешняя и лингвальная поверхности ( $t_2 = 0, 1$ ), верхняя и нижняя челюсти ( $t_1 = 0, 1$ ), правая и левая стороны ( $t_0 = 0, 1$ ). В обозначении признаков использовались выражения вида  $MGI_i$ , где  $i = t_0 + 2t_1 + 4t_2$ , например, при  $i = 3$  гингивитный индекс лингвальной поверхности верхней челюсти справа.

В табл. 1.23 представлен один из вариантов знаковой структуры факторных нагрузок, полученной по признакам  $MGI_0, \dots, MGI_7$  на молярах <sup>15</sup>, из которой видно, что номера признаков, собранные

<sup>15</sup>Молярами называются коренные зубы.

отдельно по положительным и отрицательным весам, образуют канонический дизайн  $D(8, 14, 7, 4, 3)$ . Это означает, что наибольшим образом индивиды отличаются по тому, какая сторона десны у моляров повреждена больше: щечная или лингвальная. Следующий по значимости фактор — это верхняя и нижняя челюсти: у одних больше подвержена патологии верхняя, а у других нижняя. Третий фактор отвечает за эффект их взаимодействия. Четвертый фактор соответствует различию наблюдений по правой и левой стороне.

Подобная симметричная структура матрицы факторных нагрузок встречалась при разном варьировании данными — менялись характеристики, близость к молярам или резцам, до и после обучения правильной чистке зубов, расслаивалась выборка на группы с разной стадией заболевания или по другим качественным признакам, рассматривались материалы другого подобного эксперимента.

Таблица 1.23.

$MGI_i$	Главная компонента							
$i$	1	2	3	4	5	6	7	8
0	+	+	+	+	+	+	+	+
1	+	+	+	+	-	-	-	-
2	+	+	-	-	+	+	-	-
3	+	+	-	-	-	-	+	+
4	+	-	+	-	+	-	+	-
5	+	-	+	-	-	+	-	+
6	+	-	-	+	+	-	-	+
7	+	-	-	+	-	+	+	-
$B_+$	$\Omega_3$	0123	0145	0167	0246	0257	0347	0356
$B_-$	$\emptyset$	4567	2367	2345	1357	1346	1256	1247

Были выявлены некоторые устойчивые явления. Во-первых, после „лечения“, которое заключалось в том, что индивидов обучали правильной чистке зубов, индексы гигиены приобретали более выраженный симметричный характер. В индексах воспаления более симметричная структура была характерна для больных с нормальными иммунными реакциями, в то время как у больных с разными особенностями иммунных реакций знаковая структура факторных весов приобретала вид неканонического дизайна.

## 1.7. Интегрирование дизайнов

Интегрированием дизайнов называется процедура, противоположная разделению блок-схемы на производную и остаточную [15]. Построение дизайна  $D(15, 7, 3)$  из его производного  $D(7, 3, 1)$ , т. е. проективной геометрии  $P_3^2$  из  $P_2^2$  было рассмотрено в разд. 1.2.4. В случае произвольной характеристики интегрирование дизайнов типа  $P_n^q$  осуществляется на основе рекуррентных соотношений типа Фибоначчи, которые используются в алгебре в определении импульсных последовательностей [19].

### 1.7.1. Импульсные последовательности

**Определение 9.** Последовательность  $\{\phi_j\}_{j=0}^{\infty}$  элементов конечного поля  $F_q$ , где  $q = p^r$ , называется **импульсной**, если она получена в результате рекуррентного соотношения типа Фибоначчи

$$\phi_{j+1} = \phi_j + \alpha\phi_{j-1}, \quad \alpha \in F_q. \quad (1.40)$$

Пусть  $\phi_0 = 0$ ,  $\phi_1 = 1$ . Из [19] известно, что существует параметр рекуррентности  $\alpha$ , при котором достигается максимальный период импульсной последовательности, равный  $q^2 - 1 = (q - 1)(q + 1)$ . Например, при  $q = 5$  при  $\alpha = 1, 2, 3$  соответственно с периодами 20, 4 и 24 имеем последовательности:

$$\begin{aligned} &(01123) (03314) (04432) (02241) (01123) \dots, \\ &(0113)(0113) \dots, \\ &(011424)(022343)(044131)(033212)(011424) \dots \end{aligned}$$

### 1.7.2. Параметр рекуррентности при $p \neq 2$

**Предложение 7.** При  $p \neq 2$  элемент импульсной последовательности (1.40) может быть выражен в явном виде

$$\phi_k = \frac{(1 + \sqrt{x})^k - (1 - \sqrt{x})^k}{2^k \sqrt{x}}, \quad \text{где } x = 1 + 4\alpha. \quad (1.41)$$

**Доказательство** (по индукции). Очевидно, что  $\phi_0 = 0, \phi_1 = 1$ .

$$\begin{aligned} \phi_{k+1} &= \phi_k + \alpha\phi_{k-1} = \\ &= \frac{(1 + \sqrt{x})^k - (1 - \sqrt{x})^k}{2^k \sqrt{x}} + \frac{\alpha((1 + \sqrt{x})^{k-1} - (1 - \sqrt{x})^{k-1})}{2^{k-1} \sqrt{x}} = \\ &= \frac{(1 + \sqrt{x})^{k-1}(1 + \sqrt{x} + 2\alpha) - (1 - \sqrt{x})^{k-1}(1 - \sqrt{x} + 2\alpha)}{2^k \sqrt{x}} = \\ &= \frac{(1 + \sqrt{x})^{k+1} - (1 - \sqrt{x})^{k+1}}{2^{k+1} \sqrt{x}}, \end{aligned}$$

так как  $4\alpha = x - 1$ , следовательно,

$$\begin{aligned} 1 + \sqrt{x} + 2\alpha &= 1 + \sqrt{x} + \frac{x-1}{2} = \frac{(1 + \sqrt{x})^2}{2}, \\ 1 - \sqrt{x} + 2\alpha &= 1 - \sqrt{x} + \frac{x-1}{2} = \frac{(1 - \sqrt{x})^2}{2}. \quad \square \end{aligned}$$

Нетрудно доказывается следующее утверждение.

**Лемма 4.** *Над полем  $F_q$  биномиальные коэффициенты обладают свойствами:*

- 1)  $C_q^0 = C_q^q = 1, C_q^k = 0 \quad \forall 0 < k < q$ ;
- 2)  $C_{q+1}^1 = C_{q+1}^q = 1, C_{q+1}^k = 0 \quad \forall 1 < k < q$ ;
- 3)  $(z+1)^q = z^q + 1$ .

**Теорема 12.** 1) *Если  $1 + 4\alpha$  является квадратом в поле  $F_q$ , тогда  $\phi_{q-1} = 0$ .*

2) *Если  $1 + 4\alpha$  не является квадратом в поле  $F_q$ , то  $\phi_{q+1} = 0$ , а мнимая единица  $i = \sqrt{1 + 4\alpha}$  обладает свойствами  $i^{q-1} + 1 = 0$  и  $(i+1)^q = 1 - i$ .*

3) *Если  $-4\alpha$  является квадратом в поле  $F_q$ , а число  $1 + 4\alpha$  нет, тогда  $\phi_{\frac{q+1}{2}} = 0$ .*

**Доказательство.**

1) В случае, когда  $1 + 4\alpha$  является квадратом в поле  $F_q$ , выражение  $\sqrt{1 + 4\alpha} \in F_q$  равносильно тому, что

$$(1 + \sqrt{1 + 4\alpha})^{q-1} = 1, \quad (1 - \sqrt{1 + 4\alpha})^{q-1} = 1.$$

Тогда в (1.41) получаем равенство нулю при  $k = q - 1$ .

$$\begin{aligned}
2) \quad \phi_{q+1} = 0 &\iff \frac{(1+i)^{q+1} - (1-i)^{q+1}}{2^{q+1}i} = 0, \iff \\
&\frac{1}{2^{q+1}i} \left( \sum_{k=0}^{q+1} C_{q+1}^k i^k - \sum_{k=0}^{q+1} C_{q+1}^k (-i)^k \right) = 0, \iff \\
\frac{1}{2^q i} \left( \sum_{k=1,3,5,\dots,q+1} C_{q+1}^k i^k \right) &\stackrel{\text{(по лемме 4)}}{=} \frac{1}{2^q i} (i + i^q) = 0.
\end{aligned}$$

Отсюда  $i^{q-1} = -1$  и  $i^q = -i$ . Также из леммы 4 получаем  $(1+i)^q = 1 + i^q = 1 - i$ .

3) То, что выражение  $-4\alpha = 1 - i^2$  является квадратом в поле  $F_q$  означает, что  $(\sqrt{-4\alpha})^{q-1} = (\sqrt{1-i^2})^{q-1} = 1$ . Отсюда получаем

$$\phi_{\frac{q+1}{2}} = 0 \iff (\sqrt{1+i})^{q+1} = (\sqrt{1-i})^{q+1}.$$

Умножаем обе части на  $(\sqrt{1+i})^{q+1}$ .

$$(1+i)^{q+1} = (\sqrt{1-i^2})^{q+1}.$$

Левая часть, согласно второму утверждению теоремы, равна

$$(1+i)^{q+1} = (1+i)^q(1+i) = (1-i)(1+i),$$

а правая по условию  $(\sqrt{1-i^2})^{q-1} = 1$  равна

$$(\sqrt{1-i^2})^{q+1} = (\sqrt{1-i^2})^{q-1}(1-i^2) = 1 - i^2. \quad \square$$

**Следствие.** Для того чтобы выполнялись  $\phi_{q+1} = 0$  и  $\phi_k \neq 0$  при  $k < q+1$ , необходима мнимость выражений  $\sqrt{1+4\alpha}$  и  $\sqrt{-4\alpha}$ .

### 1.7.3. Параметр рекуррентности при $p = 2$

**Предложение 8.** При  $p = 2$  элемент импульсной последовательности (1.40) может быть выражен в явном виде

$$\phi_k = (1+x)^k + x^k, \quad \text{где } x^2 + x + \alpha = 0. \quad (1.42)$$



**Доказательство.** В поле четной характеристики справедливо соотношение  $x + x = 0$ , поэтому очевидно, что  $\phi_0 = 0, \phi_1 = 1$ . Далее по индукции

$$\begin{aligned}\phi_{k+1} &= \phi_k + \alpha\phi_{k-1} = (1+x)^k + x^k + \alpha((1+x)^{k-1} + x^{k-1}) = \\ &= (1+x)^{k-1}(1+x+\alpha) + x^{k-1}(x+\alpha) = \\ &= (1+x)^{k-1}(1+x+x^2+x) + x^{k-1}(x+x^2+x) = \\ &= (1+x)^{k-1}(1+x)^2 + x^{k-1}x^2. \quad \square\end{aligned}$$

**Теорема 13.** Если для  $x$  из (2.13) справедливо  $x \in F_q$ , то  $\phi_{q-1} = 0$ , а если  $x \notin F_q$ , то мнимая единица  $x = i$  со свойством  $i^q = i + 1$  приводит к  $\phi_{q+1} = 0$ .

**Доказательство.** Если  $x \in F_q$ , тогда  $\phi_{q-1} = (1+x)^{q-1} + x^{q-1} = 1 + 1 = 0$ . Далее, выражение  $1+i = i^q$  равносильно тому, что  $(1+i)^2 = i^{2q}$ . Умножим левую часть на единицу, равную  $1 = i^{q-1}(i+1)^{q-1}$  из условия  $\alpha = i^2 + i \in F_q$ .

$$\begin{aligned}i^{q-1}(i+1)^{q-1}(1+i)^2 &= i^{2q} \iff i^{q-1}(i+1)^{q+1} = i^{2q} \iff \\ &\iff (i+1)^{q+1} = i^{q+1} \iff \phi_{q+1} = 0. \quad \square\end{aligned}$$

В поле натуральных чисел при  $\alpha = 1$  импульсные последовательности с начальными условиями  $\phi_0 = 0, \phi_1 = 1$  и  $\phi_0 = 2, \phi_1 = 1$  носят названия рядов Фибоначчи  $(0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots)$  и Люка  $(2, 1, 3, 4, 7, 11, 18, 29, 47, \dots)$ . Существует много научных и околонуточных трудов о связи пропорций, встречающихся в природе, с этими двумя рядами. Предел отношения двух последовательных элементов этих рядов равен числу Фидия  $\Phi = 1.618\dots = \frac{1+\sqrt{5}}{2}$ , лежащего в основе золотого сечения. Второе число Фидия имеет вид  $\Phi_- = 1 - \Phi = -\frac{1}{\Phi} = \frac{1-\sqrt{5}}{2} = -0.618\dots$ . Формула вычисления явного вида ряда Фибоначчи носит название формулы Бине.

$$\phi_k = \frac{\Phi^k - \Phi_-^k}{\sqrt{5}}.$$

Ряд Люка также выражается через числа Фидия.

$$\psi_k = \Phi^k + \Phi_-^k.$$

В конечном поле  $F_3$ , в котором для достижения максимального периода импульсной последовательности нужно взять параметр рекуррентности  $\alpha = 1$ , а мнимая единица удовлетворяет условию

$i^2 = -1$ , выражения для рядов Фибоначчи и Люка имеют вид

$$\phi_k = \frac{z^k - \bar{z}^k}{i}, \quad \psi_k = z^k + \bar{z}^k \quad \text{где} \quad z = \frac{1+i}{2}.$$

При вычислении суммы квадратов

$$\begin{aligned} \phi_k^2 + \psi_k^2 &= \frac{(z^k - \bar{z}^k)^2}{i^2} + (z^k + \bar{z}^k)^2 = \\ &= -(z^{2k} - 2z^k \bar{z}^k + \bar{z}^{2k}) + (z^{2k} + 2z^k \bar{z}^k + \bar{z}^{2k}) = \\ &= 4z^k \bar{z}^k = (1 - i^2)^k = i^{2k} \end{aligned}$$

обнаруживаем, что в  $F_3$  нормированные на  $i^k$  элементы рядов Фибоначчи и Люка обладают свойствами синуса и косинуса.

$$\left(\frac{\phi_k}{i^k}\right)^2 + \left(\frac{\psi_k}{i^k}\right)^2 = 1.$$

#### 1.7.4. Пример интегрирования дизайнов характеристики три

Покажем на примере поля  $F_3$  интегрирование  $D(13, 4, 1)$  из вырожденного дизайна  $D(4, 1, 0)$ , в качестве элементов которого рассматриваются четыре проективных вектора  $(x_0, x_1)$  над  $F_3$ , они же блоки. Итак, имеем проективную прямую  $P_1^3$  с точками:

$$\left(\begin{array}{c} 0 \\ 1 \end{array}\right), \left(\begin{array}{c} 1 \\ 0 \end{array}\right), \left(\begin{array}{c} 1 \\ 1 \end{array}\right), \left(\begin{array}{c} 1 \\ 2 \end{array}\right).$$

Блок  $B_1$  дизайна  $D(13, 4, 1)$  состоит элементов дизайна  $D(4, 1, 0)$ , вложенных в трехмерные вектора  $(x_0, x_1, 0)$  с дополнительной нулевой координатой.

$$B_1 = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \end{array}\right).$$

Остальные блоки вычисляются через импульсные последовательности, причем в качестве  $\phi_0$  рассматриваются четыре элемента дизайна  $D(4, 1, 0)$ , а в качестве  $\phi_1$  берутся вектора  $(x_0, x_1)$  над  $F_3$ , вложенные в трехмерный вектор  $(x_0, x_1, 1)$  с дополнительной единичной координатой. Например, по начальным условиям  $\phi_0 = (1, 0, 0)$

и  $\phi_1 = (0, 0, 1)$  строится проективный период импульсной последовательности с параметром рекуррентности  $\alpha = 1$ .

$$B_2 = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}.$$

Легко убедиться в том, что если  $\phi_1 \in B_2$ ,  $\phi_1 \neq \phi_0$ , т.е.  $\phi_1 = (1, 0, 1)$  или  $\phi_1 = (2, 0, 1)$ , то блок  $B_2$  останется тем же. Для построения блока  $B_3$  в качестве  $\phi_1$  выбирается  $(x_0, x_1, 1) \notin B_2$ , например, вектор  $(0, 1, 1)$ . По этим начальным условиям строится второй проективный период импульсной последовательности.

$$B_3 = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix}.$$

Наконец, в качестве  $\phi_1$  выбирается вектор  $(x_0, x_1, 1) \notin \{B_2, B_3\}$ . Это могут быть  $(0, 2, 1)$ ,  $(1, 2, 1)$  или  $(2, 2, 1)$ . В частности,

$$B_4 = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 0 & 2 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 \end{pmatrix}.$$

Таким образом, по одному блоку дизайна  $D(4, 1, 0)$  получены три новых блока дизайна  $D(13, 4, 1)$ . Аналогично по остальным трем блокам дизайна  $D(4, 1, 0)$  строятся оставшиеся 9 блоков дизайна  $D(13, 4, 1)$ .

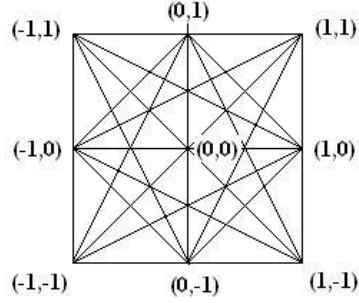


Рис. 1.10. Интерпретация дизайна  $D(9, 12, 4, 3, 1)$ .

$$\begin{aligned}
 B_5 &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & B_6 &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \\
 B_7 &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 2 & 2 & 2 \end{pmatrix}, & B_8 &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 2 \end{pmatrix}, \\
 B_9 &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}, & B_{10} &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix}, \\
 B_{11} &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 0 & 2 & 1 \end{pmatrix}, & B_{12} &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 2 & 0 & 2 & 1 \end{pmatrix}, \\
 & & B_{13} &= \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \end{pmatrix}.
 \end{aligned}$$

На рисунке 1.10 изображен дизайн  $D(9, 12, 4, 3, 1)$ , соответствующий аффинной геометрии  $E_3^2$ . Его элементами являются 9 элементов из  $D(13, 4, 1)$  вида  $(x_0, x_1, 1)$ . Блоки представлены 12 прямыми вида  $x = c$ ,  $y = c$ ,  $x + y = c$ ,  $x - y = c$ , где  $c = 0, 1, -1$  над  $F_3$ .

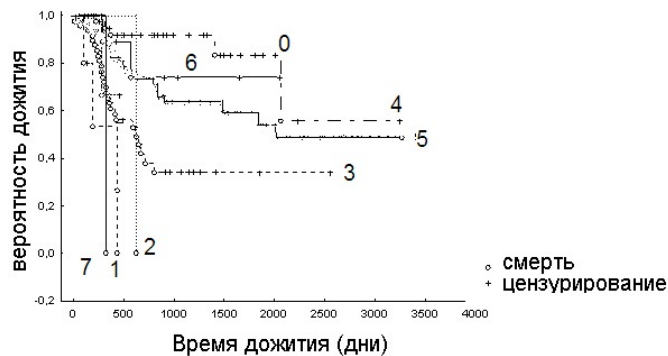


Рис. 1.11. Дожитие по элементам синдрома  $S_2(X_2, X_5, X_{16})$ .

## 1.8. Симптомно-синдромальный анализ в кривых дожития

При рассмотрении данных о длительности послеоперационного периода больных глиомой по критерию Гехана-Вилкоксона были выявлены наиболее значимые факторы, влияющих на дожитие: головная боль  $X_2$ , речевые нарушения  $X_5$ , а также взаимодействие факторов оглушения сознания  $X_1$  и судорог  $X_6$  через симптом  $X_{16}$ . Распределение больных по синдрому  $S_2(X_2, X_5, X_{16})$  представлено в табл. 1.24, где  $n_m$  — количество больных в группах  $m = X_2 + 2X_5 + 4X_{16}$ . Кривые дожития изображены на рис. 1.11.

Таблица 1.24.

$m$	0	1	2	3	4	5	6	7
$X_{16}$	0	0	0	0	1	1	1	1
$X_5$	0	0	1	1	0	0	1	1
$X_2$	0	1	0	1	0	1	0	1
$n_m$	17	82	6	8	53	96	2	8

Симптом  $X_{16} = 1$  в случае, когда или оглушение сознания  $X_1$ , или судороги  $X_6$ . Если нет ни того, ни другого, то  $X_{16} = 0$ . Согласно табл. 1.25, исключение составляют 6 человек из 273, которые имеют сочетание этих признаков.

Таблица 1.25.

	$X_1 = 0$	$X_1 = 1$	сумма
$X_6 = 0$	107	18	125
$X_6 = 1$	142	6	148
сумма	249	24	273

По кривым дожития выделяются разные с прогностической точки зрения группы:

- 1 — больные с головной болью, без речевых нарушений. Из них у 77 человек наблюдаются ясное сознание и нет судорог; только у 5 человек присутствуют и судороги, и оглушение сознания;
- 2 — шесть человек только с речевыми нарушениями;
- 3 — восемь с головной болью и речевыми нарушениями;
- 7 — восемь больных, у которых представлены все факторы риска.

К более благополучным относятся группы:

- 0 — без указанных факторов риска (17 человек);
- 4 — ясное сознание и наличие судорог у 53 больных;
- 5 — в этой группе из 96 больных с ясным сознанием и головной болью, в большинстве случаев с судорогами (87 и 9);
- 6 — двое больных без головной боли с речевыми нарушениями: у одного оглушение сознания, у другого судороги.

Выделяя наиболее отличающиеся по дожитию пары, получаем разделение больных на группы: (0,4) и (5,6), (2,3) и (1,7) при помощи

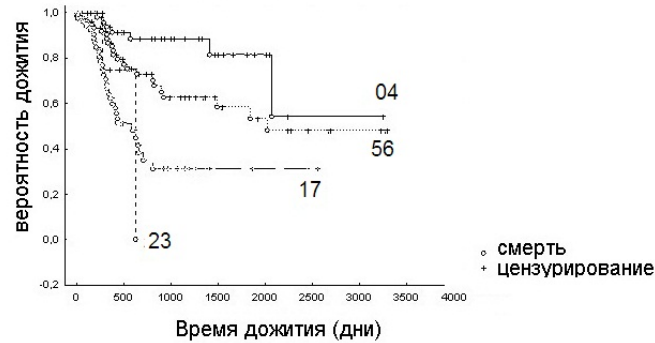


Рис. 1.12. Дожитие по элементам синдрома  $S_1$ .

обобщенного синдрома первого порядка  $S_1$ . Соответствующие этим группам кривые дожития представлены рис. 1.12.

Группа (2,3), состоящая из 14 человек, отличается прежде всего большим возрастом ( $62.79 \pm 8.95$  по сравнению с  $43.38 \pm 14.84$  у остальных, доверительный уровень вероятности соответствующего критерия однородности  $p < 0.0001$ ). У этих больных имеют место нарушения речи при ясном сознании и отсутствии судорог, чаще встречается левая сторона поражения (13 к 1, когда поражены обе стороны), 50% имеют поражения в левой теменной доле по сравнению с 17% в остальных группах. По индексу Карновского больные из группы (2,3) значительно тяжелее ( $57.14 \pm 14.37$  по сравнению с  $69.76 \pm 13.12$  у остальных больных).

Наихудший прогноз среди более молодых больных наблюдается в группе (1,7). Больные отличаются большим смещением срединных структур ( $9.00 \pm 5.06$  по сравнению с  $4.32 \pm 4.57$ ,  $p < 0.00001$ ), большим объемом новообразования ( $64880 \pm 46404$  по сравнению с  $47612 \pm 44844$ ,  $p = 0.0233$ ), соответственно большим объемом операции.

Более благополучные больные в группе (0,4) отличаются от остальных меньшим возрастом ( $40.89 \pm 14.87$  по сравнению с  $45.59 \pm 15.17$ ,  $p = 0.025$ ), высокими индексами Карновского ( $76.71 \pm 11.51$  по сравнению с  $66.40 \pm 13.08$ ,  $p < 0.000001$ ), меньшими размерами опухоли ( $412793 \pm 41206$  по сравнению с  $58595 \pm 47034$ ,  $p = 0.032$ ).

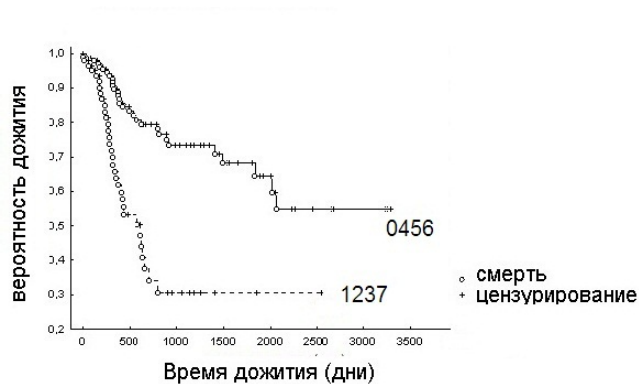


Рис. 1.13. Дожитие по элементам синдрома  $S_0$ .

На рис. 1.13 представлены кривые дожития элементов наиболее информативного синдрома нулевого порядка (нелинейного симптома), при помощи которого задается группировка на блоки (0,4,5,6) и (1,2,3,7). Больные из группы (1,2,3,7) отличается более тяжелым диагнозом, у них чаще поражена правая затылочная доля, для большего количества больных применялась химиотерапия.

В качестве второй наиболее информативной компоненты синдрома 1-го порядка (0,4), (5,6), (2,3) и (1,7) выбирается обобщенный симптом, задающий группировку вида (0,2,3,4) и (1,5,6,7). У больных из (1,5,6,7) тяжесть больше, чаще поражена теменная сторона, большее смещение срединных структур и больше размеры опухоли.

Таким образом, первой главной компонентой, определяющей по категориальным признакам время дожития, оказывается тяжесть диагноза, а второй — размеры новообразования (рис. 1.13).

## 1.9. Анализ категориальных рядов

Для иллюстрации симптомного метода выявления пенетрантной формы периодичности рассмотрим текст из шести слов с окончанием „ение“ (*пение чтение прощение видение бдение приведение*). При помощи двоичной кодировки, в которой одной букве соответ-



ствуется кодовое слово из четырех знаков (нулей и единиц), получаем дихотомический ряд  $(f_1, f_2, \dots, f_N)$  длиной в  $N = 232$  знака с учетом двух пробелов между словами и шести пробелов по краям, кодируемых четырьмя нулями. Окончанию „*енне*“ соответствует набор знаков (0101 1001 0111 0101).

По аналогии с методом выделения скрытой периодичности для метрических рядов „ГУСЕНИЦА“ - *SSA (Singular Spectrum Analysis)* [12], исследуемый категориальный ряд  $(f_1, f_2, \dots, f_N)$  преобразуется в траекторную матрицу  $X = [X_1, \dots, X_K]$ , где вектора вложения  $X_i$  при длине  $L$  „окна“ имеют вид:

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iL})^T = (f_i, \dots, f_{i+L-1})^T, \quad i = 1, \dots, K.$$

По векторам вложения для всех  $\tau = (t_1, \dots, t_m) \subseteq \{1, 2, \dots, K\}$  построим симптомы

$$X_\tau = a_1 X_1 + \dots + a_K X_K \pmod{2},$$

где  $a_j = 1$  при  $j \in \tau$  и  $a_j = 0$  при  $j \notin \tau$ , и упорядочим их согласно (1.5). Порядковый номер симптома  $X_\tau$  с набором индексов  $\tau = (t_1, \dots, t_m)$  в импульсно упорядоченном синдроме будем называть *импульсным рангом*  $R(\tau)$ . Он может быть вычислен по формуле:

$$R(\tau) = \sum_{j=1}^m 2^{t_j-1}. \quad (1.43)$$

Самым информативным с точки зрения энтропии является симптом с одинаковым числом нулей и единиц в векторе реализаций. Поэтому если в векторах вложения нулей много больше, чем единиц, то симптом  $X_\tau$ , в котором доля единиц ближе всего к одной второй, оказывается наиболее информативным. Количество единиц вычисляется при помощи выражения  $X_\tau^T \mathbf{e}_L$ , где через  $\mathbf{e}_L$  обозначен вектор из  $L$  единиц. Отметим, что требование преобладания нулей в дихотомическом ряде является ограничением метода, а задача определения допустимой пропорции между нулями и единицами требует отдельного рассмотрения. В нашем примере для большей контрастности оказалось достаточно ввести два пробела между словами вместо одного.

Дополним правило для выявления наиболее информативного симптома принципом минимума действия, который реализуется в

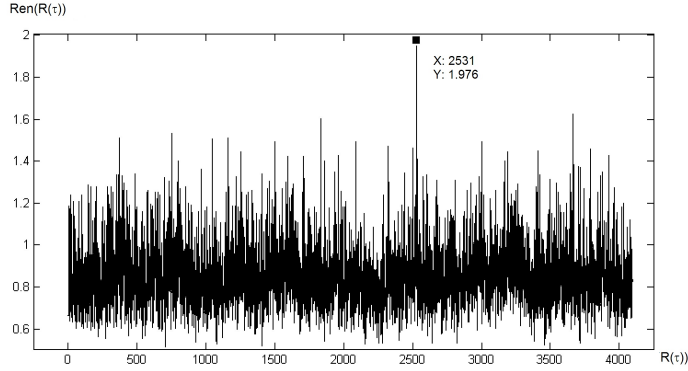


Рис. 1.14. Диаграмма рентабельности симптомов.

том, что более информативный симптом должен содержать в себе к тому же меньшее количество переключений от одного знака к другому. Оно равно числу единиц в векторе, полученном в результате произведения  $\Theta X_\tau$  над  $F_2$ , где для элементов матрицы  $\Theta$  размерности  $L - 1$  на  $L$  справедливы выражения:  $\Theta_{i,i} = 1$ ,  $\Theta_{i,i+1} = 1$  для  $i = 1, 2, \dots, L - 1$ , остальные нули. Объединим эти характеристики в *показателе рентабельности*

$$Ren(R(\tau)) = \frac{X_\tau^T \mathbf{e}_L}{(\Theta X_\tau)^T \mathbf{e}_{L-1}},$$

который зависит от ранга симптома  $R(\tau)$ . Экстремальное значение рентабельности достигается при ранге  $R(\tau) = 2531$  (рис. 1.14). Из (1.43) получаем набор  $\tau = (1, 2, 6, 7, 8, 9, 12)$  из ненулевых компонент линейной комбинации  $X_\tau$ .

Графически ряд  $(f_1, f_2, \dots, f_N)$  будем изображать в виде отрезков, соединяющих точки  $(i, f_i)$ ,  $i = 1, \dots, N$ . На рис. 1.15 одинаковые фрагменты дихотомического ряда, кодирующего текст, накрываются промежутками постоянства сдвинутого на константу  $c$  наиболее информативного симптома  $X_{1,2,6,7,8,9,12}$ .

Выделим из симптома  $X_\tau$  над  $F_2$  фрагмент  $\mathbf{e}_t = (e_1, \dots, e_t)^T$ , в частности, это могут быть компоненты, равные  $t$  единицам. Он

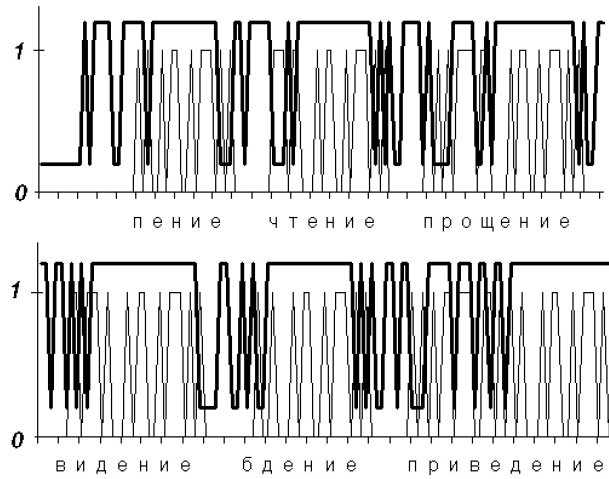


Рис. 1.15. Графики рядов  $X_1$  и  $X_{1,2,6,7,8,9,12} + c$ .

получен в результате линейного преобразования фрагмента последовательности  $F = (f_{n+1}, \dots, f_{n+k+t-1})^T$ , т. е.

$$e_j = \sum_{i=1}^k a_i f_{n+i+j-1}, \quad j = 1, \dots, t, \quad (1.44)$$

или вектор  $e_t$  выражается в матричном виде

$$\begin{bmatrix} \sum_{i=1}^k a_i f_{i+n} \\ \sum_{i=1}^k a_i f_{i+n+1} \\ \vdots \\ \sum_{i=1}^k a_i f_{i+n+t-1} \end{bmatrix} = \begin{bmatrix} f_{n+1} & f_{n+2} & \dots & f_{n+k} \\ f_{n+2} & f_{n+3} & \dots & f_{n+k+1} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n+t} & f_{n+t+1} & \dots & f_{n+t+k-1} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}.$$

При  $k = t$ , зная фрагмент  $F$ , из уравнений (1.44) можно получить коэффициенты  $(a_1, \dots, a_k)$  симптома  $X_\tau$ , который идентифицирует этот фрагмент при помощи вектора  $e_k$ . В частности, для фрагмента

$F = (0\ 0101\ 1001\ 0111\ 0101\ 00000)^T$  длины  $k = 12$ , кодирующего окончание „ение“, получаем соответствующий симптому с индексом  $\tau = (1, 2, 6, 7, 8, 9, 12)$  вектор  $(a_1, \dots, a_k) = (1100011110001)$ .

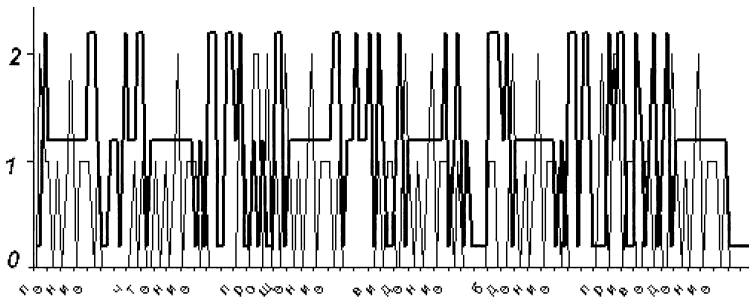


Рис. 1.16. График  $Y_1$  и  $Y_\tau$ , где  $\tau = (1, 2, \underline{4}, \underline{5}, 6, 7, \underline{8}, \underline{9}, 10, 11)$ .

