ГЕНЕТИКА, 2023, том 59, № 11, с. 1326–1340

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ

УДК 303.725.23:575.177:303.725.23

ОПИСАНИЕ ДИВЕРГЕНЦИИ СУБПОПУЛЯЦИЙ В ИЕРАРХИЧЕСКОЙ СИСТЕМЕ ПРИ АНАЛИЗЕ ИЗОНИМИИ. II. ВЕРОЯТНОСТИ НЕИЗОНИМНЫХ ВСТРЕЧ

© 2023 г. В. П. Пасеков[1], *

[1]Федеральный исследовательский центр "Информатика и управление" Российской академии наук, Москва, 119991 Россия

**e-mail: pass40@mail.ru*

Поступила в редакцию 01.05.2023 г.
После доработки 22.05.2023 г.
Принята к публикации 25.05.2023 г.

Рассматриваются типичные метапопуляции человека с иерархической подразделенностью на части (субпопуляции), соответствующие классификации на основе административно-территориального деления (скажем, село, сельсовет, район, область и так далее) или генеалогического подхода, базирующегося на этногенезе, а также на других принципах биологической классификации. Целью настоящей работы является анализ общих свойств распределения концентрации фамилии по субпопуляциям при их иерархической структуре. Внимание концентрируется на описании фамильной дивергенции субпопуляций, в качестве показателя которой рассматривается общая вероятность встреч (*Hs*) лиц с разными фамилиями, понимаемая как вероятность встречи в выбранной наугад субпопуляции, единицы наблюдения, из рассматриваемой метапопуляции с иерархической структурой. Данная вероятность является фамильным аналогом концентрации гетерозигот в метапопуляции со случайным скрещиванием в ее субпопуляциях, единицах наблюдения. Получено разложение *Hs* по уровням иерархии, обобщающее эффект Валунда в популяционной генетике. Общая вероятность неизонимных встреч в иерархически подразделенной метапопуляции меньше вероятности случайных встреч в ней на сумму средних внутригрупповых дисперсий концентрации фамилии, соответствующих отдельным уровням. Такие свойства являются чисто статистическими характеристиками иерархической структуры, а не особенностью конкретной популяционной системы и не выводятся из закономерностей той или иной модели микроэволюции. Они вычислительно формулируются одинаково для любой иерархической системы, хотя в общем случае не совпадают количественно. Полученные результаты относятся к сельским и городским иерархическим метапопуляциям как отдельным компонентам всего населения.

Ключевые слова: иерархическая структура популяций, метапопуляции, концентрации фамилии в субпопуляциях человека, характеристики неоднородности субпопуляций, разложение вероятности встреч носителей разных фамилий по уровням иерархии.

DOI: 10.31857/S0016675823110097, **EDN:** TNBXXK

Случайный генный дрейф [1] приводит к генетической дивергенции популяций с общим происхождением. Аналогично в результате случайного дрейфа фамилий [2, 3] происходит фамильная дивергенция таких популяций. Оба этих процесса протекают синхронно в одной и той же популяции с ограниченной численностью. Фамильный состав следующего поколения с численностью N мужской составляющей можно упрощенно рассматривать как результат случайной выборки с возвращением размера N из совокупности фамилий мужчин родительского поколения. Генетический состав нового поколения диплоидной популяции при отсутствии систематических давлений является результатом выборки $4N$ гамет родительского поколения ($2N$ размер популяции с учетом женщин и $4N$ размер выборки гамет, формирующих $2N$ диплоидных потомков – по две гаметы на потомка). В результате имеется полная аналогия модели фамильного дрейфа с генетической моделью диплоидной популяции со случайным скрещиванием, рассматриваемой в отношении одного аутосомного локуса с множественными аллелями. Роль фамилий играют аллели. Аналогом случайной встречи пары индивидуумов с разными фамилиями (следовательно, с разными родоначальниками) является появление гетерозиготного генотипа при случайном скрещивании, а аналогом пар однофамильцев будут гомозиготы.

Очевидно, что закономерности динамики фамильного и генетического состава популяции имеют один и тот же характер выборочного дрейфа. Когда закономерности данных процессов сходны, то по наблюдению за результатами одного из них можно делать выводы о результатах протекания другого. Однако оба процесса не идентичны. Для одной и той же диплоидной популяции случайный процесс выборочных колебаний концентраций фамилий в ряду поколений интенсивнее генного дрейфа (соответствует вчетверо меньшей выборке фамилий, чем выборка гамет). Подробнее об этом говорится в [2, 3].

Популяционно-генетический анализ оперирует с данными по генетической структуре популяций. Получение данных по фамильной структуре менее трудоемко и менее затратно. Поэтому желательно проанализировать пути экстраполяции результатов дивергенции популяций по фамильным данным в генетические выводы. При этом важно уметь описывать свойства дивергенции в типичных для населения ситуациях, и мы начнем с такого описания.

Для метапопуляций человека типична иерархическая подразделенность на части (субпопуляции), соответствующие классификации на основе административно-территориального деления (скажем, село, сельсовет, район, область и так далее) или генеалогического подхода, базирующегося на этногенезе, а также на других принципах биологической классификации. В предыдущей части [4] данной работы в качестве характеристики фамильной дивергенции в иерархически подразделенной метапопуляции служила дисперсия распределения фамилии по субпопуляциям. Теперь обратимся к другой характеристике – общей (полной) вероятности случайной встречи (столкновения) индивидуума с фиксированной фамилией с носителем какой-либо другой, которую понимаем как вероятность встречи в выбранной наугад субпопуляции, единице наблюдения, из рассматриваемой метапопуляции с иерархической структурой.

Напомним, что под *иерархической системой* понимается система с многоуровневой структурой, элементы которой связаны отношениями подчинения, причем отдельный элемент, в свою очередь, является иерархической системой (более низкого ранга). У нас иерархической системой является метапопуляция, подразделенная на субпопуляции различного уровня иерархии. Повторим, что вероятность рассматриваемой встречи аналогична доле (концентрации) гетерозигот в популяционной генетике. Это позволяет при соответствующих предположениях оценивать такой важный генетический показатель как коэффициент инбридинга популяции [5].

Итак, объектом нашего анализа служит реальная или теоретическая метапопуляция $\mathbf{s}$ с иерархической подразделенностью на субпопуляции $\{\mathbf{s}_i\}$ по уровням иерархии. При подразделенности системы *отношение подчинения означает вхождение подсистемы низкого уровня в качестве составной части в соответствующую подсистему* (*группу*) *более высокого ранга*. Например, субпопуляции (скажем, сёла) группируются в сельсоветы, районы и т.д. с уровнями иерархии 2, 3 ... соответственно, и множество субпопуляций на каждом из этих уровней представляет собой разбиение рассматриваемой метапопуляции. Здесь выполняется иерархическое подчинение одной субпопуляции другой, т.е. вхождение первой в качестве части во вторую с более высоким уровнем иерархии.

Дадим несколько слов относительно обозначений и терминологии. Под концентрациями фамилий в популяции подразумеваются концентрации однофамильцев (носителей данной фамилии). Векторы набраны полужирным шрифтом, к обозначениям фамильных аналогов популяционно-генетических характеристик (дисперсии, коэффициента фамильного инбридинга) добавлено окончание *s* (*Vs* и *Fs* соответственно). Символ тождества (≡) у нас используется в смысле равенства по определению, а символ ◀ обозначает конец доказательства.

Напомним, что идентификацию и положение субпопуляции в иерархической структуре можно задавать с помощью мультиномеров, как это было сделано нами ранее в предыдущей части [4]. Пусть номер конкретного села (первый уровень) обозначается как s_1; номер сельсовета (второй уровень), куда входит это село, как s_2; номер района (третий уровень), включающего указанные сельсовет и село, как s_3; и т.д. Тогда мультиномер $\mathbf{s}_1 \equiv (s_1, s_2, s_3, \ldots)$ однозначно определяет рассматриваемое село среди прочих сел первого уровня, вектор $\mathbf{s}_2 \equiv (s_2, s_3, s_4, \ldots)$ – идентификатор сельсовета, вектор $\mathbf{s}_i \equiv (s_i, s_{i+1}, \ldots)$ идентифицирует субпопуляцию i-го уровня среди прочих таких же субпопуляций внутри соответствующей группы следующего уровня $i + 1$. Это аналогично почтовому адресу, в котором административные составляющие заменены числами.

В результате субпопуляция $\mathbf{s}_1$ входит в надлежащую субпопуляцию $\mathbf{s}_2$, а $\mathbf{s}_i$ входит в $\mathbf{s}_{i+1}$ и т.д., т.е. субпопуляция некоторого уровня иерархии включает в себя в качестве своей части соответствующие субпопуляции более низкого уровня. Между объектами и их идентификаторами имеется взаимно однозначное соответствие, и мы иногда будем писать идентификатор вместо названия объекта (села, сельсовета и т.д.). Кроме того, повторим, что множество субпопуляций на каждом отдельно выбранном уровне иерархии представ-

ляет собой разбиение всей метапопуляции, т.е. составляет ее целиком.

Данный способ нумерации приложим также к концентрации x интересующей фамилии, которую в селе $\mathbf{s}_1$ будем обозначать как $x(\mathbf{s}_1)$, а концентрацию фамилии в субпопуляции i-го уровня как $x(\mathbf{s}_i)$. Когда некоторые из номеров $\{s_i\}$ рассматриваются как случайные величины (например, при выборе субпопуляции наугад), а другие как фиксированные, то для наглядности будем писать фиксированные величины после вертикальной черты. В частности, будем рассматривать $x(\mathbf{s}_{i-1}|\mathbf{s}_i)$ как *случайную величину*, значениями которой являются концентрации фамилии, получаемые при выборе наугад субпопуляции $(i-1)$-го уровня из фиксированной группы $\mathbf{s}_i$. *Первый аргумент* $\mathbf{s}_{i-1}$ *у* $x(\mathbf{s}_{i-1}|\mathbf{s}_i)$ *указывает на случайно выбираемую субпопуляцию уровня* $i-1$*, а второй* $(\mathbf{s}_i)$ *– на содержащую ее фиксированную группу уровня* i.

Таким образом, концентрация фамилии в фиксированной субпопуляции $\mathbf{s}_i$ обозначается как $x(\mathbf{s}_i)$, а $x(\mathbf{s}_{i-1}|\mathbf{s}_i)$ дает случайную величину, принимающую значения концентраций фамилии в субпопуляциях уровня $i-1$, выбираемых наугад из $\mathbf{s}_i$. Обозначим математическое ожидание (среднее значение) случайной величины $x(\mathbf{s}_{i-1}|\mathbf{s}_i)$ как $x(\mathbf{s}_i)$, а ее дисперсию как $Vs(x(\mathbf{s}_{i-1}|\mathbf{s}_i))$.

Цель настоящей работы – продолжение анализа в [4] *общих свойств распределения концентрации фамилии по субпопуляциям при их иерархической подразделенности. Они являются чисто статистическими характеристиками иерархической структуры, а не особенностью конкретной популяционной системы.* Анализируемые свойства присущи любой иерархически подразделенной метапопуляции и не выводятся из закономерностей той или иной модели микроэволюции. Они вычислительно формулируются одинаково для любой иерархической системы, хотя в общем случае не совпадают количественно. Теоретически это может позволить выделить специфические особенности исследуемого материала. По-прежнему внимание концентрируется на изучении дивергенции субпопуляций, в качестве показателя которой теперь рассматривается не дисперсия концентрации фамилии в субпопуляциях [4], а вероятности встреч пары лиц с неизонимными фамилиями. Обратим внимание на то, что *в рассматриваемых вероятностях учитывается порядок фамилий в паре*, что не вполне традиционно.

Структура изложения материала следующая. После напоминания системы идентификации субпопуляций в метапопуляции с иерархической структурой рассматриваются вероятности встреч носителей разных фамилий в качестве показателей дивергенции субпопуляций на разных уровнях иерархии. Далее подробно рассматривается изменчивость субпопуляций в случае многоуровневой иерархической структуры метапопуляции. Выведено разложение по уровням иерархии общей (полной) вероятности неизонимныхвстреч в выбранной наугад субпопуляции, единице наблюдения. В найденном разложении каждому отдельному уровню иерархии соответствует неотрицательный компонент, равный соответствующей средней внутригрупповой дисперсии концентрации фамилии.