Используя аппарат обобщенного обращения матриц [10], можно решить обратную задачу, связанную с восстановлением фрагмента последовательности  $F = (f_{n+1}, \dots, f_{n+2k-1})^T$  по коэффициентам  $(a_1, \dots, a_k)$  симптома  $X_\tau$ , так как вектор  $e_k$  можно иначе выразить через произведение матриц

$$\begin{bmatrix} a_1 & a_2 & \dots & a_k & 0 & 0 & \dots & 0 \\ 0 & a_1 & \dots & a_{k-1} & a_k & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & a_1 & a_2 & \dots & a_k \end{bmatrix} \text{ и } \begin{bmatrix} f_{n+1} \\ f_{n+2} \\ \vdots \\ f_{n+2k-1} \end{bmatrix}.$$

При наличии в категориальной последовательности трех градаций можно построить линейные комбинации векторов вложения над  $F_3$ . Для заданного фрагмента  $F$  вектор коэффициентов вычисляется по аналогии с (1.44). В частности (рис. 1.16), при помощи симптома  $Y_\tau = Y_1 + Y_2 - Y_4 - Y_5 + Y_6 + Y_7 - Y_8 - Y_9 + Y_{10} + Y_{11}$  в вектор из единиц преобразуется фрагмент  $(101\ 012\ 011\ 101)^T$ , которому над  $F_3$  соответствует окончание „ение“.

## 1.10. Связность категориальных рядов

### 1.10.1. О вероятности случайной классификации

Пусть имеются две популяции  $W$  и  $\bar{W}$ , идентифицируемые переменной  $Y$ . Если индивид из популяции  $W$ , то  $Y = 1$ , если индивид из популяции  $\bar{W}$ , то  $Y = 0$ . Количество всевозможных комбинаций компонент вектора  $Y$  равно  $2^n$ , где  $n$  — число индивидов. Популяции характеризуются  $m$  дихотомическими признаками, наблюдения за которыми

$$X_i = (x_{i1}, \dots, x_{in})^T, \quad i = 1, \dots, m,$$

собираются в матрицу  $X = [X_1 | \dots | X_m]$  размерности  $n \times m$  над полем  $F_2$  ранга  $r$ . Поиск наиболее информативного симптома сводится к минимизации выражения

$$\|b\| = \min \left\{ \frac{1}{n} \sum_{i=1}^n b_i, 1 - \frac{1}{n} \sum_{i=1}^n b_i \right\}, \quad (1.45)$$

где  $b = (b_1, \dots, b_n)^T = Xa - Y$ , через  $a$  обозначен вектор из  $m$  параметров над  $F_2$ . Введем в рассмотрение  $\mathcal{L}(X)$  множество всех конечно-линейных комбинаций

$$\mathcal{L}(X) = \left\{ \sum_{j=1}^m a_j X_j \pmod{2} \right\}.$$

Обозначим через  $\mathbf{0}$  и  $\mathbf{e}$  вектора из  $n$  нулей и  $n$  единиц соответственно. Поскольку рассматриваются две популяции, считаем, что  $Y \neq \mathbf{0}$  и  $Y \neq \mathbf{e}$ . Вероятность  $p_0(r, X)$  случайной классификации вектора  $Y$  по матрице  $X = X(n, m)$  ранга  $r$  может быть вычислена при помощи отношения

$$p_0(r, X) = \frac{|\mathcal{L}(X) \cup (\mathcal{L}(X) + \mathbf{e})|}{2^n - 2} \quad (1.46)$$

и зависит от того, содержит или нет линейная оболочка  $\mathcal{L}(X)$  вектор из единиц  $\mathbf{e}$ , так как  $\mathcal{L}(X) = \mathcal{L}(X) + \mathbf{e}$  при  $\mathbf{e} \in \mathcal{L}(X)$ . Будем обозначать матрицу  $X$  через  $X^\oplus$  при  $\mathbf{e} \in \mathcal{L}(X)$  и через  $X^\ominus$  при  $\mathbf{e} \notin \mathcal{L}(X)$ .

**Лемма 5.** Пусть  $Y \neq \mathbf{0}$ ,  $Y \neq \mathbf{e}$ , тогда вероятность случайной классификации по матрице данного ранга имеет вид

$$p_0(r, X^\oplus) = \frac{2^r - 2}{2^n - 2}, \quad p_0(r, X^\ominus) = \frac{2(2^r - 1)}{2^n - 2}. \quad (1.47)$$

Действительно, в случае матрицы  $X^\oplus$  ранга  $r$  мощность линейной оболочки равна  $|\mathcal{L}(X)| = 2^r - 1$ . Так как  $\mathbf{e} \in \mathcal{L}(X)$ , линейная оболочка  $\mathcal{L}(X)$  совпадает с  $\mathcal{L}(X) + \mathbf{e}$ . Исключая вектор  $\mathbf{e}$  из  $\mathcal{L}(X)$ , получаем  $p_0(r, X^\oplus) = \frac{2^r - 2}{2^n - 2}$ . В случае матрицы  $X^\ominus$  справедливо  $(\mathcal{L}(X) + \mathbf{e}) \cap \mathcal{L}(X) = \emptyset$ . Следовательно,  $p_0(r, X^\ominus) = \frac{2(2^r - 1)}{2^n - 2}$ .  $\square$

Таким образом, для вычисления вероятности случайной классификации нам требуется знать не просто количество  $M_{nm}^r$  матриц  $n \times m$  над  $F_2$  ранга  $r$ , но и таких, что  $\mathbf{e} \in \mathcal{L}(X)$ , их будем обозначать через  $M_{nm}^{r\oplus}$ , и таких, что  $\mathbf{e} \notin \mathcal{L}(X)$ , их будем обозначать через  $M_{nm}^{r\ominus}$ . Согласно [40], количество проективных подпространств размерности  $r$  из проективного пространства размерности  $m$  равно

$$M_{nm}^r = N_m^r \prod_{k=0}^{r-1} (2^n - 2^k), \quad \text{где} \quad N_m^r = \prod_{k=0}^{r-1} \frac{2^m - 2^k}{(2^r - 2^k)}.$$

**Лемма 6.** Пусть  $M_{nm}^{r\ominus}$  и  $M_{nm}^{r\oplus}$  равны количеству матриц  $X(n \times m)$  над  $F_2$  таких, что  $\mathbf{e} \notin \mathcal{L}(X)$  и  $\mathbf{e} \in \mathcal{L}(X)$  соответственно. Тогда

1.  $M_{nm}^{r\ominus} = N_m^r \prod_{k=0}^{r-1} (2^n - 2^{k+1});$
2.  $M_{nm}^{r\oplus} = N_m^r (2^r - 1) \prod_{k=1}^{r-1} (2^n - 2^k).$

Действительно, при вычислении  $M_{nm}^{r\ominus}$  обнаруживаем, во-первых, что как и в случае вычисления  $M_{nm}^r$  [40] существует  $N_m^r$  способов размещения  $r$  линейно независимых векторов по  $m$  местам. Во-вторых, в качестве первого вектора  $X_1$  может быть использован любой не  $\mathbf{e}$  из  $2^n - 2$  ненулевых векторов, в качестве второго вектора — любой из  $2^n - 2^2$  (кроме  $\mathbf{e}$ ,  $\mathbf{0}$ ,  $X_1$  и  $X_1 + \mathbf{e}(\text{mod } 2)$ ), в качестве третьего вектора — любой из  $2^n - 2^3$  (кроме  $\mathbf{e}$ ,  $\mathbf{0}$ ,  $X_1$  и  $X_1 + \mathbf{e}(\text{mod } 2)$ ,  $X_2$ ,  $X_2 + \mathbf{e}(\text{mod } 2)$ ,  $X_1 + X_2(\text{mod } 2)$ ,  $X_1 + X_2 + \mathbf{e}(\text{mod } 2)$ ) и так далее.

Для вычисления  $M_{nm}^{r\oplus}$  воспользуемся разностью

$$M_{nm}^{r\oplus} = M_{nm}^r - M_{nm}^{r\ominus} = N_m^r \prod_{k=0}^{r-1} (2^n - 2^k) - N_m^r \prod_{k=0}^{r-1} (2^n - 2^{k+1}).$$

Приводя к одинаковым сомножителям, получаем выражение

$$\begin{aligned} M_{nm}^{r\oplus} &= N_m^r \prod_{k=0}^{r-1} (2^n - 2^k) - N_m^r \prod_{k=1}^r (2^n - 2^k) = \\ &= N_m^r (2^n - 1) \prod_{k=1}^{r-1} (2^n - 2^k) - N_m^r (2^n - 2^r) \prod_{k=1}^{r-1} (2^n - 2^k) = \\ &= N_m^r \prod_{k=1}^{r-1} (2^n - 2^k) (2^n - 1 - 2^n + 2^r) = \\ &= N_m^r (2^r - 1) \prod_{k=1}^{r-1} (2^n - 2^k). \quad \square \end{aligned}$$

Покажем, что вероятность  $p_1(r, X)$  случайной классификации вектора  $Y \neq \mathbf{0}$  и  $Y \neq \mathbf{e}$  по матрице  $X = X(n, m)$  ранга  $r$  равна

$$p_1(r, X) = \frac{(2^r - 1)(2^n - 2^r + 2^{r-1} - 1)}{(2^n - 1)(2^{n-1} - 1)}. \quad (1.48)$$

Действительно, используя результаты лемм 5 и 6, получаем

$$\begin{aligned} p_1(r) &= \frac{M_{nm}^{r\oplus}}{M_{nm}^r} \cdot \frac{2^r - 2}{2^n - 2} + \frac{M_{nm}^{r\ominus}}{M_{nm}^r} \cdot \frac{2(2^r - 1)}{2^n - 2} = \\ &= \frac{2^r - 1}{2^n - 1} \cdot \frac{2^{r-1} - 1}{2^{n-1} - 1} + \frac{2^n - 2^r}{2^n - 1} \cdot \frac{2^r - 1}{2^{n-1} - 1} = \\ &= \frac{(2^r - 1)(2^n - 2^r + 2^{r-1} - 1)}{(2^n - 1)(2^{n-1} - 1)}. \quad \square \end{aligned}$$

Если  $Y = \mathbf{0}$  или  $Y = \mathbf{e}$ , то имеется всего одна популяция, и вероятность  $p_0(r, X)$  случайной классификации вектора  $Y$  по матрице  $X = X(n, m)$  ранга  $r$  равна

$$p_0(r, X) = \begin{cases} 1 & \text{при } r < m, \\ \frac{M_{nm}^{r\oplus}}{M_{nm}^r} = \frac{2^r - 1}{2^n - 1} & \text{при } r = m. \end{cases} \quad (1.49)$$

Таким образом, получено следующее утверждение.

**Теорема 14.** Пусть  $p_1(r)$  и  $p_0(r)$  имеют вид (1.48) и (1.49) соответственно. Тогда при разбиении  $Y$  на  $i+1$  популяцию, где  $i = 0, 1$ , вероятность случайной классификации равна

$$p_i = \sum_{r=1}^m p_i(r) \frac{M_{nm}^r}{2^{nm} - 1} = \frac{1}{2^{nm} - 1} \sum_{r=1}^m p_i(r) N_m^r \prod_{k=0}^{r-1} (2^n - 2^k). \quad (1.50)$$

Для вычисления критических значений числа ошибок была использована функция распределения, построенная по модельной выборке. При числе параметров ( $n = 100$ ,  $m = 10$ ), рассматриваемых в приложении, те же самые критические значения были получены при помощи приближенной формулы (1.52), основанной на предположении о независимости событий, связанных с появлением векторов в линейной оболочке матрицы  $X$ . Количество векторов  $Y_\delta$ , отличающихся от  $Y$  или от  $Y + \mathbf{e}$  не более чем в  $S$  позициях, то есть  $Y_\delta + Y \pmod{2} = \delta_s$  или  $Y_\delta + Y + \mathbf{e} \pmod{2} = \delta_s$ , где  $|\delta_s| = S$ , равно

$$b = 2 \sum_{s=0}^S C_n^s. \quad (1.51)$$

Отсюда вероятность  $P_b$  того, что хотя бы один из  $b$  векторов  $Y_\delta$  содержится в  $\mathcal{L}(X)$  или  $\mathcal{L}(X) + \mathbf{e}$ , с учетом предположения о независимости может быть приблизительно вычислена по формуле

$$P_b \approx 1 - (1 - p_0)^b \approx 1 - e^{-bp_0}. \quad (1.52)$$

### 1.10.2. Случайная классификация в случае ганкелевых матриц

Будем рассматривать существенный частный случай классификации, соответствующий обобщенной ганкелевой матрице наблюдений с шагом  $k$ , которая выражается через элементы последовательности  $x_1, \dots, x_N$  следующим образом:

$$\mathbf{X}^{(k)} = \begin{bmatrix} x_1 & x_{k+1} & x_{2k+1} & \dots & x_{(m-1)k+1} \\ x_2 & x_{k+2} & x_{2k+2} & \dots & x_{(m-1)k+2} \\ \dots & \dots & \dots & \dots & \dots \\ x_n & x_{k+n} & x_{2k+n} & \dots & x_{(m-1)k+n} \end{bmatrix}. \quad (1.53)$$



Таблица 1.26.

$r$	$\Gamma_{nm}^{\oplus}(1)$	$\Gamma_{nm}^{\ominus}(1)$	$\Gamma_{nm}^r(1)$
0	0	1	1
$[1; m-1]$	$2^{2(r-1)}$	$2 \cdot 2^{2(r-1)}$	$3 \cdot 2^{2(r-1)}$
$m$	$2^{2(r-1)}$	$2^{n+m-1} - 2 \cdot 2^{2(r-1)}$	$2^{n+m-1} - 2^{2(r-1)}$

$r$	$\Gamma_{nm}^{\oplus}(2)$	$\Gamma_{nm}^{\ominus}(2)$
0	0	1
$[1; m-1]$	$3 \cdot 2^{3r-4} - 2^{2r-3}$	$2^{3r} + 2^{3(r-1)} - 2^{2(r-1)}$
$m$	$3 \cdot 2^{3m-4} - 2^{2m-3}$	$2^{n+2(m-1)} - 3 \cdot 2^{3(m-1)} + 2^{2(m-1)}$

В частности,  $\mathbf{X}^{(1)}$  и  $\mathbf{X}^{(2)}$  имеют соответственно вид

$$\begin{bmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{m+1} \\ \dots & \dots & \dots & \dots \\ x_n & x_{n+1} & \dots & x_{m+n-1} \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} x_1 & x_3 & \dots & x_{2m-1} \\ x_2 & x_4 & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_n & x_{n+2} & \dots & x_{2m+n-2} \end{bmatrix}.$$

Обозначим через  $\Gamma_{nm}^r(k)$  количество обобщенных ганкелевых матриц с шагом  $k$ . При  $k = 1$ , согласно [36],

$$\Gamma_{nm}^r(1) = \begin{cases} 1, & r = 0, \\ 3 \cdot 2^{2(r-1)}, & 1 \leq r \leq m-1, \\ 2^{n+m-1} - 2^{2(m-1)}, & r = m. \end{cases}$$

При  $k = 2$  для  $m > 2$ ,  $n \geq 2m-1$  справедливо [37] выражение

$$\Gamma_{nm}^r(2) = \begin{cases} 1, & r = 0, \\ 21 \cdot 2^{3r-4} - 3 \cdot 2^{2r-3}, & 1 \leq r \leq m-1, \\ 2^{n+2(m-1)} - 3 \cdot 2^{3m-4} + 2^{2m-3}, & r = m. \end{cases} \quad (1.54)$$

Обозначим через  $\Gamma_{nm}^{\oplus}(k)$  и  $\Gamma_{nm}^{\ominus}(k)$  количество ганкелевых матриц  $\mathbf{X}$  с шагом  $k$ , для которых соответственно  $\mathbf{e} \in \mathcal{L}(\mathbf{X})$  и  $\mathbf{e} \notin \mathcal{L}(\mathbf{X})$ . Выражения  $\Gamma_{nm}^{\oplus}(1)$  и  $\Gamma_{nm}^{\ominus}(1)$  из табл. 1.26 были аналитически получены

Усевичем К.Д., а формулы для  $\Gamma_{nm}^{\oplus}(2)$ ,  $\Gamma_{nm}^{\ominus}(2)$  были вычислены экспериментально для разных  $n, m, r$ .

Выражение для вероятности случайной классификации получается непосредственно из (1.50), где вместо  $M_{nm}^r$  используется  $\Gamma_{nm}^r(2)$ , а в (1.48) вместо  $M_{nm}^{r\ominus}$  и  $M_{nm}^{r\oplus}$  — выражения для  $\Gamma_{nm}^{r\ominus}(2)$  и  $\Gamma_{nm}^{r\oplus}(2)$ . Например, при  $n = 100$ ,  $m = 10$ ,  $k = 2$  согласно (1.52) при уровне значимости  $\alpha = 0.05$  получено критическое число допустимых ошибок, равное  $S = 29$ .

### 1.10.3. Связность последовательностей РНК

Рассмотрим данные нуклеотидных последовательностей РНК трех микроорганизмов, которые обозначаются через  $16S Halobact$ ,  $16S Ecoli$  и  $16S Deionoc$ . При помощи кодировки аминокислот

$$a = (0, 0), \quad t = (1, 1), \quad g = (1, 0), \quad c = (0, 1)$$

получаем три последовательности  $(x_1, \dots, x_n)$ ,  $x_i \in \{0, 1\}$ . Из-за двухсимвольной кодировки будем строить ганкелеву матрицу с шагом  $k = 2$ .

В качестве метрик зависимости между двумя бинарными последовательностями  $X = (x_1, \dots, x_n)^T$  и  $Y = (y_1, \dots, y_n)^T$  рассматриваем односторонние  $J_{x|y}$  и  $J_{y|x}$  и двусторонний  $J$  коэффициенты неопределенности, метрики

$$\varrho_1(X, Y) = 1 - \|X + Y \pmod{2}\|,$$

$$\varrho_2(X, Y) = 1 - H((X + Y) \pmod{2}),$$

где через  $H(X)$ ,  $X = (x_1, \dots, x_n)^T$ ,  $x_i \in \{0, 1\}$ , обозначена оценка энтропии

$$H(x_1, \dots, x_n) = -\hat{p} \log_2(\hat{p}) - (1 - \hat{p}) \log_2(1 - \hat{p}),$$

$$\hat{p} = \frac{1}{n} \sum_{j=1}^n x_j.$$

Некоторые метрики проще в исследовании статистических свойств, некоторые удобнее в интерпретации.

В качестве двойственной к  $X = (x_1, \dots, x_n)^T$  используем последовательность приращений

$$\check{X} = (x_1, x_1 + x_2, x_2 + x_3, \dots, x_{n-1} + x_n)^T \text{ над } F_2.$$

Последовательность  $X$  может быть восстановлена из  $\check{X}$  через кумулятивную последовательность

$$X^{\leq} = \{\check{x}_j^{\leq}\} = \left\{ \sum_{i=1}^j \check{x}_i \right\},$$

причем безразлично накапливать двойственные или рассматривать двойственную к кумулятивной.

Поскольку отягощающим фактором является пенетрантность, и разные фрагменты последовательностей  $x_1, \dots, x_N$  и  $y_1, \dots, y_N$ ,  $N > n$ , связаны не одинаково, выделим из этих последовательностей некоторые интересующие нас фрагменты и в соответствии с (1.53) преобразуем их в ганкелевы матрицы

$$\mathbf{X}^{(2)} = [X_1 | \dots | X_m] \quad \text{и} \quad \mathbf{Y}^{(2)} = [Y_1 | \dots | Y_t].$$

Затем найдем максимально зависящие между собой симптомы  $X_\tau$  и  $Y_\mu$  в соответствии с метрикой  $\varrho(X_\tau, Y_\mu)$ , где  $\tau = (\tau_1, \dots, \tau_m)$ ,  $\mu = (\mu_1, \dots, \mu_t)$ .

Например, в качестве  $X = (x_1, \dots, x_{N_1})$  и  $Y = (y_1, \dots, y_{N_2})$  рассматриваются закодированные нулями и единицами нуклеотидные последовательности РНК *16S Halobact* и *16S Ecoli*,  $n = 100$ , в одной из последовательностей имеется только один вектор вложения, т.е.  $t = 1$ .

Подпоследовательности одинаковой длины  $n$  с началом в  $n_1$  или  $n_2$  будем обозначать соответственно через

$$X(n_1) = (x_{n_1}, \dots, x_{n_1+n})^T, \quad Y(n_2) = (y_{n_2}, \dots, y_{n_2+n})^T.$$

При исследовании  $X(n_1)$  при  $n_1 = 433$  и  $Y(n_2)$  было обнаружено 60 случаев, когда число ошибок  $\varrho_1(X(433), Y(n_2))$  было меньше критического 30, что составляет 4.02%, и 66 случаев (4.42%) для двойственных последовательностей. Сочетание обеих характеристик позволяет дифференцировать эти случаи на сильные и слабые в зависимости от их совместного или отдельного проявления (рис. 1.17). Тогда удается выделить наиболее значимые метрики:

$$\begin{aligned} \varrho_1(\textit{Halobac}(433), \textit{Ecoli}(475)) &= 23, \quad p = 5.64 \cdot 10^{-5}, \\ \varrho_1(\textit{Ecoli}(1), \textit{Deionoc}(3)) &= 8, \quad p = 2.22 \cdot 10^{-16}, \\ \varrho_1(\textit{Halobac}(407), \textit{Deionoc}(423)) &= 17, \quad p = 1.34 \cdot 10^{-8}. \end{aligned}$$

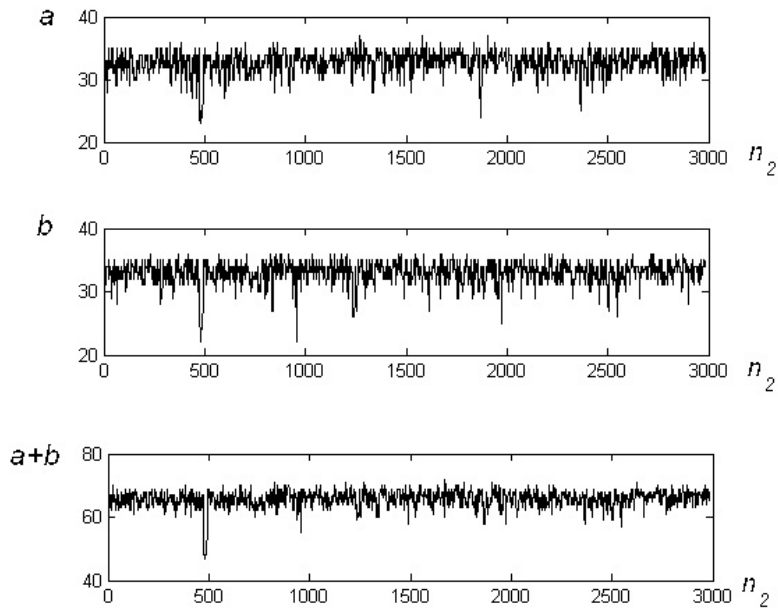


Рис. 1.17. Зависимости  $a = \varrho_1(X(n_1), Y(n_2))$ ,  $b = \varrho_1(\check{X}(n_1), \check{Y}(n_2))$  и  $a + b$  от  $n_2$  при  $n = 100$ ,  $n_1 = 433$ .

Значимости  $p$ , вычисленные с учетом свойства ганкелевости матриц наблюдений или по всем матрицам, оказались одинаковыми.

Для иллюстрации действия логического принципа двойственности [32] неслучайно были выбраны данные РНК микроорганизмов с одинаковыми фрагментами. На этом примере хорошо видно, что одних статистических критериев и принципа маловероятных событий оказывается недостаточно для выявления закономерностей. Для того чтобы не интерпретировать все более-менее статистически значимые результаты, можно остановиться только на тех фактах, которые устойчивы по отношению к двойственной задаче.

## 1.11. Приоритетные кривые в когнитивной биологии

В когнитивной биологии изучают поведение людей или животных в разных обстоятельствах или условиях эксперимента. Например, в течение 100 секунд регистрировалось поведение самца крысы после того, как к нему в клетку подсаживали другую крысу того же пола. В эксперименте участвовали две группы самцов крыс:  $n_2 = 9$  в контрольной группе  $C$  и  $n_1 = 7$  в группе  $D$  депрессированных ранее крыс тем, что к ним индивидуально в клетку несколько раз подсаживали агрессивного самца.

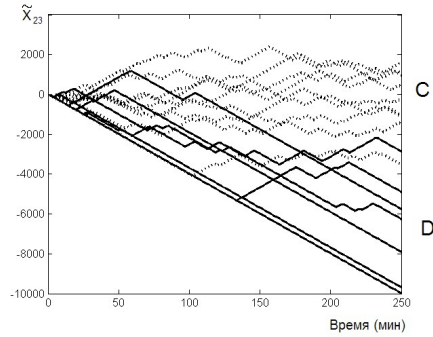


Рис. 1.18. Приоритетные кривые  $\tilde{X}_{23}$ .

Рассмотрим поведенческие акты (табл.1.27) с точки зрения основных факторов: стрессовость ситуации ( $Agr$ ), моторика ( $Mot$ ) и социальный статус ( $Soz$ ). Будем отмечать значения этих факторов в каждые 0.01 секунды и построим три дихотомических временных ряда  $X_1(t)$ ,  $X_2(t)$ ,  $X_3(t)$ , симптомы  $X_{12}(t)$ ,  $X_{13}(t)$ ,  $X_{23}(t)$  и  $X_{123}(t)$ , а затем для всех семи дихотомических рядов  $X_\tau(t)$ ,  $\tau \subseteq \{1, 2, 3\}$ , кумулятивные ряды

$$\tilde{X}_\tau(t) = \sum_{j=1}^t (2X_\tau(j) - 1),$$

представляющие собой реализации случайных процессов с мартингалными свойствами, которые назовем *приоритетными кривыми*.

Таблица 1.27.

Event	Поведенческий акт	Agr	Mot	Soz	Cod
agg-groom	агрессивный груминг партнера	1	1	1	7
aggregatio	сгучивание	0	1	1	3
allogroom	чистка тела партнера	0	0	1	1
bite	укус	1	1	1	7
body-groom	чистка своего тела	0	0	0	0
body-snif	обнюхивание тела партнера	0	0	1	1
climbing	залезание на партнера	0	1	1	3
digging	копание	0	1	0	2
eating	еда	0	0	0	0
escape	убегание	1	1	0	6
following	следование за партнером	0	1	1	3
freezing	замирание	1	0	0	4
gen-snif	обнюхивание гениталий партнера	0	0	1	1
kick	удар	1	1	1	7
mount	карапкаться на партнера	0	1	1	3
nose-snif	обнюхивание носа партнера	0	0	1	1
on-back	защитная поза на спине	1	0	0	4
opposition	оппозиция	1	0	0	4
over	положение над партнером	0	1	1	3
rotation	вращение	0	1	0	2
scramble	карабканье, схватка	1	1	1	7
scratching	почесывание	0	1	0	2
shaking	отряхивание, тряска головой	0	1	0	2
side-up	отход от партнера	1	1	0	6
sideway	притирка боком	0	0	1	1
sit-snif	сидение с принюхиванием	0	0	0	0
sitting	сидение	0	0	0	0
stretching	потягивание	0	0	0	0
substituti	вытеснение	1	1	1	7
threat	угроза	1	0	1	5
under	положение под партнером	0	1	1	3
up-right	защитная стойка	1	0	0	4
vert-def	вертикальная стойка	1	0	0	4
walking	прогулка	0	1	0	2

Таблица 1.28.

$\tau$	$\hat{p} \pm \hat{\sigma}$	
	D	C
1	$0.14 \pm 0.12$	$0.10 \pm 0.05$
2	$0.23 \pm 0.20$	$0.31 \pm 0.07$
12	$0.37 \pm 0.21$	$0.40 \pm 0.10$
3	$0.10 \pm 0.18$	$0.24 \pm 0.12$
13	$0.24 \pm 0.20$	$0.35 \pm 0.10$
23	$0.16 \pm 0.13$	$0.48 \pm 0.08$
123	$0.30 \pm 0.20$	$0.58 \pm 0.08$

Для примера на рис.1.18 представлены 40-ые точки реализации кумулятивного симптома  $\tilde{X}_{23}$  в группах  $D$  и  $C$ , на рис.1.19 усреднения по группам  $D$  и  $C$  (пунктир) симптомов  $\tilde{X}_{23}$ ,  $\tilde{X}_1$ ,  $\tilde{X}_{123}$ . Оценки  $\hat{p}$  приоритетных вероятностей  $p = P\{X_\tau = 1\}$  используются для их характеристики (табл.1.28).

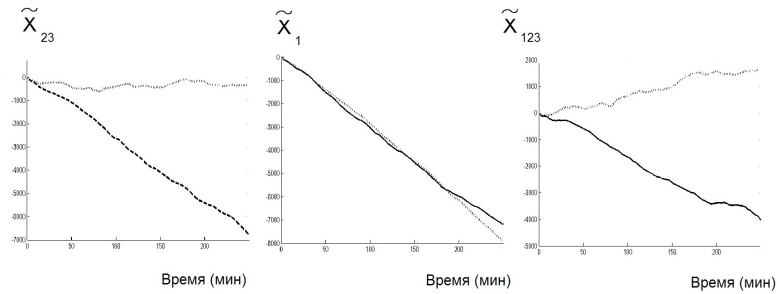


Рис. 1.19. Приоритетные кривые с усреднением в группах.

Убывающий тренд по  $\tilde{X}_{23}$  в группе  $D$  свидетельствует о выраженной при депрессии ассоциированности социального поведения с моторикой  $\hat{p} = 0.16$ . Для сравнения  $\hat{p} = 0.48$  в группе  $C$ . Приоритетные кривые по  $\tilde{X}_1$  в группах  $C$  и  $D$  не отличаются. Процесс  $\tilde{X}_{123}$  в группе  $D$  является супермартингалом  $\hat{p} = 0.3 < 0.5$ , а в группе  $C$  субмартингалом  $\hat{p} = 0.58$ . Он интерпретируется как фактор адекватной реактивности на появление нового объекта. При депрессии этот фактор существенно снижен.

## Глава 2.

# Проблемы неоднородности и неполноты данных

### 2.1. Классификация по расслоенной выборке

Пусть имеется две нормально распределенные популяции  $W_1$  и  $W_2$  с одинаковой известной ковариационной матрицей  $\Sigma$  и векторами средних  $\mu_1$  и  $\mu_2$ . Задача дискриминантного анализа состоит в том, чтобы решить, к какой из этих популяций отнести вектор наблюдений  $X = (x_1, \dots, x_p)^T$ .

Классификация осуществляется через построение линейной дискриминантной функции

$$f(X) = \alpha_1 x_1 + \dots + \alpha_p x_p = \alpha^T X,$$

где  $\alpha = (\alpha_1, \dots, \alpha_p)^T$  вектор параметров, вычисление критического значения  $c$  и введение правила, согласно которому, если  $f(X) > c$ , то  $X$  относится к популяции  $W_1$ , иначе к  $W_2$ . Эта задача актуальна для диагностики и отдаленного прогнозирования. Если  $f(X) \leq c$  для  $X \in W_1$  или  $f(X) \geq c$  для  $X \in W_2$ , то имеет место ошибочная классификация. Естественно необходимо, чтобы доля этой ошибки



была как можно меньше. Для упорядочения признаков по степени их значимости для классификации используется пошаговая процедура. В случаях, когда линейные методы работают плохо, используют метод опорных векторов или нейронные сети. Но в этих методах остается скрытой структура классификации, а экспериментатору важно знать, за счет изменения каких признаков индивид относится к той или иной популяции.

Одной из распространенных причин неразделимости многомерных выборок при помощи гиперплоскостей является их неоднородность и связанное с нею различие в структуре значимых факторов для отдельных подвыборок. Оказалось, при достаточно большом объеме выборки можно снизить вероятность ошибочной классификации, если популяции расслоить и построить отдельные линейные дискриминантные функции внутри слоев. Для выбора наилучшего расслаивающего фактора могут быть использованы корреляционные, симптомно-синдромальные или эвристические методы.

### 2.1.1. Вероятность ошибочной классификации

Если наблюдение из популяции  $W \sim \mathcal{N}(\mu, \Sigma)$ , то  $z = \alpha^T X$  имеет нормальное распределение  $\mathcal{N}(\zeta, \sigma)$ , где  $\zeta = \alpha^T \mu$ ,  $\sigma^2 = \alpha^T \Sigma \alpha$ .