ПОДРАЗДЕЛЕННЫЕ МЕТАПОПУЛЯЦИИ

В случае изучения популяций человека классификация и группировка данных часто производятся на основе административно-территориального деления, имеющего иерархический характер (скажем, село, сельсовет, район, область и др.), генеалогического подхода на основе этногенеза, лингвистических данных и пр. Получаемая группировка субпопуляций приближенно является иерархической. Иерархическая структура метапопуляции отражается на ее свойствах, в частности на распределении фамилий в популяциях человека, где типична опора на официальные данные иерархического характера, сбор и обработку материалов в соответствии с ними. Настоящая статья мотивирована анализом фамильных данных с ориентацией на популяционную генетику. Использование фамилий для получения выводов о генетической структуре популяций основывается на существующих параллелях в передаче на популяционном уровне потомкам фамилий и аутосомных аллелей (см., например, [2, 3]). Плодотворность такого использования продемонстрирована в ряде работ [6–8] (изонимные браки) [9] (фундаментальная монография), в том числе в исследованиях популяций России [10] (медико-генетические аспекты).

Повторим, что у нас объектами являются субпопуляции, их совокупность образует метапопуляцию (иерархическую систему), подчинение означает принадлежность одной субпопуляции другой в качестве ее части. При этом у самой метапопуляции наивысшей уровень иерархии, и все субпопуляции являются ее частями. В роли числового признака субпопуляции может рассматриваться любая числовая характеристика, но мы ограничимся преимущественно дискретными признаками.

На популяционном уровне в качестве признака субпопуляции рассматривается среднее значение соответствующей характеристики для множества входящих в нее субпопуляций более низкого уровня. Вычисление средних значений основано на суммировании, и желательно выяснить результат разбиений метапопуляции или субпопуляции на более мелкие группы для соответствующего среднего значения признака. При его вычислении суммирование происходит по входящим в объект базовым единицам самого низкого первого уровня,

которые у нас считаются заданными (служат начальными данными). Варьирование характера разбиения приводит лишь к перегруппировке слагаемых, а результат остается неизменным. Таким образом, среднее значение признака рассматриваемого объекта не зависит от характера разбиения последнего, что подробнее рассматривается далее.

Важным примером характеристики субпопуляции, состоящей из индивидуумов с дискретным признаком T, является средняя величина T, принимающего значения 1 и 0 в зависимости от его категории (1 для интересующей категории и 0 в противном случае; скажем, интересующей категорией может быть фамилия, аллель, гетерозигота и т.п.). Покажем, что среднее значение T в произвольной субпопуляции $\mathbf{s}_i$ представляет собой не что иное как концентрацию в ней носителей интересующей категории. Эта концентрация также не зависит от характера подразделенности $\mathbf{s}_i$.

Замечание 1. *Пусть субпопуляция $\mathbf{s}_i$ состоит из индивидуумов с дискретным признаком T, кодируемым единицей при интересующей категории и кодируемым нулем в противном случае.*

Тогда среднее значение T в $\mathbf{s}_i$ равно концентрации носителей рассматриваемой категории в данной субпопуляции (доле ее носителей; иначе говоря, вероятности случайно встретить такого носителя в субпопуляции при одинаковых шансах встретить любого индивидуума).

Доказательство. Среднее значение T находится как сумма значений T (по условию это 1 и 0) с весами, равными вероятностям, с которыми T принимает данные значения, т.е. равно

$$1\times Pr\{T=1\}+0\times Pr\{T=0\}=Pr\{T=1\},$$

где $Pr\{T=...\}$ обозначает вероятность наблюдения соответствующего значения T. Вероятность $Pr\{T=1\}$ интерпретируется как вероятность при выборе наугад из $\mathbf{s}_i$ индивидуума получить носителя интересующей категории. При одинаковых шансах выбрать любого индивидуума вероятность $Pr\{T=1\}$ равна доле (концентрации) носителей этой категории в $\mathbf{s}_i$. ◀

Далее рассматривается только популяционный уровень организации, т.е. сами носители остаются за кадром, а анализируемыми объектами являются субпопуляции, характеризуемые средними значениями (концентрациями) интересующих показателей, например фамилии. Для субпопуляций самого низкого первого уровня иерархии значения концентрации фамилии полагаем заданными исходно (рассматриваем их как начальные данные).

Обратим внимание на то, что даже когда предметом изучения является единственная иерархически подразделенная метапопуляция, выводы относительно ее свойств можно делать в вероятностных терминах в результате трактовки характеристик субпопуляции как результатов ее случайного выбора из соответствующего множества субпопуляций, частей разбиения метапопуляции. При этом поскольку концентрация является средним значением для T, то к ней приложимы известные свойства средних значений (математических ожиданий).

Рассмотрим метапопуляцию $\mathbf{s}_j$. Покажем, что *концентрация $T(\mathbf{s}_j)$ носителей интересующей категории дискретного признака T в $\mathbf{s}_j$ равна средней величине для его концентраций в составляющих разбиение $\mathbf{s}_j$ субпопуляциях $\{\mathbf{s}_i\}$ и не зависит от разбиения.*

Замечание 2. *Пусть метапопуляция $\mathbf{s}_j$ с общей численностью $N(\mathbf{s}_j)$ разбита на какие-либо группы $\{\mathbf{s}_i\}$ с численностями $\{N(\mathbf{s}_i)\}$. Положим численность носителей рассматриваемой категории T в метапопуляции $\mathbf{s}_j$ обозначена как $T_N(\mathbf{s}_j)$.*

Тогда концентрация $T(\mathbf{s}_j)\equiv T_N(\mathbf{s}_j)/N(\mathbf{s}_j)$ носителей данной категории T в метапопуляции $\mathbf{s}_j$, произвольно разбитой на группы $\{\mathbf{s}_i\}$, находится как

$$T(\mathbf{s}_j)=\sum_{\mathbf{s}_i\in\mathbf{s}_j}T(\mathbf{s}_i)\frac{N(\mathbf{s}_i)}{N(\mathbf{s}_j)}\equiv$$
$$\equiv\sum_{\mathbf{s}_i\in\mathbf{s}_j}T(\mathbf{s}_i)Pr(\mathbf{s}_i|\mathbf{s}_j)=E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}_j\},$$
$$0\le Pr(\mathbf{s}_i|\mathbf{s}_j)\equiv\frac{N(\mathbf{s}_i)}{N(\mathbf{s}_j)}\le 1,\quad\sum_{\mathbf{s}_i}Pr(\mathbf{s}_i|\mathbf{s}_j)=1,$$

т.е. как математическое ожидание случайной величины $T(\mathbf{s}_i|\mathbf{s}_j)$, концентрации T в $\mathbf{s}_i$ при выборе наугад $\mathbf{s}_i$ из $\mathbf{s}_j$ с вероятностью $Pr(\mathbf{s}_i|\mathbf{s}_j)\equiv N(\mathbf{s}_i)/N(\mathbf{s}_j)$. Здесь E является символом операции получения математического ожидания, нижний индекс у E указывает на используемую при усреднении переменную.

Значение $T(\mathbf{s}_j)$ не зависит от разбиения метапопуляции $\mathbf{s}_j$, и $T(\mathbf{s}_j)$ не меняется при его изменении.

Доказательство. Концентрация $T(\mathbf{s}_j)$ рассматриваемых носителей в разбитой на группы $\{\mathbf{s}_i\}$ метапопуляции $\mathbf{s}_j$ определяется как отношение их численности $T_N(\mathbf{s}_j)$ в $\mathbf{s}_j$, складывающейся из численностей в подгруппах $\{T_N(\mathbf{s}_i)\}$, к общей численности $N(\mathbf{s}_j)$ метапопуляции $\mathbf{s}_j$, складывающейся из численностей ее подгрупп $\{N(\mathbf{s}_i)\}$. Для субпопуляций $\mathbf{s}_i$ и $\mathbf{s}_j=\{\mathbf{s}_i\}$ имеем

$$T_N(\mathbf{s}_i)\equiv T(\mathbf{s}_i)N(\mathbf{s}_i),$$
$$T_N(\mathbf{s}_j)=\sum_{\mathbf{s}_i\in\mathbf{s}_j}T_N(\mathbf{s}_i)=T(\mathbf{s}_i)N(\mathbf{s}_i),$$

$$T(\mathbf{s}_j)\equiv T_N(\mathbf{s}_j)/N(\mathbf{s}_j)=\sum_{\mathbf{s}_i\in\mathbf{s}_j}T(\mathbf{s}_i)\frac{N(\mathbf{s}_i)}{N(\mathbf{s}_j)}.$$

Следовательно, концентрация $T(\mathbf{s}_j)$ в $\mathbf{s}_j$ равняется среднему взвешенному значению для концентраций $\{T(\mathbf{s}_i)\}$ в составляющих разбиение $\mathbf{s}_j$ субпопуля-

циях $\{\mathbf{s}_i\}$ с весами, равными относительным численностям субпопуляций $\{N(\mathbf{s}_i)/N(\mathbf{s}_j)\}$. Очевидно, эти веса в сумме по $\{\mathbf{s}_i\}$ дают единицу и неотрицательны. Их значения $\{N(\mathbf{s}_i)/N(\mathbf{s}_j)\}$ можно интерпретировать как вероятности $\{Pr(\mathbf{s}_i|\mathbf{s}_j)\}$ попасть на индивидуума из соответствующей субпопуляции $\mathbf{s}_i$ при выборе наугад индивидуума из $\mathbf{s}_j$ (как вероятности выбора субпопуляции $\mathbf{s}_i$ из $\mathbf{s}_j$) при одинаковых шансах каждого быть выбранным.

Таким образом, формула для $T(\mathbf{s}_j)$ показывает, что $T(\mathbf{s}_j)$ является математическим ожиданием для $T(\mathbf{s}_i|\mathbf{s}_j)$ при выборе наугад соответствующей субпопуляции $\mathbf{s}_i$ из $\mathbf{s}_j$:

$$T(\mathbf{s}_j) \equiv \sum_{\mathbf{s}_i \in \mathbf{s}_j} T(\mathbf{s}_i)\frac{N(\mathbf{s}_i)}{N(\mathbf{s}_j)} = \\ = \sum_{\mathbf{s}_i \in \mathbf{s}_j} T(\mathbf{s}_i) Pr(\mathbf{s}_i|\mathbf{s}_j) = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}_j\}.$$

Очевидно, приведенный вывод верен для любого разбиения $\mathbf{s}_j$ на субпопуляции $\{\mathbf{s}_i\}$. При этом ее общая численность $N(\mathbf{s}_j)$ и общее количество носителей $T_N(\mathbf{s}_j)$ в ней не изменятся в результате выбора другого разбиения. Следовательно, их отношение (концентрация носителей) не зависит от разбиения $\mathbf{s}_j$ и находится по такой же формуле. ◄

Типичным признаком при изучении изонимии служит концентрация x какой-либо фамилии (точнее, носителей данной фамилии). Как и для признака T, среднее значение $x(\mathbf{s}_j)$ для концентрации фамилии в некоторой метапопуляции $\mathbf{s}_j$ равно средней величине для концентраций в составляющих ее разбиение субпопуляциях. Скажем, если $\mathbf{s}_j$ разбита на $\{\mathbf{s}_1\}$, то $x(\mathbf{s}_j) = E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_j\}$, а если рассматривается разбиение на $\{\mathbf{s}_i\}$, то $x(\mathbf{s}_j) = E_{\mathbf{s}_i}\{x(\mathbf{s}_i)|\mathbf{s}_j\}$.

Это свойство средних значений признаков в подразделенных популяциях рассматривается далее с более общих позиций как выражение полного математического ожидания.

СВОЙСТВА СРЕДНИХ ЗНАЧЕНИЙ В ИЕРАРХИЧЕСКОЙ СИСТЕМЕ

Пусть дана какая-либо метапопуляция $\mathbf{s}$ с иерархической структурой подразделенности. На первом уровне $\mathbf{s}$ разбита на субпопуляции $\{\mathbf{s}_1\}$, а на i-м уровне на $\{\mathbf{s}_i\}$. Понятно, что каждая из субпопуляций $\mathbf{s}_i$, в свою очередь, подразделена на соответствующие подмножества субпопуляций первого уровня. Положим, что рассматривается случайная переменная величина T (ею может быть концентрация носителей какой-нибудь категории признака T). Пусть значения концентрации $\{T(\mathbf{s}_1)\}$ *заданы для всех субпопуляций первого уровня иерархии* $\{\mathbf{s}_1\}$, а значения $T(\mathbf{s}_i)$ на более высоких уровнях находятся по определению как

$$T(\mathbf{s}_i) \equiv E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_i\} = \sum_{\mathbf{s}_1} T(\mathbf{s}_1) Pr(\mathbf{s}_1|\mathbf{s}_i), \\ i = 2, 3, \ldots; \quad T(\mathbf{s}) \equiv E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}\}. \qquad (1)$$

Таким образом, на каждом из уровней иерархии i значение $T(\mathbf{s}_i)$ является средним значением для $T(\mathbf{s}_1)$ при выборе наугад $\mathbf{s}_1$ из $\mathbf{s}_i$.

Рассмотрим некоторые из свойств математического ожидания для иерархических систем, формулируемые для произвольной случайной величины T. Напомним широко используемую далее *формулу полного математического ожидания* (частным случаем которой является полученная выше формула для $T(\mathbf{s})$ в предыдущем замечании):

$$E\{T\} = E_A\{E_T\{T|A\}\} = \underset{i}{E} E_T\{T|A_i\} Pr(A_i).$$

Здесь случайное событие A принадлежит множеству событий $\{A_i\}$, составляющих *полную систему* несовместимых случайных событий, реализующихся с вероятностями $\{Pr(A_i)\}$ и таких, что обязательно происходит одно из них; $E_T\{T|A_i\}$ означает условное (условие пишем после вертикальной черты) математическое ожидание для случайной величины T (нижний индекс у E указывает на переменную, которая служит для усреднения) при условии реализации соответствующего случайного события A_i.

У нас при выборе наугад субпопуляции $\mathbf{s}_i$ из $\mathbf{s}_j$ роль полной системы случайных событий играет множество субпопуляций $\{\mathbf{s}_i\}$, представляющих собой разбиение рассматриваемой популяции $\mathbf{s}_j$. Аналогично при случайном выборе $\mathbf{s}_1$ из $\mathbf{s}_j$ полную систему образуют субпопуляции $\{\mathbf{s}_1\}$. Тогда для случайной величины T формула полного математического ожидания выглядит следующим образом:

$$E\{T(\mathbf{s}_j)\} \equiv E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_j\} = \\ = E_{\mathbf{s}_i}\{E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_i\}|\mathbf{s}_j\} = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}_j\}, \qquad (2)$$

т.е. *среднее значение T у некоторой группы $\mathbf{s}_j$ является средним для математических ожиданий* $E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_i\} \equiv T(\mathbf{s}_i)$ *у разбивающих ее подгрупп* $\{\mathbf{s}_i\}$. Это согласуется с полученными ранее формулами для среднего значения случайной величины T или для концентрации x в подразделенной популяции.

В частности, при $j = i + 1$ получаем рекуррентное уравнение

$$T(\mathbf{s}_{i+1}) = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)\}|\mathbf{s}_{i+1}\}\}. \qquad (3)$$

Если в (2) вместо $\mathbf{s}_j$ рассматривать всю метапопуляцию $\mathbf{s}$, разбитую на $\{\mathbf{s}_i\}$, по формуле полного математического ожидания и определению (1)

$$T(\mathbf{s}) = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}\} = E_{\mathbf{s}_i}\{E_{\mathbf{s}_1}\{T(\mathbf{s}_1)\}|\mathbf{s}_i\}|\mathbf{s}\} = \\ = \sum_{\mathbf{s}_i} T(\mathbf{s}_i) Pr(\mathbf{s}_i|\mathbf{s}), \quad i = 1, 2, \ldots, \qquad (4)$$

т.е. $E_{\mathbf{s}_i}\{E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_i\}|\mathbf{s}\} = E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}\}$.

Мы имеем два выражения для одного и того же значения $T(\mathbf{s})$ – по определению (1) и по формуле (4), т.е.

$$T(\mathbf{s}) \equiv E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}\} = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}\},$$
$$i = 1, 2, 3, \ldots, \quad T(\mathbf{s}_i) \equiv E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}_i\}. \qquad (5)$$

Эта формула показывает в явном виде равенство полного (общего) среднего значения $T(\mathbf{s})$ средней величине для математических ожиданий отдельных частей $\{T(\mathbf{s}_i)\}$ одного и того же уровня иерархии i, какими бы ни были эти части и выбранный уровень (i принимает значения, соответствующие отдельным уровням иерархии метапопуляции $\mathbf{s}$). Поскольку разбиение $\mathbf{s}$ на $\{\mathbf{s}_i\}$ может быть любым, то $T(\mathbf{s})$ не зависит от характера подразделенности $\mathbf{s}$.

Аналогичная формула верна для $T(\mathbf{s}_j)$, когда $\mathbf{s}_j$ разбита на субпопуляции $\{\mathbf{s}_i\}$. Если $\mathbf{s}$ состоит из $\{\mathbf{s}_j\}$, а $\mathbf{s}_j$ из $\{\mathbf{s}_i\}$, то

$$E_{\mathbf{s}_j}\{E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}_j\}|\mathbf{s}\} = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}\}. \qquad (6)$$

Повторим, что здесь случайная величина T принимает значения $\{T(\mathbf{s}_i)\}$ с вероятностями $\{Pr(\mathbf{s}_i|\mathbf{s}_j)\}$ выбора наугад соответствующей субпопуляции $\mathbf{s}_i$ на уровне иерархии i из субпопуляции $\mathbf{s}_j$ на уровне j. Среднее значение $E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}_j\}$ для $T(\mathbf{s}_i|\mathbf{s}_j)$ обозначается как $T(\mathbf{s}_j)$ и характеризует составную часть $\mathbf{s}_j$ метапопуляции $\mathbf{s}$. Это среднее значение само является случайной величиной $T(\mathbf{s}_j|\mathbf{s})$, если $\mathbf{s}_j$ наугад выбирается из фиксированной метапопуляции $\mathbf{s} \equiv \{\mathbf{s}_j\}$. Согласно формуле полного математического ожидания полное (общее) среднее значение рассматриваемой случайной величины $T(\mathbf{s}_i|\mathbf{s}_j)$ для всей метапопуляции $\mathbf{s} \equiv \{\mathbf{s}_j\}$ равно математическому ожиданию для средних значений в ее отдельных частях $\{\mathbf{s}_j\}$, что формально отображается формулами (5)–(6).

Признак T определен для субпопуляций на любом уровне иерархии. При этом его базовыми значениями являются значения $\{T(\mathbf{s}_1)\}$, заданные для субпопуляций $\{\mathbf{s}_1\}$ первого уровня. Через них выражаются величины T для субпопуляций последующих уровней как средние значения в соответствующих группировках базовых значений. Поэтому обозначение $T(\mathbf{s}_i)$ подразумевает вычисление при условии значений $\{T(\mathbf{s}_1)\}$, а средние величины T для ряда субпопуляций одного и того же уровня не зависят от выбранного уровня и совпадают с величиной для всей метапопуляции $\mathbf{s}$:

$$E_{\mathbf{s}_1}\{T(\mathbf{s}_1)|\mathbf{s}\} = E_{\mathbf{s}_i}\{T(\mathbf{s}_i)|\mathbf{s}\} =$$
$$= \ldots E_{\mathbf{s}_j}\{T(\mathbf{s}_j)|\mathbf{s}\} = \ldots = T(\mathbf{s}),$$

что говорит о независимости значения $T(\mathbf{s})$ от разбиений $\mathbf{s}$, в том числе неиерархических.

ПРИМЕРЫ СЛУЧАЙНЫХ ВЕЛИЧИН ПРИ ИЗУЧЕНИИ ИЗОНИМИИ

Далее в качестве случайных величин будем рассматривать такие признаки субпопуляции как концентрация фамилии x и вероятность Hs неизонимных встреч (столкновений) в субпопуляции. Проанализируем также свойства дисперсии $Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_j))$, т.е. дисперсии стоящего в скобках аргумента (случайной величины $x(\mathbf{s}_i|\mathbf{s}_j)$).

КОНЦЕНТРАЦИЯ ФАМИЛИИ

Рассмотрим подразделенную субпопуляцию $\mathbf{s}_2$. Положим, что в качестве T рассматривается концентрация x носителей интересующей фамилии, $x(\mathbf{s}_2)$ обозначает концентрацию фамилии во всей фиксированной подразделенной субпопуляции $\mathbf{s}_2$. Она находится по определению (1) как среднее значение для концентраций $\{x(\mathbf{s}_1|\mathbf{s}_2)\}$ в субпопуляциях единицах наблюдения первого уровня иерархии $\{\mathbf{s}_1\}$ внутри $\mathbf{s}_2$:

$$x(\mathbf{s}_2) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_2\} = \sum_{\mathbf{s}_1} x(\mathbf{s}_1)\, Pr(\mathbf{s}_1|\mathbf{s}_2)\}.$$

Среднее значение концентрации фамилии $x(\mathbf{s}_3)$ для распределения $\{x(\mathbf{s}_1|\mathbf{s}_3)\}$ в фиксированной субпопуляции $\mathbf{s}_3$ находится по определению (1) и по формуле полного математического ожидания (4) как

$$x(\mathbf{s}_3) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1|\mathbf{s}_3)\} = E_{\mathbf{s}_2}\{x(\mathbf{s}_2)|\mathbf{s}_3\} =$$
$$= E_{\mathbf{s}_2}\{E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_2\}|\mathbf{s}_3\} = \sum_{\mathbf{s}_2} x(\mathbf{s}_2) Pr(\mathbf{s}_2|\mathbf{s}_3).$$

Согласно (4) для произвольного уровня иерархии j (при $j > i$, когда $\mathbf{s}_i$ входит в $\mathbf{s}_j$ как составная часть, элемент разбиения $\mathbf{s}_j$)

$$x(\mathbf{s}_j) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_j\} = \sum_{\mathbf{s}_1} x(\mathbf{s}_1) Pr(\mathbf{s}_1|\mathbf{s}_j) =$$
$$= \sum_{\mathbf{s}_i} x(\mathbf{s}_i) Pr(\mathbf{s}_i|\mathbf{s}_j), \quad \mathbf{s}_j = \{\mathbf{s}_i\}.$$

Если здесь вместо $\mathbf{s}_j$ взять всю метапопуляцию $\mathbf{s}$, то ясно, что концентрация фамилии $x(\mathbf{s})$ не зависит от использования при ее подсчетах уровня разбиения $\mathbf{s}$ (как мы видели для произвольной случайной величины T в (5)).

Для иерархически подразделенной метапопуляции $\mathbf{s}$ мы имеем при заданных значениях $\{x(\mathbf{s}_1)\}$:

$$x(\mathbf{s}_i) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_i\} =$$
$$= \sum_{\mathbf{s}_1} x(\mathbf{s}_1) Pr(\mathbf{s}_1|\mathbf{s}_i), \quad \mathbf{s}_i = \{\mathbf{s}_1\};$$

$$x(\mathbf{s}_j) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_j\} = E_{\mathbf{s}_i}\{x(\mathbf{s}_i)|\mathbf{s}_j\} =$$
$$= E_{\mathbf{s}_i}\{E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_i\}|\mathbf{s}_j\} = \sum_{\mathbf{s}_i} x(\mathbf{s}_i)Pr(\mathbf{s}_i|\mathbf{s}_j),$$
$$\mathbf{s}_j = \{\mathbf{s}_i\}, \quad \mathbf{s}_i = \{\mathbf{s}_1\};$$
$$x(\mathbf{s}) \equiv E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}\} =$$
$$= E_{\mathbf{s}_i}\{x(\mathbf{s}_i)|\mathbf{s}\} = \sum_{\mathbf{s}_i} x(\mathbf{s}_i)Pr(\mathbf{s}_i|\mathbf{s}).$$

При $i = j - 1$ получаем еще одну запись рекуррентного уравнения $x(\mathbf{s}_j) \equiv E_{\mathbf{s}_{j-1}}\{x(\mathbf{s}_{j-1})|\mathbf{s}_j\}$, которую можно исследовать с целью получения выражения для $x(\mathbf{s}_j)$ на произвольном уровне иерархии j в зависимости от начальных значений $\{x(\mathbf{s}_1)\}$.