Обозначим через  $\zeta_i = \alpha^T \mu_i$  средние значения дискриминантной функции  $z = \alpha^T X$  в случае популяции  $W_i$ ,  $i = 1, 2$ . Имеет смысл выбрать такой вектор  $\alpha$ , чтобы  $\zeta_1$  и  $\zeta_2$  были как можно дальше удалены друг от друга. Для этого введем расстояние Махаланобиса

$$\Delta^2 = \frac{(\zeta_1 - \zeta_2)^2}{\sigma^2} = \frac{(\alpha^T(\mu_1 - \mu_2))^2}{\alpha^T \Sigma \alpha}.$$

В 1936 году Фишер показал, что *максимальное расстояние Махаланобиса доставляет вектор  $\alpha$ , который является решением системы*

$$\Sigma \alpha = \mu_1 - \mu_2. \quad (2.1)$$

ПРИМЕР. Рассмотрим две популяции:  $W_1$  - больные бронхитом ( $n_1 = 66$ ),  $W_2$  - больные пневмонией ( $n_2 = 67$ ), которые характеризуются признаками:  $x_1$  серосодержащие белки в восстановительной форме,  $x_2$  скорость оседания эритроцитов. Средние значения с ошибкой равны  $\mu_{11} = 7.1 \pm 0.7$  в  $W_1$  и  $\mu_{12} = 5.8 \pm 0.8$  в  $W_2$ ,  $\mu_{21} = 32.9 \pm 4.4$  в  $W_1$  и  $\mu_{22} = 46.0 \pm 5.4$  в  $W_2$ . Вычисляем оценку ковариационной матрицы

$$\hat{\Sigma} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \begin{bmatrix} 0.57 & -0.18 \\ -0.18 & 24 \end{bmatrix},$$

где через  $S_1^2$  и  $S_2^2$  обозначены несмещенные оценки ковариационных матриц внутри популяций. Решая уравнение (2.1) относительно  $\alpha$ , вычисляем коэффициенты  $\alpha_1$  и  $\alpha_2$  дискриминантной функции  $z = -1.97x_1 + 0.53x_2$ . Среднее значение дискриминантной функции у больных бронхитом равно  $\zeta_1 = \alpha^T \mu_1 = 3.45$ , у больных с пневмонией  $\zeta_2 = \alpha^T \mu_2 = 12.95$ . Граничное значение равно  $c = (\zeta_1 + \zeta_2)/2 = 8.22$ . На рис. 2.1 уравнение классифицирующей прямой имеет вид  $x_2 = 3.7x_1 + 15.5$ . Например, для больного из  $W_2$ , у которого  $x_1 = 5.5$ ,  $x_2 = 40$ , получаем  $z = 10.37 > c = 8.22$ . Этот больной классифицируется правильно, так как  $z$  находится ближе к  $\zeta_2 = 12.95$ , чем к  $\zeta_1 = 3.45$ .

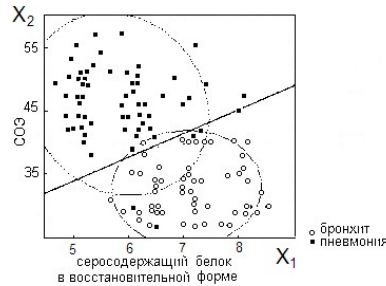


Рис. 2.1. Классификация больных по  $x_1$  и  $x_2$ .

Для вычисления вероятности ошибочной классификации применяют формулу полной вероятности

$$\mathcal{P} = q_1 P(2|1) + q_2 P(1|2), \quad (2.2)$$

где  $P(2|1)$  определяется как вероятность ошибочного отнесения индивида из  $W_1$  к популяции  $W_2$ ,  $P(1|2)$  по аналогии (рис. 2.2),  $q_i$  — априорные вероятности того, что объект принадлежит популяции  $W_i$ :  $q_1 = P(W_1)$ ,  $q_2 = P(W_2)$ ,  $q_1 + q_2 = 1$ . Через  $P(W_1|X)$  и  $P(W_2|X)$  обозначаются апостериорные вероятности того, что объект, характеризуемый вектором наблюдений  $X$ , принадлежит к  $W_i$ , а через  $P(X|W_i)$  — условная вероятность получения вектора наблюдений  $X$ , если известно, что объект относится к популяции  $W_i$ .

Заменяя  $P(X|W_i)$  на плотности  $f_i(x)$ , вычисляем апостериорные вероятности по формуле Байеса

$$P(W_i|X) = \frac{q_i f_i(x)}{q_1 f_1(x) + q_2 f_2(x)}.$$

Наблюдение  $X$  относится к  $W_1$ , если  $P(W_1|X) > P(W_2|X)$ , что равносильно  $q_1 f_1(x) \geq q_2 f_2(x)$ , иначе к  $W_2$ . При этом, согласно [39] и [35], минимизируется (2.2). Если  $W_i \sim \mathcal{N}(\mu_i, \Sigma)$ , а вектор  $\alpha$  есть решение системы (2.1) и  $\zeta_i = \mu_i^T \alpha$ , то классификация к  $W_1$  означает выполнение неравенства

$$x^T \alpha \geq \frac{\zeta_1 + \zeta_2}{2} + \ln \frac{q_2}{q_1}.$$

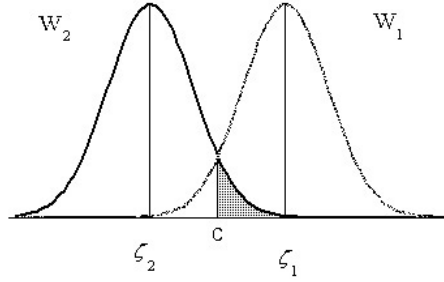


Рис. 2.2. Плотности дискриминантных функций.

Для определенности будем считать, что  $\zeta_1 > \zeta_2$ <sup>1</sup>. Обозначим через  $\Phi(x)$  функцию стандартного нормального распределения, а также

$$K = \ln \frac{q_2}{q_1}, \quad T = \frac{\zeta_1 + \zeta_2}{2} + K. \quad (2.3)$$

Тогда, используя  $\sigma = \Delta$  и  $\Delta^2 = \zeta_1 - \zeta_2$ , получим

$$\begin{aligned} P(2|1) &= \Phi\left(\frac{T - \zeta_1}{\sigma}\right) = \Phi\left(\frac{\frac{\zeta_1 + \zeta_2}{2} + K - \zeta_1}{\Delta}\right) = \\ &= \Phi\left(\frac{\frac{\zeta_2 - \zeta_1}{2} + K}{\Delta}\right) = \Phi\left(\frac{-\frac{\Delta^2}{2} + K}{\Delta}\right) = \Phi\left(-\frac{\Delta}{2} + \frac{K}{\Delta}\right), \\ P(1|2) &= 1 - \Phi\left(\frac{T - \zeta_2}{\sigma}\right) = 1 - \Phi\left(\frac{\Delta}{2} + \frac{K}{\Delta}\right) = \Phi\left(-\frac{\Delta}{2} - \frac{K}{\Delta}\right). \end{aligned}$$

<sup>1</sup>Иначе можно просто поменять знак у дискриминантной функции.

### 2.1.2. Расслоение популяции

Пусть  $q = P(W_1)$  и  $1 - q = P(W_2)$  — априорные вероятности. По некоторому дихотомическому признаку имеет место расслоение популяции на  $S^1 = W_1^1 + W_2^1$  и  $S^2 = W_1^2 + W_2^2$  с вероятностями  $s = P(S^1)$ ,  $1 - s = P(S^2)$ . Обозначим через  $x = P(W_1^1)$ , тогда

$$P(W_2^1) = s - x, \quad P(W_1^2) = q - x, \quad P(W_2^2) = 1 - s - q + x.$$

**Предложение 9.** *Если расстояния Махаланобиса инвариантны относительно расслоения популяции, то вероятность случайной классификации при расслоении не увеличится.*

Введем отдельные обозначения (2.3) по нерасслоенной ( $i = 0$ ) и расслоенным ( $i = 1, 2$ ) выборкам:

$$u_i = \frac{K_i}{\Delta_i} - \frac{\Delta_i}{2}, \quad v_i = -\frac{K_i}{\Delta_i} - \frac{\Delta_i}{2},$$

где  $K_0 = \ln\left(\frac{1}{q} - 1\right)$ ,  $K_1 = \ln\left(\frac{s}{x} - 1\right)$ ,  $K_2 = \ln\left(\frac{1-s}{q-x} - 1\right)$ .

Вычисляем вероятности ошибочной классификации:

$$\begin{aligned} \mathcal{P}_0 &= q\Phi(u_0) + (1-q)\Phi(v_0), \\ s\mathcal{P}_1 &= x\Phi(u_1) + (s-x)\Phi(v_1), \\ (1-s)\mathcal{P}_2 &= (q-x)\Phi(u_2) + (1-s-q+x)\Phi(v_2). \end{aligned}$$

Эффективность расслоения определяется через разность

$$\begin{aligned} P(x) &= \mathcal{P}_0 - (s\mathcal{P}_1 + (1-s)\mathcal{P}_2) = \\ &= x(\Phi(u_0) - \Phi(u_1)) + (q-x)(\Phi(u_0) - \Phi(u_2)) + \\ &+ (s-x)(\Phi(v_0) - \Phi(v_1)) + (1-s-q+x)(\Phi(v_0) - \Phi(v_2)). \end{aligned}$$

Если признаки, определяющие расслоение общей популяции на  $W_1^1$ ,  $W_2^1, W_1^2, W_2^2$ , независимы, то  $x = qs$ , следовательно,  $K_0 = K_1 = K_2$ ,  $u_0 = u_1 = u_2$ ,  $v_0 = v_1 = v_2$ ,  $P(sq) = 0$ . Дифференцируя по  $x$ , можно показать, что  $P'(sq) = 0$ ,  $P''(sq) > 0$ , т.е. в точке  $x = qs$  функция  $P(x)$  имеет минимум. Это означает, что улучшение классификации можно ожидать при большей зависимости признаков.

## 2.2. Эргодичность и закон Харди—Вайнберга

Закон Харди—Вайнберга известен в биологии как закон генетической устойчивости [23],[20],[24], согласно которому при панмиксии (независимое скрещивание) и достаточно большой численности популяции частоты генов остаются неизменными от поколения к поколению, а частоты гамет стремятся к некоторым стационарным значениям, соответствующим равновесному состоянию популяции.

Известны две статистические модели закона Харди—Вайнберга: модель с линейным условием перехода и квадратичной структурой пространства состояний [22] и, наоборот, модель с квадратичными переходными уравнениями и линейной структурой пространства состояний [23],[20]. Для этих двух моделей любая форма обобщения является проблемой, поскольку в них в связи с жесткой привязкой к условиям биологического эксперимента пространство состояний представляется в виде искаженной аффинной геометрии.<sup>2</sup>

Одной из важных задач является построение индуктивной (допускающей переход от частного случая к общему) стохастической модели эволюции распределений генотипов на основе дискретного марковского процесса с пространством состояний в виде конечной аффинной геометрии.

Этот марковский процесс существует не сам по себе, а в совокупности с целым семейством попутных марковских процессов, заданных на внутренних аффинных геометриях меньшего порядка. При этом вероятность состояния процесса над конечной аффинной геометрией наибольшего порядка является сквозной ссылкой для упорядоченного в соответствии со структурой двойственной проективной геометрии пакета частот попутных распределений, обеспечивая их прямо-двойственную сцепленность. Переход к этому пакету частот позволяет получить в общем виде стационарное распределение, соответствующее в генетике состоянию равновесия популяции. Распределение аллелей в локусе генотипа может быть получено суммированием вероятностей состояний, неразличимых с биологической точки зрения.

---

<sup>2</sup>Под „искаженностью“ этой конечной аффинной геометрии подразумевается неразличимость точек с одинаковыми суммами парных компонент.

### 2.2.1. Подстановки вектора распределения

Обозначим точки аффинной [25] геометрии  $E_k^r$  порядка  $k$  и характеристики  $r$  через  $(x_1, \dots, x_k)$ , где  $x_j = 0, 1, \dots, r-1$ . Будем рассматривать сначала случай  $r = 2$ , выделяя затем анализ изменений, вызванных увеличением характеристики, в самостоятельное исследование. Упорядочим  $2^k$  точек  $E_k^2$  согласно правилу

$$m = m(x_1, \dots, x_k) = x_1 + 2x_2 + 2^2x_3 + \dots + 2^{k-1}x_k. \quad (2.4)$$

Распределение случайного вектора  $X = (X_1, \dots, X_k)^T$  со значениями  $(x_1, \dots, x_k)$  из аффинной геометрии  $E_k^2$  обозначим через вектор  $P = (p_0, \dots, p_n)^T$ ,  $n = 2^k - 1$ , с компонентами

$$p_m = P\{X_1 = x_1, \dots, X_k = x_k\}. \quad (2.5)$$

Пусть точка  $m = m(y_1, \dots, y_k)$  получена в результате координатного сложения точек  $m_1 = m_1(x_1, \dots, x_k)$  и  $m_2 = m_2(\widetilde{x}_1, \dots, \widetilde{x}_k)$  по модулю 2, т.е.  $y_i = x_i + \widetilde{x}_i \pmod{2}$  для  $i = 1, \dots, k$ . Вычисленный в соответствии с (2.4) индекс  $m = m(y_1, \dots, y_k)$  будем считать равным сумме индексов  $m_1$  и  $m_2$  и обозначать через

$$m = m_1 \oplus m_2. \quad (2.6)$$

*Подстановку распределения*, обусловленную прибавлением к индексам всех компонент вектора  $P$  одного и того же индекса  $m$  над полем  $F_q$ ,  $q = 2^k$ , будем обозначать через

$$P(m) = (p_{0 \oplus m}, \dots, p_{n \oplus m})^T. \quad (2.7)$$

Очевидно, что вектор распределения  $P = (p_0, \dots, p_n)^T$  является тривиальной нулевой подстановкой  $P(0) = (p_{0 \oplus 0}, \dots, p_{n \oplus 0})^T$ .

Например, при  $k = 2$  подстановки распределения имеют вид:

$$\begin{aligned} P(0) &= (p_{0 \oplus 0}, p_{1 \oplus 0}, p_{2 \oplus 0}, p_{3 \oplus 0}) = (p_0, p_1, p_2, p_3), \\ P(1) &= (p_{0 \oplus 1}, p_{1 \oplus 1}, p_{2 \oplus 1}, p_{3 \oplus 1}) = (p_1, p_0, p_3, p_2), \\ P(2) &= (p_{0 \oplus 2}, p_{1 \oplus 2}, p_{2 \oplus 2}, p_{3 \oplus 2}) = (p_2, p_3, p_0, p_1), \\ P(3) &= (p_{0 \oplus 3}, p_{1 \oplus 3}, p_{2 \oplus 3}, p_{3 \oplus 3}) = (p_3, p_2, p_1, p_0). \end{aligned}$$

### 2.2.2. Компоненты вектора перераспределения

Структура конечной аффинной геометрии предполагает наличие внутри нее двойственной структуры проективной геометрии (разд. 1.3.4). Для распределения на точках аффинной геометрии в качестве двойственной проективной структуры выступает вектор *перераспределения*  $Q$ , компонентами которого являются вероятности распределений на аффинных геометриях всех более низких порядков, относящиеся к одной точке.

**Определение 10.** Пусть  $\Omega = \{1, 2, \dots, k\}$  — множество индексов компонент случайного вектора  $X = (X_1, \dots, X_k)^T$ ,  $\Theta$  — множество всех подмножеств вида  $\tau = \{\tau_1, \dots, \tau_s\} \subseteq \Omega$ ,  $s \leq k$ . Для любой точки  $m = m(x_1, \dots, x_k)$  вектор  $c$

$$Q(m) = \{q_\tau(m)\}_{\tau \in \Theta} \quad (2.8)$$

с  $2^k$  компонентами вида

$$q_\tau(m) = P\{X_{\tau_1} = x_{\tau_1}, \dots, X_{\tau_s} = x_{\tau_s}\} \quad (2.9)$$

будем называть *вектором перераспределения*.

В случае пустого множества  $\emptyset$  и полного индекса  $\Omega$  имеем соответственно тривиальную компоненту  $q_\emptyset(m) = 1$  и вероятность

$$q_\Omega(m) = q_{12\dots k}(m) = p_m = P\{X_1 = x_1, \dots, X_k = x_k\}. \quad (2.10)$$

Для нулевой точки будем использовать обозначение  $Q = Q(0)$ . При  $m = 0$  в случае  $k = 1$  имеем  $\Omega = \{1\}$ ,  $\Theta = \{\emptyset, 1\}$ , вектора распределения и перераспределения имеют вид:

$$\begin{aligned} P &= (P\{X_1 = 0\}, P\{X_1 = 1\})^T = (p_0, p_1)^T, \\ Q &= (q_\emptyset, q_1)^T, \text{ где} \\ q_\emptyset &= 1 = p_0 + p_1, \quad q_1 = P\{X_1 = 0\} = p_0. \end{aligned}$$

При  $k = 2$  и  $m = 0$  аналогично  $\Omega = \{1, 2\}$ ,  $\Theta = \{\emptyset, 1, 2, 12\}$ , вектор распределения равен

$$\begin{aligned} P &= (p_0, p_1, p_2, p_3)^T = (P\{X_1 = 0, X_2 = 0\}, P\{X_1 = 1, X_2 = 0\}, \\ &\quad P\{X_1 = 0, X_2 = 1\}, P\{X_1 = 1, X_2 = 1\})^T, \end{aligned}$$

вектор перераспределения имеет вид  $Q = (q_0, q_1, q_2, q_{12})^T$ , где

$$\begin{aligned} q_0 &= 1 = p_0 + p_1 + p_2 + p_3, \\ q_1 &= P\{X_1 = 0\} = p_0 + p_2, \\ q_2 &= P\{X_2 = 0\} = p_0 + p_1, \\ q_{12} &= P\{X_1 = 0, X_2 = 0\} = p_3. \end{aligned} \quad (2.11)$$

Между распределением и перераспределением существует взаимно-однозначное соответствие  $Q(m) = CP(m)$ . При  $k = 1$  матрица  $C = C_0$  и вектор перераспределения имеют соответственно вид:

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad (2.12)$$

$$Q = CP \iff \begin{bmatrix} q_0 \\ q_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} p_0 + p_1 \\ p_0 \end{bmatrix}.$$

При  $k = 2$  используется прямое внешнее умножение матриц

$$C = C_0 \otimes C_0 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Непосредственным умножением матриц получаем  $CP =$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} p_0 + p_1 + p_1 + p_3 \\ p_0 + p_2 \\ p_0 + p_1 \\ p_0 \end{bmatrix} = Q,$$

где  $Q = (q_0, q_1, q_2, q_{12})^T$  с компонентами из (2.11).

Очевидно, что в общем случае нужно использовать  $k$ -кратное прямое внешнее умножение матриц.

**Лемма 7.** Пусть  $C = \bigotimes_{j=1}^k C_j = C_1 \otimes \dots \otimes C_k$  —  $k$ -кратное прямое умножение матриц  $C_j = C_0$  из (2.12), тогда  $Q(m) = CP(m)$ . Обратное преобразование имеет вид  $C^{-1} = \bigotimes_{j=1}^k D_j$ , где

$$D_j = C_0^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}.$$



### 2.2.3. Импульсная упорядоченность компонент вектора перераспределения

Пусть имеются базовое множество  $M_k$  элементов  $a_1, \dots, a_k$  и рекуррентным образом построенная импульсная система подмножеств  $\{\Theta_j\}_{j=0}^k$ , в которой  $\Theta_0 = \emptyset$ ,  $\Theta_1 = \{\emptyset, a_1\}$ ,

$$\Theta_j = (\Theta_{j-1}, a_j \star \Theta_{j-1}), \text{ где } a_j \star \Theta_{j-1} = \{a_j \cup \tau \mid \tau \subseteq \Theta_{j-1}\}.$$

Упорядоченность множеств в  $\Theta_j$  будем также называть *импульсной*. Благодаря этой структуре оказывается возможным индуктивный способ построения векторов перераспределения. Например, при  $m = 0$  и базовых множествах  $\{1\}, \{1, 2\}, \{1, 2, 3\}$  имеем соответственно векторы перераспределения  $Q$ , индексы компонент которых импульсно упорядочены.

$$\begin{aligned} & (q_\emptyset, q_1), \\ & (q_\emptyset, q_1, q_2, q_{12}), \\ & (q_\emptyset, q_1, q_2, q_{12}, q_3, q_{13}, q_{23}, q_{123}). \end{aligned}$$

**Предложение 10.** Пусть  $M_k = \{1, 2, \dots, k\}$ . Из импульсного построения перераспределения  $Q$  для набора его индексов

$$I_k = (i_0, \dots, i_{2^k-1}) = (\emptyset, 1, 2, (12), 3, (13), \dots, (M_k))$$

имеем  $I_k = (I_{k-1}, I_{k-1} \cup k)$ . Тогда  $i_j \cup i_{2^k-1-j} = M_k$ .

**Доказательство** (по индукции). При  $k = 1$  в  $I_1 = (\emptyset, 1)$  объединение первого и последнего элементов исчерпывает множество  $M_1 = \{1\}$ . При  $k = 2$  в  $I_2 = (\emptyset, 1, 2, (12))$ , множество  $M_2 = \{1, 2\}$  исчерпывается объединением первого и последнего или второго и предпоследнего. Пусть  $1 < r < k$  и  $\forall j < 2^{r-1}$  справедливо

$$i_j \cup i_{2^{r-1}-j} = M_r.$$

Перейдем к  $I_{r+1} = (I_r, I_r \cup (r+1))$ . Для  $j < 2^r$  получаем

$$i_j \cup i_{2^{r+1}-1-j} = i_j \cup (i_{2^r-1-j} \cup r+1) = M_r \cup r+1 = M_{r+1}. \quad \square$$

### 2.2.4. Переходные уравнения

Обозначим через  $p_0, p_1, p_2, p_3$  вероятности гамет у родителей. В табл. 2.1 указаны варианты гамет  $(X_1, X_2)$  у потомков в зависимости от сочетания гамет родителей. Нулем и единицей кодируются два разных аллеля. Вероятность появления, например, гаметы 00 у потомков вычисляется по формуле полной вероятности

$$p_0^* = p_0^2 + \frac{1}{2}p_0p_1 + \frac{1}{2}p_0p_2 + \frac{1}{4}p_0p_3 + \frac{1}{2}p_0p_1 + \frac{1}{4}p_1p_2 + \frac{1}{2}p_0p_2 + \frac{1}{4}p_1p_2 + \frac{1}{4}p_0p_3 = p_0^2 + p_0p_1 + p_0p_2 + \frac{1}{2}p_0p_3 + \frac{1}{2}p_1p_2 = p_0 - \frac{\Delta}{2},$$

где  $\Delta = p_0p_3 - p_1p_2$ . Вычислим все вероятности распределения:

$$p_0^* = p_0 - \frac{\Delta}{2}, \quad p_1^* = p_1 + \frac{\Delta}{2}, \quad p_2^* = p_2 + \frac{\Delta}{2}, \quad p_3^* = p_3 - \frac{\Delta}{2}. \quad (2.13)$$

Таблица 2.1.

Генотипы отца	Генотипы матери			
	00	10	01	11
00	00	00,10	00,01	00,01, 10,11
10	00,10	10	00,01, 10,11	10,11
01	00,01	00,01 10,11	01	01,11
11	00,01 10,11	10,11	11,01	11

Заметим, что распределение компонент останется тем же.

$$q_1^* = P\{X_1 = 0\}^* = p_0^* + p_2^* = p_0 + p_2,$$

$$q_2^* = P\{X_2 = 0\}^* = p_0^* + p_1^* = p_0 + p_1.$$

Интересующая нас марковская цепь с состояниями на  $E_k^2$  имеет сложную структуру переходной матрицы, зависящей, с одной стороны, от вектора распределения на предыдущем шаге, что означает

квадратичную нелинейность марковской цепи. С другой стороны, переходная матрица зависит от параметров  $(\alpha_1, \dots, \alpha_k)$ , которые были названы в соответствии с приложением этой задачи в генетике *параметрами предпочтения*, поскольку имеют смысл вероятности появления  $i$ -го аллеля,  $0 \leq \alpha_i \leq 1$ . Пусть

$$A(x_i) = \begin{bmatrix} 1 & \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} \\ \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} & 0 \end{bmatrix}, \quad (2.14)$$

локальная матрица предпочтений, соответствующая координате  $x_i$  точки  $m = m(x_1, \dots, x_k)$ , где  $\bar{\alpha}_i = 1 - \alpha_i$ . С помощью прямого внешнего произведения объединим локальные матрицы в одну матрицу предпочтений вида

$$A_m = A(x_1) \otimes A(x_2) \otimes \dots \otimes A(x_k) = \bigotimes_{j=1}^k A(x_k). \quad (2.15)$$

Пусть  $P^{(N)} = (p_0^{(N)}, \dots, p_n^{(N)})^T$  — распределение на  $N$ -м шаге марковской цепи,  $P(m)^{(N)}$  — его подстановка вида (2.5),  $n = 2^k - 1$ . Предметом нашего исследования является марковская цепь, у которой вероятность состояния  $m = m(x_1, \dots, x_k)$  на  $(N+1)$ -м шаге определяется по формуле

$$p_m^{(N+1)} = \left( P(m)^{(N)} \right)^T A_m P(m)^{(N)}. \quad (2.16)$$

Выясним, как вероятность  $p_m^{(N+1)}$  выражается через параметры перераспределения. При помощи матриц  $C_0$  и  $C$  из леммы 7 введем матрицы

$$B_m = (C^{-1})^T A_m C^{-1} = \bigotimes_{j=1}^k B(x_k), \quad \text{где} \quad (2.17)$$

$$B(x_i) = (C_0^{-1})^T A(x_i) C_0^{-1} = \begin{bmatrix} 0 & \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} \\ 1 - \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} & 1 - 2\bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} \end{bmatrix}.$$

При отсутствии предпочтения, то есть при  $\alpha_i = 0.5$ , имеем

$$A = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}.$$

**Лемма 8.** Пусть имеются  $A_m$  из (2.15) и  $B_m$  из (2.17). Тогда

$$P^T A_m P = Q^T B_m Q. \quad (2.18)$$

### 2.2.5. Индексный бином

На примере случая  $k = 3$  рассмотрим, как могут быть преобразованы компоненты вектора распределения на первом шаге марковской цепи (2.16) и (2.18).

$$\begin{aligned} p_0^* &= Q^T B Q = \frac{q_0 q_{12} + q_1 q_2 + q_2 q_1 + q_{12} q_0}{4} = \frac{q_{12} + q_1 q_2}{2}, \\ \frac{q_0 q_{123} + q_1 q_{23} + q_2 q_{13} + q_{12} q_3 + q_3 q_{12} + q_{13} q_2 + q_{23} q_1 + q_{123} q_0}{8} &= \\ &= \frac{q_0 q_{123} + q_1 q_{23} + q_2 q_{13} + q_3 q_{12}}{4}. \end{aligned}$$

**Определение 11.** Пусть  $q_\tau(m)$  — компоненты (2.9) вектора перераспределения (2.8),  $\tau \subseteq M = \{1, 2, \dots, k\}$ ,  $\bar{\tau}$  его дополнение. *Индексным биномом* будем называть выражение вида

$$\mathcal{B}(m) = \frac{1}{2^k} \sum_{\tau \subseteq M} q_\tau(m) q_{\bar{\tau}}(m). \quad (2.19)$$

При необходимости указания  $N$ -го шага марковской цепи будем использовать обозначение  $\mathcal{B}(m)^{(N)}$ . Так как при  $\alpha_i = 0.5$  матрицы  $A_m$  из (2.15) одинаковы для любого индекса состояния  $m$  из (2.4), т.е.  $A_m = A$ , а матрица  $B_m$  из (2.17) имеет на побочной диагонали константы  $2^{-k}$ , остальные ее элементы равны нулю, из (2.10), (2.16), (2.18) и структуры матрицы  $B_m$  получаем

$$\begin{aligned} q_M(m)^{(N+1)} &= p_m^{(N+1)} = \left(P_m^{(N)}\right)^T A P_m^{(N)} = \\ &= \left(Q(m)^{(N)}\right)^T B_m Q(m)^{(N)} = \mathcal{B}(m)^{(N)}. \end{aligned}$$

Через  $\tilde{\mathcal{B}}(m)$  обозначим усеченный индексный бином, не содержащий в качестве  $\tau$  все множество  $M$  и не имеющий одинаковых слагаемых,

$$\tilde{\mathcal{B}}(m) = \frac{1}{2^{k-1}} \sum_{\tau \subseteq M, \{1\} \in \tau, \tau \neq M} q_\tau(m) q_{\bar{\tau}}(m).$$

Для  $\mathcal{B} = \mathcal{B}(m)$ ,  $\tilde{\mathcal{B}} = \tilde{\mathcal{B}}(m)$ ,  $q_\tau = q_\tau(m)$  и  $q_{\bar{\tau}} = q_{\bar{\tau}}(m)$  имеет место следующее выражение:

$$\mathcal{B} = \frac{1}{2^k} \sum_{\tau \subseteq M} q_\tau q_{\bar{\tau}} = \frac{1}{2^{k-1}} \sum_{\tau \subseteq M, \{1\} \in \tau} q_\tau q_{\bar{\tau}} = \frac{q_M}{2^{k-1}} + \tilde{\mathcal{B}}. \quad (2.20)$$

### 2.2.6. Инвариантность распределения компонент вектора

Пусть  $\mathcal{K}_i = \mathcal{K}_i(x_i)$  — это множество из  $2^{k-1}$  индексов  $m = x_1 + 2x_2 + \dots + 2^{k-1}x_k$ , соответствующих значениям в точках  $(X_1, \dots, X_k)$  аффинной геометрии  $E_k^2$  при компоненте  $X_i = x_i \in \{0, 1\}$ . Для краткости обозначений пусть  $x_i = 0$  и  $\mathcal{K}_i = \mathcal{K}_i(0)$ . Очевидно,

$$P\{X_i = 0\} = \sum_{m \in \mathcal{K}_i} p_m. \quad (2.21)$$

**Лемма 9.** Для любого  $\tau \subseteq \{1, 2, \dots, k\}$  верно

$$\sum_{m \in \mathcal{K}_i} q_\tau(m) q_{\bar{\tau}}(m) = q_i = P\{X_i = 0\}.$$

**Доказательство.** Не умаляя общности, считаем  $i = 1$ . Обозначим через  $P(1, \dots, t) = P\{X_1 = 0, \dots, X_t = x_t\}$  и  $P(t+1, \dots, k) = P\{X_{t+1} = x_{t+1}, \dots, X_k = x_k\}$ .

$$\begin{aligned} \sum_{m \in \mathcal{K}_1} q_\tau(m) q_{\bar{\tau}}(m) &= \sum_{m \in \mathcal{K}_1} P(1, \dots, t) P(t+1, \dots, k) = \\ &= \sum_{x_2=0}^1 \dots \sum_{x_k=0}^1 P(1, \dots, t) P(t+1, \dots, k) = P\{X_1 = 0\}. \quad \square \end{aligned}$$

**Теорема 15.** Справедливо равенство  $q_i^{(n+1)}(m) = q_i^{(n)}(m)$ .

**Доказательство.** Пусть  $M_k = \{1, 2, \dots, k\}$ . Из (2.21) получаем

$$\begin{aligned} q_i^{(n+1)}(m) &= P\{X_i^{(n+1)} = x_i\} = \\ &= \sum_{m \in K_i} p_m^{(n+1)} = \sum_{m \in K_i} P^{(n)}(m)^T A_k P^{(n)}(m) = \\ &= \sum_{m \in K_i} \frac{1}{2^k} \sum_{\tau \subseteq M_k} q_\tau(m)^{(n)} q_{\bar{\tau}}(m)^{(n)} = \\ &= \frac{1}{2^k} \sum_{\tau \subseteq M_k} \sum_{m \in K_i} q_\tau(m)^{(n)} q_{\bar{\tau}}(m)^{(n)} = q_i(m)^{(n)}. \quad \square \end{aligned}$$

## 2.2.7. Стационарное распределение при отсутствии предпочтения

**Теорема 16.** *При всех параметрах предпочтения  $\alpha_i = 0.5$  стационарное распределение марковской цепи с переходными уравнениями (2.16) имеет вид распределения случайного вектора  $X$  с независимыми компонентами.*

$$\lim_{N \rightarrow \infty} p_m^{(N)} = \lim_{N \rightarrow \infty} q_{12\dots k}^{(N)}(m) = q_1(m)q_2(m) \dots q_k(m), \quad (2.22)$$

где  $q_j(m) = P\{X_j = x_j\}$ ,  $j = 1, \dots, k$ ,  $m = 0, 1, \dots, 2^k - 1$ .

**Доказательство.** Обозначим через  $q_\tau = q_\tau(m)$ ,  $q_{\bar{\tau}} = q_{\bar{\tau}}(m)$  и воспользуемся индукцией. При  $k = 2$  из (2.20) и (2.20) с учетом теор. 15 получаем

$$q_{12}^{(N)} = \frac{1}{2} \left( q_{12}^{(N-1)} + q_1 q_2 \right). \quad (2.23)$$

Переходя к пределам по  $N$  в левой и правой части уравнения (2.23), получаем  $\lim_{N \rightarrow \infty} q_{12}^{(N)} = q_1 q_2$ . Далее по индукции. Пусть для любого  $k < n$  справедливо утверждение

$$\lim_{N \rightarrow \infty} q_{12\dots k}^{(N)}(m) = q_1(m)q_2(m) \dots q_k(m). \quad (2.24)$$

Покажем, что  $\lim_{N \rightarrow \infty} q_{12\dots n}^{(N)}(m) = q_1(m)q_2(m) \dots q_n(m)$ . Согласно (2.19) и (2.20), при  $k = n$  и  $M = (12\dots n)$  имеем

$$q_M^{(N)} = \frac{q_M^{(N-1)}}{2^{n-1}} + \frac{1}{2^{n-1}} \sum_{\tau \subseteq M, \{1\} \in \tau, \tau \neq M} q_\tau^{(N-1)} q_{\bar{\tau}}^{(N-1)}. \quad (2.25)$$

Обозначим через  $L = \lim_{N \rightarrow \infty} q_M^{(N)}$  и перейдем к пределам по  $N$  в обеих частях уравнения (2.25).

$$L = \frac{L}{2^{n-1}} + \frac{1}{2^{n-1}} (2^{n-1} - 1) q_1 q_2 \dots q_n,$$

так как  $\lim_{N \rightarrow \infty} q_\tau^{(N-1)} q_{\bar{\tau}}^{(N-1)} = q_1 q_2 \dots q_n$  по индукционному предположению, а количество слагаемых в сумме по  $\tau$  равно  $2^{n-1} - 1$ . Отсюда получаем, что  $L = q_1 q_2 \dots q_n$ , и утверждение доказано.