КЛАССИФИКАЦИЯ ДИСПЕРСИИ КОНЦЕНТРАЦИИ

Рассмотрим внутригрупповую дисперсию $Vs_{in}(\mathbf{s}_2) \equiv Vs(x(\mathbf{s}_1|\mathbf{s}_2))$ концентрации $x(\mathbf{s}_1)$ по субпопуляциям $\{\mathbf{s}_1\}$ внутри $\mathbf{s}_2$, т.е. дисперсию распределения $\{x(\mathbf{s}_1)\}$ значений x у субпопуляций $\{\mathbf{s}_1\}$, являющихся единицами наблюдения и содержащихся в $\mathbf{s}_2$. Она является случайной при выборе наугад $\mathbf{s}_2$ из $\mathbf{s}_3$. Признак Vs_{in} определен для субпопуляций с уровнем не ниже второго (говорить о внутригрупповой дисперсии для $\mathbf{s}_1$ не имеет смысла, так как в ней нет субпопуляций, различия между которыми характеризуются дисперсией). Формально можно определить значение $Vs(x(\mathbf{s}_n|\mathbf{s}_n))$ равным нулю.

Начальные значения для случайной величины x заданы для субпопуляций первого уровня иерархии как $\{x(\mathbf{s}_1)\}$. Эти значения $\{x(\mathbf{s}_1)\}$ внутри $\mathbf{s}_2$ определяют начальные значения для $Vs(x(\mathbf{s}_1|\mathbf{s}_2)) = Vs_{in}(x(\mathbf{s}_1|\mathbf{s}_2))$, внутригрупповой дисперсии Vs_{in} для $\mathbf{s}_2$. Повторим, что данные значения случайны, если $\mathbf{s}_2$ наугад выбирается из фиксированной субпопуляции $\mathbf{s}_3$.

Напомним классификацию дисперсий в случае метапопуляции $\mathbf{s}_3$ с тремя уровнями иерархии. Здесь выделяют три типа дисперсий. Под *общей* (*полной*) *дисперсией* $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}_3))$ метапопуляции понимают дисперсию распределения концентрации $\{x(\mathbf{s}_1)\}$ признака по субпопуляциям первого уровня иерархии, содержащимся в $\mathbf{s}_3$. Рассматриваемая метапопуляция $\mathbf{s}_3$ разбита на группы (субпопуляции) $\{\mathbf{s}_2\}$ и дисперсия $Vs_{betw}(x(\mathbf{s}_2|\mathbf{s}_3))$ распределения концентрации $\{x(\mathbf{s}_2)\}$ по этим группам называется *межгрупповой дисперсией* в метапопуляции. Наконец, каждая группа состоит из соответствующих субпопуляций $\{\mathbf{s}_1\}$ с концентрациями $\{x(\mathbf{s}_1)\}$. Дисперсия $Vs(x(\mathbf{s}_1|\mathbf{s}_2))$ распределения концентраций $\{x(\mathbf{s}_1)\}$ внутри $\mathbf{s}_2$ называется *внутригрупповой дисперсией* $Vs_{in}(x(\mathbf{s}_1|\mathbf{s}_2))$ группы $\mathbf{s}_2$. Метапопуляция $\mathbf{s}_3$ в целом характеризуется *средней внутригрупповой дисперсией* $E_{\mathbf{s}_2}\{Vs_{in}(x(\mathbf{s}_2|\mathbf{s}_3))\}$.

Ремарка 3. Когда у метапопуляции количество уровней иерархии более трех, приведенная классификация дисперсии становится расплывчатой. Отметим, что дисперсия Vs полностью определяется ее аргументом, а индексы *tot*, *betw*, *in* служат только для облегчения ориентации в отношении аспекта рассмотрения данной дисперсии. Без них свободно можно было бы обойтись без ущерба для логики изложения.

При многоуровневой иерархии метапопуляцию $\mathbf{s}$ по-прежнему можно охарактеризовать, скажем, в отношении концентрации x некоторого признака *полной дисперсией* $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s})$, являющейся дисперсией распределения признака по субпопуляциям самого низкого, первого, уровня иерархии, содержащимся в $\mathbf{s}$.

Еще одной характеристикой субпопуляции $\mathbf{s}$ является *межгрупповая дисперсия* $\mathbf{s}$ *на уровне* i, т.е. дисперсия $Vs(x(\mathbf{s}_i|\mathbf{s}))$ распределения концентрации фамилии по группам уровня i, на которые разбита исходная субпопуляция $\mathbf{s}$. Для каждого из промежуточных уровней у $\mathbf{s}$ будет своя межгрупповая дисперсия.

Другой характеристикой $\mathbf{s}$ служит *внутригрупповая дисперсия* $Vs_{in}(\mathbf{s}_i|\mathbf{s}_j)$ *субпопуляции* $\mathbf{s}_j$, *на уровне* i, т.е. дисперсия распределения $\{x(\mathbf{s}_i|\mathbf{s}_j)\}$ значений x у субпопуляций $\mathbf{s}_i$ в $\mathbf{s}_j$. Внутригрупповую дисперсию можно интерпретировать когда i равно единице как полную дисперсию $\mathbf{s}_j$, т.е. когда в $\mathbf{s}_j$ субпопуляции $\{\mathbf{s}_i\}$ являются единицей наблюдения.

Напомним, что внутригрупповая дисперсия носит случайный характер при выборе наугад $\mathbf{s}_j$ из $\mathbf{s}$. Тогда вся субпопуляция $\mathbf{s}$ характеризуется математическим ожиданием для $\{Vs_{in}(\mathbf{s}_i|\mathbf{s})\}$, иначе говоря, *средней внутригрупповой дисперсией на уровне* i *для субпопуляций* j*-го уровня иерархии в* $\mathbf{s}$:

$$E_{\mathbf{s}_j}\{Vs_{in}(\mathbf{s}_i|\mathbf{s}_j)|\mathbf{s}\} = \sum_{\mathbf{s}_j} Vs_{in}(\mathbf{s}_i|\mathbf{s}_j)Pr(\mathbf{s}_j|\mathbf{s}).$$

Вспомним, что при анализе концентрации фамилии x ее среднее значение в субпопуляции (метапопуляции) высокого уровня, скажем, в $\mathbf{s}_k$ выражается согласно (5) через значения у входящих в нее субпопуляций уровнем ниже как $x(\mathbf{s}_k) = E_{\mathbf{s}_j}\{x(\mathbf{s}_j)|\mathbf{s}_k\}$. Казалось бы, если в (5) положить $T(\mathbf{s}_j) = Vs_{in}(\mathbf{s}_i|\mathbf{s}_j)$, то аналогично $Vs_{in}(\mathbf{s}_i|\mathbf{s}_k) = E_{\mathbf{s}j}\{Vs_{in}(\mathbf{s}_i|\mathbf{s}_j)|\mathbf{s}_k\}$. Однако это равенство неверно, что мы подробно проанализируем далее. Суть в том, что перетасовка данных при группировке субпопуляций $\{\mathbf{s}_j\}$ не приводит к изменению результатов для концентрации фамилии как показано ранее, в отличие от дисперсии концентрации.

ВЕРОЯТНОСТИ ВСТРЕЧ (СТОЛКНОВЕНИЙ) В ПОДРАЗДЕЛЕННОЙ МЕТАПОПУЛЯЦИИ

Рассмотрим концентрацию (вероятность) $Hs(\mathbf{s})$ встреч (столкновений) индивидуума с данной фамилией с носителем какой-либо другой в метапопуляции $\mathbf{s}$, определяемую как среднее значение Hs для групп одного уровня иерархии. Согласно свойствам средних значений (5) результат усреднения не зависит от того, какое при этом используется разбиение $\mathbf{s}$ на группы на данном уровне и от самого уровня. Поскольку концепция встреч Hs скорее умозрительная, чем дающая легко доступный метод определения Hs в произвольной популяции (за исключением случая брачных пар), то возникает проблема как задать Hs на базовом уровне.

Здесь приходится прибегать к косвенным методам, основанным на дополнительных предположениях. Таким предположением является допущение о случайном характере встреч в группах $\{\mathbf{s}_1\}$. Оно разумно, когда разбиение метапопуляции состоит из территориально малых групп (базовых субпопуляций $\{\mathbf{s}_1\}$ самого низкого уровня иерархии), в каждой из которых существованием каких-либо препятствий для встреч (столкновений) индивидуумов и нарушениями внешних границ при встречах можно пренебречь.

При случайной встрече двух индивидуумов комбинирование ее участников в пары происходит независимо. В результате независимости участников случайных столкновений вероятность встречи равна произведению вероятностей встретить каждого из участников по отдельности, равных их концентрациям. Так, вероятность $Hs(x(\mathbf{s}_1))$ встречи индивидуума с интересующей фамилией (с концентрацией $x(\mathbf{s}_1)$ ее носителей в $\mathbf{s}_1$) и индивидуума с какой-либо иной с учетом порядка участников, очевидно, находится как $Hs_0(x(\mathbf{s}_1)) \equiv x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$. Эта формула аналогична выражению для концентрации гетерозигот по закону Харди–Вайнберга в популяционной генетике при учете порядка аллелей в гетерозиготе, поскольку в обоих случаях наблюдаем случайные объединения в парах.

Метапопуляция, состоящая из неподразделенных субпопуляций, характеризуется средней величиной для значений Hs по субпопуляциям. Например, общая (полная) вероятность $Hs(x(\mathbf{s}_1|\mathbf{s}_2))$ встречи индивидуумов с разными фамилиями в *подразделенной* метапопуляции $\mathbf{s}_2$ (скажем, в сельсовете $\mathbf{s}_2$, подразделенном на села) определяется как среднее значение $Hs(x(\mathbf{s}_1))$ для входящих в нее субпопуляций (сел $\{\mathbf{s}_1\}$), под которой понимается вероятность при выборе наугад субпопуляции $\mathbf{s}_1$ из рассматриваемой метапопуляции. Тем самым подразумевается, что *встречи происходят только внутри отдельных сел* по аналогии с рядом моделей популяционной генетики, в которых случайное скрещивание идет внутри отдельной элементарной популяции после формирования ее генетического состава к моменту скрещивания с учетом всех факторов динамики. Чтобы подчеркнуть, что когда речь идет о средней вероятности встреч в метапопуляции, а не в отдельной неподразделенной субпопуляции, то среднее значение вероятности встреч называем *общей* (*полной*) *вероятностью* $Hs(x(\mathbf{s}_2))$.

Очевидно, при произвольном разбиении метапопуляции $\mathbf{s}$ на $\{\mathbf{s}_1\}$ у нас нет аргументов в пользу случайного характера встреч в каждой из субпопуляций $\{\mathbf{s}_1\}$ и соответственно в пользу вероятности встреч в виде $Hs(\mathbf{s}_1) = Hs_0(\mathbf{s}_1) \equiv x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$. При разбиении метапопуляции $\mathbf{s}$ на $\{\mathbf{s}_1\}$, когда в $\{\mathbf{s}_1\}$ встречи случайны,

$$Hs(\mathbf{s}) = \sum_{\mathbf{s}_1} Hs_0(\mathbf{s}_1) Pr(\mathbf{s}_1|\mathbf{s}) =$$
$$= \sum_{\mathbf{s}_1} x(\mathbf{s}_1)(1 - x(\mathbf{s}_1)) Pr(\mathbf{s}_1|\mathbf{s}).$$

Здесь и далее для краткости мы пишем $Hs(\mathbf{s})$ вместо $Hs(x(\mathbf{s}))$.

Повторим, что роль вероятности $Hs(\mathbf{s}_1) = Hs(x(\mathbf{s}_1))$ случайной встречи рассматриваемых индивидуумов с учетом порядка их фамилий в неподразделенной субпопуляции $\mathbf{s}_1$ с концентрацией интересующей фамилии $x(\mathbf{s}_1)$ играет $x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$, а вероятность встречи $Hs(\mathbf{s}_2)$ в метапопуляции $\mathbf{s}_2 = \{\mathbf{s}_1\}$ определяется как среднее значение $Hs(\mathbf{s}_1)$ при выборе наугад $\mathbf{s}_1$ из метапопуляции $\mathbf{s}_2$. Таким образом, полной вероятностью встречи $Hs(\mathbf{s}_2)$ в метапопуляции $\mathbf{s}_2$ будет

$$Hs(\mathbf{s}_2) \equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}_2\} = E_{\mathbf{s}_1}\{x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))|\mathbf{s}_2\}.$$

Напомним, что $Pr(\mathbf{s}_1|\mathbf{s}_2)$ обозначает вероятность случайного выбора села $\mathbf{s}_1$ из фиксированного сельсовета $\mathbf{s}_2$; $E_{\mathbf{s}_1}$ – символ операции получения математического ожидания (среднего значения) случайной величины, стоящей в фигурных скобках; нижний индекс $\mathbf{s}_1$ у $E_{\mathbf{s}_1}$ указывает на служащую для усреднения переменную $\mathbf{s}_1$. По определению вероятность $Hs(\mathbf{s}_i)$ рассматриваемой встречи для произвольного уровня иерархии i задается формулой

$$Hs(\mathbf{s}_i) \equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}_i\} = \sum_{\mathbf{s}_1} Hs(\mathbf{s}_1) Pr(\mathbf{s}_1|\mathbf{s}_i) =$$
$$= \sum_{\mathbf{s}_1} x(\mathbf{s}_1)(1 - x(\mathbf{s}_1)) Pr(\mathbf{s}_1|\mathbf{s}_i).$$

На примере вероятности (концентрации) Hs мы видим, что в качестве характеристики субпопуляций может служить не только концентрация фамилии x, но и функция Hs от x. В каждой субпопуляции $\mathbf{s}_1$ значение Hs находится как функция $x(1 - x)$ от концентрации x в $\mathbf{s}_1$. Казалось бы есте-

ственно предположить, что среднее значение Hs по субпопуляциям $\mathbf{s}_1$, составляющим $\mathbf{s}_i$, т.е. $Hs(x(\mathbf{s}_i))$, находится как такая же функция $x(1-x)$ теперь от значения x в $\mathbf{s}_i$. Однако $Hs(x(\mathbf{s}_i))$ не равно $x(\mathbf{s}_i)(1-x(\mathbf{s}_i))$, что на первый взгляд может показаться парадоксальным.

ВЕРОЯТНОСТЬ ВСТРЕЧ В ПОДРАЗДЕЛЕННОЙ МЕТАПОПУЛЯЦИИ И КОЭФФИЦИЕНТ ФАМИЛЬНОГО ИНБРИДИНГА

Обратимся к анализу вероятности (доли, концентрации) встреч Hs, являющейся аналогом концентрации гетерозигот в популяционной генетике. Как уже говорилось, эту концентрацию можно уподобить вероятности соответствующего столкновения в некоторой совокупности частиц. Напомним, что здесь подразумевается отыскание $Hs(\mathbf{s}_i)$ в подразделенной популяции при условии заданных значений вероятности столкновений в субпопуляциях первого уровня.

На первом уровне иерархии вероятность встречи $Hs(\mathbf{s}_1)$ в субпопуляции $\mathbf{s}_1$ двух индивидуумов (с заданной фамилией с концентрацией $x(\mathbf{s}_1)$ и какой-либо другой с учетом их порядка) находится в предположении случайных столкновений как $Hs(\mathbf{s}_1) \equiv Hs_0(\mathbf{s}_1) = x(\mathbf{s}_1)(1-x(\mathbf{s}_1))$ аналогично закону Харди–Вайнберга. Для субпопуляций более высокого уровня i используется обозначение $Hs(\mathbf{s}_i)$, в котором зависимость от значений $Hs(\mathbf{s}_1)$ неявна. Концентрация $Hs(\mathbf{s}_i)$ находится как среднее значение для $\{Hs(\mathbf{s}_1)\}$ в субпопуляциях $\{\mathbf{s}_1\}$, составляющих разбиение $\mathbf{s}_i$. Когда $Hs(\mathbf{s}_1) = x(\mathbf{s}_1)(1-x(\mathbf{s}_1))(1-F(\mathbf{s}_1))$, для вероятности $Hs(\mathbf{s}_i)$ используется обозначение $Hs(\mathbf{s}_i|Fs_1)$, явным образом указывающее на вид $Hs(\mathbf{s}_1)$.

По определению значение Hs для каждой субпопуляции уровня иерархии i выше единицы равно средней величине для значений Hs у всех входящих в $\mathbf{s}_i$ субпопуляций на первом уровне (или на одном из других предшествующих i уровней) согласно (5), а на первом уровне в предположении случайного характера встреч равно $x(\mathbf{s}_1)(1-x(\mathbf{s}_1))$:

$$\begin{gathered} Hs(\mathbf{s}_1) \equiv Hs_0(\mathbf{s}_1) \equiv x(\mathbf{s}_1)(1-x(\mathbf{s}_1)), \\ Hs(\mathbf{s}_i) \equiv E_{\mathbf{s}_1}\{Hs_0(\mathbf{s}_1)|\mathbf{s}_i\} = \\ = E_{\mathbf{s}_1}\{x(\mathbf{s}_1)(1-x(\mathbf{s}_1))|\mathbf{s}_i\} = \\ = \sum_{\mathbf{s}_1} x(\mathbf{s}_1)(1-x(\mathbf{s}_1))\,Pr(\mathbf{s}_1|\mathbf{s}_i), \quad i = 2, 3, \ldots . \end{gathered} \tag{7}$$

Иначе говоря, *полная вероятность случайной встречи* (*концентрация* $Hs(\mathbf{s}_i)$) *в* $\mathbf{s}_i$ *находится как среднее значение* $Hs(\mathbf{s}_1)$ *при выборе наугад субпопуляции* $\mathbf{s}_1$ *из* $\mathbf{s}_i$. По формуле полного математического ожидания (5)

$$Hs(\mathbf{s}_j) \equiv E_{\mathbf{s}_2}\{Hs(\mathbf{s}_2)|\mathbf{s}_j\} = E_{\mathbf{s}_i}\{Hs(\mathbf{s}_i)|\mathbf{s}_j\}, \quad j > i > 1.$$

В частности, значения Hs на соседних уровнях иерархии связывает рекуррентное соотношение

$$\begin{gathered} Hs(\mathbf{s}_{i+1}) \equiv E_{\mathbf{s}_i}\{Hs(\mathbf{s}_i)|\mathbf{s}_{i+1}\}, \text{ или} \\ Hs(\mathbf{s}_i) \equiv E_{\mathbf{s}_{i-1}}\{Hs(\mathbf{s}_{i-1})|\mathbf{s}\}; \\ Hs(\mathbf{s}_1) \equiv x(\mathbf{s}_1)(1-x(\mathbf{s}_1)). \end{gathered} \tag{8}$$

Отсюда вытекает согласующаяся с (6) запись формулы полного математического ожидания для Hs

$$E_{\mathbf{s}_i}\{Hs(\mathbf{s}_i)|\mathbf{s}_{i+1}\} = E_{\mathbf{s}_i}\{E_{\mathbf{s}_{i-1}}\{Hs(\mathbf{s}_{i-1})|\mathbf{s}_i\}|\mathbf{s}_{i+1}\}. \tag{9}$$

Повторим, что (8) означает, что полная вероятность рассматриваемых встреч в субпопуляции $\mathbf{s}_i$ является средней величиной для вероятностей встреч в субпопуляциях $\{\mathbf{s}_{i-1}\}$, вместе составляющих разбиение $\mathbf{s}_i$. Выясним зависимость вероятности подобной встречи от фамильной дивергенции между субпопуляциями, рассмотренной нами ранее в первой части [4] в терминах дисперсии концентрации рассматриваемой фамилии.

Результат 4 (*фамильный аналог эффекта Валунда*). *Пусть дана метапопуляции* $\mathbf{s}$, *разбитая на субпопуляции* $\{\mathbf{s}_1\}$ *с концентрациями* $\{x(\mathbf{s}_1)\}$ *рассматриваемой фамилии в них и с вероятностями рассматриваемых неизонимных встреч* $\{x(\mathbf{s}_1)(1-x(\mathbf{s}_1))\}$.

Тогда общая (*полная*) *вероятность* $Hs(\mathbf{s})$ *случайной встречи двух индивидуумов* (*с фиксированной фамилией и с какой-либо иной при учете порядка фамилий*) *в наугад выбранной из* $\mathbf{s}$ *субпопуляции* $\mathbf{s}_1$ *имеет вид*

$$\begin{gathered} Hs(\mathbf{s}) \equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}\} = \\ = \sum_{\mathbf{s}_1} x(\mathbf{s}_1)(1-x(\mathbf{s}_1))\,Pr(\mathbf{s}_1|\mathbf{s}) = \\ = x(\mathbf{s})(1-x(\mathbf{s})) - Vs_{betw}(x(\mathbf{s}_1|\mathbf{s})), \\ Vs_{betw}(x(\mathbf{s}_1|\mathbf{s})) \equiv Vs(x(\mathbf{s}_1|\mathbf{s})) = \\ = \sum_{\mathbf{s}_1} (x(\mathbf{s}_1)-x(\mathbf{s}))^2 Pr(\mathbf{s}_1|\mathbf{s}). \end{gathered} \tag{10}$$

Здесь $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))$ *обозначает межгрупповую дисперсию концентраций* $\{x(\mathbf{s}_1|\mathbf{s})\}$ *рассматриваемой фамилии в субпопуляциях* (*группах*) $\{\mathbf{s}_1\}$, *составляющих разбиение метапопуляции* $\mathbf{s}$; $Pr(\mathbf{s}_1|\mathbf{s})$ *дает вероятность случайного выбора субпопуляции* $\mathbf{s}_1$ *из метапопуляции* $\mathbf{s}$.

Иным образом общую (*полную*) *вероятность* $Hs(\mathbf{s})$ *можно представить как*

$$\begin{gathered} Hs(\mathbf{s}) = x(\mathbf{s})(1-x(\mathbf{s}))\left(1 - \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))}{x(\mathbf{s})(1-x(\mathbf{s}))}\right) = \\ = x(\mathbf{s})(1-x(\mathbf{s}))(1-Fs(\mathbf{s})), \\ Fs(\mathbf{s}) \equiv \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))}{x(\mathbf{s})(1-x(\mathbf{s}))} = \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))}{Hs_0(\mathbf{s})} \geq 0, \end{gathered} \tag{11}$$

где $Fs(\mathbf{s})$ является фамильным аналогом коэффициента инбридинга С. Райта в популяционной генетике.

Доказательство. Пусть неподразделенная субпопуляция $\mathbf{s}_1$ рассматривается как единица наблюдения. Например, в сельском населении это может быть село, а метапопуляция $\mathbf{s}$ соответствует, скажем, сельсовету. Дисперсия $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))$ относится к распределению концентрации заданной фамилии по селам $\{\mathbf{s}_1\}$ внутри фиксированного сельсовета $\mathbf{s}$. Данная дисперсия $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}))$ характеризует фамильную дивергенцию сел внутри сельсовета.

Напомним, что для любой случайной величины x выполняется равенство $E\{x^2\} = (E\{x\})^2 + V(x)$, так как $V(x) \equiv E\{x^2\} - (E\{x\})^2$. Для дальнейшего эту формулу с учетом равенства $E_{\mathbf{s}_i}\{x(\mathbf{s}_i)|\mathbf{s}_{i+1}\} = x(\mathbf{s}_{i+1})$ удобно представить как

$$E_{\mathbf{s}_i}\left\{x^2(\mathbf{s}_i)\middle|\mathbf{s}_{i+1}\right\} = \left(E_{\mathbf{s}_i}\left\{x(\mathbf{s}_i)\middle|\mathbf{s}_{i+1}\right\}\right)^2 + \\ + Vs\left(x\left(\mathbf{s}_i\middle|\mathbf{s}_{i+1}\right)\right) = x^2(\mathbf{s}_{i+1}) + Vs\left(x\left(\mathbf{s}_i\middle|\mathbf{s}_{i+1}\right)\right). \quad (12)$$

Как уже говорилось, для субпопуляции $\mathbf{s}_1$ первого уровня иерархии вероятность случайной встречи рассматриваемой пары индивидуумов с учетом их порядка равна $x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$ подобно закону Харди–Вайнберга в популяционной генетике. Значение $Hs(\mathbf{s})$ для $\mathbf{s}$ по определению находится как среднее значение $Hs(\mathbf{s}_1)$ при случайном выборе $\mathbf{s}_1$ из $\mathbf{s}$, т.е.

$$Hs(\mathbf{s}) = E_{\mathbf{s}_1}\left\{Hs(\mathbf{s}_1)\middle|\mathbf{s}\right\} = E_{\mathbf{s}_1}\left\{x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))\middle|\mathbf{s}\right\} = \\ = E_{\mathbf{s}_1}\left\{x(\mathbf{s}_1)\middle|\mathbf{s}\right\} - E_{\mathbf{s}_1}\left\{x^2(\mathbf{s}_1)\middle|\mathbf{s}\right\} = \\ = x(\mathbf{s}) - \left(Vs\left(x\left(\mathbf{s}_1\middle|\mathbf{s}\right)\right) + \left(E_{\mathbf{s}_1}\left\{x(\mathbf{s}_1)\middle|\mathbf{s}\right\}\right)^2\right) = \\ = x(\mathbf{s})(1 - x(\mathbf{s})) - Vs\left(x\left(\mathbf{s}_1\middle|\mathbf{s}\right)\right)$$

согласно (12). Очевидно, данное выражение $Hs(\mathbf{s})$ можно переписать с использованием коэффициента фамильного инбридинга как $x(\mathbf{s})(1 - x(\mathbf{s}))(1 - Fs(\mathbf{s}))$, т.е. формула (11) верна. ◄

Согласно доказанному результату в метапопуляции $\mathbf{s}$, разбитой на несколько неподразделенных субпопуляций $\{\mathbf{s}_1\}$, в среднем вероятность указанной встречи в субпопуляции $\mathbf{s}_1$ из $\mathbf{s}$ меньше на величину межгрупповой дисперсии $Vs(x(\mathbf{s}_1|\mathbf{s}))$, чем вероятность независимых сочетаний фамилий двух индивидуумов во всей метапопуляции $\mathbf{s}$. Эта формула является практически полным аналогом эффекта Валунда в популяционной генетике (см., например, [1]). Чем больше фамильная дивергенция субпопуляций $\{\mathbf{s}_1\}$, тем меньше общая вероятность $Hs(\mathbf{s})$.