### 2.2.8. Распределение при частичной независимости

**Теорема 17.** Пусть вектор  $X = (X_1, \dots, X_k, X_{k+1}, \dots, X_{2k})^T$  состоит из компонент независимых случайных векторов

$$X' = (X_1, \dots, X_k)^T \quad \text{и} \quad X'' = (X_{k+1}, \dots, X_{2k})^T,$$

через  $P = P^{(N)}$ ,  $P_{X'} = P_{X'}^{(N)}$  и  $P_{X''} = P_{X''}^{(N)}$  обозначены распределения векторов  $X$ ,  $X'$  и  $X''$  на  $N$ -м шаге, через  $p_m, p'_{m_1}, p''_{m_2}$  соответствующие вероятности состояний векторов  $X$ ,  $X'$  и  $X''$ , где  $m_1 = x_1 + 2x_2 + \dots + 2^{k-1}x_k$ ,  $m_2 = x_{k+1} + 2x_{k+2} + \dots + 2^{k-1}x_{2k}$  и  $m = m_1 + 2^k m_2$ . Тогда 1)  $P = P_{X'} \otimes P_{X''}$ ; 2)  $p_m = p'_{m_1} p''_{m_2}$ .

**Доказательство.** 1) Проиллюстрируем это утверждение в частном случае, который естественным образом может быть распространен на большие размерности. Если распределения независимых векторов  $X_1$  и  $X_2$  имеют вид

$$P(X_1) = (q_1, p_1), \quad P(X_2) = (q_2, p_2),$$

то для совместного распределения справедливо

$$P(X_1, X_2) = (q_1 q_2, p_1 q_2, q_1 p_2, p_1 p_2) = (q_1, p_1) \otimes (q_2, p_2).$$

2) Воспользуемся формулой (2.16) и получим

$$p'_{m_1} = P_{X'}^T A_{m_1} P_{X'}, \quad p''_{m_2} = P_{X''}^T A_{m_2} P_{X''}.$$

Согласно (2.15), матрица  $A_m$  имеет вид  $A_m = A_{m_1} \otimes A_{m_2}$ . Отсюда

$$\begin{aligned} p_m &= P_X^T A_m P_X = (P_{X'} \otimes P_{X''})^T (A_{m_1} \otimes A_{m_2}) (P_{X'} \otimes P_{X''}) = \\ &= (P_{X'}^T A_{m_1} P_{X'}) \cdot (P_{X''}^T A_{m_2} P_{X''}) = p'_{m_1} p''_{m_2}, \end{aligned}$$

так как по одному из свойств внешнего умножения для любых матриц  $(A \otimes B)(C \otimes D) = AC \otimes BD$ , а для матриц единичного порядка внешнее умножение совпадает с обычным.  $\square$

### 2.2.9. Обобщенная стохастическая модель закона Харди—Вайнберга

В основе данной стохастической модели эволюции распределений генотипов лежит, во-первых, представление о конечно-геометрической структуре генотипов, во-вторых, необязательным является требование независимости распределений генов в гаметах, в третьих, вводится параметрическое описание нарушения равновесия популяции. Генотип, состоящий из  $k$  локусов, представляется в виде вектора

$$(X_1, Y_1, X_2, Y_2, \dots, X_k, Y_k)^T \quad (2.26)$$

размерности  $2k$ , где компоненты  $X_i$  и  $Y_j$  локуса  $(X_i, Y_i)$  независимы.

ПРИМЕР. Проиллюстрируем независимость аллелей в локусе на примере распределении генотипов  $AA, AG, GG$  ( $n_0 = 83, n_1 = 139$  и  $n_2 = 59$  индивидов соответственно) гена фермента катаболизма моноаминов катехол-орто-метилтрансферазы *COMT*. Оценка вероятности успеха (аллель  $G$ ) максимального правдоподобия биномиального распределения для двух независимых испытаний имеет вид  $\hat{p} = \frac{n_0 + n_1/2}{n_0 + n_1 + n_2} = 0.5427, \hat{q} = 1 - \hat{p} = 0.4572$ , значимость статистики согласия  $\chi^2$  равна  $P = 0.95$ . Оценки абсолютных частот генотипов практически совпадают с наблюдаемыми частотами и равны соответственно 82.76, 139.48, 58.76.

Каждая из компонент принимает  $r$  значений ( $r$  аллелей). Это структура конечной аффинной геометрии  $E_{2k}^r$ . Случай характеристики  $r = 2$  описан нами подробно.

При  $r > 2$  некоторым изменениям подвергается переходная матрица, но структура вектора перераспределения остается той же. Компонентам  $(X_1, X_2, \dots, X_k)^T$  и отдельно  $(Y_1, Y_2, \dots, Y_k)^T$  из вектора (2.26) соответствуют в генетике понятия гамет. Из теоремы 17 заключаем, что для получения вида стационарного распределения достаточно рассматривать динамику распределений гамет. Эволюция распределений генотипов при панмиксии описывается при помощи нелинейной марковской цепи над конечной аффинной геометрией с уравнениями перехода вида (2.16).

Рассмотрим процедуру построения переходных уравнений (2.16) на примере трех ( $r = 3$ ) аллелей и двух ( $k = 2$ ) гамет. Гаметы имеют структуру  $(X_1, X_2)^T$  со значениями компонент 0,1,2 и кодируются элементами поля  $F_9$  в виде  $m = m(x_1, x_2) = 3x_2 + x_1 \in \{0, 1, \dots, 8\}$ . Варианты гамет с тремя аллелями у потомков в зависимости от



Таблица 2.2.

$x_1$	0	1	2	0	1	2	0	1	2
$x_2$	0	0	0	1	1	1	2	2	2
$m$	0	1	2	3	4	5	6	7	8
0	0	0,1	0,2	0,3	0,3 1,4	0,3 2,5	0,6	0,6 1,7	0,6 2,8
1	0,1	1	1,2	1,4 0,3	1,4	1,4 2,5	1,7 0,6	1,7	1,7 2,8
2	0,2	1,2	2	2,5 0,3	2,5	2,5	2,8 0,6	2,8 1,7	2,8
3	0,3	0,3 1,4	0,3 2,5	3	3,4	3,5	3,6	3,6 4,7	3,6 5,8
4	1,4 0,3	1,4	1,4 2,5	3,4	4	4,5	4,7 3,6	4,7	4,7 5,8
5	2,5 0,3	2,5 1,4	2,5	3,5	4,5	5	5,8 3,6	5,8 4,7	5,8
6	0,6	0,6 1,7	0,6 2,8	3,6	3,6 4,7	3,6 5,8	6	6,7	6,8
7	1,7 0,6	1,7	1,7 2,8	4,7 3,6	4,7	4,7 5,8	6,7	7	7,8
8	2,8 0,6	2,8 1,7	2,8	5,8 3,6	5,8	5,8	6,8	7,8	8

гамет родителей при панмиксии представлены в табл. 2.2. Например, при скрещивании носителей гамет 0 и 4 возможно появление одного из четырех видов гамет: 0, 1, 3 и 4.

При отсутствии предпочтения в новом поколении компонента  $p_0^*$  вектора распределения по формуле полной вероятности имеет вид

$$\begin{aligned}
 p_0^* &= p_0^2 + p_0p_1 + p_0p_2 + p_0p_3 + p_0p_6 + \\
 &+ \frac{1}{2} \left( p_0p_4 + p_0p_5 + p_0p_7 + p_0p_8 + p_1p_3 + p_2p_3 + p_1p_6 + p_2p_6 \right) = \\
 &= \frac{1}{2} \left( p_0 + (p_0 + p_1 + p_2)(p_0 + p_3 + p_6) \right),
 \end{aligned}$$

для остальных, складывая в индексы над  $F_9$ , получаем

$$\begin{aligned} p_m^* &= \frac{1}{2} \left( p_m + (p_m + p_{m\oplus 1} + p_{m\oplus 2})(p_m + p_{m\oplus 3} + p_{m\oplus 6}) \right) = \\ &= \frac{1}{2} \left( P\{X_1 = x_1, X_2 = x_2\} + P\{X_2 = x_2\}P\{X_1 = x_1\} \right). \end{aligned}$$

Итак, при увеличении числа аллелей  $r$  локальная матрица предпочтений отличается от (2.14)  $(r - 1)$ -кратным повторением второй строки и второго столбца

$$A(x_i) = \begin{bmatrix} 1 & \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} & \dots & \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} \\ \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} & 0 & \dots & 0 \\ & & \dots & \\ \bar{\alpha}_i^{x_i} \alpha_i^{1-x_i} & 0 & \dots & 0 \end{bmatrix}.$$

Общая матрица предпочтений также как в (2.15) является внешним произведением локальных матриц.

Вектор перераспределения  $Q(m)$  имеет тот же вид, что и в случае характеристики  $r = 2$  в (2.8). Матрица  $C$  преобразования вектора распределения в вектор перераспределения  $Q(m) = CP(m)$  (лемма 7) также является  $k$ -кратным внешним произведением матриц  $C_0$ , в которых при увеличении характеристики увеличивается количество столбцов, а именно  $(r - 1)$  раз повторяется второй столбец. В случае  $r = 3$  при  $\alpha_i = 0.5$  для любого  $m = 3x_2 + x_1$  получаем

$$A_m = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \end{bmatrix}, \quad C_0 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Переходные уравнения (2.18) остаются без изменений, следовательно, при  $r > 2$  справедливы теоремы 16 и 17. Заметим, что при увеличении числа аллелей в аффинной геометрии, описывающей распределение генотипов и гамет, характеристика увеличивается, а в проективной, описывающей вектор перераспределения, характеристика два сохраняется.

Таким образом, этот метод позволяет получить частоту генотипа в любом поколении при любом количестве аллелей при произвольном начальном распределении и исследовать нарушение равновесия при панмиксии через параметры предпочтения.

### 2.3. Эргодический метод компенсации пропусков

Повторные выборки в биометрии относятся к так называемым лонгитюдным данным. Это могут быть результаты исследований в течение длительного времени, краткосрочные наблюдения за эффектом лечения или действия фармакологического препарата.

Для статистического анализа такого рода данных используются планы с расщепленными блоками [29], в которых проверка гипотез осуществляется только при отсутствии пропусков. Но на практике не всегда удается получить данные от каждого индивида во всех временных точках, поскольку индивиды имеют склонность по разным причинам выбывать из эксперимента.

Большое количество работ по биостатистике посвящено проблеме искусственной замены пропущенных наблюдений, влекущей за собой, в основном, необоснованное увеличение числа степеней свободы. Наиболее эффективным [41] признается LOCF метод (The Last Observation Carried Forward method) замены пропусков последним имеющимся наблюдением. Однако при наличии тренда параметры оцениваются с большим смещением.

В данной работе речь идет об обобщении метода Repeated Measures ANOVA, основанного на выделении не зависящих от фактора времени (расщепленных блоков) компонент наблюдений. В случае полных данных в качестве последних рассматриваются индивидуальные средние. Если данные неполные, то при наличии тренда на индивидуальное среднее оказывает влияние то, в какие моменты времени у индивида имелись наблюдения. Чтобы исключить это влияние, в индивидуальные средние вводится поправка на смещение в виде бесконечной суммы величин, полученных в результате определяемого далее *перекрестного усреднения* разностей между наблюдениями и индивидуальными средними. Необходимым условием достижения несмещенности при улучшении эргодичности<sup>3</sup> является наличие полных данных хотя бы в одной временной точке, например, в первой или последней, что для реальных экспериментов вполне допустимо.

---

<sup>3</sup>Под улучшением эргодичности в данном случае понимается уменьшение расстояния между ожидаемыми значениями среднего по времени и среднего по повторностям. В случае полных данных это расстояние равно нулю.

### 2.3.1. Планы с расщепленными блоками в условиях неполных данных

В статистическом анализе клинических данных необходимость использования планов с расщепленными блоками диктуется условиями проводимого эксперимента.

ПРИМЕР 1. При исследовании эффективности фармакологической помощи в выполнении программы отказа от употребления наркотиков участвующим в программе индивидам случайным образом назначались следующие препараты: блокатор опиатных рецепторов налтрексон, антидепрессант фремакс или плацебо в разных комбинациях. Каждые две недели больные должны были являться для контроля за их состоянием на сеанс психотерапии. Однако графики их посещения оказались чрезвычайно разнообразными — кто-то сразу перестал появляться, кто-то продержался месяц или два, у кого-то оказался длительный перерыв. Необходимо было исследовать то, каким образом препараты влияют на динамику физиологических и психологических характеристик (депрессия, тревожность, индекс глобального функционирования и так далее). При использовании стандартных методов дисперсионного анализа с расщепленными блоками в статистическую обработку попадают данные только тех индивидов, у которых есть наблюдения во всех точках. При этом выборка не только сокращается, но и перестает быть презентабельной, поскольку менее благополучные больные с большим количеством отказов в ней не оказываются.

Однофакторная модель дисперсионного анализа для зависимых данных с фиксированными эффектами „препарата“  $A$  и „времени“  $B$  имеет вид

$$x_{ijt} = \mu + \alpha_i + e_{ij}^1 + \beta_t + \gamma_{it} + e_{ijt}. \quad (2.27)$$

При этом используются следующие обозначения:  $x_{ijt}$  — наблюдение в  $i$ -й группе ( $i = 1, \dots, I$ ) в  $t$ -й момент времени ( $t = 1, \dots, T$ ) индивида  $j = 1, \dots, \nu_i$ , где  $\nu_i$  — число индивидов  $i$ -й группы в самый представительный момент времени,  $n = \sum_{i=1}^I \nu_i$ ;  $\mu$  — генеральное среднее;  $\alpha_i$ ,  $\beta_t$  и  $\gamma_{it}$  — фиксированные эффекты факторов  $A$ ,  $B$  и их взаимодействия;  $e_{ij}^1$  — ошибка наблюдений, вызванная разнообразием индивидов;  $e_{ijt}$  — общая ошибка наблюдений. Ошибки  $e_{ij}^1$  и  $e_{ijt}$  предполагаются независимыми, нормально распределенными с  $\mathbf{E}e_{ijt}^1 = \mathbf{E}e_{ijt} = 0$  и с дисперсиями  $\sigma_1^2$  и  $\sigma^2$  соответственно.

Пусть  $m_{it}$  — это количество индивидов в  $i$ -й группе и в  $t$ -й точке. У каждого  $j$ -го индивида из  $i$ -й группы имеется множество  $N_{ij}$  наблюдений,  $\text{card}(N_{ij}) = n_{ij} \leq T$ . В  $t$ -й точке и в  $i$ -й группе имеется

множество  $M_{it}$  наблюдений,  $\text{card}(M_{it}) = m_{it}$ . Число наблюдений в  $t$ -й и во всех точках равны соответственно

$$m_{\cdot t} = \sum_{i=1}^I m_{it}, \quad m_{\cdot\cdot} = \sum_{i=1}^I \sum_{t=1}^T m_{it}.$$

В дальнейшем будем использовать следующие обозначения для средних: общего

$$x_{\dots} = \frac{1}{m_{\cdot\cdot}} \sum_{i=1}^I \sum_{t=1}^T \sum_{j \in M_{it}} x_{ijt} = \frac{1}{m_{\cdot\cdot}} \sum_{i=1}^I \sum_{j=1}^{\nu_i} n_{ij} x_{ij\cdot}, \quad (2.28)$$

индивидуального и внутригруппового

$$x_{ij\cdot} = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} x_{ijt}, \quad x_{i\cdot\cdot} = \frac{1}{m_i} \sum_{j=1}^{\nu_i} n_{ij} x_{ij\cdot}, \quad (2.29)$$

в момент  $t$  и для  $i$ -й группы в момент  $t$  соответственно

$$x_{\cdot\cdot t} = \frac{1}{m_{\cdot t}} \sum_{i=1}^I m_{it} x_{i\cdot t}, \quad \text{и} \quad x_{i\cdot t} = \frac{1}{m_{it}} \sum_{j \in M_{it}} x_{ijt}. \quad (2.30)$$

Для оценки параметров модели (2.27) наблюдения представляются в виде суммы двух компонент  $x_{ijt} = z_{ij} + y_{ijt}$  с  $\mathbf{E}z_{ij} = \mu + \alpha_i$ ,  $\mathbf{E}y_{ijt} = \beta_t + \gamma_{it}$ . Параметры оцениваются по методу наименьших квадратов (МНК). Для единственности решения систем нормальных уравнений предполагается, что дифференциальные эффекты обладают следующими свойствами:

$$\sum_{i=1}^I u_i \alpha_i = 0, \quad \sum_{t=1}^T v_t \beta_t = 0, \quad \text{где} \quad \sum_{i=1}^I u_i = 1, \quad \sum_{t=1}^T v_t = 1, \\ \sum_{i=1}^I w_{it} \gamma_{it} = 0, \quad \sum_{t=1}^T w_{it} \gamma_{it} = 0, \quad t = 1, \dots, T, \quad i = 1, \dots, I, \quad (2.31)$$

$$\text{где} \quad \sum_{t=1}^T \sum_{i=1}^I w_{it} = 1, \quad u_i = \frac{m_i}{m_{\cdot\cdot}}, \quad v_t = \frac{m_{\cdot t}}{m_{\cdot\cdot}}, \quad w_{it} = \frac{m_{it}}{m_{\cdot\cdot}}.$$

В условиях полных данных в качестве  $z_{ij} = x_{ij\cdot}$  берутся индивидуальные средние. При наличии пропусков в индивидуальные средние нужно вводить такое смещение  $H_{ij}$ , чтобы  $\mathbf{E}(x_{ijt} - z_{ij}) = \beta_t + \gamma_{it}$ , где  $z_{ij} = x_{ij\cdot} - H_{ij}$ .

### 2.3.2. Проблема устранения индивидуального влияния

**Определение 12.** Для любого  $j = 1, \dots, \nu_i$  из группы  $i = 1, \dots, I$ , где через  $\nu_i$  обозначено число индивидов в  $i$ -й группе, зададим последовательность  $\{A_{ij}(k)\}_{k=1}^{\infty}$  *перекрестных усреднений*, где

$$\begin{aligned} A_{ij}(1) &= \frac{1}{n_{ij}} \sum_{\tau \in N_{ij}} \frac{1}{m_{i\tau}} \sum_{l \in M_{i\tau}} (x_{il\tau} - x_{il.}), \\ A_{ij}(k+1) &= \frac{1}{n_{ij}} \sum_{\tau \in N_{ij}} \frac{1}{m_{i\tau}} \sum_{l \in M_{i\tau}} A_{il}(k). \end{aligned} \quad (2.32)$$

Назовем *индивидуальным смещением* ряд, состоящий из элементов последовательности перекрестных усреднений, вида

$$H_{ij} = \sum_{k=1}^{\infty} A_{ij}(k) \quad (2.33)$$

и используем его для построения несмещенных оценок. Таким образом, не зависящая от времени компонента приобретает вид

$$z_{ij} = x_{ij.} - \sum_{k=1}^{\infty} A_{ij}(k). \quad (2.34)$$

Для проверки  $\mathbf{E}z_{ij} = \mu + \alpha_i$  используем эргодические свойства последовательности перекрестных усреднений.

### 2.3.3. Эргодичность перекрестного усреднения

**Определение 13.** Матрица инцидентности полноты наблюдений  $J = J(n, T)$  имеет вид  $J = [J_1^T, \dots, J_I^T]^T$ , где  $J_i$  — локальная матрица инцидентности размерности  $\nu_i \times T$ ,  $i = 1, \dots, I$ , с относящимися к  $i$ -й группе элементами

$$J_i(j, t) = \begin{cases} 1, & \text{если } x_{ijt} \text{ — полное наблюдение,} \\ 0, & \text{если } x_{ijt} \text{ — пропуск.} \end{cases}$$

Соберем  $A_{ij}(k)$  из (2.32) в векторе  $A_i(k) = (A_{i1}(k), \dots, A_{i\nu_i}(k))^T$  и пусть

$$A(k) = (A_1(k)^T, \dots, A_I(k)^T)^T. \quad (2.35)$$

Обозначим через  $\Lambda_{\nu_i}$ ,  $i = 1, \dots, I$ , диагональную матрицу порядка  $\nu_i$  с элементами на диагонали  $\frac{1}{n_{ij}}$ , где  $n_{ij}$  — количество полных наблюдений у  $j$ -го индивида,  $j = 1, \dots, \nu_i$ , из  $i$ -й группы. Через  $\Lambda_{iT}$  обозначим диагональную матрицу порядка  $T$  с элементами на диагонали  $\frac{1}{m_{it}}$ , где  $m_{it}$  — количество полных наблюдений в  $i$ -й группе и  $t$ -й момент времени. Далее построим стохастические матрицы вида

$$P(i) = \Lambda_{\nu_i} J_i \Lambda_{iT} J_i^T$$

и единую стохастическую матрицу

$$P = \begin{bmatrix} P(1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & P(2) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & P(I) \end{bmatrix}.$$

Несложно убедиться в том, что рекуррентное уравнение (2.32) может быть записано в матричном виде

$$\begin{aligned} A_i(k+1) &= P(i)A_i(k) = P(i)^k A_i(1); \\ A(k+1) &= PA(k) = P^k A(1). \end{aligned} \quad (2.36)$$

Предположение о наличии полных наблюдений хотя бы в один из моментов времени соответствует тому, что все элементы матрицы  $P(i)$  положительны. Предельная матрица  $\mathcal{P}_i = \lim_{k \rightarrow \infty} P(i)^k$  в этом случае имеет одинаковые строки  $\pi_i = (\pi_{i1}, \dots, \pi_{i\nu_i})^T$ , соответствующие левому собственному вектору, такому что  $\pi_i^T \mathcal{P}_i = \pi_i^T$ . Непосредственным умножением вектора  $\pi_i$  на  $P(i)$  можно убедиться в том, что

$$\pi_{ij} = \frac{n_{ij}}{m_i}. \quad (2.37)$$

Общий стохастический стационарный вектор имеет вид

$$\pi = \left( \frac{m_1 \cdot \pi_1}{m_{..}}, \dots, \frac{m_I \cdot \pi_I}{m_{..}} \right)^T.$$

### 2.3.4. Ошибки для компонент наблюдений $z_{ij}$ и $y_{ijt}$

Неполнота данных приводит к тому, что в зависимости от количества наблюдений у индивидов могут быть получены разные внутригрупповые средние ошибок

$$e_{i..}(j) = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} e_{i..t},$$

где  $e_{i..t} = \frac{1}{m_{it}} \sum_{l \in M_{it}} e_{ilt}$ . Обозначим через

$$\delta_{ij} = e_{ij.} - e_{i..}(j) \quad (2.38)$$

разность между средней индивидуальной ошибкой

$$e_{ij.} = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} e_{ij.t}$$

и соответствующей внутригрупповой средней ошибкой.

**Теорема 18.** Пусть  $y_{ijt} = x_{ijt} - x_{ij.} + H_{ij}$ ,  $z_{ij} = x_{ij.} - H_{ij}$ , где смещение  $H_{ij}$  определено в (2.33), квадратная матрица  $Q(i)$  порядка  $\nu_i$  имеет вид

$$Q(i) = \lim_{N \rightarrow \infty} \sum_{k=0}^N P^k(i), \quad (2.39)$$

модифицированные ошибки имеют вид:

$$\begin{aligned} \mathcal{E}_{ijt} &= e_{ij.t} - e_{i..} - \sum_{j_0=1}^{\nu_i} Q_{jj_0}(i) \delta_{ij_0}, \\ \mathcal{E}_{ij}^1 &= e_{ij.}^1 + e_{i..} + \sum_{j_0=1}^{\nu_i} Q_{jj_0}(i) \delta_{ij_0}. \end{aligned} \quad (2.40)$$

Тогда в предположении о том, что имеются полные данные хотя бы в один момент времени, справедливы выражения

$$z_{ij} = \mu + \alpha_i + \mathcal{E}_{ij}^1 \quad \text{и} \quad y_{ijt} = \beta_t + \gamma_{it} + \mathcal{E}_{ijt}. \quad (2.41)$$



**Доказательство.** Как уже отмечалось, предположение о наличии хотя бы одного столбца из единиц в матрице инцидентности  $J$  необходимо для существования стационарного распределения. Обозначим через

$$\beta_{\cdot}(i, j) = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \beta_t, \quad \text{и} \quad \gamma_{i\cdot}(j) = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \gamma_{it}.$$

Из модели (2.27) получаем, что

$$x_{ijt} - x_{ij\cdot} = \beta_t + \gamma_{it} + e_{ijt} - (\beta_{\cdot}(i, j) + \gamma_{i\cdot}(j) + e_{ij\cdot}), \quad (2.42)$$

откуда появляется возможность разбить  $A_{ij}(k)$  и  $y_{ijt}$  на три слагаемые  $A_{ij}(k) = A_{ij}^{\beta}(k) + A_{ij}^{\gamma}(k) + A_{ij}^e(k)$  и  $y_{ijt} = y_{ijt}^{\beta} + y_{ijt}^{\gamma} + y_{ijt}^e$ , относящиеся соответственно к компонентам  $\beta_t$ ,  $\gamma_{it}$  и  $e_{ijt}$ . Рассмотрим первое слагаемое:

$$\begin{aligned} A_{ij}^{\beta}(1) &= \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \frac{1}{m_{it_0}} \sum_{j_0 \in M_{it_0}} (\beta_{t_0} - \beta_{\cdot}(i, j_0)) = \\ &= \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \beta_{t_0} - \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \frac{1}{m_{it_0}} \sum_{j_0 \in M_{it_0}} \frac{1}{n_{ij_0}} \sum_{t_1 \in N_{ij_0}} \beta_{t_1}. \end{aligned}$$

Пусть  $\mathcal{N} = (\mathcal{N}(1), \dots, \mathcal{N}(i))$  — матрица размерности  $n$  на  $T$ , состоящая из матриц  $\mathcal{N}(i) = \Lambda_{\nu_i} J$  размерности  $\nu_i$  на  $T$ . Обозначим через  $b = (\beta_1, \dots, \beta_T)^T$ . Непосредственно убеждаемся в том, что

$$A^{\beta}(1) = \mathcal{N}b - P\mathcal{N}b.$$

Следовательно, для любого  $t$  выражению  $y_{ijt}^{\beta} - \beta_t$  соответствует  $(i, j)$ -я компонента построенного согласно правилу (2.35) вектора

$$-\mathcal{N}b + (\mathcal{N}b - P\mathcal{N}b) + (P\mathcal{N}b - P^2\mathcal{N}b) + \dots = -P\mathcal{N}b.$$

Так как  $\pi^T \mathcal{N}b =$

$$= \sum_{i=1}^I \frac{m_{i\cdot}}{m_{\cdot\cdot}} \sum_{j=1}^{\nu_i} \frac{n_{ij}}{m_{i\cdot}} \frac{1}{n_{ij}} \sum_{t \in N_{ij}} b_t = \frac{1}{m_{\cdot\cdot}} \sum_{t=1}^T \sum_{i=1}^I \sum_{j \in M_{it}} \beta_t =$$

$= \frac{1}{m_{..}} \sum_{t=1}^T m_{\cdot t} \beta_t = 0$  по условию ограничения на параметры (2.31),

имеем  $y_{ijt}^\beta = \beta_t$ . Аналогично покажем, что  $y_{ijt}^\gamma = \gamma_t$ . Из компонент

$$\begin{aligned} A_{ij}^\gamma(1) &= \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \frac{1}{m_{it_0}} \sum_{j_0 \in M_{it_0}} (\gamma_{it_0} - \gamma_{i \cdot}(j_0)) = \\ &= \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \gamma_{it_0} - \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \frac{1}{m_{it_0}} \sum_{j_0 \in M_{it_0}} \frac{1}{n_{ij_0}} \sum_{t_1 \in N_{ij_0}} \gamma_{it_1} \end{aligned}$$

построим вектор  $A_i^\gamma(1) = \mathcal{N}(i)c_i - P\mathcal{N}(i)c_i$ , где  $c_i = (\gamma_{i1}, \dots, \gamma_{iT})^T$ . Выражению  $y_{ijt}^\gamma - \gamma_t$  соответствует компонента вектора

$$\begin{aligned} -\mathcal{N}(i)c_i + \mathcal{N}(i)c_i - P(i)\mathcal{N}(i)c_i + P(i)\mathcal{N}(i)c_i - \\ -P(i)^2\mathcal{N}(i)c_i + \dots = -\mathcal{P}_i\mathcal{N}(i)c_i. \end{aligned}$$

Из ограничений на параметры следует, что  $\pi_i^T \mathcal{N}(i)c_i = 0$ , так как

$$\sum_{j=1}^{\nu_i} \frac{n_{ij}}{m_{i \cdot}} \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \gamma_{it} = \frac{1}{m_{i \cdot}} \sum_{t=1}^T \sum_{j \in M_{it}} \gamma_{it} = \frac{1}{m_{i \cdot}} \sum_{t=1}^T m_{it} \gamma_{it}.$$

Несколько сложнее представлена компонента  $y_{ijt}^e$ .

$$A_{ij}^e(1) = \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} e_{i \cdot t_0} - \frac{1}{n_{ij}} \sum_{t_0 \in N_{ij}} \frac{1}{m_{it_0}} \sum_{j_0 \in M_{it_0}} e_{ij_0 \cdot}.$$

Обозначим через  $U_i = \{e_{i \cdot}(j)\}_{j=1}^{\nu_i}$ ,  $V_i = \{e_{ij \cdot}\}_{j=1}^{\nu_i}$  векторы порядка  $\nu_i$ . Тогда  $A_i^e(1) = U_i - P(i)V_i$ . Следовательно,

$$\begin{aligned} -V_i + \sum_{k=1}^{\infty} A_i^e(k) &= \\ = -V_i + U_i - P(i)V_i + P(i)(U_i - P(i)V_i) + \dots &= \\ = Q(i)(U_i - V_i) - \mathcal{P}(i)V_i, \end{aligned}$$

где  $Q(i)$  из (2.39). Из (2.37) все строки стационарной матрицы  $\mathcal{P}(i)$  равны  $\{\pi_{ij}\}_{j=1}^{\nu_i}$ , где  $\pi_{ij} = \frac{n_{ij}}{m_{i \cdot}}$ , откуда  $\pi_i V_i = \frac{1}{m_{i \cdot}} \sum_{j=1}^{\nu_i} n_{ij} e_{ij \cdot} = e_{i \cdot}$ ,

$$y_{ijt}^e = e_{ijt} - e_{i \cdot} - \sum_{j_0=1}^{\nu_i} Q_{jj_0}(i) \delta_{ij_0}. \quad (2.43)$$

Наконец,  $z_{ij} = x_{ijt} - y_{ijt} = e_{ij}^1 + e_{i \cdot} + \sum_{j_0=1}^{\nu_i} Q_{jj_0}(i) \delta_{ij_0}$ .  $\square$

### 2.3.5. Дисперсионные матрицы ошибок

Пусть  $D(i, j; i_1, j_1) = \frac{1}{n_{ij}n_{i_1j_1}} \sum_{t \in N_{ij} \cap N_{i_1j_1}} \frac{1}{m_{it}}$  коэффициент совместного присутствия индивидов  $j$  и  $j_1$ ,  $\mathbf{I}_A$  индикатор множества.

**Лемма 10.** *Справедливы следующие равенства.*

1.  $\mathbf{E}e_{ij} \cdot e_{i_1j_1} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}};$
2.  $\mathbf{E}e_{i..}(j)e_{i_1j_1t} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, t \in N_{ij} \cap N_{i_1j_1}]}}{n_{ij}m_{it}};$
3.  $\mathbf{E}e_{i..}(j)e_{i_1j_1} = \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1);$
4.  $\mathbf{E}e_{i..}(j)e_{i_1..}(j_1) = \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1);$
5.  $\mathbf{E}\delta_{ij}\delta_{i_1j_1} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}} - \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1);$
6.  $\mathbf{E}\delta_{ij}e_{i_1j_1t} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}} - \frac{\sigma^2 \mathbf{I}_{[i=i_1, t \in N_{ij} \cap N_{i_1j_1}]}}{n_{ij}m_{it}};$
7.  $\mathbf{E}e_{i..}\delta_{ij} = 0.$

**Доказательство.** 1.  $\mathbf{E}e_{ij} \cdot e_{i_1j_1} = \frac{1}{n_{i_1j_1}} \sum_{t \in N_{i_1j_1}} \mathbf{E}e_{ij} \cdot e_{i_1j_1t} =$

$$= \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}}.$$

$$\begin{aligned} 2. \mathbf{E}e_{i..}(j)e_{i_1j_1t} &= \frac{\mathbf{I}_{[i=i_1]}}{n_{ij}} \sum_{t_1 \in N_{ij}} \mathbf{E}e_{i \cdot t_1} e_{i_1j_1t} = \\ &= \sum_{t_1 \in N_{ij}} \frac{\mathbf{I}_{[i=i_1]}}{m_{it_1}n_{ij}} \sum_{j_2 \in M_{it_1}} \mathbf{E}e_{ij_2t_1} e_{i_1j_1t} = \sum_{t_1 \in N_{ij}} \frac{\sigma^2 \mathbf{I}_{[i=i_1, j_1 \in M_{it_1}, t_1=t]}}{n_{ij}m_{it_1}} = \\ &= \sum_{t_1 \in N_{ij} \cap N_{i_1j_1}} \frac{\sigma^2 \mathbf{I}_{[i=i_1, t_1=t]}}{n_{ij}m_{it_1}} = \frac{\sigma^2 \mathbf{I}_{[i=i_1, t \in N_{ij} \cap N_{i_1j_1}]}}{n_{ij}m_{it}}. \end{aligned}$$

$$3. \mathbf{E}e_{i..}(j)e_{i_1j_1} = \frac{1}{n_{i_1j_1}} \sum_{t \in N_{i_1j_1}} \mathbf{E}e_{i..}(j)e_{i_1j_1t} =$$

$$= \frac{1}{n_{i_1 j_1}} \sum_{t \in N_{i_1 j_1}} \frac{\sigma^2 \mathbf{I}_{[i=i_1, t \in N_{ij} \cap N_{i_1 j_1}]}}{n_{ij} m_{it}} = \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1).$$

4. Используем то, что  $\mathbf{E} e_{i \cdot t} e_{i_1 \cdot t_1} = \mathbf{I}_{[i=i_1, t=t_1]} \frac{\sigma^2}{m_{it}}$ .

$$\begin{aligned} \mathbf{E} e_{i \cdot \cdot}(j) e_{i_1 \cdot \cdot}(j_1) &= \mathbf{E} \left( \frac{1}{n_{ij}} \sum_{t \in N_{ij}} e_{i \cdot t} \right) \left( \frac{1}{n_{i_1 j_1}} \sum_{t_1 \in N_{i_1 j_1}} e_{i_1 \cdot t_1} \right) = \\ &= \frac{\mathbf{I}_{[i=i_1]}}{n_{ij} n_{i_1 j_1}} \sum_{t \in N_{ij} \cap N_{i_1 j_1}} \frac{\sigma^2}{m_{it}} = \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1). \end{aligned}$$

$$\begin{aligned} 5. \mathbf{E} \delta_{ij} \delta_{i_1 j_1} &= \mathbf{E} (e_{ij \cdot} - e_{i \cdot \cdot}(j)) (e_{i_1 j_1 \cdot} - e_{i_1 \cdot \cdot}(j_1)) = \\ &= \mathbf{E} e_{ij \cdot} e_{i_1 j_1 \cdot} - \mathbf{E} e_{ij \cdot} e_{i_1 \cdot \cdot}(j_1) - \mathbf{E} e_{i \cdot \cdot}(j) e_{i_1 j_1 \cdot} + \mathbf{E} e_{i \cdot \cdot}(j) e_{i_1 \cdot \cdot}(j_1) = \\ &= \frac{\sigma^2 \mathbf{I}_{[i=i_1, j=j_1]}}{n_{ij}} - \sigma^2 \mathbf{I}_{[i=i_1]} D(i, j; i, j_1). \end{aligned}$$

6. Следует из соотношения  $\mathbf{E} \delta_{ij} e_{i_1 j_1 t} = \mathbf{E} (e_{ij \cdot} - e_{i \cdot \cdot}(j)) e_{i_1 j_1 t}$ .