Изложенное показывает важность дисперсии распределения концентрации рассматриваемой фамилии по субпопуляциям для анализа их дивергенции. Это вызывает интерес к изучению компонентов дисперсии концентрации. К данному случаю приложимо правило сложения дисперсий, которое в нашей ситуации выглядит следующим образом.

Замечание 5 (*правило сложения дисперсий*). *Пусть метапопуляция $\mathbf{s}$ разбита на субпопуляции $\{\mathbf{s}_2\}$, каждая из которых состоит из соответствующих неподразделенных групп $\{\mathbf{s}_1\}$, являющихся единицами наблюдения с концентрациями $\{x(\mathbf{s}_1)\}$ рассматриваемой фамилии в них.*

Тогда общая (полная) дисперсия $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}))$ распределения концентрации фамилии по единицам наблюдения $\{\mathbf{s}_1\}$ во всей метапопуляции $\mathbf{s}$ равна сумме межгрупповой дисперсии $Vs_{betw}(x(\mathbf{s}_2|\mathbf{s}))$, характеризующей дивергенцию концентраций фамилии $\{x(\mathbf{s}_2)\}$ между субпопуляциями $\{\mathbf{s}_2\}$, и средней внутригрупповой дисперсии $W(x(\mathbf{s}_1|\mathbf{s}_2)|\mathbf{s}) \equiv E_{\mathbf{s}_2}\{Vs_{in}(x(\mathbf{s}_1|\mathbf{s}_2))|\mathbf{s}\}$, характеризующей среднюю фамильную дивергенцию концентраций $x(\mathbf{s}_1)$ субпопуляций $\{\mathbf{s}_1\}$ внутри отдельных субпопуляций из множества $\{\mathbf{s}_2\}$:

$$Vs_{tot}\left(x\left(\mathbf{s}_1\middle|\mathbf{s}\right)\right) = Vs_{betw}\left(x\left(\mathbf{s}_2\middle|\mathbf{s}\right)\right) + \\ + E_{\mathbf{s}_2}\left\{Vs_{in}\left(x\left(\mathbf{s}_1\middle|\mathbf{s}_2\right)\right)\middle|\mathbf{s}\right\}. \quad (13)$$

Доказательство. По определению полная (общая) дисперсия $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}))$ популяции $\mathbf{s}$ равна $E_{\mathbf{s}_1}\{(x(\mathbf{s}_1) - E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}\}|\mathbf{s})^2\}$. Вспомним, что $E_{\mathbf{s}_2}\{x(\mathbf{s}_2)|\mathbf{s}\} = x(\mathbf{s})$, $E_{\mathbf{s}_1}\{x(\mathbf{s}_1)|\mathbf{s}_i\} = x(\mathbf{s}_i)$, т.е. $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s})) = E_{\mathbf{s}_1}\{(x(\mathbf{s}_1) - x(\mathbf{s}))^2|\mathbf{s}\}$. Представим $(x(\mathbf{s}_1) - x(\mathbf{s}))^2$ в виде

$$\left(x(\mathbf{s}_1) - x(\mathbf{s}_2) + x(\mathbf{s}_2) - x(\mathbf{s})\right)^2 = \\ = \left(x(\mathbf{s}_1) - x(\mathbf{s}_2)\right)^2 + \left(x(\mathbf{s}_2) - x(\mathbf{s})\right)^2 + \\ + 2\left(x(\mathbf{s}_1) - x(\mathbf{s}_2)\right)\left(x(\mathbf{s}_2) - x(\mathbf{s})\right).$$

Математическое ожидание правой части равно сумме средних значений ее отдельных слагаемых, которые находим с применением формулы полного математического ожидания (2). При этом учтем, что за знак E можно выносить независящие от усредняемой величины сомножители.

$$E_{\mathbf{s}_1}\left\{\left(x(\mathbf{s}_1) - x(\mathbf{s}_2)\right)^2\middle|\mathbf{s}\right\} = \\ = E_{\mathbf{s}_2}\left\{E_{\mathbf{s}_1}\left\{\left(x(\mathbf{s}_1) - x(\mathbf{s}_2)\right)^2\middle|\mathbf{s}_2\right\}\middle|\mathbf{s}\right\} = \\ = E_{\mathbf{s}_2}\left\{Vs_{in}\left(x\left(\mathbf{s}_1\middle|\mathbf{s}_2\right)\right)\middle|\mathbf{s}\right\},$$

т.е. первое слагаемое равно средней внутригрупповой дисперсии.

Для второго слагаемого находим

$$E_{\mathbf{s}_1}\left\{\left(x(\mathbf{s}_2) - x(\mathbf{s})\right)^2\middle|\mathbf{s}\right\} = \\ = E_{\mathbf{s}_2}\left\{E_{\mathbf{s}_1}\left\{\left(x(\mathbf{s}_2) - x(\mathbf{s})\right)^2\middle|\mathbf{s}_2\right\}\middle|\mathbf{s}\right\} = \\ = E_{\mathbf{s}_2}\left\{\left(x(\mathbf{s}_2) - x(\mathbf{s})\right)^2 E_{\mathbf{s}_1}\{1|\mathbf{s}_2\}|\mathbf{s}\right\} = Vs_{betw}\left(x\left(\mathbf{s}_2\middle|\mathbf{s}\right)\right),$$

т.е. получаем межгрупповую дисперсию.

Наконец, покажем, что среднее значение третьего слагаемого равно нулю:

$$E_{\mathbf{s}_1}\{2(x(\mathbf{s}_1)-x(\mathbf{s}_2))(x(\mathbf{s}_2)-x(\mathbf{s}))|\mathbf{s}\} =$$
$$= E_{\mathbf{s}_2}\{2(x(\mathbf{s}_2)-x(\mathbf{s}))\,E_{\mathbf{s}_1}\{(x(\mathbf{s}_1)-x(\mathbf{s}_2))|\mathbf{s}_2\}|\mathbf{s}\} = 0,$$

так как среднее отклонение от средней величины $E_{\mathbf{s}_1}\{(x(\mathbf{s}_1)-x(\mathbf{s}_2))|\mathbf{s}_2\}$ всегда равняется нулю.

В итоге получаем требуемое выражение для полной дисперсии

$$Vs_{tot}(x(\mathbf{s}_1|\mathbf{s})) = Vs_{betw}(x(\mathbf{s}_2|\mathbf{s})) + \\ + E_{\mathbf{s}_2}\{Vs_{in}(x(\mathbf{s}_1|\mathbf{s}_2))|\mathbf{s}\}. \blacktriangleleft$$

Данное правило приложимо к любой совокупности, разбитой на группы ее элементов, характеризуемых числовым признаком. В частном случае метапопуляции с произвольным количеством уровней иерархии общая дисперсия концентрации фамилии в ней равна сумме межгрупповой дисперсии распределения концентации по субпопуляциям (группам) одного и того же произвольно выбранного уровня k плюс средняя внутригрупповая дисперсия концентраций в них. Скажем, в разбитой на субпопуляции $\{\mathbf{s}_1\}$ метапопуляции $\mathbf{s} = \mathbf{s}_m$ с m уровнями иерархии правило сложения дисперсий выполняется, когда вместо субпопуляций $\{\mathbf{s}_2\}$ рассматриваются $\{\mathbf{s}_k\}$:

$$Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}_m)) = Vs_{betw}(x(\mathbf{s}_k|\mathbf{s}_m)) + \\ + E_{\mathbf{s}_k}\{Vs_{in}(x(\mathbf{s}_1|\mathbf{s}_k))|\mathbf{s}_m\}, \quad 2 < k < m. \tag{14}$$

Выше мы считали, что встречи индивидуумов в субпопуляциях $\{\mathbf{s}_1\}$ низшего уровня иерархии (единицах наблюдения) случайны. Благодаря этому $Hs(\mathbf{s}_1) = Hs_0(\mathbf{s}_1) = x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$. В реальных исследованиях данное условие может не выполняться (скажем, при существовании скрытой подразделенности в $\mathbf{s}_1$ или из-за иных причин), и желательно рассмотреть такую ситуацию, поскольку для нее приведенный аналог результата Валунда неверен.

В абстрактном случае, скажем, в совокупности сталкивающихся частиц нескольких типов предположение о случайной природе их столкновений не гарантировано. Распространенным выражением для вероятности *Hs неслучайных* столкновений частиц рассматриваемого типа с концентрацией x с иными типами является используемая в популяционной генетике вероятность столкновений вида

$$Hs = x(1-x)(1-Fs) = Hs_0(1-Fs),$$
$$0 \le Hs \le 1, \quad 0 \le |Fs| \le 1.$$

При $Fs > 0$ происходит своего рода отталкивание, и разноименные частицы сталкиваются реже, чем чисто случайно. Обратим внимание, что здесь неявно предполагается независимость Fs от типа сталкивающихся частиц (в противном случае рассматриваемая вероятность столкновения зависит от концентраций всех типов). В случае только двух типов (рассматриваемого и искусственного объединения всех прочих типов) коэффициент Fs имеет смысл корреляции между сталкивающимися типами.

Следствие 6. *Рассмотрим метапопуляцию* $\mathbf{s}_2$, *разбитую на субпопуляции* $\{\mathbf{s}_1\}$ с *концентрациями* $\{x(\mathbf{s}_1)\}$ *носителей рассматриваемой фамилии в них. Пусть вероятность* $Hs(\mathbf{s}_1)$ *их встреч в* $\mathbf{s}_1$ *с носителями иной фамилии неслучайна и формулируется как*

$$Hs(\mathbf{s}_1) = x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))(1 - Fs(\mathbf{s}_1)) = \\ = Hs_0(\mathbf{s}_1)(1 - Fs(\mathbf{s}_1)), \quad \mathbf{s}_2 = \{\mathbf{s}_1\},$$

где $Hs_0(\mathbf{s}_1)$ *обозначает вероятность* $x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))$ *чисто случайной встречи в субпопуляции* $\mathbf{s}_1$ *на первом уровне иерархии.*

Тогда общая (*полная*) *вероятность* $Hs(\mathbf{s}_2|Fs_1)$ *рассматриваемой встречи в наугад выбранной из* $\mathbf{s}_2$ *субпопуляции* $\mathbf{s}_1$ *при независимости концентрации x рассматриваемой фамилии и коэффициента фамильного инбридинга* $Fs(\mathbf{s}_1)$ *в субпопуляциях* $\{\mathbf{s}_1\}$ *имеет вид*

$$Hs(\mathbf{s}_2|Fs_1) = (x(\mathbf{s}_2)(1 - x(\mathbf{s}_2)) - \\ - Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2)))(1 - E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}) \\ Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2)) = Vs(x(\mathbf{s}_1|\mathbf{s}_2)) = \\ = \sum_{\mathbf{s}_1}(x(\mathbf{s}_1) - x(\mathbf{s}_2))^2 Pr(\mathbf{s}_1|\mathbf{s}_2). \tag{15}$$

Здесь $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2))$ *обозначает межгрупповую дисперсию распределения концентрации* $\{x(\mathbf{s}_1|\mathbf{s}_2)\}$ *рассматриваемой фамилии по субпопуляциям* (*группам*) $\{\mathbf{s}_1\}$, *составляющим разбиение метапопуляции* $\mathbf{s}_2$; $Pr(\mathbf{s}_1|\mathbf{s}_2)$ *дает вероятность случайного выбора субпопуляции* $\mathbf{s}_1$ *из метапопуляции* $\mathbf{s}_2$.

Иным образом общую (*полную*) *вероятность* $Hs(\mathbf{s}_2|Fs_1)$ *можно представить как*

$$Hs(\mathbf{s}_2|Fs_1) = Hs_0(\mathbf{s}_2)\left(1 - \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2))}{x(\mathbf{s}_2)(1-x(\mathbf{s}_2))}\right)\times \\ \times (1 - E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}) = Hs_0(\mathbf{s}_2)(1 - Fs(\mathbf{s}_2))\times \\ \times (1 - E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}), \\ Fs(\mathbf{s}_2) \equiv \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2))}{x(\mathbf{s}_2)(1-x(\mathbf{s}_2))} \ge 0, \tag{16}$$

где $Fs(\mathbf{s}_2)$ *является фамильным аналогом случайного коэффициента инбридинга* F_{ST}, *а средний коэффициент фамильного инбридинга* $E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}$ *для субпопуляций* $\{\mathbf{s}_1\}$ *в* $\mathbf{s}_2$ *служит аналогом коэффициента* F_{IS} *неслучайного инбридинга С. Райта в популяционной генетике.*

Доказательство. Общая вероятность $Hs(\mathbf{s}_2)$ рассматриваемых встреч в метапопуляции $\mathbf{s}_2$ по определению находится как математическое ожидание

для значений $Hs(\mathbf{s}_1)$ в субпопуляциях $\{\mathbf{s}_1\}$, составляющих разбиение $\mathbf{s}_2$:

$$Hs(\mathbf{s}_2) \equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}_2\} = \\ = E_{\mathbf{s}_1}\{x(\mathbf{s}_1)(1-x(\mathbf{s}_1))(1-Fs(\mathbf{s}_1))|\mathbf{s}_2\}.$$

При независимости x и Fs данное выражение равно произведению математических ожиданий сомножителей и

$$Hs(\mathbf{s}_2) = E_{\mathbf{s}_1}\{x(\mathbf{s}_1)(1-x(\mathbf{s}_1))|\mathbf{s}_2\}\times \\ \times E_{\mathbf{s}_1}\{1-Fs(\mathbf{s}_1)|\mathbf{s}_2\} = x(\mathbf{s}_2)(1-x(\mathbf{s}_2))\times \\ \times(1-Fs(\mathbf{s}_2))(1-E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}.$$

Здесь коэффициент фамильного инбридинга $Fs(\mathbf{s}_2)$ характеризует подразделенность метапопуляции $\mathbf{s}_2$ (фамильную дивергенцию субпопуляций $\mathbf{s}_1$ внутри $\mathbf{s}_2$) и по определению равен отношению дисперсии распределения концентрации фамилии по субпопуляциям $\{\mathbf{s}_1\}$ к вероятности *случайных* встреч $Hs_0(\mathbf{s}_2) \equiv x(\mathbf{s}_2)(1-x(\mathbf{s}_2))$ в $\mathbf{s}_2$:

$$Fs(\mathbf{s}_2) \equiv \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2))}{x(\mathbf{s}_2)(1-x(\mathbf{s}_2))} = \frac{Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_2))}{Hs_0(\mathbf{s}_2)},$$
$$Hs_0(\mathbf{s}_2) \equiv x(\mathbf{s}_2)(1-x(\mathbf{s}_2)).$$

Он соответствует коэффициенту случайного инбридинга F_{ST} в популяционной генетике, отражающему генетическую дивергенцию *между* субпопуляциями.

Коэффициент $E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}$ является аналогом неслучайного коэффициента инбридинга F_{IS} в генетике популяций и отражает усредненное нарушение случайного характера встреч (закона Харди–Вайнберга) *внутри* субпопуляций, служащих единицей наблюдения. ◀

Повторим, что здесь неявно предполагается независимость $Fs(\mathbf{s}_1)$ и концентрации x рассматриваемой фамилии. Коэффициенты $\{Fs(\mathbf{s}_1)\}$ для субпопуляций первого уровня заданы изначально и по ним находятся средние коэффициенты неслучайного фамильного инбридинга для субпопуляций более высоких уровней. Для единообразия записи вероятности встреч метапопуляции $\mathbf{s}_2$ можно использовать коэффициент общего инбридинга F_{IT} в $\mathbf{s}_2$. В популяционной генетике он учитывает одновременно влияние как дивергенции между субпопуляциями, так и нарушения закона Харди–Вайнберга внутри них на гетерозиготность H метапопуляции $\mathbf{s}_2$ (и на корреляцию между гаметами генотипа):

$$H(\mathbf{s}_2) = x(\mathbf{s}_2)(1-x(\mathbf{s}_2))(1-F_{IT}),$$
$$1-F_{IT} \equiv (1-F_{ST})(1-F_{IS}).$$

РАЗЛОЖЕНИЕ ПО УРОВНЯМ ИЕРАРХИИ ВЕРОЯТНОСТИ ВСТРЕЧ ПРИ ИЕРАРХИЧЕСКОЙ ПОДРАЗДЕЛЕННОСТИ МЕТАПОПУЛЯЦИИ

Рассмотренный случай подразделенности метапопуляции на неподразделенные группы первого уровня можно считать простейшим случаем иерархической структуры. Теперь обратимся к подразделенной метапопуляции $\mathbf{s}_m$ с произвольным количеством уровней иерархии m. Если $\mathbf{s} = \mathbf{s}_m$, то дисперсию распределения концентрации по субпопуляциям $\{\boldsymbol{s}_k\}$ на уровне k в метапопуляции $\mathbf{s}_m$ обозначим как $Vs_{betw}(x(\boldsymbol{s}_k|\mathbf{s}_m))$, $k < m$. В частности, $Vs(x(\mathbf{s}_1|\mathbf{s}_m)) = Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_m))$, причем $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_m))$ одновременно является общей дисперсией $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}_m))$ для распределения концентрации фамилии по всем субпопуляциям, служащим единицей наблюдения.

Проанализируем выражение для вероятности встреч Hs в метапопуляции в зависимости от количества ее уровней иерархии m. Дадим решение этой задачи при независимости x и Fs путем последовательного увеличения m и индукции.

В предшествующем следствии мы нашли, что при фиксированной метапопуляции $\mathbf{s} = \mathbf{s}_2$

$$Hs(\mathbf{s}_2) \equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}_2\} = \\ = (x(\mathbf{s}_2)(1-x(\mathbf{s}_2)) - Vs(x(\mathbf{s}_1|\mathbf{s}_2)))(1-E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}).$$

Теперь зафиксируем $\mathbf{s}_3$, подставим в рекуррентное уравнение (8), определяющее $Hs(\mathbf{s}_3)$, найденное значение $Hs(\mathbf{s}_2)$, учтем эффект Валунда (11) и заменим математическое ожидание произведения произведением математических ожиданий сомножителей в силу предполагаемой независимости x и Fs:

$$Hs(\mathbf{s}_3|Fs_1) \equiv E_{\mathbf{s}_2}\{Hs(\mathbf{s}_2)|\mathbf{s}_3\} = \\ = E_{\mathbf{s}_2}\{(x(\mathbf{s}_2)(1-x(\mathbf{s}_2)) - Vs(x(\mathbf{s}_1|\mathbf{s}_2)))|\mathbf{s}_3\} \\ (1-E_{\mathbf{s}_2}\{E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_2\}|\mathbf{s}_3\}) = \\ = (x(\mathbf{s}_3)(1-x(\mathbf{s}_3)) - Vs(x(\mathbf{s}_2|\mathbf{s}_3)) - \\ -E_{\mathbf{s}_2}\{Vs(x(\mathbf{s}_1|\mathbf{s}_2))|\mathbf{s}_3\})(1-E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_3\}).$$

Ремарка 7. Заметим, что $Vs(x(\mathbf{s}_2|\mathbf{s}_3))$ можно записать как $E_{\mathbf{s}_3}\{Vs(x(\mathbf{s}_2|\mathbf{s}_3))|\mathbf{s}_3\}$, поскольку метапопуляция $\mathbf{s}_3$ единственна. Поэтому

$$E_{\mathbf{s}_2}\{Hs(x(s_2|s_3)\} = \left(x(\mathbf{s}_3)(1-x(\mathbf{s}_3)) - \sum_{i=1}^{2} E_{\mathbf{s}_{i+1}}\{Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_3\}\right)(1-E_{\mathbf{s}_1}\{Fs(x(\mathbf{s}_1)|\mathbf{s}_3\}).$$

Найденное выражение для $Hs(\mathbf{s}_3)$ и ремарка подсказывают, что для субпопуляций j-го уровня

$$Hs(\mathbf{s}_j) = \left(x(\mathbf{s}_j)(1 - x(\mathbf{s}_j)) - \sum_{i=1}^{j-1} E_{\mathbf{s}_{i+1}} \{ Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_j \} \right) (1 - E_{\mathbf{s}_1} \{ Fs(\mathbf{s}_1)|\mathbf{s}_j \}).$$

С другой стороны, поскольку $Hs(\mathbf{s}_j) = (x(\mathbf{s}_j)(1 - x(\mathbf{s}_j)) - Vs(x(\mathbf{s}_1|\mathbf{s}_j)))(1 - E_{\mathbf{s}_1}\{Fs(\mathbf{s}_1)|\mathbf{s}_j\})$ согласно фамильному аналогу эффекта Валунда (11), где роль метапопуляции $\mathbf{s}$ играет $\mathbf{s}_j$, то одновременно мы видим разложение дисперсии $Vs(x(\mathbf{s}_1|\mathbf{s}_j))$ на сумму членов $E_{\mathbf{s}_{i+1}}\{Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_j\}$.

Результат 8. *Рассмотрим иерархически подразделенную метапопуляцию $\mathbf{s}_m$ с уровнями иерархии $i = 1, 2, \ldots, m$. Положим, что в ее субпопуляциях $\{\mathbf{s}_1\}$ концентрации интересующей фамилии равны $\{x(\mathbf{s}_1)\}$ и встречи индивидуумов происходят случайно с вероятностями $\{Hs(\mathbf{s}_1) = Hs_0(\mathbf{s}_1) \equiv x(\mathbf{s}_1)(1 - x(\mathbf{s}_1))\}$.*

Пусть из $\mathbf{s}_m$ наугад выбрана субпопуляция $\mathbf{s}_1$ первого уровня. Тогда вероятность $Hs(\mathbf{s}_m)$ встречи в $\mathbf{s}_1$ пары индивидуумов с интересующей фамилией и с какой-нибудь иной с учетом их порядка (полная вероятность встречи) допускает следующее разложение:

$$\begin{aligned} Hs(\mathbf{s}_m) &\equiv E_{\mathbf{s}_1}\{Hs(\mathbf{s}_1)|\mathbf{s}_m\} = \\ &= x(\mathbf{s}_m)(1 - x(\mathbf{s}_m)) - Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_m)) = \\ &= Hs_0(\mathbf{s}_m) - \sum_{i=1}^{m-1} E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}. \end{aligned} \quad (17)$$

Таким образом, общая (полная) вероятность $Hs(\mathbf{s}_m)$ равна разности между значением вероятности $Hs_0(\mathbf{s}_m) \equiv x(\mathbf{s}_m)(1 - x(\mathbf{s}_m))$ случайной встречи такой пары во всей метапопуляции $\mathbf{s}_m$ и разложением межгрупповой дисперсии $Vs_{betw}(x(\mathbf{s}_1|\mathbf{s}_m))$ на средние внутригрупповые дисперсии $E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}$, соответствующие отдельным уровням иерархии i.