7.  $\mathbf{E} e_{i \cdot \cdot} \delta_{ij} = \mathbf{E} e_{i \cdot \cdot} (e_{ij \cdot} - e_{i \cdot \cdot}(j)) = 0$ , так как, с одной стороны,

$$\mathbf{E} e_{i \cdot \cdot} e_{ij \cdot} = \frac{1}{m_i} \sum_{l=1}^{\nu_i} n_{il} \mathbf{E} e_{il} e_{ij \cdot} = \frac{1}{m_i} \sum_{l=1}^{\nu_i} n_{il} \left( \frac{\sigma^2 \mathbf{I}_{l=j}}{n_{ij}} \right) = \frac{\sigma^2}{m_i}, \text{ с другой стороны,}$$

$$\begin{aligned} \mathbf{E} e_{i \cdot \cdot} e_{i \cdot \cdot}(j) &= \mathbf{E} e_{i \cdot \cdot} \left( \frac{1}{n_{ij}} \sum_{t \in N_{ij}} e_{i \cdot t} \right) = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \mathbf{E} e_{i \cdot \cdot} e_{i \cdot t} = \\ &= \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \mathbf{E} \left( \frac{1}{m_i} \sum_{s=1}^T m_{is} e_{i \cdot s} \right) e_{i \cdot t} = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \frac{1}{m_i} \sum_{s=1}^T m_{is} \mathbf{E} e_{i \cdot s} e_{i \cdot t} = \\ &= \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \frac{1}{m_i} \sum_{s=1}^T m_{is} \frac{\sigma^2 \mathbf{I}_{s=t}}{m_{it}} = \frac{1}{n_{ij}} \sum_{t \in N_{ij}} \frac{\sigma^2}{m_i} = \frac{\sigma^2}{m_i}. \end{aligned}$$

□

Рассмотрим  $\mathcal{E}_{ij}^1$  и  $\mathcal{E}_{ijt}$  из моделей (2.41)  $z_{ij} = \mu + \alpha_i + \mathcal{E}_{ij}^1$  и  $y_{ijt} = \beta_t + \gamma_{it} + \mathcal{E}_{ijt}$  как компоненты векторов  $\mathcal{E}_i^1 = (\mathcal{E}_{i1}^1, \dots, \mathcal{E}_{i\nu_i}^1)^T$  размерности  $\nu_i$  и  $\mathcal{E}_{it} = (\mathcal{E}_{ij_1 t}, \dots, \mathcal{E}_{ij_{m_{it}} t})^T$  размерности  $m_{it}$ ,  $j_k \in M_{it}$ . Вектора  $\delta_i = (\delta_{i1}, \dots, \delta_{i\nu_i})^T$ ,  $e_i = (e_{i \cdot \cdot}, \dots, e_{i \cdot \cdot})^T$ ,  $e_i^1 = (e_{i1}^1, \dots, e_{i\nu_i}^1)^T$

имеют размерность  $\nu_i$ . Вектор  $e_{it} = (e_{ij_1t}, \dots, e_{ij_{m_{it}}t})^T$  имеет размерность  $m_{it}$ , как и  $\tilde{e}_i$  и  $\tilde{\delta}_i$ , которые представляют собой соответственно  $e_i$  и  $\delta_i$  с компонентами из множества  $M_{it}$ . Обозначим через

$$R_i = \frac{1}{\sigma^2} \mathbf{E} \delta_i \delta_i^T = \left\{ \frac{\mathbf{I}_{[j=j_1]}}{n_{ij}} - D(i, j; i, j_1) \right\}_{j, j_1=1}^{\nu_i},$$

$$W_i = \frac{1}{\sigma^2} \mathbf{E} \delta_i e_{it}^T = \left\{ \frac{\mathbf{I}_{[j=j_1]}}{n_{ij}} - \frac{\mathbf{I}_{[t \in N_{ij} \cap N_{ij_1}]}}{n_{ij} m_{it}} \right\}_{j=1, j_1=1}^{\nu_i},$$

через  $R_{it}$ ,  $W_{it}$ ,  $Q_{it}$  матрицы порядка  $m_{it}$ , полученные из соответствующих множеству  $M_{it}$  строк и столбцов вышеопределенных матриц  $R_i$ ,  $W_i$  и матрицы  $Q_i$  из (2.39).

**Теорема 19.** Пусть  $\mathcal{E}_i^1 = e_i^1 + e_i + Q_i \delta_i$  и  $\mathcal{E}_{it} = e_{it} - \tilde{e}_i - Q_{it} \tilde{\delta}_i$ . Тогда ковариационные матрицы ошибок  $\Omega_i = \mathbf{E} \mathcal{E}_i^1 (\mathcal{E}_i^1)^T$  и  $\Sigma_{it} = \mathbf{E} \mathcal{E}_{it} \mathcal{E}_{it}^T$  имеют соответственно вид:

$$\Omega_i = \sigma_1^2 \mathbf{I}_{\nu_i} + \sigma^2 \left( Q_i R_i Q_i^T + \frac{1}{m_i} \right), \quad (2.44)$$

$$\Sigma_{it} = \sigma^2 \left( \mathbf{I}_{m_{it}} - W_{it}^T Q_{it}^T - Q_{it} W_{it} + Q_{it} R_i Q_{it}^T - \frac{1}{m_i} \right). \quad (2.45)$$

**Доказательство.**

$$\begin{aligned} \Omega_i &= \mathbf{E} \mathcal{E}_i^1 (\mathcal{E}_i^1)^T = \mathbf{E} (e_i^1 + e_i + Q_i \delta_i) (e_i^1 + e_i + Q_i \delta_i)^T = \\ &= \mathbf{E} e_i^1 (e_i^1)^T + \mathbf{E} e_i e_i^T + \mathbf{E} e_i \delta_i^T Q_i^T + Q_i \mathbf{E} \delta_i e_i^T + Q_i \mathbf{E} \delta_i \delta_i^T Q_i^T = \\ &= \sigma_1^2 \mathbf{I}_{\nu_i} + \sigma^2 \left( \frac{1}{m_i} + Q_i R_i Q_i^T \right). \end{aligned}$$

$$\begin{aligned} \Sigma_{it} &= \mathbf{E} \mathcal{E}_{it} \mathcal{E}_{it}^T = \mathbf{E} (e_{it} - \tilde{e}_i - Q_{it} \tilde{\delta}_i) (e_{it} - \tilde{e}_i - Q_{it} \tilde{\delta}_i)^T = \\ &= \mathbf{E} (e_{it} - \tilde{e}_i) (e_{it} - \tilde{e}_i)^T - \\ &- \mathbf{E} (e_{it} - \tilde{e}_i) \tilde{\delta}_i^T Q_{it}^T - Q_{it} \mathbf{E} \tilde{\delta}_i (e_{it} - \tilde{e}_i)^T + Q_{it} \mathbf{E} \tilde{\delta}_i \tilde{\delta}_i^T Q_{it}^T = \\ &= \sigma^2 \left( \mathbf{I}_{m_{it}} - \frac{1}{m_i} - W_{it}^T Q_{it}^T - Q_{it} W_{it} + Q_{it} R_i Q_{it}^T \right). \quad \square \end{aligned}$$

### 2.3.6. Значимость эффектов времени и взаимодействия факторов

Предположим, что в каждый момент времени  $t = 1, \dots, T$  имеются представители всех  $I$  групп. Обозначим вектор параметров через  $\Theta = (\beta_1, \gamma_{11}, \dots, \gamma_{\iota, 1}, \dots, \beta_{T-1}, \gamma_{1, T-1}, \dots, \gamma_{\iota, T-1})$  с  $I(T-1)$  компонентами. В случае коррелированных ошибок соберем матрицы  $\Sigma_{it}$  из (2.45) в объединенную матрицу  $\Sigma$ . Вектор наблюдений имеет вид  $Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{I1}, \dots, Y_{IT})^T$ , где  $Y_{it}$  вектор-строки с  $m_{it}$  компонентами  $\{y_{ijt}\}_{j=1}^{m_{it}}$ . Построим соответствующую модели  $y_{ijt} = \beta_t + \gamma_{it} + \mathcal{E}_{ijt}$  из (2.41) и ограничениям (2.31) матрицу плана  $H$  размерности  $m_{..}$  на  $I(T-1)$  (табл. 2.3), в которой использованы обозначения  $\iota = I-1$ ,  $\tau = T-1$ ,

$$q_{it} = -\frac{m_{it}}{m_{iT}}, \quad q_{.t} = -\frac{m_{.t}}{m_{.T}}, \quad p_{it} = -\frac{m_{it}}{m_{It}}, \quad p_{i.} = -\frac{m_{i.}}{m_{I.}}, \quad r_{it} = \frac{m_{it}}{m_{IT}}.$$

Таблица 2.3.

	$\beta_1$	$\gamma_{11}$	$\gamma_{21}$	$\dots$	$\gamma_{\iota, 1}$	$\dots$	$\beta_\tau$	$\gamma_{1, \tau}$	$\gamma_{2, \tau}$	$\dots$	$\gamma_{\iota, \tau}$
$Y_{11}$	1	1	0	$\dots$	0	$\dots$	0	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_{1, \tau}$	0	0	0	$\dots$	0	$\dots$	1	1	0	$\dots$	0
$Y_{1, T}$	$q_{.1}$	$q_{11}$	0	$\dots$	0	$\dots$	$q_{. \tau}$	$q_{1, \tau}$	0	$\dots$	0
$Y_{21}$	1	0	1	$\dots$	0	$\dots$	0	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_{2, \tau}$	0	0	0	$\dots$	0	$\dots$	1	0	1	$\dots$	0
$Y_{2, T}$	$q_{.1}$	0	$q_{21}$	$\dots$	0	$\dots$	$q_{. \tau}$	0	$q_{2, \tau}$	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_{\iota, 1}$	1	0	0	$\dots$	1	$\dots$	0	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_{\iota, \tau}$	0	0	0	$\dots$	0	$\dots$	1	0	0	$\dots$	1
$Y_{\iota, T}$	$q_{.1}$	0	0	$\dots$	$q_{\iota, 1}$	$\dots$	$q_{. \tau}$	0	0	$\dots$	$q_{\iota, \tau}$
$Y_{I, 1}$	1	$p_{11}$	$p_{21}$	$\dots$	$p_{\iota, 1}$	$\dots$	0	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y_{I, \tau}$	0	0	0	$\dots$	0	$\dots$	1	$p_{1, \tau}$	$p_{2, \tau}$	$\dots$	$p_{\iota, \tau}$
$Y_{I, T}$	$q_{.1}$	$r_{11}$	$r_{21}$	$\dots$	$r_{\iota, 1}$	$\dots$	$q_{. \tau} v$	$r_{1, \tau}$	$r_{2, \tau}$	$\dots$	$r_{\iota, \tau}$

В соответствии с (2.31) вычисляются граничные параметры:

$$\begin{aligned}\beta_T &= -\sum_{t=1}^{T-1} \frac{m_{.t}}{m_{.T}} \beta_t = \sum_{t=1}^{T-1} q_{.t} \beta_t, \\ \gamma_{iT} &= -\sum_{t=1}^{T-1} \frac{m_{it}}{m_{iT}} \gamma_{it} = \sum_{t=1}^{T-1} q_{it} \gamma_{it}, \quad \gamma_{It} = -\sum_{i=1}^{I-1} \frac{m_{it}}{m_{It}} \gamma_{it} = \sum_{i=1}^{I-1} p_{it} \gamma_{it}, \\ \gamma_{IT} &= \sum_{i=1}^{I-1} \sum_{t=1}^{T-1} \frac{m_{it}}{m_{IT}} \gamma_{it} = \sum_{i=1}^{I-1} \sum_{t=1}^{T-1} r_{it} \gamma_{it}.\end{aligned}$$

Для проверки отсутствия эффекта взаимодействия  $\gamma_{it} = 0$  используем усеченную модель

$$E y_{ijt} = \beta_t. \quad (2.46)$$

Из условия минимизации квадратичной формы

$$Q_{2e}(\Theta) = (Y - H\Theta^T)^T \Lambda^{-1} (Y - H\Theta^T), \quad (2.47)$$

где  $\Lambda$  такое, что  $\Sigma = \sigma^2 \Lambda$ , получаем оценку вектора параметров  $\hat{\Theta} = (H^T \Lambda^{-1} H)^{-1} H^T \Lambda^{-1} Y$  и несмещенную оценку дисперсии

$$\hat{\sigma}^2 = Q_{2e}(\hat{\Theta}) / \mu_{2e}. \quad (2.48)$$

Степени свободы  $\mu_{2e}$  вычисляются через разность между числом независимых наблюдений  $m_{..} - n$  и рангом  $I(T-1)$  матрицы  $H$ , т.е.

$$\mu_{2e} = m_{..} - n - I(T-1). \quad (2.49)$$

Для усеченной модели (2.51) с ограничениями (2.31) удалением соответствующих столбцов из матрицы  $H$  строим матрицу плана  $H_*$ , вводим вектор параметров  $\Theta_* = (\beta_1, \dots, \beta_{T-1})^T$  и из условия минимизации квадратичной формы

$$Q_{2e*}(\Theta_*) = (Y - H_* \Theta_*^T)^T \Lambda^{-1} (Y - H_* \Theta_*^T) \quad (2.50)$$

получаем оценку  $\hat{\Theta}_* = (H_*^T \Lambda^{-1} H_*)^{-1} H_*^T \Lambda^{-1} Y$ . Для проверки гипотезы отсутствия эффекта взаимодействия в числителе статистики  $F$  вычисляем разности  $Q_{2e*}(\hat{\Theta}_*) - Q_{2e}(\hat{\Theta})$  и соответствующее число степеней свободы

$$\begin{aligned}\mu_{2e*} - \mu_{2e} &= m_{..} - n - T + 1 - \\ &- (m_{..} - n - I(T-1)) = IT - T - I + 1.\end{aligned}$$

Отсюда получаем статистику Фишера вида

$$F = \frac{(Q_{2e*}(\hat{\Theta}_*) - Q_{2e}(\hat{\Theta}))(m_{..} - n - I(T - 1))}{Q_{2e}(\hat{\Theta})(IT - T - I + 1)}.$$

Для проверки отсутствия эффекта времени  $\beta_t = 0$  используем усеченную модель

$$Ey_{ijt} = \gamma_{it}, \quad (2.51)$$

удаляя в матрице плана столбцы, относящиеся к параметрам  $\beta_t$ ,  $t = 1, \dots, T - 1$ . По усеченной модели оцениваем параметры  $\Theta$  и вычисляем  $\widetilde{Q}_{2e}$ . Статистика  $F$ , определяемая равенством

$$F = \frac{\widetilde{Q}_{2e}(\widetilde{\Theta}) - Q_{2e}(\hat{\Theta})}{Q_{2e}(\hat{\Theta})} \cdot \frac{\mu_{2e}}{\widetilde{\mu}_{2e} - \mu_{2e}},$$

при отсутствии эффекта времени имеет распределение Фишера с числом степеней свободы  $\mu_{2e}$  из (2.49)

$$\widetilde{\mu}_{2e} - \mu_{2e} = m_{..} - n - (T - 1)(I - 1) - (m_{..} - n - I(T - 1)).$$

### 2.3.7. Проверка значимости главного эффекта

Если коррелированностью ошибок пренебречь нельзя, то необходимо оценивать параметры модели  $\mathbf{E}z_{ij} = \mu + \alpha_i$  при ковариационной матрице ошибок  $\Omega_i$  из (2.44), которая зависит от параметров  $\sigma^2$  и  $\sigma_1^2$ . Объединим матрицы  $\Omega_i$  в матрице  $\Omega = \sigma_1^2 R(\sigma_1^2)$  и компоненты  $z_{ij}$  в векторе  $Z = (z_{11}, \dots, z_{1\nu_1}, \dots, z_{I1}, \dots, z_{I\nu_I})$ . В качестве оценки  $\sigma^2$  можно использовать выражение (2.48). Для оценки дисперсии  $\sigma_1^2$  используем рекурсивную процедуру, связанную с минимизацией выражения

$$\begin{aligned} \mathcal{L}(\sigma_1^2, \Theta) &= \left( \frac{(Z - H\Theta)^T R^{-1}(\sigma_1^2)(Z - H\Theta)}{n - I} - \sigma_1^2 \right)^2, \\ \hat{\Theta} &= (H^T R^{-1}(\sigma_1^2)H)^{-1} H^T R^{-1}(\sigma_1^2)Z, \\ \hat{\sigma}_1^2 &= \frac{(Z - H\Theta)^T R^{-1}(\sigma_1^2)(Z - H\Theta)}{n - I}. \end{aligned}$$

Обозначим через  $R(\sigma_1^2) = R$ . Для основной и усеченной моделей соответственно вводим векторы параметров  $\Theta = (\mu, \alpha_1, \dots, \alpha_{I-1})$  и



$\Theta_* = \mu$  и построим матрицы планов  $H$  и  $H_*$ . Оценим параметры

$$\hat{\Theta} = (H^T R^{-1} H)^{-1} H^T R^{-1} Z, \quad \hat{\mu} = (H_*^T R^{-1} H_*)^{-1} H_*^T R^{-1} Z,$$

вычислим невязки

$$\begin{aligned} R_1^2 &= (Z - H_* \hat{\mu})^T R^{-1} (Z - H_* \hat{\mu}), \\ R_0^2 &= (Z - H \hat{\Theta})^T R^{-1} (Z - H \hat{\Theta}) \end{aligned}$$

и построим статистику Фишера

$$F = \frac{(R_1^2 - R_0^2)(n - r)}{R_0^2(r - 1)},$$

которая имеет распределение Фишера с числом степеней свободы  $n - 1 - (n - r) = r - 1$  и  $n - r$ .

Переход к двухфакторной модели может быть осуществлен перестройкой матриц плана  $H$  и  $H_*$ . Можно показать, что смещение для двухфакторной модели с числом градаций  $r$  и  $s$  имеет тот же вид, что и для однофакторной с разбиением на  $rs$  групп.

ПРИМЕР 1 (продолжение). На основе данных о депрессии наркоманов в процессе выполнения программы по отказу от употребления наркотиков [38] выясним, как влияет антидепрессант фрамакс на снижение уровня депрессии. По назначению препаратов будем обозначать подгруппы: налтрексон и фрамакс  $NF$ , налтрексон и антидепрессант-плацебо  $NP$ , плацебо-блокатор опиатных рецепторов и фрамакс  $PF$  и плацебо-плацебо  $PP$ .

На рис.2.3 изображены средние уровни депрессии в разных группах и в разные моменты времени по индивидам, у которых имеются наблюдения во всех восьми временных точках ( $n = 75$ ). Эффект времени выражен значимо ( $p < 10^{-3}$ ). Значимость эффекта взаимодействия равна  $p = 0.2$ . В группах  $PF$ ,  $PP$ ,  $NF$  и  $NP$  средние уровни депрессии равны 6.7 и 6.25, 9.2 и 5.5 соответственно ( $p = 0.07$ ). Объяснить самый низкий уровень депрессии при назначении налтрексона без антидепрессанта и самый высокий уровень при назначении налтрексона с антидепрессантом довольно сложно, тем более, что динамика депрессии анализируется только у тех, кто дошел до благополучного завершения программы отказа от употребления наркотиков.

Применим эргодический метод компенсации пропусков. При использовании в вычислениях псевдообратных матриц можно не исключать индивидов, которые не явились ни на один из контрольных приемов. Значимости эффектов времени и взаимодействия фактора группы и фактора времени остались на том

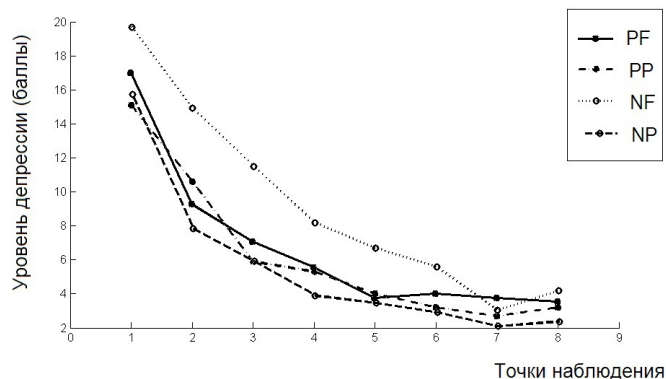


Рис. 2.3. Динамика депрессии при удалении пропусков ( $n = 75$ ).

же уровне, что и в предыдущем методе (рис. 2.4). В группах *PF*, *PP*, *NF* и *NP* средние уровни депрессии равны 8.9 и 11.9, 12.2 и 10.9 соответственно ( $p = 0.0002$ ). Во-первых, они существенно выше, чем у тех, кто программу выполнил. Во-вторых, наименьший уровень депрессии достигается в группе больных, получавших антидепрессант без вызывающего дополнительно депрессию налтрексона, что вполне объяснимо. Таким образом, результаты, полученные на основе этого метода, получают более адекватное обоснование, чем при избирательном сужении выборки при удалении неполных данных.

**ПРИМЕР 2.** Это приложение касается исследования эффективности лечебной физкультуры (ЛФ) для больных хронической сердечной недостаточностью (ХСН) на основе данных теста шестиминутной ходьбы (ТШХ), означающего количество метров, пройденных в спокойном режиме за 6 минут. ТШХ измеряется у больных в разные моменты времени: при поступлении в стационар, при выписке и через три месяца после лечения. Сложность анализа этих данных заключалась в том, что из 387 больных только у 19 имелись данные во всех трех точках, и из них 13 человек, которые занимались лечебной физкультурой (ЛФ) [18]. У женщин с функциональным классом ХСН не выше второго уровня ТШХ соответственно для не занимающихся и занимающихся ЛФ равны 315 и 340 ( $p = 0.2$ , значимость эффекта взаимодействия с фактором времени  $p_\gamma = 0.27$ ), у мужчин 388 и 413 ( $p = 0.11$ ,  $p_\gamma = 0.14$ ), в третьем и четвертом классе ХСН у женщин 257 и 265 ( $p = 0.65$ ,  $p_\gamma = 0.6$ ), у мужчин 310 и 345 ( $p = 0.24$ ,  $p_\gamma = 0.85$ ).

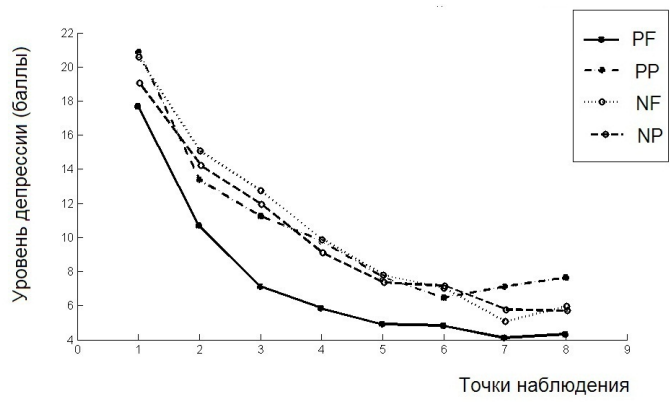


Рис. 2.4. Динамика депрессии с учетом пропусков ( $n = 332$ ).

Таким образом, можно заметить, что в разных группах при ЛФ показатели ТШХ выше, за исключением женщин с высоким функциональным классом ХСН, у которых ТШХ не зависит от занятий ЛФ. Но в целом, на прогрессирование ТШХ большее влияние оказывает лечение в стационаре, а эффект ЛФ дольше держится у мужчин с меньшим уровнем ХСН.

## Глава 3.

# Роль масштабирования в управлении биосистем

### 3.1. Частичное обращение функций

#### 3.1.1. Определение частично обратных функций

Идея частичного обращения функций и применения его в анализе медико-биологических систем принадлежит А.Г.Барту [10].

Особенности оператора обращения связаны с решением уравнения  $y = f(x)$  при  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ . Неоднозначность решения требует сужения  $\mathcal{X}^- \subseteq \mathcal{X}$  области определения функции; отсутствие решений приводит к необходимости расширения множества значений  $\mathcal{Y} \subseteq \mathcal{Y}^+$ .

**Определение 14.** Обобщенным обратным к  $f : \mathcal{X} \rightarrow \mathcal{Y}$  называется отображение  $f^- : \mathcal{Y}^+ \rightarrow \mathcal{X}$ , что  $\forall y \in \mathcal{Y} \subseteq \mathcal{Y}^+$  и  $\forall x \in \mathcal{X}^- \stackrel{def}{=} f^-(\mathcal{Y}^+)$ , выполнено хотя бы одно из четырех определяющих соотношений:

$$\begin{array}{ll} O_1. ff^-(y) = y; & O_2. f^-ff^-(y) = f^-(y); \\ O_3. f^-f(x) = x; & O_4. ff^-f(x) = f(x). \end{array}$$

Обозначим через  $f_\alpha$  частичную функцию для  $f$  или сужение  $f$  на  $\{\mathcal{X}_\alpha\}$ , где  $\{\mathcal{X}_\alpha\}$  означает разбиение множества  $\mathcal{X}$  на области мономорфности.

**Определение 15.** Пусть  $f_\alpha^-$  — обобщенное обратное отображение к  $f_\alpha$ . Если выполнение одного из определений  $O_1$ – $O_4$  влечет за собой выполнение всех остальных, то  $f_\alpha^-$  называется частично-обратным отображением.

### 3.1.2. Параметризация частично обратных функций

Рассмотрим измеримую функцию  $f(x)$ . Для каждого  $y = f(x)$  определим прообразы: соответственно левый и правый

$$\overleftarrow{\mathcal{X}}_y = \{x : f(x) \geq y\}, \quad \overrightarrow{\mathcal{X}}_y = \{x : f(x) \leq y\},$$

вещественный  $\tilde{\mathcal{X}}_y = \overleftarrow{\mathcal{X}}_y \cap \overrightarrow{\mathcal{X}}_y$  и полный

$$\mathcal{X}_y = \{x : \forall \delta > 0, X_\delta \cap \overleftarrow{\mathcal{X}}_y \neq \emptyset, X_\delta \cap \overrightarrow{\mathcal{X}}_y \neq \emptyset\},$$

где  $X_\delta = (x - \delta, x + \delta)$ . Обозначим через  $R$  минимальный отрезок, содержащий полный прообраз  $\mathcal{X}_y$ . Пусть  $\mu$  —  $\sigma$ -конечная мера, эквивалентная мере Лебега. Если  $0 < \mu(R) < \infty$ , то всякому  $\alpha \in [0, 1]$  можно сопоставить точку  $z_\alpha \in \bar{R}$ , такую что

$$\alpha = \frac{\mu((-\infty, z_\alpha] \cap R)}{\mu(R)},$$

и всякому  $y = f(x)$  точку  $x_\alpha = \sup\{(x : x \leq z_\alpha) \cap \mathcal{X}\} = f_\alpha^-(y)$ . Функция  $f_\alpha^-(y)$  является частично обратной функцией к  $f(x)$  [10]. Заметим, что частично обратной будет также функция  ${}_a f^-(y)$ , которой сопоставляется точка  ${}_a x = \inf\{(x : x \geq z_\alpha) \cap \mathcal{X}\}$ .

Введем обозначения:

$$\begin{aligned} f_{00}^-(y) &= \inf(x : f(x) \geq y), & f_{01}^-(y) &= \inf(x : f(x) \leq y), \\ f_{10}^-(y) &= \sup(x : f(x) \geq y), & f_{11}^-(y) &= \sup(x : f(x) \leq y). \end{aligned}$$

**Определение 16.** Крайними левой и правой обратными называются  $\overset{-1}{\lceil} f(y) = \max\{f_{00}^-(y), f_{01}^-(y)\}$  и  $\overset{+1}{\lceil} f(y) = \min\{f_{10}^-(y), f_{11}^-(y)\}$  соответственно.

Для примера вычислим левую обратную функции  $f(x) = \lfloor x \rfloor$ . Пусть  $k < y < k + 1$ , тогда  $\overset{-1}{\lceil} f(y) = \lceil y \rceil$ , так как  $\inf(x : f(x) \geq y) = k + 1$ ,  $\inf(x : f(x) \leq y) = -\infty$ ,  $\max\{k + 1, -\infty\} = k + 1$ .

### 3.1.3. Оценка параметров обобщенного геометрического распределения

Применим частичное обращение к функции числа успехов  $\xi(n)$  от  $n$  независимых испытаний Бернулли с вероятностью успеха  $p$ . Поскольку функция  $k = \xi(n)$  монотонная, определения левой и правой обратных значительно упростятся. Левая обратная

$$\xi_0^-(k) = \min\{n : \xi(n) \geq k\}$$

означает число испытаний до  $k$ -го успеха, правая обратная

$$\xi_1^-(k) = \max\{n : \xi(n) \leq k\}$$

— число неудач до  $k + 1$  успеха. В простейшем случае  $k = 0$  левая обратная  $\xi_0^-(0) = \min\{n : \xi(n) \geq 0\}$  равна нулю, а правая обратная

$$\xi_1^-(0) = \max\{n : \xi(n) \leq 0\}$$

как число неудач до первого успеха имеет геометрическое распределение

$$P\{\xi_1^-(0) = j\} = p(1-p)^j, \quad j = 0, 1, \dots$$

Частично обратные из-за дискретности распределения могут быть введены двояким образом:  $\xi_\alpha^-(0) = \lfloor \alpha \xi_1^-(0) \rfloor$  и  ${}_\alpha \xi^-(0) = \lceil \alpha \xi_1^-(0) \rceil$ . Найдем их законы распределения.

**Теорема 20.**

$$\begin{aligned} P\{\xi_\alpha^-(0) = j\} &= q^{\lceil \frac{j}{\alpha} \rceil} - q^{\lceil \frac{j+1}{\alpha} \rceil}, \quad j = 0, 1, 2, \dots, \\ P\{{}_\alpha \xi^-(0) = j\} &= q^{\lfloor \frac{j-1}{\alpha} \rfloor + 1} - q^{\lfloor \frac{j}{\alpha} \rfloor + 1}, \quad j = 1, 2, \dots, \\ P\{{}_\alpha \xi^-(0) = 0\} &= 1 - q. \end{aligned}$$

**Доказательство.** Для нецелого  $x$  имеет место

$$P\{\xi_1^-(0) < x\} = \sum_{j=0}^{\lfloor x \rfloor} P\{\xi_1^-(0) = j\} = 1 - q^{\lfloor x \rfloor + 1},$$

для целого  $x = n$  получаем

$$P\{\xi_1^-(0) < n\} = \sum_{j=0}^{n-1} P\{\xi_1^-(0) = j\} = 1 - q^n,$$

поэтому в общем случае  $P\{\xi_1^-(0) < x\} = 1 - q^{\lfloor x \rfloor}, \implies$

$$\begin{aligned} P\{\xi_\alpha^-(0) = j\} &= P\{\lfloor \alpha \xi_1^-(0) \rfloor = j\} = P\left\{\frac{j}{\alpha} \leq \xi_1^-(0) \leq \frac{j+1}{\alpha}\right\} = \\ &= q^{\lfloor \frac{j}{\alpha} \rfloor} - q^{\lfloor \frac{j+1}{\alpha} \rfloor}. \end{aligned}$$

Для нецелого  $x$  имеет место

$$P\{\xi_1^-(0) \leq x\} = \sum_{j=0}^{\lfloor x \rfloor} P\{\xi_1^-(0) = j\} = 1 - q^{\lfloor x \rfloor + 1},$$

для целого  $x = n$  получаем

$$P\{\xi_1^-(0) \leq n\} = \sum_{j=0}^n P\{\xi_1^-(0) = j\} = 1 - q^{n+1},$$

поэтому в общем случае  $P\{\xi_1^-(0) \leq x\} = 1 - q^{\lfloor x \rfloor + 1}, \implies$

$$\begin{aligned} P\{\alpha \xi^-(0) = 0\} &= P\{\xi_1^-(0) = 0\} = 1 - q, \\ P\{\lfloor \alpha \xi_1^-(0) \rfloor = j\} &= P\left\{\frac{j-1}{\alpha} < \xi_1^-(0) \leq \frac{j}{\alpha}\right\} = \\ &= q^{\lfloor \frac{j-1}{\alpha} \rfloor + 1} - q^{\lfloor \frac{j}{\alpha} \rfloor + 1}, \quad j > 0. \quad \square \end{aligned}$$

Распределение величины  $\xi_\alpha^-(0)$  в [31] было названо *обобщенным геометрическим*. Его будем обозначать через  $\beta_-^B(\cdot | p, \alpha)$ , а модифицированное обобщенное геометрическое распределение величины  $\alpha \xi^-(0)$  через  $\beta_-^A(\cdot | p, \alpha)$ .

На рис.3.1 изображены две схемы случайного блуждания по целочисленной решетке, приводящие при  $\alpha = \frac{2}{2r+1}$  к этим распределениям. Способ чередования вероятностей можно интерпретировать через действие тренированной и нетренированной системы. В тренированной большая  $1 - q^{r+1}$  и меньшая  $1 - q^r$  вероятности успеха чередуются в регулярном порядке, а в нетренированной — сначала наименьшая вероятность успеха  $1 - q$ , а затем выбор масштаба  $r$  (настроились) и чередование меньшей и большей вероятностей успеха  $1 - q^r$  и  $1 - q^{r+1}$ .

В случае, когда задана рациональная структура параметра частичности  $\alpha = \frac{s}{m}$ , где  $m = sr + t$ , и  $s$  известно, оценка параметров обобщенного геометрического распределения  $\beta_-^B(\cdot | p, \frac{s}{m})$  строится

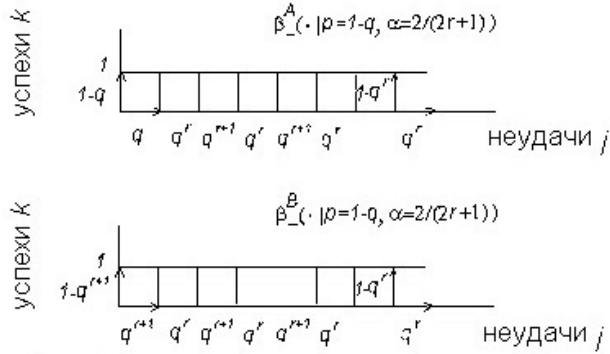


Рис. 3.1. Схемы  $\beta_-(j|p, \alpha)$ -блуждания по решетке.

довольно просто.<sup>1</sup> Пусть имеется вариационный ряд  $N_0, N_1, \dots, N_k$  с общим числом наблюдений  $N_0 + N_1 + \dots + N_k = N$ . Обозначим через  $k_1 = \lceil \frac{k+1}{s} \rceil$  и представим вариационный ряд в виде матрицы размерности  $k_1$  на  $s$  с элементами

$$\nu_{ji} = \frac{N_{sj+i}}{N},$$

вычисляя соответственно суммы по строкам и столбцам:

$$\nu_{j\cdot} = \sum_{i=0}^{s-1} \nu_{ji}, \quad \nu_{\cdot i} = \sum_{j=0}^{k_1-1} \nu_{ji},$$

$j = 0, 1, \dots, k_1 - 1, i = 0, 1, \dots, s - 1$ . Если  $\xi \sim \beta_-^B(\cdot|p, \alpha), \alpha = \frac{s}{m}$ , то случайная величина  $\eta = \lfloor \frac{\xi}{m} \rfloor$  имеет геометрическое распределение с параметром неудачи  $q^m$ .

$$\begin{aligned} P\{\eta = j\} &= P\left\{\left\lfloor \frac{\xi}{m} \right\rfloor = j\right\} = P\left\{j \leq \frac{\xi}{m} < j+1\right\} = \\ &= P\{mj \leq \xi < m(j+1)\} = \\ &= q^{mj} - q^{m(j+1)} = q^{mj}(1 - q^m), \quad j = 0, 1, \dots \end{aligned}$$

<sup>1</sup>Вопрос аналитической оценки параметров распределения  $\beta_-^A(\cdot|p, \alpha)$  остается открытым.



Вычисляя выборочное группированное среднее  $\bar{x}_* = \sum_{j=0}^{k_1-1} j\nu_j$ , из оценки максимального правдоподобия параметра геометрического распределения получаем оценку

$$\hat{q}^m = \frac{\bar{x}_*}{\bar{x}_* + 1}. \quad (3.1)$$

Для оценивания параметра  $m$  обозначим первые  $s$  вероятностей распределения  $\beta_-^B(\cdot|p, \frac{s}{m})$  через  $\pi_i = P\{\xi = i\}$ ,  $i = 0, 1, \dots, s-1$ .

$$\begin{array}{cccc|c} \pi_0 & \pi_1 & \dots & \pi_{s-1} & 1 - q^m \\ \pi_0 q^m & \pi_1 q^m & \dots & \pi_{s-1} q^m & (1 - q^m)q^m \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \hline \frac{\pi_0}{1 - q^m} & \frac{\pi_1}{1 - q^m} & \dots & \frac{\pi_{s-1}}{1 - q^m} & 1 \end{array}$$

Оценкой выражения  $\frac{\pi_i}{1 - q^m}$  является  $\nu_i$ ,  $\hat{q}^m$  берем из (3.1), откуда

$$\hat{\pi}_i = \nu_i(1 - \hat{q}^m) = \frac{\nu_i}{1 + \bar{x}_*}. \quad (3.2)$$

Выражения (3.2) для оценки структуры параметра частичности являются избыточными, за исключением случая  $\alpha = 2/(2r + 1)$ . Пусть  $Q = q^{2r+1}$ . К объединенному по парам вариационному ряду  $\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_{k_1}$  применим метод максимального правдоподобия.

$$\begin{aligned} \ln \mathcal{L}(Q) &= \sum_{i=0}^{k_1-1} \tilde{N}_i \ln(1 - Q) + \sum_{i=0}^{k_1} i \tilde{N}_i \ln(Q), \\ \frac{d\mathcal{L}(Q)}{dQ} &= - \sum_{i=0}^{k_1-1} \frac{\tilde{N}_i}{1 - Q} + \sum_{i=0}^{k_1} \frac{i \tilde{N}_i}{Q} = 0. \end{aligned}$$

Решая нормальное уравнение и используя оценку (3.2) параметра  $\pi_0 = 1 - q^{r+1}$ , нетрудно показать, что

$$\hat{Q} = \frac{\sum_{i=0}^K i \tilde{N}_i}{\sum_{i=0}^K i \tilde{N}_i + \sum_{i=0}^{K-1} \tilde{N}_i}, \quad \hat{r} = \frac{\ln \hat{Q} - \ln(1 - \hat{\pi}_0)}{2 \ln(1 - \hat{\pi}_0) - \ln \hat{Q}}. \quad (3.3)$$

Отсюда  $\hat{q} = (1 - \hat{\pi}_0)^{\frac{1}{r+1}}$  и  $\hat{m} = 2\hat{r} + 1$ .

Таблица 3.1.

$j$	0	1	2	3	4	5
$n_j^1$	7	3	10	6	2	12
$n_j^2$	25	9	2	1	2	1

ПРИМЕР. Перед операцией АКШ (аорто-коронарное шунтирование), при которой в качестве кондукта используется либо вена, либо лучевая артерия (ЛА), для ЛА проводится холодовая проба (ХП). Рука на некоторое фиксированное время кладется на лед, и после того как рука убрана при помощи аппарата УЗИ (ультра-звукового исследования) измеряется диаметр ЛА и время в минутах до момента возвращения диаметра ЛА к исходному состоянию.

Обозначим через  $n_j^i$  количество больных с временем  $j$  восстановления ЛА в  $i = 1, 2$  группах больных (табл. 3.1). Известно, что в процессе операции АКШ в 1-ой группе был вазоспазм ЛА, во 2-й не был. Вазоспазм ЛА является опасным явлением и может служить причиной смерти больного.

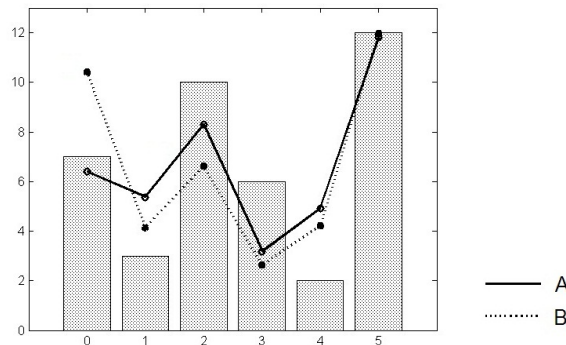


Рис. 3.2. Апробация  $A$  и  $B$  моделей  $\beta_-(j|p, \alpha)$ .

Используя агрегированный вариационный ряд  $(\tilde{N}_0, \tilde{N}_1, \tilde{N}_2) = (10, 16, 14)$  и оценку (3.3), получаем  $\hat{Q} = 0.6286$ ,  $\hat{q} = Q^{\frac{1}{3}} = 0.8566$ . Значимость статистики хи-квадрат согласно с  $\beta_-^B(\cdot|\hat{p} = 0.1434, \alpha = 2/3)$  равна  $P = 0.07$ , при укрупнении ячеек  $P = 0.18$ . Однако для модели  $\beta_-^A(\cdot|\hat{p} = 0.1434, \alpha = 2/3)$  имеем лучшие значимости  $P = 0.18$  и с укрупнением ячеек  $P = 0.66$ . Небольшим варьировани-

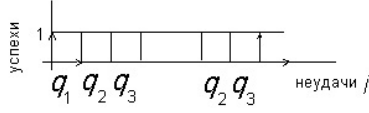


Рис. 3.3. Схема  $\beta_-^C(j|p, \alpha)$ -блуждания по решетке.

ем по параметру  $p$  получаем наибольшие значимости  $P = 0.225$  и  $P = 0.73$  при значении параметра  $\hat{p} = 0.16$ . На рис. 3.2 представлена гистограмма распределения времени восстановления ЛА у больных с вазоспазмом, а также полигоны частот обобщенных геометрических распределений типов  $A$  и  $B$ .

Итак, в менее благополучной с прогностической точки зрения группе со слабым тономусом ЛА получено удовлетворительное согласие с распределением  $\beta_-^A(\cdot|p, \alpha)$ , описывающим нетренированную обучаемую систему с невысокой вероятностью  $\hat{p} = 0.16$  восстановления ЛА в одном испытании. В более благополучной группе имеет место согласие с распределением  $\beta_-^B(\cdot|\hat{p} = 0.55, \alpha = 1)$ , которое соответствует тренированной системе.

### 3.1.4. Обобщение положительного биномиального распределения

Одним из способов обобщения негативного биномиального распределения [10] является переход к сумме  $k$  независимых обобщенных геометрически распределенных случайных величин  $\xi_\alpha^-(0)$  и  ${}_\alpha\xi^-(0)$ . В зависимости от вида частично обратной используем обозначения  $\beta_-^{*B}(j|k, p, \alpha)$  или  $\beta_-^{*A}(j|k, p, \alpha)$ .

Обобщенное положительное биномиальное распределение определяется через фидуциальное тождество

$$\beta_+^{\leq}(k|n, p, \alpha) = 1 - \beta_-^{\leq}(n|k, p, \alpha). \quad (3.4)$$

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 положительное биномиальное распределение  $\beta_+^{*C}(\cdot|n, p_1, p_2, p_3)$ , фидуциальным образом (3.4) построенное из свертки  $C$ -обобщения геометрического  $\beta_-^C(\cdot|p_1, p_2, p_3)$  с блужданием на решетке, изображенным на рис. 3.3.

## 3.2. Модель реинтрантного бинома в радиобиологии

При исследовании изменения вероятности формирования интерфазных опухолевых клеток с одиночными или множественными аномалиями<sup>2</sup> на ядре под воздействием облучения была предпринята попытка найти подходящую статистическую модель.

Таблица 3.2.

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

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-гау in vivo дозами 5-45 Гр<sup>3</sup>. В другой серии опытов были исследованы ядерные аномалии у отдаленных потомков этих трансплантатов. Суспензированные клетки прививались путем инъекции в переднюю камеру глаза другого животного сразу после облучения in vitro. Распределения числа аномалий при разных дозах облучения представлены в табл. 3.2.

<sup>2</sup>Аномалиями считались межъядерные мосты, ядерные протрузии и гантелевидные ядра.

<sup>3</sup>Грей (обозначение Гр или 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 расчете на один килограмм массы.

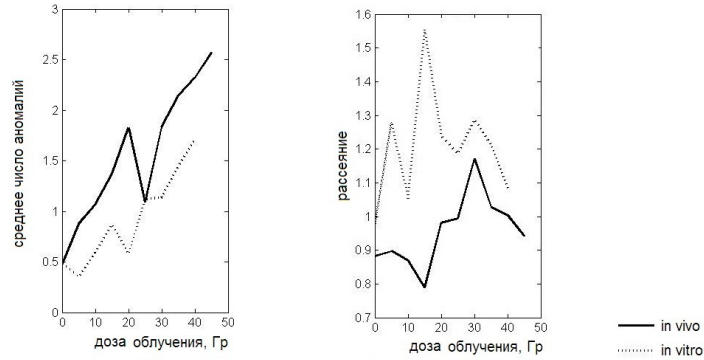


Рис. 3.4. Динамика среднего числа аномалий и рассеяния.

Согласно первичной статистике, между средним числом аномалий и дозой облучения существует положительная корреляция. Кроме того, при увеличении дозы облучения (рис. 3.4) распределение изменяет характер рассеяния<sup>4</sup>, т.е. из недорассеянного становится перерассеянным. Поэтому в качестве модели не могут быть использованы ни недорассеянное положительное, ни перерассеянное отрицательное биномиальные распределения. Выясним, можно ли для этих целей привлечь сложные распределения.

### 3.2.1. Реинтрантный бином

**Определение 17.** *Сложным* называется распределение случайной суммы независимых одинаково распределенных случайных величин

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

где  $\tau$  и  $\eta_j$  независимые случайные величины [22].

<sup>4</sup>Рассеяние  $e$  — это отношение дисперсии к среднему. Пуассоновское распределение имеет рассеяние, равное единице. Распределение с  $e < 1$  называется недорассеянным, с  $e > 1$  перерассеянным.

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

$$h(\nu) = f_1(f_0(\nu)), \quad (3.5)$$

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  $\xi_1 = \eta_1 + \dots + \eta_\tau$  имеет вид

$$h_1(\nu) = f_1(f_0(\nu)) = ((p_0\nu + q_0)^{n_0} p_1 + q_1)^{n_1}, \quad (3.6)$$

а соответствующий закон распределения  $\beta(\cdot|p_1, p_0; n_1, n_1)$  будем называть *реинтрантным*, поскольку в [22] сложно-биномиальным называется распределение с пуассоновской случайной величиной  $\tau$ .<sup>5</sup>

Используя свойства производных производящих функций (3.6)

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

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  $e < 1$ , 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 11.

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

<sup>5</sup>По аналогии могут быть построены суперпозиции производящих функций любого порядка: второго  $h_2(\nu) = f_2(f_1(f_0(\nu)))$ , третьего  $h_3(\nu) = f_3(f_2(f_1(f_0(\nu))))$  и так далее.

**Доказательство.** Поскольку индекс  $i$  в параметрах  $p_i$  и  $n_i$  в формуле  $\beta(\cdot|n_i; p_i)$  один и тот же, 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  $k_1 = 0$ ,

$$0 < k_0 \leq n_0 \iff 0 < \frac{k_0}{n_0} \leq 1 \implies 1 \leq k_1 \leq n_1,$$

$$n_0 < k_0 \leq 2n_0 \iff 1 < \frac{k_0}{n_0} \leq 2 \implies 2 \leq k_1 \leq n_1,$$

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

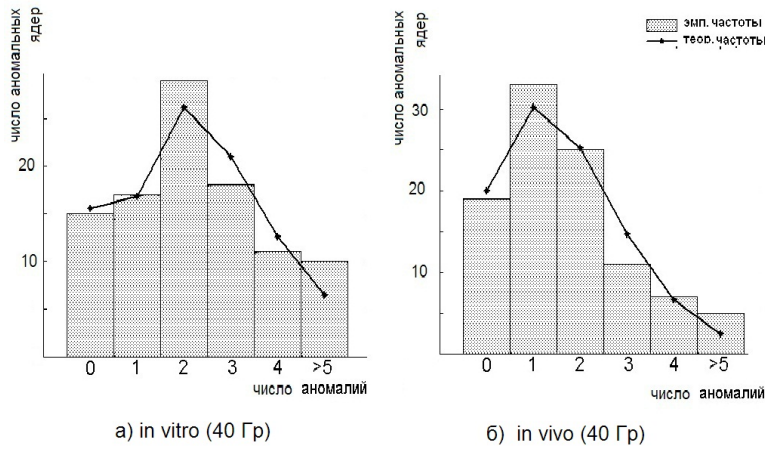


Рис. 3.5. Оценка адекватности модели  $\beta(\cdot|p_1, p_0; n_1, n_1)$ .

### 3.2.2. 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  $n_1$ , 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. 3.5 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.

Следует обратить внимание на более симметричную при увеличении внешней реинтрантной компоненты  $m_1$  форму распределения по аналогии с симметризацией плотности гамма-распределения при увеличении параметра экстенсивности.



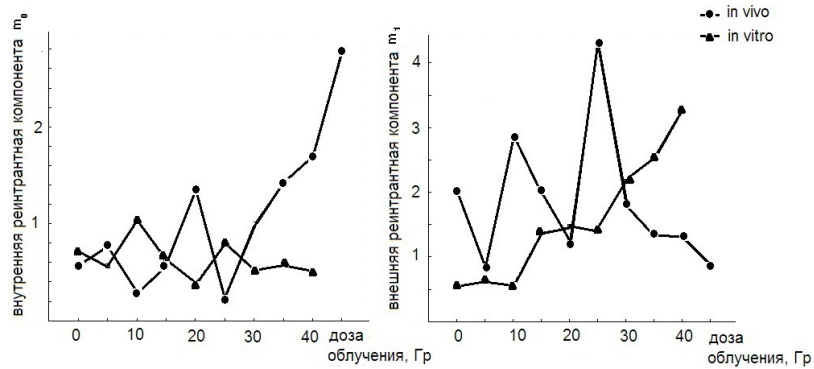


Рис. 3.6. Динамика средних реинтрантных компонент.

Динамика средних реинтрантных компонент в зависимости от дозы облучения в разных экспериментах представлена на рис. 3.6. 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  $m_1$ .

Значения реинтрантных компонент ( $\hat{m}_0 = 0.64$ ,  $\hat{m}_1 = 1.55$ ) у потомков оказались в среднем меньше, чем у клеток ( $\hat{m}_0 = 1.07$ ,  $\hat{m}_1 = 1.86$ ). Это можно объяснить тем, что потомки не были под облучением непосредственно. Скорее всего, более слабые клетки не смогли дать потомства. Это следует из того, что при повышении дозы облучения имплантируемых клеток процент успешных прививок уменьшался, а формирование привитых трансплантатов за-

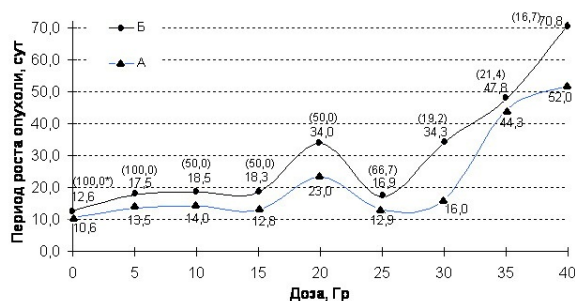


Рис. 3.7. Характеристики прививаемости.

медлялось. На рис. 3.7 отображены характеристики прививаемости опухолевых клеток: А) время от прививки до начала роста опухоли; Б) время от прививки до заполнения растущей опухолью глазной полости. Та и другая увеличиваются по мере увеличения дозы облучения. В скобках указаны проценты успешных прививок. При дозе 45 Гр все прививки оказались неудачными.

### 3.2.3. Сложно-обобщенная биномиальная модель

Для изучения структуры воздействия радиации используем расширение модели. Биномиальные распределения внутренней (*in vivo*) или внешней (*in vitro*) компоненты сложного распределения заменяются обобщенным распределением  $\beta_+^C(\cdot|n, p_1, p_2, p_3)$  из раздела 3.1.4, которое характеризует работу нетренированной системы с более общими параметрами, не связанными частично с  $\alpha$ . 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

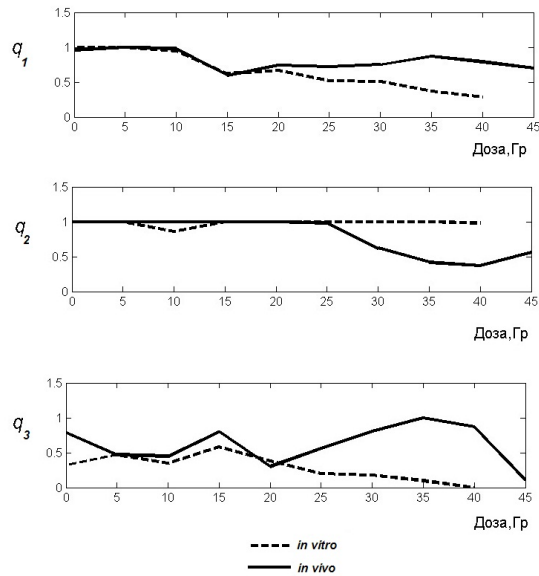


Рис. 3.8. Динамика вероятностей репарации.

репарации) вероятности образования аномалии уменьшаются и изменяются 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 3.3.

Графики зависимости вероятностей репарации  $q_1, q_2, q_3$  от дозы облучения представлены на рис. 3.8. До уровня 20-25 Гр параметры *in vivo* и *in vitro* довольно схожи. 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 более жизнеспособных потомков оказывается более высокой вероятностью вторичной репарации  $q_2$ .

Таблица 3.3.

доза (in vivo)	$n_0$	$p_1$	$p_2$	$p_3$	$F_{\chi^2}$	$n_0$	$p_1$	$p_2$	$p_3$	$F_{\chi^2}$
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)	$n_1$	$p_1$	$p_2$	$p_3$	$F_{\chi^2}$	$n_1$	$p_1$	$p_2$	$p_3$	$F_{\chi^2}$
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

### 3.3. Кривые саногенеза

Одной из важных медико-биологических проблем является описание распада саногенетических или защитных факторов (кривой саногенеза), имеющих место при хроническом заболевании. В качестве аналитической модели кривой саногенеза в [10] была предложена функция  $S(t) = e^{-\eta t} \cos \tau t$  (рис.3.9), которую отличают по крайней мере два преимущества. С одной стороны, модель включает в себя взаимодействие убывающей экспоненты с периодически восстанавливающейся тригонометрической функцией, что интерпретируется как распад защитных сил в организме с периодически подключающейся его компенсацией. С другой стороны, параметр частичности для двойной обратной этой кривой может быть использован как параметр управления, поскольку в зависимости от него меняется время первого достижения этой кривой нулевого значения.

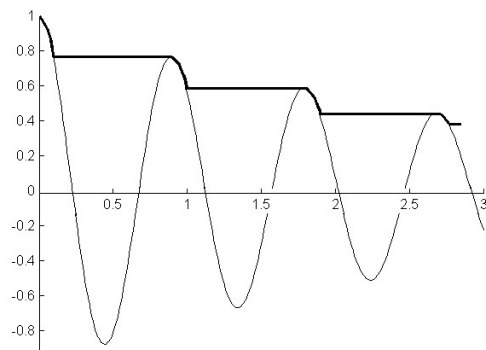


Рис. 3.9. Функция  $S(t) = e^{-\eta t} \cos \tau t$  и ее двойная правая обратная.

#### 3.3.1. Оценка параметров КМНС процесса

Модель  $S(t)$  может быть использована в анализе данных типа времени жизни или в оценивании параметров корреляционной функции комплексного марковского стационарного процесса (КМНС).

В качестве вещественной и мнимой части могут служить наблюдения по некоррелированным<sup>6</sup> признакам:  $x_j(t) = u_j(t) + iv_j(t)$ ,  $j = 1, \dots, n$ ,  $t = 1, \dots, k_j$ , где  $k_j$  число первых точек наблюдения у  $j$ -го индивида, с нулевыми математическими ожиданиями и ковариационной функцией

$$\mathcal{B}(t) = \mathbf{E}x_j(k)x_j(k+t)^* = \sigma^2 e^{-\eta|t| - i\tau t}, \quad \eta > 0.$$

Плотность распределения [11] комплексного  $n$ -мерного случайного вектора  $X$  с вектором средних  $\mu = (\mu_1, \dots, \mu_n)^T$  ковариационной матрицей  $\Sigma$  имеет вид

$$f(X) = \frac{1}{\pi^n |\Sigma|} \exp(-(\overline{X - \mu})^T \Sigma^{-1} (X - \mu)).$$

Обозначим через  $r = e^{-\eta - i\tau} = e^{-\eta}(\cos \tau - i \sin \tau)$ . Если рассматривать случайный вектор с компонентами, относящимися к одному индивиду в разные моменты времени  $X = (X_1, \dots, X_m)$ ,  $\mathbf{E}X_t = 0$ ,  $\mathbf{E}X_t X_{t+k}^* = \sigma^2 r^k$ , то корреляционную матрицу  $R = \frac{1}{\sigma^2} \Sigma$  можно представить в виде

$$R = \begin{bmatrix} 1 & r & r^2 & \dots & r^{m-1} \\ r^* & 1 & r & \dots & r^{m-2} \\ r^{*2} & r^* & 1 & \dots & r^{m-3} \\ \dots & \dots & \dots & \dots & \dots \\ r^{*m-1} & r^{*m-2} & r^{*m-3} & \dots & 1 \end{bmatrix}. \quad (3.7)$$

Ее определитель равен  $(1 - rr^*)^{m-1}$ . Для обратной матрицы, как в этом можно непосредственно убедиться, справедливо выражение

$$R^{-1} = \frac{1}{1 - rr^*} \begin{bmatrix} 1 & -r & 0 & 0 & \dots & 0 \\ -r^* & 1 + rr^* & -r & 0 & \dots & 0 \\ 0 & -r^* & 1 + rr^* & -r & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (3.8)$$

Соберем повторные наблюдения в один вектор

$$X = (x_{11}, \dots, x_{1k_1}, \dots, x_{n1}, \dots, x_{nk_n})^T$$

<sup>6</sup>В [10] в качестве компонент КМНС процесса предлагалось использовать значения первой и второй главных компонент. В силу того, что в разных временных точках факторы интерпретируются часто неодинаково, от этой идеи пришлось отказаться.

длиной  $N = \sum_{j=1}^n k_j$ . Корреляционная матрица этого вектора, как и ее обратная, имеет блочную структуру (рис.3.10) с матрицами (3.7), для обратной (3.8), порядка  $k_j$ ,  $j = 1, \dots, n$ , по главной диагонали. Определитель ковариационной матрицы равен  $\sigma^{2N}(1 - rr^*)^{N-n}$ , а функцию правдоподобия можно записать в виде

$$\mathcal{L}(\tau, \eta, \sigma^2) = (\pi\sigma^2)^{-N}(1 - rr^*)^{-N+n} \exp(Q),$$

где  $Q = \frac{-A_1 - rr^*A_2 + rA_3 + r^*A_3^*}{\sigma^2(1 - rr^*)}$ .

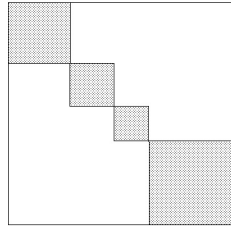


Рис. 3.10. Блочная структура корреляционной матрицы.

Статистики  $A_1$ ,  $A_2$  и  $A_3$  вида

$$A_1 = \sum_{j=1}^n \sum_{i=1}^{k_j} x_j(i)^* x_j(i), \quad A_2 = \sum_{j=1}^n \sum_{i=2}^{k_j-1} x_j(i)^* x_j(i),$$

$$A_3 = \sum_{j=1}^n \sum_{i=1}^{k_j-1} x_j(i)^* x_j(i+1)$$

соответственно имеют математические ожидания:

$$\mathbf{E}A_1 = \sum_{j=1}^n \sum_{i=1}^{k_j} \mathbf{E}x_j(i)^* x_j(i) = N\sigma^2,$$

$$\mathbf{E}A_2 = \sum_{j=1}^n \sum_{i=2}^{k_j-1} \mathbf{E}x_j(i)^* x_j(i) = (N - 2n)\sigma^2,$$

$$\mathbf{E}A_3 = \sum_{j=1}^n \sum_{i=1}^{k_j-1} \mathbf{E}x_j(i)^* x_j(i+1) = (N - n)\sigma^2 e^{-\eta+i\tau}.$$

Преобразуем параметр  $\eta = -\ln \theta$ ,  $\implies rr^* = \theta^2$ ,

$$\ln \mathcal{L} = -N \ln(\pi\sigma^2) - (N - n) \ln(1 - \theta^2) - \frac{f(\theta, \tau)}{\sigma^2(1 - \theta^2)},$$

где  $f(\theta, \tau) = A_1 - 2\theta Z(\tau) + A_2\theta^2$ ,  
 $Z(\tau) = \operatorname{Re}A_3 \cos \tau + \operatorname{Im}A_3 \sin \tau$ .

Соответствующие математические ожидания имеют вид:

$$\begin{aligned} \mathbf{E}f(\theta, \tau) &= N(1 - \theta^2)\sigma^2, & \mathbf{E}f'_\theta(\theta, \tau) &= -2n\theta\sigma^2, \\ \mathbf{E}f''_\theta(\theta, \tau) &= 2(N - 2n)\sigma^2, \\ \mathbf{E}Z(\tau) &= (N - n)\theta\sigma^2, & \mathbf{E}Z'_\tau(\tau) &= 0, \\ \mathbf{E}Z''(\tau) &= -(N - n)\theta\sigma^2. \end{aligned}$$

Вычислим частные производные первого порядка от логарифма функции  $\mathcal{L}$ :

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial \tau} &= \frac{2\theta Z'(\tau)}{\sigma^2(1 - \theta^2)}, \\ \frac{\partial \ln \mathcal{L}}{\partial \sigma^2} &= -\frac{N}{\sigma^2} + \frac{f(\theta, \tau)}{\sigma^4(1 - \theta^2)}, \\ \frac{\partial \ln \mathcal{L}}{\partial \theta} &= \frac{2\theta(N - n)}{1 - \theta^2} - \frac{f'_\theta(\theta, \tau)}{\sigma^2(1 - \theta^2)} - \frac{2\theta f(\theta, \tau)}{\sigma^2(1 - \theta^2)^2}. \end{aligned}$$

Убедимся, что математические ожидания этих производных равны нулю. В [10] показано, что оценка параметра  $\tau$  удовлетворяет уравнению  $\operatorname{tg} \hat{\tau} = \operatorname{Im}A_3/\operatorname{Re}A_3$ , а в случае полных данных ( $k_j = k$ ) и при справедливости  $A_1(k - 2) = A_2k$  оценки остальных параметров имеют соответственно вид

$$\hat{\eta} = -\ln \frac{Z(\hat{\tau})k}{A_1(k - 1)} \quad \text{и} \quad \hat{\sigma}^2 = \frac{A_1}{kn}. \quad (3.9)$$

В случае неполных данных  $\hat{\sigma}^2 = \frac{f(\hat{\theta}, \hat{\tau})}{N(1 - \hat{\theta}^2)}$ , а  $\hat{\theta} = e^{-\hat{\eta}}$  является решением кубического уравнения

$$A_2(N - n)\theta^3 - Z(\hat{\tau})(N - 2n)\theta^2 - (NA_2 + nA_1)\theta + Z(\hat{\tau})N = 0.$$



### 3.3.2. Вычисление дисперсии оценок

Вычислим частные производные второго порядка от логарифма функции правдоподобия  $\mathcal{L}$  и их математические ожидания.