Доказательство. Воспользуемся индукцией. Выше мы видели, что разложение (17) для $Hs(\mathbf{s}_m)$ верно при некотором m. Покажем, что тогда оно справедливо для $Hs(\mathbf{s}_{m+1})$. Согласно рекуррентному уравнению (8)

$$\begin{aligned} Hs(\mathbf{s}_{m+1}) &\equiv E_{\mathbf{s}_m}\{Hs(\mathbf{s}_m)|\mathbf{s}_{m+1}\} = \\ &= E_{\mathbf{s}_m}\{x(\mathbf{s}_m)(1 - x(\mathbf{s}_m)) - \\ &- \sum_{i=1}^{m-1} E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}|\mathbf{s}_{m+1}\} = \\ &= E_{\mathbf{s}_m}\{x(\mathbf{s}_m)(1 - x(\mathbf{s}_m))\} - \\ &- E_{\mathbf{s}_m}\left\{\sum_{i=1}^{m-1} E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}|\mathbf{s}_{m+1}\right\} = \\ &= x(\mathbf{s}_{m+1})(1 - x(\mathbf{s}_{m+1})) - Vs(x(\mathbf{s}_m|\mathbf{s}_{m+1})) - \\ &- \sum_{i=1}^{m-1} E_{\mathbf{s}_m}\{E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}|\mathbf{s}_{m+1}\}. \end{aligned}$$

Рассмотрим отдельное слагаемое в сумме по i. Согласно (6), где роль i играет $i + 1$, а $T(\mathbf{s}_{i+1}) = Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1})$, имеем

$$\begin{aligned} &E_{\mathbf{s}_m}\{E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}|\mathbf{s}_{m+1}\} = \\ &= E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_{m+1}\}. \end{aligned}$$

Поскольку в соответствии со сделанной ремаркой $Vs_{in}(x(\mathbf{s}_m|\mathbf{s}_{m+1})) = E_{\mathbf{s}_{m+1}}\{Vs_{in}(x(\mathbf{s}_m|\mathbf{s}_{m+1}))|\mathbf{s}_{m+1}\}$, этот член можно включить в рассматриваемую сумму, верхний предел которой увеличится на единицу и станет равен $(m - 1) + 1 = m$:

$$\begin{aligned} Hs(\mathbf{s}_{m+1}) &\equiv E_{\mathbf{s}_m}\{Hs(\mathbf{s}_m)|\mathbf{s}_{m+1}\} = x(\mathbf{s}_{m+1})(1 - x(\mathbf{s}_{m+1})) - \\ &- \sum_{i=1}^{m} E_{\mathbf{s}_{i+1}}\{Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_{m+1}\}. \end{aligned}$$

Полученное выражение совпадает с (17), если вместо m поставить $m + 1$. ◀

Из (17) отчетливо видно, что вероятность Hs для уровня выше первого меньше значения, соответствующего случайным встречам, и чем больше сумма средних внутригрупповых дисперсий, тем больше увеличивается указанный разрыв. Полученное разложение является некоторого рода обобщением эффекта Валунда [10]. Заметим, что здесь не требуется независимость концентрации фамилии и коэффициента фамильного инбридинга.

Следствие 9. *В иерархически подразделенной метапопуляции $\mathbf{s}_m$ с m уровнями иерархии полная дисперсия концентрации фамилии $Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}_m))$ допускает разложение*

$$\begin{aligned} Vs_{tot}(x(\mathbf{s}_1|\mathbf{s}_m)) &= \sum_{i=1}^{m-1} E_{\mathbf{s}_{i+1}}\{Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\} = \\ &= \sum_{i=1}^{m-1} (Vs(x(\mathbf{s}_i|\mathbf{s}_m)) - Vs(x(\mathbf{s}_{i+1}|\mathbf{s}_m))). \end{aligned} \quad (18)$$

Доказательство первого равенства содержится в предыдущем результате, т.е. там попутно получено разложение дисперсии, найденное нами ранее в [4] более длинным способом.

Второе равенство вытекает из правила сложения дисперсий (14), согласно которому $E_{\mathbf{s}_k}\{Vs_{in}(x(\mathbf{s}_n|\mathbf{s}_k))|\mathbf{s}_m\} = Vs_{tot}(x(\mathbf{s}_n|\mathbf{s}_m)) - Vs_{betw}(x(\mathbf{s}_k|\mathbf{s}_m)) \geq 0$. В нашем случае при $n = i$, $k = i + 1$

$$\begin{aligned} &E_{\mathbf{s}_{i+1}}\{Vs_{in}(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\} = Vs(x(\mathbf{s}_i|\mathbf{s}_m)) - \\ &- Vs(x(\mathbf{s}_{i+1}|\mathbf{s}_m)), \quad n < k < m. \end{aligned}$$

Таким образом, полная дисперсия $Vs(x(\mathbf{s}_1|\mathbf{s}_m))$ равна сумме приращений межгрупповой дисперсии при переходе от уровня $i + 1$ к предыдущему i.

Подобно тому, как пройденный путь складывается из приращений по шагам, так и при суммировании приращений дисперсии по i получим полную дисперсию $Vs(x(\mathbf{s}_1|\mathbf{s}_m))$, поскольку все члены в сумме сокращаются, кроме $Vs(x(\mathbf{s}_1|\mathbf{s}_m))$, а $Vs(x(\mathbf{s}_m|\mathbf{s}_m))$ разумно определить как нуль. ◄

Так как левая часть $E_{\mathbf{s}_{i+1}}\{Vs(x(\mathbf{s}_i|\mathbf{s}_{i+1}))|\mathbf{s}_m\}$ последней формулы неотрицательна (как среднее значение неотрицательной дисперсии), то при переходе к более высокому уровню $i + 1$ межгрупповая дисперсия (дисперсия концентрации фамилии в субпопуляциях одного и того же уровня) не увеличивается. Это интуитивно ожидается, поскольку при группировке данных разброс значений не возрастает, и дисперсия средних значений для сгруппированных данных не может быть больше дисперсии для исходных данных.

ЗАКЛЮЧЕНИЕ

Пройдемся по основным линиям данной работы – ее мотивации, подходу к анализу, результатам. Интерес к фамильной структуре мотивирован вниманием к генеалогии, родственным связям и др. аспектам, так или иначе связанным с генетикой населения. Настоящая статья ориентирована на популяционно-генетическую сторону изучения распределения фамилий, и эта ориентация приводит к использованию и обобщениям традиционных методов анализа генетической структуры по данным о распределении концентраций маркеров в ряде популяций.

Выбор такого подхода обусловлен тем, что популярные модели процессов выборочного генного дрейфа Райта–Фишера и дрейфа фамилий в популяциях ограниченной численности при стандартных предположениях совпадают (отличаются лишь интенсивностью, см. [2, 3]). Следовательно, это должно позволить при соответствующих поправках экстраполировать результаты одного из них на ожидаемые результаты для другого. Природа процессов дрейфа с динамикой генетического и фамильного состава популяции обусловлена случайными колебаниями количества потомков у отдельных индивидуумов.

Если у некоторого родителя-мужчины больше сыновей, чем в среднем, то в следующем поколении будет больше носителей его фамилии и его генов. На популяционном уровне это приводит к случайным колебаниям фамильного и генетического составов популяции от поколения к поколению, представляющим сущность процессов дрейфа. Оба процесса приводят к дивергенции родственных популяций (генетической и фамильной). Очевидно, требуется развивать методы, характеризующие дивергенцию популяций (неоднородность системы популяций в целом) для типичных для человека популяционных структур.

Данные по фамилиям часто доступны в иерархически структурированном виде, например в соответствии с административно-территориальной подразделенностью. В качестве показателей дивергенции служат межгрупповая дисперсия распределения концентрации (кодоминантного аллеля, фамилии) в группах (субпопуляциях), нарушения закона Харди–Вайнберга с эффектом Валунда, коэффициент случайного инбридинга. Рассмотрены фамильные аналоги этих показателей и даны их обобщения для иерархически подразделенной метапопуляции с произвольным количеством уровней иерархии. Данные обобщения вычислительно универсальны в том смысле, что не зависят от характера разбиений метапопуляции на каждом из уровней, от типа микроэволюционного процесса, приведшего к текущему состоянию метапопуляции, от существования миграций и пр. Неоднородность (дивергенция) в иерархических системах характеризуется алгоритмически одинаковыми показателями, различающимися между системами количественно.

В настоящей статье в качестве основного показателя фамильной дивергенции субпопуляций рассматривается общая вероятность встреч (столкновений) Hs лиц с разными фамилиями (аналог концентрации гетерозигот), понимаемая как вероятность встречи в выбранной наугад субпопуляции, единицы наблюдения, из рассматриваемой метапопуляции с иерархической структурой субпопуляций. Получено разложение Hs по уровням иерархии, обобщающее эффект Валунда в популяционной генетике. Общая вероятность неизонимных встреч в иерархически подразделенной метапопуляции меньше вероятности случайных встреч в ней на сумму средних внутригрупповых дисперсий концентрации фамилии, соответствующих отдельным уровням.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

СПИСОК ЛИТЕРАТУРЫ

1. *Ли Ч.* Введение в популяционную генетику. М.: Мир, 1978. 555 с.
2. *Пасеков В.П.* К анализу случайных процессов изонимии. I. Структура изонимии // Генетика. 2021. Т. 57. № 10. С. 1194–1204. https://doi.org/10.31857/S001667582110009X
3. *Пасеков В.П.* К анализу случайных процессов изонимии. II. Динамика дивергенции популяций // Генетика. 2021. Т 57. № 11. С. 1318–1329. https://doi.org/10.31857/S0016675821110114
4. *Пасеков В.П.* Описание дивергенции субпопуляций в иерархической системе при анализе изони-

мии. I. Дисперсия как показатель дивергенции // Генетика. 2022. Т. 58. № 6. С. 713–727.

5. *Wright S.* The interpretation of population structure by *F* statistics with special regard to systems of mating // Evolution. 1965. V. 19. P. 395–420.
6. *Гинтер Е.К., Зинченко Р.А., Ельчинова Г.И. и др.* Роль факторов популяционной динамики в распространении наследственной патологии в российских популяциях // Мед. генетика. 2004. Т. 3. № 12. С. 548–555.
7. *Ревазов А.А., Парадеева Г.М., Русакова Г.И.* Пригодность русских фамилий в качестве квазигенетического маркера // Генетика. 1986. Т. 22. № 4. С. 699–703.
8. *Crow J.F., Mange A.P.* Measurement of inbreeding from the frequency of marriages between persons of the same surname // Social Biology. 1982. V. 29. № 1/2. P. 101–105.
9. *Lasker W.G.* Surnames and Genetic Structure. Cambridge: Cambr. Univ. Press, 1985. 148 p.
10. *Сорокина И.Н., Чурносов М.И., Балтуцкая И.В. и др.* Антропогенетическое изучение населения центральной России. М.: Изд-во РАМН, 2014. 336 с.

Description of Divergence of Subpopulations in the Hierarchical System under the Analysis of Isonymy. II. Probabilities of Non-Isonymic Encounters

V. P. Passekov[a, *]

[a] *Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, 119991 Russia*
**e-mail: pass40@mail.ru*

Metapopulations with typical for human populations hierarchical subdivision into parts (subpopulations) are considered, corresponding to classifications based on administrative-territorial division (say, village, village council, district, region, and so on) or on a genealogical approach based on ethnogenesis, as well as on other principles of biological classification. The purpose of this work is to analyze the general properties of the distribution of the surname concentration over subpopulations in their hierarchical structure. Attention is focused on the description of the surname divergence of subpopulations, as an indicator of which the total probability of two person with different surnames encounter is considered, which is understood as the probability of the encounter in a randomly chosen subpopulation, a unit of observation, from metapopulation with a hierarchical structure. Its decomposition by hierarchy levels is obtained, which generalizes the Wahlund effect in population genetics. The total probability of non-isonimic encounters in a hierarchically subdivided metapopulation is less than the probability of random encounters in it by the sum of the average surname concentration intragroup variances corresponding to separate levels. Such properties are purely statistical characteristics of the hierarchical structure, and not a feature of a particular population system, and are not derived from the regularities of one or another model of microevolution. They are computationally formulated in the same way for any hierarchical system, although in the general case they do not coincide quantitatively. The results obtained refer to rural and urban hierarchical metapopulations as separate components of the entire population.

Keywords: hierarchical structure of populations, metapopulations, surname concentrations in human subpopulations, characteristics of heterogeneity of subpopulations, decomposition of the probability of encounter carriers of different surnames by hierarchy levels.