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}}{\partial \tau^2} &= \frac{2\theta Z''(\tau)}{\sigma^2(1-\theta^2)}, & \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \tau^2} &= -\frac{2\theta^2(N-n)}{1-\theta^2}, \\ \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \tau \partial \sigma^2} &= 0, & \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \tau \partial \theta} &= 0, \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \sigma^2} &= \frac{N}{\sigma^4} - \frac{2f(\theta, \tau)}{\sigma^6(1-\theta^2)}, \\ \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \sigma^2} &= \frac{N}{\sigma^4} - \frac{2N(1-\theta^2)\sigma^2}{\sigma^6(1-\theta^2)} = -\frac{N}{\sigma^4}.\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \theta} &= \frac{f'_\theta(\theta, \tau)(1-\theta^2) + 2\theta f(\theta, \tau)}{\sigma^4(1-\theta^2)^2}, \\ \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \theta} &= \frac{-2n\theta\sigma^2(1-\theta^2) + 2\theta N(1-\theta^2)\sigma^2}{\sigma^4(1-\theta^2)^2} = \frac{2\theta(N-n)}{\sigma^2(1-\theta^2)}.\end{aligned}$$

Покажем, что

$$\mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} = -\frac{2(N-n)(1+\theta^2)}{(1-\theta^2)^2}.$$

В выражении для первой производной

$$\frac{\partial \ln \mathcal{L}}{\partial \theta} = \frac{2(N-n)\theta\sigma^2 - f'_\theta(\theta, \tau)}{\sigma^2(1-\theta^2)} - \frac{2\theta f(\theta, \tau)}{\sigma^2(1-\theta^2)^2}$$

дифференцирование первого слагаемого приводит к выражению

$$\frac{(1-\theta^2)(2(N-n)\sigma^2 - f''_\theta) - (-2\theta)(2(N-n)\theta\sigma^2 - f'_\theta)}{\sigma^2(1-\theta^2)^2},$$

математическое ожидание которого равно

$$\begin{aligned}\frac{(1-\theta^2)(2(N-n) - 2(N-2n)) + 2\theta(2(N-n)\theta + 2n\theta)}{(1-\theta^2)^2} &= \\ &= \frac{2(n(1-\theta^2) + 2N\theta^2)}{(1-\theta^2)^2}.\end{aligned}$$

(3.10)

Пусть  $U(f, \theta) = (f + \theta f')(1 - \theta^2)^2 - 2\theta(1 - \theta^2)(-2\theta)f$ , тогда

$$\begin{aligned} \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} &= -\frac{2U(f, \theta)}{\sigma^2(1 - \theta^2)^4} = \\ &= -\frac{2}{\sigma^2(1 - \theta^2)^3} \{(f + \theta f')(1 - \theta^2) + 4\theta^2 f\}, \\ \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} &= -\frac{2}{\sigma^2(1 - \theta^2)^3} \{(N(1 - \theta^2)\sigma^2 + \\ &+ \theta(-2n\theta\sigma^2))(1 - \theta^2) + 4\theta^2 N(1 - \theta^2)\sigma^2\}. \end{aligned}$$

Сокращая на  $\sigma^2$  и  $(1 - \theta^2)$ , получаем

$$\begin{aligned} &-\frac{2}{(1 - \theta^2)^2} \{N(1 - \theta^2) - 2n\theta^2 + 4N\theta^2\} = \\ &= -\frac{2}{(1 - \theta^2)^2} \{N + 3N\theta^2 - 2n\theta^2\}. \end{aligned} \quad (3.11)$$

Складывая (3.10) и (3.11), получаем выражение

$$\begin{aligned} \mathbf{E} \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} &= \frac{2}{(1 - \theta^2)^2} \{n - n\theta^2 + 2N\theta^2 - N - 3N\theta^2 + 2n\theta^2\} = \\ &= -\frac{2(N - n)(1 + \theta^2)}{(1 - \theta^2)^2}. \end{aligned}$$

Собираем компоненты информационной матрицы

$$\begin{aligned} I &= - \begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \tau^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \tau \partial \sigma^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \tau \partial \theta} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \tau} & \frac{\partial^2 \ln \mathcal{L}}{\partial (\sigma^2)^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \sigma^2 \partial \tau} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \tau} & \frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \sigma^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2} \end{bmatrix} = \\ &= \begin{bmatrix} \frac{2\theta^2(N-n)}{1-\theta^2} & 0 & 0 \\ 0 & \frac{N}{\sigma^4} & -\frac{2\theta(N-n)}{\sigma^2(1-\theta^2)} \\ 0 & -\frac{2\theta(N-n)}{\sigma^2(1-\theta^2)} & \frac{2(N-n)(1+\theta^2)}{(1-\theta^2)^2} \end{bmatrix} \end{aligned}$$

и получаем асимптотические дисперсии оценок:

$$\begin{aligned} \mathbf{D}\hat{\tau} &= \frac{1 - \theta^2}{2\theta^2(N - n)}, \quad \mathbf{D}\hat{\sigma}^2 = \frac{\sigma^4(1 + \theta^2)}{(N - (N - 2n)\theta^2)}, \\ \mathbf{D}\hat{\theta} &= \frac{N(1 - \theta^2)}{2(N - n)(N - (N - 2n)\theta^2)}. \end{aligned}$$

### 3.3.3. Пример из гематологии

В качестве 1-го и 2-го факторов рассмотрим некоррелированные признаки: протромбиновое время  $PTT$  и протеин  $C$  (или  $S$ ) в группах больных, которым назначение аспирина снято менее чем за три дня до операции (группа 0), от 4 до 8 дней (группа 1) и более 9 дней (группа 2). В анализ включаются все больные, у которых есть наблюдения за день до операции, во время операции и на следующий день после операции. Различия в оценках параметров  $\tau$  оказалось незначимыми, а по параметрам  $\theta = e^{-\tau}$  имеется значимое отличие группы 0 от групп 1 и 2 (табл. 3.4).

Таблица 3.4.

признаки	группа 0	группа 1	группа 2
$PTT, C$	0.2554	0.4576	0.4374
$PTT, S$	0.2202	0.3967	0.6242

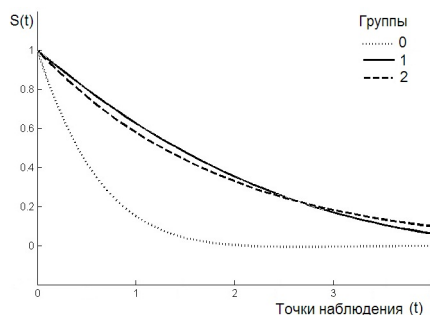


Рис. 3.11. Кривые саногенеза по признакам  $PTT$  и  $C$ .

При построении доверительных интервалов параметров кривой саногенеза, построенной по признакам  $PTT$  и  $C$ , в группе 0 отличие по параметру  $\theta_0 = e^{-\tau_0} \in [0.1010; 0.4098]$  нельзя объяснить случайностью, так как  $\hat{\theta}_1 = 0.4576$  и  $\hat{\theta}_2 = 0.4374$  соответственно в группах 1 и 2. На рис. 3.11 представлены кривые саногенеза, по которым видно, насколько существенно отличаются корреляционные связи в группе 0 с поздней отменой аспирина. Для признаков  $PTT$  и  $S$  все аналогично,  $\theta_0 = e^{-\tau_0} \in [0.0636; 0.3769]$ ,  $\hat{\theta}_1 = 0.3967$  и  $\hat{\theta}_2 = 0.6248$ .

## 3.4. Модели гамма-распределения

### 3.4.1. Синонимия распределений

Одной из распространенных проблем в статистическом анализе биометрических данных является ограничение возможностей методов, основанных на проверке гипотез о равенстве средних. Особенно актуально это становится при исследовании данных типа „мышца“ — выборочных наблюдений, в которых значимый вклад в дисперсию вносят факторы сжатия-растяжения [30].

Например, как данные типа „мышца“ можно рассматривать результаты морфологических измерений высоты клеток секреторного эпителия альвеолы молочной железы (лактоцитов) у лактирующих мышей [21], так как форма этих клеток постоянно меняется в зависимости от стадии выработки секрета. В экспериментах по исследованию морфологии лактоцитов при введении лабораторным мышам препаратов простагландина  $F_{2\alpha}$  или тиреолиберина были обнаружены значимые эффекты увеличения высоты эпителия по сравнению с контролем, хотя влияние этих препаратов на лактацию различно. В ситуации, когда проверка гипотез о равенстве средних по сути оказывается бесполезной, продолжение исследования обычно сопровождается проверкой гипотез об одинаковости функций распределения и при значимом их отличии — поиском адекватной статистической модели.

Замечено, что для данных типа „мышца“ характерно удовлетворительное согласие эмпирического распределения с гамма-распределением. Особенностью последнего является возможность его аппроксимации степенным гамма-распределением с достаточно широким спектром степенных параметров [30], т. е. при удовлетворительном согласии с гамма-распределением можно получить удовлетворительное согласие сразу с целым семейством синонимичных степенных гамма распределений, плотность которых имеет вид

$$f(x|\kappa, \beta, \lambda) = \frac{\kappa}{\beta^\lambda \Gamma(\lambda)} x^{\kappa\lambda-1} e^{-\frac{x^\kappa}{\beta}}. \quad (3.12)$$

Гамма распределение и распределение Вейбулла являются существенными частными случаями при  $\kappa = 1$  и  $\lambda = 1$  соответственно. Для сравнительного анализа приходится использоваться либо комплекс оценок, либо оценки параметров *номинативного* распределения,

которое можно было бы назвать характерным представителем [5]. В качестве последнего рассматривается распределение с минимальной дифференциальной энтропией.

При воздействии простагландина  $F_{2\alpha}$  и в контрольной группе для каждого отдельно взятого животного закон распределения высоты лактоцитов удовлетворяет семейству синонимичных степенных гамма распределений [30]. В исследовании влияния тиреолиберина имело место такое же соответствие. Неопределенность, связанная с одинаковым эффектом увеличения средней высоты лактоцитов при воздействии простагландина  $F_{2\alpha}$  и тиреолиберина, устраняется при помощи оценки параметров номинативных распределений. Получено, что указанные препараты значимо влияют на оценки параметров  $\lambda$ ,  $\beta$  и  $\kappa$ , причем противоположным образом, что, в общем, может быть объяснено особенностями их фармакологического действия.

### 3.4.2. Информационные метрики

Выражение, равное основной компоненте функции правдоподобия с обратным знаком, называется *дифференциальной энтропией*,

$$H_{ii} = - \int_{-\infty}^{\infty} \ln f_i(x) f_i(x) dx. \quad (3.13)$$

Через внесение разнообразия в индексы функций определим *смешанную дифференциальную энтропию*

$$H_{ij} = - \int_{-\infty}^{\infty} \ln f_j(x) f_i(x) dx. \quad (3.14)$$

С помощью  $H_{ij}$  выражается [17] расстояние между непрерывными распределениями  $F_i(x)$  и  $F_j(x)$ :

$$I_1 = \int_{-\infty}^{+\infty} (f_i(x) - f_j(x)) \ln \frac{f_i(x)}{f_j(x)} dx = (H_{ij} - H_{ii}) + (H_{ji} - H_{jj}).$$

Разность  $H_{ij} - H_{ii}$  (или  $H_{ji} - H_{jj}$ ) есть отклонение смешанной энтропии от истинной, а разность  $H_{ij} - H_{jj}$  (или  $H_{ji} - H_{ii}$ ) как от-

клонение смешанной энтропии от гипотетической называется расстоянием Кульбака-Лейблера.

Рассмотрим дополнительную метрику разнообразия распределений  $F_i(x)$  и  $F_j(x)$

$$I_2 = \int_{-\infty}^{+\infty} (f_i(x) - f_j(x)) \ln(f_i(x)f_j(x)) dx = (H_{ii} - H_{jj}) + (H_{ij} - H_{ji}),$$

состоящую из отклонения дифференциальных энтропий друг от друга и из отклонения смешанных энтропий друг от друга. Очевидно,  $I_2 = (H_{ij} - H_{jj}) - (H_{ji} - H_{ii})$ , то есть метрика  $I_2$  не является расстоянием в обычном смысле и равна нулю, когда одинаковы разности между дифференциальными энтропиями распределений и смешанными энтропиями, обусловленными их ошибочным предположением. Расстояние  $I_1$  будем называть *кинетической* метрикой, а характеристику  $I_2$  *потенциальной* метрикой разнообразия распределений.

### 3.4.3. Параметризация синонимичных степенных гамма-распределений

Пусть случайная величина  $\xi$  имеет гамма-распределение  $\gamma(\alpha, \lambda)$  с параметром масштаба  $\alpha$  (интенсивности) и параметром формы  $\lambda$  (экстенсивности). Тогда случайная величина  $X = \xi^{\frac{1}{\kappa}}$ ,  $\kappa > 0$ , имеет *степенное гамма-распределение*  $\gamma_p(\alpha, \lambda, \kappa)$  с плотностью

$$f_p(x|\alpha, \lambda, \kappa) = \frac{\kappa \alpha^\lambda}{\Gamma(\lambda)} x^{\kappa \lambda - 1} e^{-\alpha x^\kappa}, \quad (3.15)$$

где через  $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$  обозначена гамма-функция.

Для величин  $X_1$  и  $X_2$ , имеющих распределение  $\gamma_p(\alpha_1, \lambda_1, \kappa_i)$  и  $\gamma_p(\alpha_2, \lambda_2, \kappa_2)$  соответственно, смешанная энтропия  $H_{12}$  имеет вид

$$H_{12} = -(\ln \kappa_2 + \lambda_2 \ln \alpha_2 - \ln \Gamma(\lambda_2)) - (\kappa_2 \lambda_2 - 1) \mathbf{E} \ln X_1 + \alpha_2 \mathbf{E} X_1^{\kappa_2}, \quad (3.16)$$

где логарифмический и степенной моменты равны соответственно

$$\mathbf{E} \ln X_1 = \frac{1}{\kappa_1} (\psi(\lambda_1) - \ln \alpha_1), \quad \mathbf{E} X_1^{\kappa_2} = \frac{\Gamma(\lambda_1 + \frac{\kappa_2}{\kappa_1})}{\Gamma(\lambda_1) \alpha_1^{\frac{\kappa_2}{\kappa_1}}}. \quad (3.17)$$

Через  $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$  обозначена дигамма-функция или производная от логарифма гамма-функции.

**Определение 18.** *Распределение с плотностью  $f_2(x)$  синонимично с односторонним уровнем синонимии  $\delta_1$  распределению с плотностью  $f_1(x)$ , то есть  $f_2(x) \stackrel{\delta_1}{\succ} f_1(x)$ , если  $H_{12} - H_{11} < \delta_1$ . Аналогично определяем  $f_2(x) \stackrel{\delta_2}{\prec} f_1(x)$ . Если  $H_{12} - H_{11} < \delta_1$  и  $H_{21} - H_{22} < \delta_2$  одновременно, то распределения с плотностями  $f_1(x)$  и  $f_2(x)$  синонимичны  $f_1(x) \stackrel{\delta}{\asymp} f_2(x)$  с уровнем синонимии  $\delta = \delta_1 + \delta_2$ .*

Пусть  $X_1$  имеет распределение  $f_1(x) = \gamma_p(x|\alpha_1, \lambda_1, \kappa_1)$ . Посредством минимизации  $H_{12} - H_{11} \rightarrow \min$  найдем параметры синонимичного степенного гамма-распределения  $f_2(x) = \gamma_p(x|\alpha_2, \lambda_2, \kappa_2)$ .

**Предложение 12.** *Система нормальных уравнений для смешанной дифференциальной энтропии равносильна системе уравнений, связывающей степенные и логарифмические моменты степенных гамма распределений, а также их ковариации, то есть*

$$\Leftrightarrow \begin{cases} \frac{\partial H_{12}}{\partial \alpha_2} = 0 \\ \frac{\partial H_{12}}{\partial \kappa_2} = 0 \\ \frac{\partial H_{12}}{\partial \lambda_2} = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{\lambda_2}{\alpha_2} = \mathbf{E}X_2^{\kappa_2} = \mathbf{E}X_1^{\kappa_2}; \\ \alpha_2^{-1} = \mathbf{cov}(X_2^{\kappa_2}, \ln X_2^{\kappa_2}) = \mathbf{cov}(X_1^{\kappa_2}, \ln X_1^{\kappa_2}); \\ \mathbf{E} \ln X_1 = \mathbf{E} \ln X_2. \end{cases} \quad (3.18)$$

**Доказательство.** Дифференцирование  $H_{12}$  по параметру  $\alpha_2$  приводит к выражению

$$\frac{\partial H_{12}}{\partial \alpha_2} = -\frac{\lambda_2}{\alpha_2} + \mathbf{E}X_1^{\kappa_2} = 0 \Leftrightarrow \mathbf{E}X_1^{\kappa_2} = \frac{\lambda_2}{\alpha_2} = \mathbf{E}X_2^{\kappa_2}, \quad (3.19)$$

и первое утверждение доказано. При дифференцировании по параметрам  $\kappa_2$  и  $\lambda_2$  воспользуемся выражением

$$H_{12} = - \int_0^{\infty} \ln f_2(x|\alpha_2, \lambda_2, \kappa_2) f_1(x|\alpha_1, \lambda_1, \kappa_1) dx.$$

Вычислим частную производную

$$\begin{aligned}
\frac{\partial H_{12}}{\partial \kappa_2} &= - \int_0^{\infty} \frac{\partial f_2(x|\alpha_2, \lambda_2, \kappa_2)}{\partial \kappa_2} \frac{f_1(x|\alpha_1, \lambda_1, \kappa_1)}{f_2(x|\alpha_2, \lambda_2, \kappa_2)} dx = \\
&= - \int_0^{\infty} \left( \frac{1}{\kappa_2} + \lambda_2 \ln x - \frac{\alpha_2}{\kappa_2} \ln x^{\kappa_2} \cdot x^{\kappa_2} \right) f_1(x) dx = \\
&= - \frac{1}{\kappa_2} - \frac{\lambda_2}{\kappa_2} \mathbf{E} \ln X_1^{\kappa_2} + \frac{\alpha_2}{\kappa_2} \mathbf{E}(\ln X_1^{\kappa_2} \cdot X_1^{\kappa_2}) = 0.
\end{aligned}$$

Отсюда получаем

$$\alpha_2^{-1} = \mathbf{E}(\ln X_1^{\kappa_2} \cdot X_1^{\kappa_2}) - \frac{\lambda_2}{\kappa_2} \cdot \mathbf{E} \ln X_1^{\kappa_2}.$$

При равенстве степенных моментов (3.19) имеем

$$\alpha_2^{-1} = \mathbf{cov}(X_1^{\kappa_2}; \ln X_1^{\kappa_2}).$$

$$\begin{aligned}
\frac{\partial H_{12}}{\partial \lambda_2} &= - \int_0^{\infty} (\ln \alpha_2 + \psi(\lambda_2) + \kappa_2 \ln x) f_1(x) dx = \\
&= - \ln \alpha_2 + \psi(\lambda_2) - \kappa_2 \mathbf{E} \ln X_1 = 0 \iff \\
&\iff \mathbf{E} \ln X_1 = \frac{\psi(\lambda_2) - \ln \alpha_2}{\kappa_2} = \mathbf{E} \ln X_2. \quad \square
\end{aligned}$$

Система нормальных уравнений имеет тривиальное точное решение при  $\kappa_2 = \kappa_1$ .

**Следствие 1.** Пусть имеются случайные величины  $X_1$  и  $X_2$ , распределенные по  $\gamma_p(\alpha_1, \lambda_1, \kappa_1)$  и  $\gamma_p(\alpha_2, \lambda_2, \kappa_2)$ ,  $f_2(x) \stackrel{\delta_1}{\succ} f_1(x)$ . Тогда параметры  $\alpha_2$  и  $\lambda_2$  для любого  $\kappa_2 > 0$  и  $\theta = \frac{\kappa_2}{\kappa_1}$  выражаются через  $\alpha_1, \lambda_1$  и  $\kappa_1$  следующим образом:

$$\lambda_2 = (\theta (\psi(\lambda_1 + \theta) - \psi(\lambda_1)))^{-1}, \quad (3.20)$$

$$\alpha_2 = \frac{\lambda_2 \alpha_1^\theta \Gamma(\lambda_1)}{\Gamma(\lambda_1 + \theta)}. \quad (3.21)$$



**Доказательство.** Выражение (3.21) получаем из (3.18).

$$\mathbf{E}X_2^{\kappa_2} = \frac{\lambda_2}{\alpha_2} = \mathbf{E}X_2^{\kappa_2} \stackrel{(3.18)}{=} \mathbf{E}X_1^{\kappa_2} \stackrel{(3.17)}{=} \frac{\Gamma\left(\lambda_1 + \frac{\kappa_2}{\kappa_1}\right)}{\alpha_1^{\frac{\kappa_2}{\kappa_1}} \Gamma(\lambda_1)}.$$

Формулу (3.20) получаем, подставляя смешанный логарифмический момент

$$\mathbf{E}(X_1^{\kappa_2} \cdot \ln X_1^{\kappa_2}) = \frac{\kappa_2}{\kappa_1} \cdot \frac{\Gamma\left(\lambda_1 + \frac{\kappa_2}{\kappa_1}\right)}{\alpha_1^{\frac{\kappa_2}{\kappa_1}} \Gamma(\lambda_1)} \left( \psi\left(\lambda_1 + \frac{\kappa_2}{\kappa_1}\right) - \ln \alpha_1 \right),$$

а также  $\mathbf{E}X_1^{\kappa_2}$  и  $\mathbf{E} \ln X_1^{\kappa_2}$  из (3.17) в выражение

$$\begin{aligned} \alpha_2^{-1} &= \mathbf{cov}(X_2^{\kappa_2}, \ln X_2^{\kappa_2}) \stackrel{(3.18)}{=} \mathbf{cov}(X_1^{\kappa_2}, \ln X_1^{\kappa_2}) = \\ &= \mathbf{E}(X_1^{\kappa_2} \cdot \ln X_1^{\kappa_2}) - \mathbf{E}X_1^{\kappa_2} \mathbf{E} \ln X_1^{\kappa_2}. \quad \square \end{aligned}$$

Таким образом, для оценки параметров номинативного распределения сначала по выборочным данным оцениваются параметры гамма-распределения  $\gamma_p(\alpha_1, \lambda_1, 1)$ , а затем, из синонимичных степенных гамма-распределений  $\gamma_p(\alpha_2, \lambda_2, \kappa_2)$ , согласно след. 1, выбирается распределение с минимальной дифференциальной энтропией  $H_{22}$ . В [30] было замечено, что в случае  $\lambda > 1$  для степенного параметра номинативного распределения справедливо ограничение  $0.75 < \kappa \leq 1$ . При достаточно больших  $\lambda$  минимальность  $H_{22}$  соответствует максимуму для метрики  $I_2$ , откуда было получено предельное значение  $\kappa = 0.75$  [5].

### 3.5. Применение синонимичных распределений в физиологии

Исследование влияния препаратов простагландина  $F_{2\alpha}$  и тиреолиберина было выполнено на 38 лактирующих белых мышах (питомник РАМН „Рапполово“), массой 30–35 г, соответственно по 15 и 23 животных в каждом эксперименте. Каждая самка с потомством содержалась в отдельной клетке со свободным доступом к пище и воде в стандартных условиях вивария при  $20 \pm 2^\circ\text{C}$ . Количество мышат в помете составляло от 2 до 12.

В первом эксперименте для изучения влияния простагландина  $F_{2\alpha}$  на становление лактации пяти животным первой опытной группы  $P_1$  его вводили однократно на 3–4-й день лактации в последний молозивный период лактогенеза, который начинается во время беременности и к 3–4-му дню лактации завершается. При изучении влияния простагландина на вскармливание потомства пяти животным второй опытной группы  $P_2$  вещество вводили в период лактопоза (10–11-й день лактации). В контрольной группе  $P_0$  пять животных вскармливали потомство без каких-либо препаратов.

Во втором эксперименте десяти мышам (группа  $T_1$ ) со 2-го по 12-й день лактации с трехдневной отменой препарата с 6 по 8-й день вводился тиреолиберин в концентрации  $10^{-6}$  ммоль/мл, другим четырем (группа  $T_2$ ) в концентрации  $10^{-10}$  ммоль/мл. Тиреолиберин в концентрации  $10^{-10}$  ммоль/мл в отличие от  $10^{-6}$  ммоль/мл не влияет на повышение уровня гормонов, вырабатываемых в щитовидной железе. Контрольную группу будем обозначать через  $T_0$  (пять мышей). В дни введения тиреолиберина четырем мышам вводился физиологический раствор (группа  $T_{ss}$ ).

С помощью системы интерактивного анализа изображения „ВидеоТест“ (ИСТА-ВидеоТест, Россия) проводился морфометрический анализ гистологических препаратов (полутонкие срезы) молочной железы мыши: измеряли максимальный ( $d_1$ ), минимальный ( $d_2$ ) диаметры альвеолы и высоту секреторного эпителия ( $W$ ) как расстояние между точками, находящимися на базальной и апикальной мембранах. Под диаметром альвеолы понимали величину  $d = \frac{d_1+d_2}{2}$ . На каждом предметном стекле располагали по 5–6 полутонких срезов фиксированной ткани молочной железы. Количественные показатели в микрометрах измерялись в 10 случайно выбранных полях зрения. В каждом поле зрения измеряли диаметр всех имеющихся альвеол.

В первом эксперименте на одном предметном стекле общее количество измерений для  $d_1$  и  $d_2$  составляло от 50 до 60. На каждом срезе случайным образом проводили 40 измерений высоты лактоцитов, т. е. на одном предметном стекле общее количество измерений для  $W$  составляло от 200 до 240. Во втором эксперименте в каждом препарате обычно находилось от 15 до 35 альвеол, высота  $W$  измерялась у 50 лактоцитов в разных альвеолах.

### 3.5.1. Измерения и анализ прибавки веса

Изменение общего веса потомства у каждой самки определяли путем ежедневного трехкратного с усреднением взвешивания мышат на электронных весах „Sartorius-1602 MP“ с точностью измерения 10 мг. Через  $S(t)$  будем обозначать средний вес одного мышонка в  $t$ -й день. При исследовании влияния тиреолиберина в связи с отменой препарата с 6 по 8-й день данные о взвешивании отсутствуют. При введении тиреолиберина наблюдается более выраженная динамика привесов  $\frac{S(t)}{S(t-1)}$  в первые 5 дней этого эксперимента представлена на рис. 3.12, доверительный уровень вероятности эффекта взаимодействия факторов времени и препарата равен  $p = 0.00002$ .

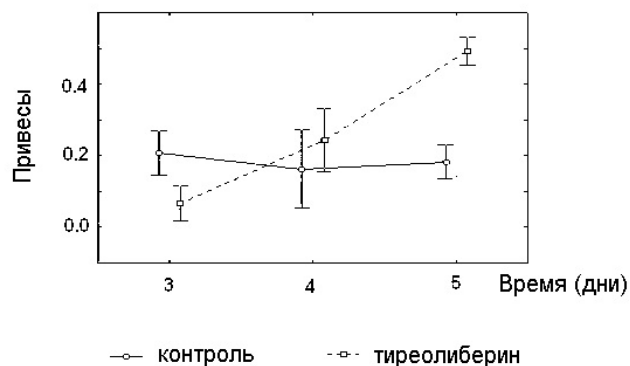


Рис. 3.12. Зависимость привесов от тиреолиберина.

Поскольку увеличение среднего веса одного мышонка носит экспоненциальный характер  $S(t) = \sigma e^{\mu t} + \varepsilon$ , где  $\varepsilon$  — случайная ошибка, в качестве итоговой характеристики будем рассматривать параметр интенсивности привесов  $\mu = (\ln S(t))'$ . Его оценкой является логарифм усредненной по времени характеристики привесов  $\frac{S(t+1)}{S(t)}$ . При введении тиреолиберина  $\mu$  вычисляется по первым пяти дням, простагландина  $F_{2\alpha}$  в период лактогенеза — с 3-го по 7-й день, в период лактопоза использовалось последнее отношение весов.

### 3.5.2. Проверка однородности морфологических данных

По данным о высоте секреторного эпителия  $W$  у каждого животного по отдельности вычислялись оценки максимального правдоподобия параметров масштаба  $B$  и формы  $\Lambda$  гамма-распределения и проверялось его согласие. Параметры  $\beta$  и  $\lambda$  синонимичных степенных гамма-распределений по заданному  $\kappa$  вычислялись по формулам (3.20) и (3.21). Параметр  $\kappa$  номинативного распределения оценивался при помощи программы, написанной на *VisualC++* 6.0, из условия минимизации дифференциальной энтропии

$$H(\kappa, \beta, \lambda) = \lambda + \ln \Gamma(\lambda) - \lambda \psi(\lambda) - \ln \kappa + \frac{\psi(\lambda) + \ln \beta}{\kappa} \quad (3.22)$$

степенного гамма-распределения, синонимичного гамма-распределению. После того как получен параметр  $\kappa$  номинативного распре-

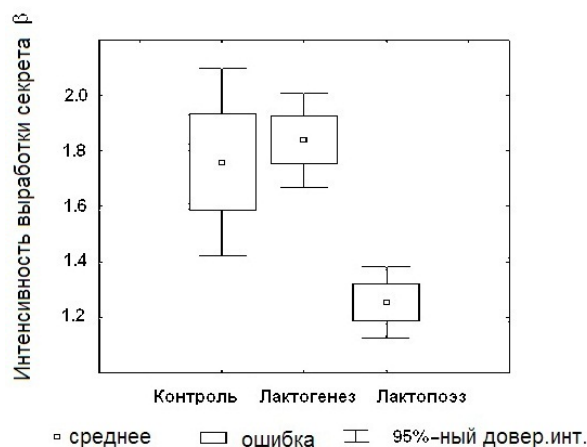


Рис. 3.13. Зависимость  $\beta$  от простагландина.

деления, оценка параметра  $\beta$  может быть пересчитана с учетом всех повторных наблюдений. Из (3.18) следует, что

$$\beta = \mathbf{E}(W^\kappa - \mathbf{E}W^\kappa)(\ln W^\kappa - \mathbf{E} \ln W^\kappa).$$

Для каждого наблюдения  $W$  рассмотрим его преобразование

$$Z = \kappa(W^\kappa - \mathbf{E}W^\kappa)(\ln W - \mathbf{E} \ln W), \quad (3.23)$$

где в качестве оценок математических ожиданий  $\mathbf{E}W^\kappa$  и  $\mathbf{E} \ln W$  могут рассматриваться соответственно среднее арифметическое от  $W^\kappa$  и логарифм среднего гармонического от  $W$ . Поскольку среднее значение  $Z$  является оценкой параметра  $\beta$ , возникает возможность сравнивать параметры интенсивности выработки секрета на выборках большего объема.

При исследовании взаимосвязи между признаками использовались коэффициенты корреляции Пирсона  $R$ , Спирмена  $R_s$  и Кендалла  $R_\tau$ . Значимости отличия по средним показателям вычислялись по критериям Фишера (однофакторный дисперсионный анализ) и Манна—Уитни. Динамику привесов исследовали при помощи дисперсионного анализа для планов с расщепленными блоками [29].

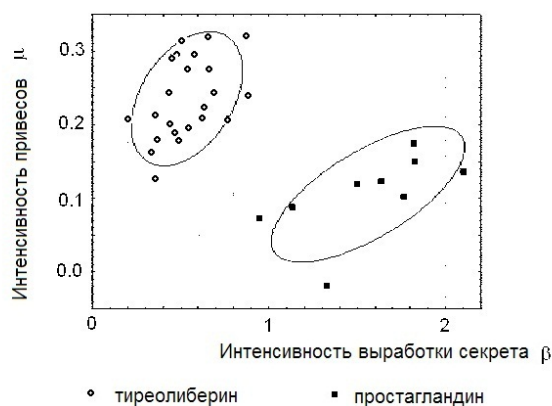


Рис. 3.14. Зависимость привесов от интенсивности  $\beta$ .

### 3.5.3. Результаты и обсуждения

При проверке гипотез однородности морфологических данных в эксперименте с простагландином получаем значимое увеличение диаметра альвеолы при введении простагландина:  $130.97 \pm 3.6$  в

группах  $P_1, P_2$  по сравнению с группой  $P_0$  контроля  $113.45 \pm 4.83$ , а также значимое отличие средней высоты эпителия во всех трех группах:  $20.33 \pm 0.41$  (введение простагландина в период лактопоза),  $29.18 \pm 0.55$  (в период лактогенеза),  $17.93 \pm 0.74$  (контрольная группа).

Во втором эксперименте между группами  $T_0$  и  $T_{ss}$  значимых отличий по морфологическим характеристикам  $W, d_1$  и  $d_2$  не было обнаружено ( $p > 0.3$ ). При уменьшении концентрации тиреолиберина диаметры альвеол увеличиваются с  $58.93 \pm 1.10$  до  $77.91 \pm 1.70$ , причем, значимо.

По высоте эпителия группы  $T_1$  и  $T_2$  с разными концентрациями тиреолиберина значимо не отличаются:  $8.08 \pm 0.19$  при концентрации  $10^{-6}$  ммоль/мл и  $7.62 \pm 0.26$  при  $10^{-10}$  ммоль/мл. Объединяя однородные данные, получаем среднюю высоту эпителия в группе введения тиреолиберина  $7.95 \pm 0.16$ , которая значимо выше средней высоты эпителия  $5.73 \pm 0.13$  в контроле, а также средний диаметр альвеолы  $67.8 \pm 1.13$ , который значимо выше среднего диаметра альвеолы  $48.66 \pm 0.87$  в группе контроля  $T_0$ .

Таким образом, в обоих экспериментах введение препаратов тиреолиберина и простагландина  $F_{2\alpha}$  привело к *одинаково значимому увеличению средней высоты эпителия и диаметра альвеолы*.

#### 3.5.4. Интерпретация параметров номинативного распределения

Сопоставим результаты измерений средней высоты эпителия  $\bar{w}$ , усредненного диаметра альвеолы  $\bar{d}$  и оценку логарифмической производной  $\mu$  среднего веса, приходящегося на одного мышонка, с оценками параметров  $\kappa, \lambda$  и  $\beta$  номинативных распределений, полученных из исследования эмпирических распределений высоты  $W$  каждого животного (табл. 3.5).

Положительные корреляции (рис. 3.14) между интенсивностью привесов  $\mu$  и параметром интенсивности  $\beta$  дают все основания для интерпретации параметра  $\beta$  как характеристики интенсивности выработки секрета. В первом эксперименте корреляция Пирсона равна  $R = 0.62, p = 0.035$ , во втором  $R = 0.47, p = 0.01$ . Значимости по Кендаллу равны соответственно  $p = 0.012$  и  $p = 0.022$ .

Параметр  $\lambda$  также отвечает за увеличение высоты эпителия, но не связанного с выработкой секрета. Это хорошо видно из резуль-

Таблица 3.5.

Код группы	$\bar{w}$	$\kappa$	$\lambda$	$\beta$	$\bar{d}$	$\mu$
Тиреолиберин (контроль)						
1	5.102	0.752	6.921	0.481	67.62	0.179
1	5.141	0.750	15.610	0.217	45.64	0.209
1	6.343	0.751	8.403	0.468	49.91	0.188
1	6.090	0.751	8.596	0.444	48.39	0.203
1	6.158	0.751	10.751	0.359	45.94	0.179
Тиреолиберин (физ. раствор)						
2	5.783	0.751	11.188	0.329	50.36	0.161
2	5.129	0.753	6.080	0.549	45.59	0.198
2	5.565	0.751	10.072	0.355	51.27	0.128
2	6.409	0.751	11.111	0.358	46.55	0.212
Тиреолиберин ( $10^{-6}$ ммоль/мл)						
3	6.750	0.753	6.466	0.635	54.95	0.223
3	11.909	0.751	10.233	0.618	47.28	0.209
3	10.971	0.751	8.517	0.697	64.03	0.243
3	9.514	0.752	6.948	0.766	65.58	0.209
3	6.223	0.753	5.849	0.660	69.84	0.319
3	6.280	0.752	6.818	0.571	56.89	0.294
3	8.533	0.754	5.622	0.871	57.40	0.319
3	7.381	0.751	9.522	0.463	60.79	0.294
3	5.935	0.752	7.806	0.479	64.23	0.297
3	7.271	0.751	8.697	0.501	55.29	0.315
Тиреолиберин ( $10^{-10}$ ммоль/мл)						
4	8.411	0.751	8.931	0.545	67.50	0.277
4	7.250	0.753	6.453	0.672	77.11	0.275
4	7.744	0.754	5.172	0.879	81.08	0.237
4	7.076	0.751	10.093	0.424	85.68	0.242
Простагландин $F_{2\alpha}$ (лактопозэ)						
6	23.086	0.752	7.783	1.333	128.70	-0.018
6	21.705	0.752	6.578	1.504	118.79	0.120
6	16.572	0.751	8.520	0.950	143.61	0.070
Простагландин $F_{2\alpha}$ (контроль)						
5	18.487	0.755	4.827	1.817	105.39	0.176
5	12.462	0.753	5.777	1.127	121.51	0.089
Простагландин $F_{2\alpha}$ (лактогенез)						
7	28.067	0.753	5.710	2.103	137.07	0.137
7	23.023	0.753	6.338	1.630	122.49	0.122
7	28.908	0.752	6.718	1.827	130.64	0.149
7	36.751	0.751	8.366	1.759	136.06	0.103

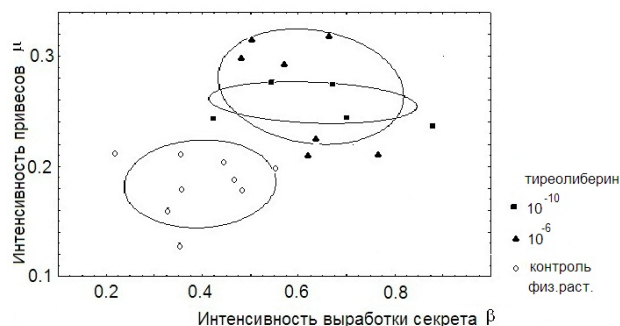


Рис. 3.15. Зависимость  $\beta$  от привесов в эксперименте 2.

татов микроскопического исследования ультраструктурных изменений лактоцитов. Введение простагландина  $F_{2\alpha}$  привело к изменению формы клеток: они приобрели более вытянутую цилиндрическую форму. Отмечались характерные признаки набухания в структуре митохондрий. В цитоплазме лактоцитов располагались крупные фагосомы, о которых известно, что эти органеллы не включаются непосредственно в секреторный процесс и появляются лишь в те периоды жизни железистого эпителия, когда возникает необходимость значительного разрушения и удаления клеточного материала при инволюции органа. Таким образом, ультраструктурные изменения свидетельствовали о возможном снижении синтетической активности и начальной стадии гибели секреторной клетки. Тиреолиберин, наоборот, приводит к уменьшению параметра  $\lambda$ , что интерпретируется как повышение синтетической активности.

Параметр  $\kappa$  интерпретируется как эластичность. Если ввести порядковый признак с градациями: 0 — введение простагландина в период лактопоза, 1 — в период лактогенеза, 2 — отсутствие влияния препарата, то можно обнаружить его значимые корреляции с параметром  $\kappa$  ( $R_r = 0.59$ ,  $p = 0.022$ ). Это означает, что чем меньше времени прошло с момента введения простагландина, тем меньше эластичность.

Введение тиреолиберина, наоборот, приводит к увеличению показателя эластичности ( $R_s = 0.42$ ,  $p = 0.048$ ), значимость по критерию Манна—Уитни равна  $p = 0.0508$ .



## 3.6. Двумерное гамма-распределение

К модели двумерного гамма-распределения пришлось обратиться в связи с тем, что по анализу усредненных иммунологических характеристик не удалось выявить никаких значимых различий в работе иммунитета онкологических больных с разными формами прогрессирования заболевания. Оказалось, что эффект лечения в большей степени зависит не от того, каким ресурсом обладает иммунитет, а от того, насколько интенсивно этот ресурс задействован.

### 3.6.1. Аддитивная модель

Рассмотрим три независимые гамма-распределенные случайные величины  $X_0, X_1, X_2$  с параметром масштаба, равным 1, и параметрами экстенсивности, равными  $\lambda_0, \lambda_1, \lambda_2$ , и построим из них две случайные величины

$$Y_1 = X_0 + X_1, \quad Y_2 = X_0 + X_2,$$

которые за счет аддитивности гамма-распределения останутся гамма-распределенными с параметрами формы  $\lambda_0 + \lambda_1$  и  $\lambda_0 + \lambda_2$ , но окажутся коррелированными с коэффициентом корреляции

$$\rho = \rho(Y_1, Y_2) = \frac{\lambda_0}{\sqrt{(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2)}},$$

так как  $\mathbf{E}Y_1Y_2 - \mathbf{E}Y_1\mathbf{E}Y_2 = \mathbf{E}(X_0 + X_1)(X_0 + X_2) -$

$$-(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2) = \mathbf{D}X_0 + (\mathbf{E}X_0)^2 + \lambda_0\lambda_1 + \lambda_0\lambda_2 + \lambda_1\lambda_2 -$$
$$-(\lambda_0 + \lambda_1)(\lambda_0 + \lambda_2) = \lambda_0.$$

Посмотрим на это с другой стороны. Предположим, что имеются две гамма-распределенные случайные величины  $Y_0$  и  $Y_1$  с единичными параметрами масштаба, равными единице, с параметрами формы

$$\Lambda_1 = \lambda_0 + \lambda_1, \quad \Lambda_2 = \lambda_0 + \lambda_2$$

и с коэффициентом корреляции  $\rho$ . Тогда следующим образом могут быть получены параметры скрытых экстенсивностей:

$$\lambda_0 = \rho\sqrt{\Lambda_1\Lambda_2}, \quad \lambda_1 = \Lambda_1 - \rho\sqrt{\Lambda_1\Lambda_2}, \quad \lambda_2 = \Lambda_2 - \rho\sqrt{\Lambda_1\Lambda_2}. \quad (3.24)$$

### 3.6.2. Плотность двумерного гамма-распределения

Совместная плотность независимых гамма-распределенных случайных величин  $X_0, X_1, X_2$  имеет вид

$$f_3(x_0, x_1, x_2) = C_1 x_0^{\lambda_0-1} x_1^{\lambda_1-1} x_2^{\lambda_2-1} e^{-(x_0+x_1+x_2)}, \quad (3.25)$$

$$C_1 = \frac{1}{\Gamma(\lambda_0)\Gamma(\lambda_1)\Gamma(\lambda_2)}, \quad x_0, x_1, x_2 > 0.$$

Введем новые переменные

$$\begin{cases} u = x_0 + x_1 \\ v = x_0 + x_2 \\ t = x_0 \end{cases} \iff \begin{cases} x_0 = t \\ x_1 = u - t \\ x_2 = v - t \end{cases}$$

с Якобианом преобразования

$$|J| = \begin{vmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Таким образом, обозначая через  $T = \min(u, v)$ , получаем выражение для совместной плотности величин  $Y_1$  и  $Y_2$

$$f_2(u, v) = C_1 \int_0^T t^{\lambda_0-1} (u-t)^{\lambda_1-1} (v-t)^{\lambda_2-1} e^{-(u+v-t)} dt. \quad (3.26)$$

Пусть  $u < v$ . Разложим экспоненту в ряд  $e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}$ , вынесем за знак интеграла  $C_2 = C_2(u, v) = \frac{1}{n!} e^{-(u+v)}$ ,  $C = C_1 C_2$ .

$$f_2(u, v) = C \sum_{n=0}^{\infty} \int_0^u t^{n+\lambda_0-1} (u-t)^{\lambda_1-1} (v-t)^{\lambda_2-1} dt.$$

Если заменить  $t$  на  $su$ , то получим  $f_2(u, v) =$

$$\begin{aligned} &= C \sum_{n=0}^{\infty} u^{r(n)-1} v^{\lambda_2-1} \int_0^1 s^{n+\lambda_0-1} (1-s)^{\lambda_1-1} \left(1 - \frac{su}{v}\right)^{\lambda_2-1} ds = \\ &= C \sum_{n=0}^{\infty} C_3(n) C_4(n) {}_2F_1\left(\alpha, \beta; \gamma; \frac{u}{v}\right), \text{ где } r(n) = n + \lambda_0 + \lambda_1, \\ {}_2F_1(\alpha, \beta; \gamma; z) &= \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1} (1-t)^{\gamma-\beta-1} (1-tz)^{-\alpha} dt \end{aligned}$$

гипергеометрическая функция Гаусса,

$$\begin{aligned} C_3(n) &= u^{n+\lambda_0+\lambda_1-1} v^{\lambda_2-1}, \quad C_4(n) = \frac{\Gamma(\beta)\Gamma(\gamma-\beta)}{\Gamma(\gamma)}, \\ \alpha &= -\lambda_2 + 1, \quad \beta = \lambda_0 + n, \quad \gamma = \lambda_0 + \lambda_1 + n. \end{aligned}$$

Выражение совместной плотности может использоваться для оценки параметров  $\lambda_0, \lambda_1, \lambda_2$  по методу максимального правдоподобия.

### 3.6.3. Оценки по методу моментов

Используя выражения для плотности (3.26) и для моментов гамма-распределения  $\gamma(1, \lambda)$ , которые имеют вид

$$\mu_2 = \lambda, \quad \mu_3 = 2\lambda, \quad \mu_4 = 3\lambda(\lambda + 2),$$

нетрудно показать, что центральные моменты двумерного гамма-распределения (3.26) с единичными параметрами масштаба, необходимые в задаче вычисления асимптотической дисперсии оценок параметров как функций от моментов [16], равны

$$\begin{aligned} \mu_{11} &= \lambda_0, \\ \mu_{12} &= \mu_{21} = 2\lambda_0, \\ \mu_{22} &= 2\lambda_0^2 + 6\lambda_0 + (\lambda_0 + \lambda_1)((\lambda_0 + \lambda_2)). \end{aligned} \tag{3.27}$$

Пусть имеется выборка наблюдений  $(x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})$  с плотностью распределения (3.26). Обозначим через  $\bar{x}_i, i = 0; 1$ , выборочное среднее, через  $m_{11}$  второй выборочный смешанный центральный момент.

Из (3.24) оценки по методу моментов можно представить в виде:  
 $\hat{\lambda}_0 = m_{11}$ ,  $\hat{\lambda}_i = H(\bar{x}_i, m_{11}) = \bar{x}_i - m_{11}$ ,  $i = 0; 1$ .

Так как  $\mu_2(m_{11}) = \frac{\mu_{22} - \mu_{11}^2}{n} + o\left(\frac{1}{n}\right)$ , из (3.27) получаем

$$\mathbf{D}\hat{\lambda}_0 = \mu_2(m_{11}) = \frac{\lambda_0^2 + 6\lambda_0 + (\lambda_0 + \lambda_1)((\lambda_0 + \lambda_2))}{n} + o\left(\frac{1}{n}\right). \quad (3.28)$$

Согласно [16], функция от моментов асимптотически нормальна с дисперсией, совпадающей с главным членом формулы

$$\mathbf{D}H(\bar{x}_i, m_{11}) = \mu_2(\bar{x}_i) - 2\mu_{11}(\bar{x}_i, m_{11}) + \mu_2(m_{11}) + o\left(\frac{1}{n}\right).$$

Используя (3.28), а также то, что  $\mu_2(\bar{x}_i) = \frac{\lambda_0 + \lambda_i}{n}$ , вычисляя

$$\mu_{11}(\bar{x}_i, m_{11}) = \frac{n-1}{n^2} \mu_{21} = \frac{2\lambda_0}{n} + o\left(\frac{1}{n}\right),$$

при  $i = 0; 1$  получаем выражение для дисперсии оценок

$$\mathbf{D}\hat{\lambda}_i = \frac{1}{n} \left( (\lambda_0 + \lambda_i)(2\lambda_0 + \lambda_1 + \lambda_2) + \lambda_0(\lambda_0 + 2) \right) + o\left(\frac{1}{n}\right).$$

### 3.6.4. Пример из иммунологии

В обследовании участвовали 160 больных, которые проходили лечение и состояли на учете в отделении биотерапии и трансплантации костного мозга НИИ онкологии им. проф. Н.Н.Петрова с 1998 по 2006 гг. В процессе вакцинотерапии и дальнейшего наблюдения больных осуществляли мониторинг состояния иммунной системы.

Относительное и абсолютное количество клеток, экспрессирующих CD3, CD4, CD8, CD20, CD16, CD25, CD38, CD71, CD95, HLA DR антигены определяли методом непрямой иммунофлюоресценции с использованием моноклональных антител. Концентрацию сывороточных иммуноглобулинов (Ig) M, G, A оценивали методом радиальной иммунодиффузии по Манчини, уровень циркулирующих иммунных комплексов (ЦИК) — методом осаждения полиэтиленгликолем. Для оценки функциональной активности Т-клеток изучали спонтанную активность мононуклеаров, пролиферативный ответ на митоген конканавалин

А (КонА) и фитогемагглютинин (ФГА), спонтанную НСТ-активность фагоцитов, фагоцитоз моноцитами и нейтрофилами частиц латекса (ФЧ-фагоцитарное число), интенсивность которого выражалась через индексный показатель (ФИ) активности этих клеток.

Таблица 3.6.

<i>IgG</i>	Группа	
	адьювантная	лечебная
$B_0$	1.27	2.4
$\Lambda_0$	10.37	5.58
$B_1$	1.33	1.93
$\Lambda_1$	9.99	6.83
$\rho$	0.23	0.50
$\lambda_0$	2.33	3.09
$\lambda_1$	8.04	2.45
$\lambda_2$	7.66	3.74

Исследования проводили перед вакцинацией (нулевая точка), перед каждой вакцинацией (точки с 1-й по 5-ю), через 1, 3, 6 и 12 месяцев после завершения вакцинации (с 6-й по 9-ю точки). Исходно больных подразделяли на адьювантную и лечебную группы, а по итогам наблюдения — на группы с прогрессированием заболевания и с ремиссией. Адьювантный (adjuvans) означает вспомогательный (о методе, лекарственном средстве и т.п.). В адьювантной группе вакцинация проводилась в профилактической целью [7].

Таким образом, могут быть рассмотрены пары положительно коррелированных признаков, относящихся к разным временным точкам, для которых не отвергается гипотеза согласия с гамма-распределением, и оценены параметры формы  $\lambda_0$ ,  $\lambda_1$ ,  $\lambda_2$ , интерпретируемые как базовая, апоптозно-эффекторная (АЭЭ) и регенерационно-стимулированная экстенсивность (РСЭ).

Например, в табл. 3.6 представлены оценки параметров по методу моментов двумерного гамма-распределения концентрации иммуноглобулинов *IgG* в адьювантной и лечебной группах в нулевой и первой точках.

Оценки параметров позволили выявить значимые отличия в динамике иммуноглобулина класса G, циркулирующих иммунных ко-

мплексов и клеток-киллеров в разных прогностических группах. Отличие наблюдается не по базовой экстенсивности, а по АЭЭ и РСЭ, которые значимо ниже в группе больных с метастазами.

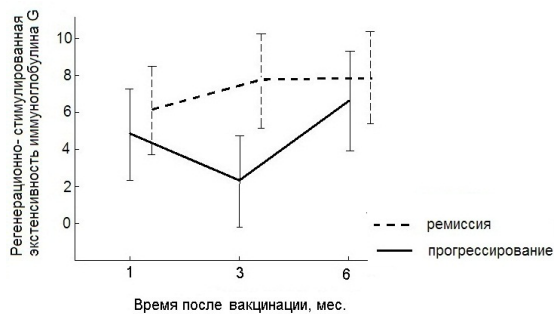


Рис. 3.16. Динамика РСЭ после вакцинации.

На рис. 3.16–3.17 представлены значения регенерационно-стимулированной и апоптозно-эффекторной экстенсивности, построенные по приращениям иммуноглобулина класса G в точках 0 и 1, 1 и 2, 2 и 3, в группах больных, у которых заболевание прогрессирует, и у больных с устойчивой ремиссией.

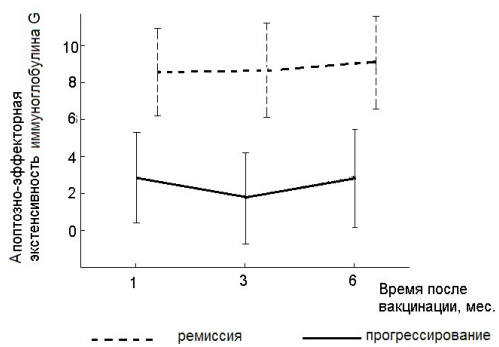


Рис. 3.17. Динамика АЭЭ после вакцинации.

## Заключение

Автономность, сцепленность и надежность синтеза компонент любой биосистемы являются ее естественными атрибутами, которые трудно поддаются прямому описанию. Косвенным образом они аккумулируются в феноменах, исследуемых в этой работе. Чтобы не наступать на одни и те же грабли, важно было найти математические структуры, которые непримитивно бы соотносились с действительностью, позволили бы четче выстраивать алгоритм статистического анализа и открывали бы самые существенные стороны изучаемых явлений.

При сегодняшнем своеобразном подведении итогов мне бы хотелось выразить глубокую признательность учителям: профессору Ермакову С.М., благодаря которому существует биометрическая специализация на кафедре статистического моделирования математико-механического факультета; основоположнику петербургской биометрической школы О.М.Калинину; автору первой книги „Анализ медико-биологических систем“ и моему научному руководителю А.Г.Барту; В.В.Некруткину, избавившему не одно поколение студентов от поверхностного отношения к математике, дружеское участие, внимание и помощь которого были всегда вовремя; самому авторитетному и известному в биометрических кругах биостатистику В.Н.Солнцеву.

Большая часть рассмотренных задач возникла в процессе совместной деятельности с исследователями в области биологии и медицины. Наиболее плодотворным оказалось сотрудничество с выдающимся специалистом в области генетики и радиобиологии профессором Института Цитологии Ю.Б.Вахтиным, с заслуженным кардиологом профессором ФГУ ФЦСКЭ Б.Б.Бондаренко и со многими другими: Е.П.Ивановой, Б.В.Мартыновым, Е.Н.Парийской, А.Г.Марковым, Е.В.Вербицкой и т.д.

Биометрическая тематика широко используется в учебном процессе. Замечательные результаты были получены при активном участии выпускников: И.Смирнова, И.Щербаковой, Е.Подхалюзинной, А.Коробейникова, К.Усевича и П.Грачевой. Талант, энергия, ответственность и целеустремленность учеников позволяют надеяться на то, что математические проблемы биостатистики не останутся без внимания, и можно с уверенностью сказать, что продолжение следует.

# Литература

- [1] **Алексеева Н.П.** Комбинаторный анализ двух форм скрытой периодичности категориальных последовательностей // Вестник СПбГУ. 2007. Сер.1. Вып.3. С. 55-64.
- [2] **Алексеева Н.П., Иванова Е.П., Бондаренко Б. Б. и др.** О способе улучшения прогнозирования вазоспазма лучевой артерии на основе симптомного расслоения популяций // Ученые записки СПбГМУ им.акад.И.П.Павлова. 2009. Т. XV. № 4. С. 59-62.
- [3] **Алексеева Н.П., Алексеев А.О., Вахтин Ю.Б.** Проблема индуктивности закона Харди-Вайнберга и прямо-двойственная сцепленность марковских процессов на конечных геометриях // Вестник С.-Петербур.ун-та. 2005. Сер.1. Вып.1. С. 3-12.
- [4] **Алексеева Н.П., Алексеев А.О., Вахтин Ю.Б., Кравцов В.Ю., Кузоватов С.Н., Скорикова Т.И.** Исследование динамики роста числа ядерных аномалий рабдомиосаркомы РА-23 при увеличении дозы острого редкоизирующего облучения на основе модели реинтрантно-биномиального распределения // Цитология. 2008. Т.50. № 6. С. 526-531.
- [5] **Алексеева Н.П., Алексеев А.О.** Кеплерово отношение в номинативных степенных и экспоненциальных гамма распределениях // Математические модели. Теория и приложения. Сб. науч. статей под ред. Чиркова М.К. — Изд-во С.-Петерб. гос. ун-та. Вып.7. 2006. С. 33-49.
- [6] **Алексеева Н.П., Алексеев А.О., Парийская Е.Н., Марков А.Г.** Реконструктивно-логический метод исследова-



ния воздействия лекарственных препаратов на морфологию клеток секреторного эпителия // Вестник С.-Петерб.ун-та. 2007. Сер 2. Вып.3. С. 93-100.

- [7] **Алексеева Н.П., Щербакова И.С., Щекина Л.А.** Модель двумерного гамма-распределения в исследовании стимулирования иммунитета у онкологических больных // СПИД, рак и общественное здоровье. 2009. Т. 13. С.6-7.
- [8] **Абрамовиц М. и Стиган И.** Справочник по специальным функциям. — М.:Наука. 1979. 832с.
- [9] **Барт А.Г., Клочкова Н.П., Некруткин В.В.** Симметрии в стоматологических данных. // Сб. Статистические методы в клинических испытаниях. — Изд-во С.-Петерб. гос. ун-та. 1996. С. 165-219.
- [10] **Барт А.Г.** Анализ медико-биологических систем. Метод частично обратных функций. — Изд-во С.-Петерб. гос. ун-та. 2003. 280с.
- [11] **Бриллинджер Д.** Временные ряды. Обработка данных и теория. — М.: Мир. 1980. 536с.
- [12] **Голяндина Н.Э.** Метод „ГУСЕНИЦА“ - *SSA*: анализ временных рядов. — Изд-во С.-Петерб. гос. ун-та. 2004. 78с.
- [13] **Дюге Д.** Теоретическая и прикладная статистика. — М.: Мир. 1972.
- [14] **Игнациус Г.И.** Проективная геометрия. — М.: Знание. 1966.
- [15] **Клочкова Н.П.** Интегрирование дизайнов // Вестник С.-Петерб. ун-та. 1996. Сер.1. Вып.1. С.39-43.
- [16] **Крамер Г.** Математические методы статистики. — М.: Наука. 1967. 648с.
- [17] **Кульбак С.** Теория информации и статистика. — М.: Мир. 1975. 408с.
- [18] **Кутузова А.Э., Петрова Н.Н., Алексеева Н.П. и др.** Комплексная оценка состояния больного с хронической сердечной недостаточностью. — Изд-во СПбГМУ. 2005. 25с.

- [19] **Лидл Р., Нидеррайтер Г.** Конечные поля. — М.: Мир. 1988. Т.1,2. 818с.
- [20] **Любич Ю.И.** Основные понятия и теоремы эволюционной генетики свободных популяций // Успехи математических наук. 1971. Т.XXVI. Вып.5. С.51-116.
- [21] **Марков А.Г., Парийская Е.Н., Толкунов Ю.А.** Исследование влияния протагандина  $F_{2\alpha}$  на клетки молочной железы у мышей // Рос. физиол. журн. им. И.М.Сеченова. 2002. Т.88. № 4, С.503-509.
- [22] **Феллер В.** Введение в теорию вероятностей и ее приложения. Т.1. — М.: Мир. 1984. 528с.
- [23] **Фолконер Д.С.** Введение в генетику количественных признаков. — М.: Агропромиздат. 1985. 486с.
- [24] **Хедрик Ф.** Генетика популяций. — М.: Техносфера. 2003. 588с.
- [25] **Холл М.** Комбинаторика.— М.: Мир. 1970. 424с.
- [26] **Шеффе Г.** Дисперсионный анализ. — М.:Наука. 1980. 511с.
- [27] **Энслейн К., Рэлстон Э., Уилф Г.С.** Статистические методы для ЭВМ. — М.:Наука. 1986. 464с.
- [28] **Яглом А.М., Яглом И.М.** Вероятность и информация. — М.:Наука. 1957. 159с.
- [29] **A.A.Affi, S.P.Asen.** Statistical Analysis. A computer Oriented Approach. — Academic Press. New York · San Francisco · London. 1979. 486p.
- [30] **Alexeyeva N.P., A.O.Alexeyev A.O.** Synonymy of power gamma distributions in the statistical model of „muscles“ // Simulation. 2005. P. 39-43.
- [31] **N.Alexeyeva.** The generalized geometric distribution and the associated Galton-Watson model with a medical statistical application // 6th St.Petersburg workshop on simulation. — SPb. 2009. Vol.2. P. 953-959.

- [32] **N.Alexeyeva, A.Alexeyev, E.Verbitskaya, E.Krupitsky.** Final geometry and a logic principle of projectivity in the statistical analysis of the medical data. // Bulletin of the international statistical institute *LXII-2007*. — Lisbon, Portugal. 2009. P. 3021-3024.
- [33] **N. Alexeyeva, P. Gracheva and others.** The finitely geometric symptom analysis in the glioma survival // The 2nd International Conference on BMEI'09.— China, 2009. 10.1109/BMEI.2009.5305560.
- [34] **N.Alexeyeva, A.Alexeyev, P.Gracheva and others.** Symptom and syndrome analysis of categorial series, logical principles and forms of logic. // Proceedings, 3rd International Conference on BMEI — China. 2010. P. 2603-2606.
- [35] **T.W.Anderson** An Introduction to Multivariate Statistical Analysis. — Wiley Publications in Statistics. 1958.
- [36] **Jorgen Cheply.** Exponential sums and rank of persymmetric matrices over  $F_2$  // arXiv:0711.1306v1[math.NT]. 8 Nov 2007.
- [37] **J. Cherly.** Exponential sums and rank of double persymmetric matrices over  $F_2$  // arXiv:0711.1937v1 [math.NT]. 13 Nov 2007.
- [38] **Krupitsky E.M., Verbitskaya E.V. and others.** Naltrexone for heroin dependence treatment in St. Petersburg, Russia // J. Subst Abuse Treat. 2004. P. 285-294.
- [39] **Rao, C. Radhakrishna.** Linear statistical inference and its applications. — Wiley. N.Y. 1965. 574p.
- [40] **Theresa Migler, Kent E. Morrison, Mitchell Ogle.** How Much Does a Matrix of Rank  $k$  Weigh? // MATHEMATICS MAGAZINE. OCTOBER 2006. VOL. 79. №4, P. 262-271.
- [41] **William R. Myers.** Handling missing data in clinical trials: an overview. // Drug Information Journal. 2000. Vol. 34. P. 525-533.
- [42] **C. J. W. Zorn.** Generalized Estimating Equation Models for Correlated Data: A Review with Applications // American Journal of Political Science. 2001. Vol. 45. № 2. P. 470-490.

# Оглавление

<b>1. Проективные методы</b>	
<b>в анализе данных</b>	<b>5</b>
1.1. Конечные геометрии . . . . .	5
1.1.1. Проективные преобразования . . . . .	5
1.1.2. Теорема Дезарга . . . . .	7
1.1.3. Аксиомы проективной геометрии . . . . .	10
1.1.4. Конечные поля . . . . .	11
1.2. Конечные геометрии и дизайны . . . . .	14
1.2.1. Определение и свойства дизайнов . . . . .	14
1.2.2. Теорема Зингера . . . . .	15
1.2.3. Аффинная геометрия . . . . .	16
1.2.4. Изоморфные и неизоморфные	
подстановки дизайнов . . . . .	19
1.2.5. Первый исключительный изоморфизм . . . . .	20
1.3. Бинарные признаки и $P_n^2$ . . . . .	24
1.3.1. Определение симптома . . . . .	24
1.3.2. Проективные свойства синдрома . . . . .	25
1.3.3. Автоморфизмы синдрома . . . . .	27
1.3.4. Двойственность симптомов и синдромов . . . . .	29
1.4. Второй исключительный	
изоморфизм $SL_3^{F_2} \cong PSL_2^{F_7}$ . . . . .	31
1.4.1. Группа $PSL_2^{F_7}$ и подстановки на $P_1^7$ . . . . .	31
1.4.2. Подстановки инвариантных относительно	
$PSL_2^{F_7}$ дизайнов $D(8, 14, 7, 4, 3 P_1^7)$ . . . . .	32
1.4.3. Линейно-логарифмическое соответствие	
между $P_1^7$ и $F_8$ . . . . .	35
1.4.4. Двойственность $\alpha$ - и $\beta$ -подстановок . . . . .	37

1.4.5.	Сдвиговый параметр . . . . .	40
1.4.6.	Теорема когерентности . . . . .	41
1.5.	Информативность симптомов . . . . .	44
1.5.1.	Распределение синдрома . . . . .	44
1.5.2.	Информационные главные компоненты . . . . .	45
1.5.3.	Генотип АПФ и прогноз артериальной гипертензии . . . . .	49
1.5.4.	Симптомы и эффекты взаимодействия в дисперсионном анализе . . . . .	50
1.5.5.	Энтропия дизайна . . . . .	53
1.5.6.	Разнообразии блоков . . . . .	55
1.6.	Обобщенный синдром . . . . .	57
1.6.1.	Порядковые дизайны $D^1(7, 3, 1 F_8 \setminus \{0\})$ . . . . .	58
1.6.2.	Субпорядковые дизайны $D^1(7, 3, 1 F_8 \setminus \{0\})$ . . . . .	60
1.6.3.	О сумме порядковых дизайнов $D^1(7, 3, 1 F_8 \setminus \{0\})$ . . . . .	62
1.6.4.	Неавтоморфные подстановки дизайна $D(7, 3, 1)$ . . . . .	63
1.6.5.	Комбинаторная структура нечетных дизайнов $D^1(7, 3, 1 F_8 \setminus \{0\})$ . . . . .	64
1.6.6.	Интерпретация дизайна $D(15, 7, 3)$ на додекаэдре . . . . .	65
1.6.7.	Симметрии в стоматологических данных . . . . .	67
1.7.	Интегрирование дизайнов . . . . .	69
1.7.1.	Импульсные последовательности . . . . .	69
1.7.2.	Параметр рекуррентности при $p \neq 2$ . . . . .	69
1.7.3.	Параметр рекуррентности при $p = 2$ . . . . .	71
1.7.4.	Пример интегрирования дизайнов характеристики три . . . . .	73
1.8.	Симптомно-синдромальный анализ в кривых дожития . . . . .	76
1.9.	Анализ категориальных рядов . . . . .	79
1.10.	Связность категориальных рядов . . . . .	84
1.10.1.	О вероятности случайной классификации . . . . .	84
1.10.2.	Случайная классификация в случае ганкелевых матриц . . . . .	87
1.10.3.	Связность последовательностей РНК . . . . .	89
1.11.	Приоритетные кривые в когнитивной биологии . . . . .	92

<b>2. Проблемы неоднородности и неполноты данных</b>	<b>95</b>
2.1. Классификация по расслоенной выборке . . . . .	95
2.1.1. Вероятность ошибочной классификации . . . . .	96
2.1.2. Расслоение популяции . . . . .	99
2.2. Эргодичность и закон Харди—Вайнберга . . . . .	100
2.2.1. Подстановки вектора распределения . . . . .	101
2.2.2. Компоненты вектора перераспределения . . . . .	102
2.2.3. Импульсная упорядоченность компонент вектора перераспределения . . . . .	104
2.2.4. Переходные уравнения . . . . .	105
2.2.5. Индексный бином . . . . .	107
2.2.6. Инвариантность распределения компонент вектора . . . . .	108
2.2.7. Стационарное распределение при отсутствии предпочтения . . . . .	109
2.2.8. Распределение при частичной независимости . . . . .	110
2.2.9. Обобщенная стохастическая модель закона Харди—Вайнберга . . . . .	111
2.3. Эргодический метод компенсации пропусков . . . . .	114
2.3.1. Планы с расщепленными блоками в условиях неполных данных . . . . .	115
2.3.2. Проблема устранения индивидуального влияния . . . . .	117
2.3.3. Эргодичность перекрестного усреднения . . . . .	117
2.3.4. Ошибки для компонент наблюдений $z_{ij}$ и $y_{ijt}$ . . . . .	119
2.3.5. Дисперсионные матрицы ошибок . . . . .	122
2.3.6. Значимость эффектов времени и взаимодействия факторов . . . . .	125
2.3.7. Проверка значимости главного эффекта . . . . .	127

<b>3. Роль масштабирования в управлении биосистем</b>	<b>131</b>
3.1. Частичное обращение функций . . . . .	131
3.1.1. Определение частично обратных функций . . . . .	131
3.1.2. Параметризация частично обратных функций . . . . .	132
3.1.3. Оценка параметров обобщенного геометрического распределения . . . . .	133
3.1.4. Обобщение положительного биномиального распределения . . . . .	138
3.2. Модель реинтрантного бинома в радиобиологии . . . . .	139
3.2.1. Реинтрантный бином . . . . .	140
3.2.2. Оценка параметров $\beta(\cdot p_1, p_0; n_1, n_1)$ . . . . .	143
3.2.3. Сложно-обобщенная биномиальная модель . . . . .	145
3.3. Кривые саногенеза . . . . .	148
3.3.1. Оценка параметров КМНС процесса . . . . .	148
3.3.2. Вычисление дисперсии оценок . . . . .	152
3.3.3. Пример из гематологии . . . . .	154
3.4. Модели гамма-распределения . . . . .	155
3.4.1. Синонимия распределений . . . . .	155
3.4.2. Информационные метрики . . . . .	156
3.4.3. Параметризация синонимичных степенных гамма-распределений . . . . .	157
3.5. Применение синонимичных распределений в физиологии . . . . .	160
3.5.1. Измерения и анализ прибавки веса . . . . .	162
3.5.2. Проверка однородности морфологических данных . . . . .	163
3.5.3. Результаты и обсуждения . . . . .	164
3.5.4. Интерпретация параметров номинативного распределения . . . . .	165
3.6. Двумерное гамма-распределение . . . . .	168
3.6.1. Аддитивная модель . . . . .	168
3.6.2. Плотность двумерного гамма-распределения . . . . .	169
3.6.3. Оценки по методу моментов . . . . .	170
3.6.4. Пример из иммунологии . . . . .	171

Научное издание

*Нина Петровна Алексеева*

**Анализ медико-биологических систем:  
реципрокность, эргодичность, синонимия**

**Монография**

Зав. редакцией *Г.И. Чердниченко*



Подписано в печать с оригинала-макета 14.01.2012.  
Ф-т 60 × 84/16. Печать офсетная. Усл. печ. л. 10,93.  
Уч.-изд. л. 8,72. Тираж 300. Заказ N

Редакция оперативной подготовки изданий  
Издательства С.-Петербургского университета.  
199034, С.-Петербург, Университетская наб., 7/9.

ЦОП типографии Издательства СПбГУ.  
199034, С.-Петербург, наб. Макарова, 6.