

А. Т. Фоменко

**НЕКОТОРЫЕ СТАТИСТИЧЕСКИЕ  
ЗАКОНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ  
ПЛОТНОСТИ ИНФОРМАЦИИ В ТЕКСТАХ  
СО ШКАЛОЙ**

1. В работе предьявляются результаты серии статистических исследований, в процессе которых были обнаружены новые статистические инварианты типа «законов сохранения».

Под текстом  $X$  со шкалой  $P$  будем понимать текст, снабженный фиксированным планом-программой  $P$  так, что  $X$  допускает единственное разбиение по пунктам  $P$ , т. е. каждому пункту  $t \in P$  плана-программы соответствует некоторая часть текста  $X(t)$  и, наоборот, каждая фраза (слово) текста отнесена к одному и только одному пункту  $t \in P$ . Такой план-программу будем называть шкалой или параметром текста  $X$ ; текст  $X$  параметризован параметром  $t$ ; имеем:  $X = \bigcup_{t \in P} X(t)$ ;  $X(t_1) \cap X(t_2) = \emptyset$ , если  $t_1 \neq t_2$ ;

$t_1, t_2 \in P$ . Хорошим примером таких текстов  $X$  являются тексты исторического характера, например, монографии, хроники, вузовские учебники и т. д. Роль параметра в этих текстах играет время  $t$ , т. е. даты описываемых событий. В то же время понятие текста со шкалой является значительно более общим, чем приведенный пример: в частности, параметр  $t$  может пробегать более сложные области определения, чем натуральные числа; например,  $t$  может иметь непрерывную область изменения. Это имеет место в том случае, если в качестве текста со шкалой берется описание какого-либо физического непрерывного процесса, параметризованное временем; тогда в ка-

честве  $X(t)$  можно взять описание мгновенного состояния этого процесса, составленное непрерывно следящей системой. Далее, для приложений, рассматриваемых в настоящей статье, большую роль играет также следующий пример: шкала  $P$  состоит из последовательности целочисленных отрезков  $[A_j, B_j]$ , таких, что  $[A_j, B_j] \cap [A_\alpha, B_\alpha] = \emptyset$ , если  $\alpha \neq j$  и  $\cup [A_j, B_j] = [A, B]$ ; здесь  $t$  пробегает последовательность  $\{[A_j, B_j]\}$  (подробности см. ниже в пункте 3).

2. Пусть  $X$  — текст со шкалой  $P$  и  $X = \cup_{t \in P} X(t)$ .

Пусть параметр  $t$ , для простоты, пробегает целые числа (положительные или отрицательные) от  $A$  до  $B$ ;  $t \in [A, B]$  (например, в тексте  $X$  описан период времени от  $A$  до  $B$ ; в этом случае  $t, A, B$  могут измеряться, например, в годах, месяцах, днях, часах). Сопоставим каждому  $t \in [A, B]$  набор  $f_1(t), f_2(t), \dots, f_N(t)$  формальных информационных характеристик подмножества  $X(t)$  текста  $X$ , соответствующего  $t$ . ( $X(t)$  — подмножество  $X$ , описывающее год  $t \in [A, B]$ ). Характеристики  $f_i(t)$  могут быть весьма разнообразны, например:

$$(1) f_1(t) = \frac{1}{V} \cdot v(t),$$

где  $v(t)$  — число строк (или страниц, или знаков), составляющих  $X(t)$ ; тогда  $f_1(t) \geq 0$ ,  $\sum_{t=A}^B f_1(t) = 1$ ; здесь

$V$  — полный объем  $X$ . В полный объем  $V$  не включаются иллюстрации, схемы, библиография. Условие нормирования

$\sum_{t=A}^B f_1(t) = 1$  удобно при усреднениях и сравнениях, являясь дискретным аналогом непрерывной модели:

$$\int_A^B f_1(x) dx = 1.$$

$$(2) f_2(t) = \frac{1}{S} \cdot s(t),$$

где  $s(t)$  — (число упоминаний года  $t$  в  $X$ ), а  $S$  — общее число упоминаний дат (лет) в  $X$ .

$$(3) f_3(t) = \frac{1}{M} \cdot m(t),$$

где  $m(t)$  — число имен персонажей, упомянутых в  $X(t)$ ,  
 $M$  — полное число всех имен в  $X$ .

$$(4) f_4(t) = \frac{m(t)}{M},$$

где  $m_i(t)$  — число упоминаний какого-то одного конкретного имени  $m_i$  в  $X(t)$ ;  $M_i$  — общее число упоминаний  $m_i$  в  $X$ .

Ниже будут приведены другие важные примеры функций  $f_i(t)$ , которые будем называть информативными (или частотными) функциями  $X$ .

3. Подсчет  $f_i(t)$  сводится к формальному подсчету в  $X(t)$  указанной числовой характеристики. Иногда в конкретных тестах  $X$  подмножества  $X(t)$  устроены так, что однозначно определено только разбиение  $X = \cup_j X(A_j, B_j)$ , где  $X(A_j, B_j)$  — часть  $X$ , описывающая целый временной отрезок  $[A_j, B_j]$ , без детализирования событий по отдельным годам  $t \in [A_j, B_j]$ ;  $[A, B] = \cup_j [A_j, B_j]$ . В этом случае следует рассмотреть

укрупненную шкалу (для одного и того же текста  $X$  может существовать несколько различных шкал, которые, таким образом, можно варьировать — в зависимости от поставленной задачи), в которой  $t$  пробегает последовательность отрезков  $\{[A_j, B_j]\}$ , и функция  $f_i(t)$  подсчитывается так:  $f_i(t) = \frac{1}{f_i(A, B)} \cdot \frac{f_i(A_j, B_j)}{B_j - A_j}$  (среднее значение на  $[A_j, B_j]$ ), где  $t \in [A_j, B_j]$ , а  $f_i(A_j, B_j)$  и  $f_i(A, B)$  — значения, которые функция  $f_i$  принимает на отрезках  $[A_j, B_j]$  и  $[A, B]$ , соответственно. Например, для  $f_1$ ,  $f_1(A, B) =$  число строк, или страниц, составляющих  $X(A, B)$ ;  $V = f_1(A, B) =$  (полный объем текста  $X$ ); здесь  $X = \cup_j X(A_j, B_j)$ ;  $f_1(A, B) = \sum_j f_1(A_j, B_j)$ . Итак, каждая информативная функция  $f_i$  изображается графиком, определенным на отрезке  $[A, B]$ ;  $f_i \geq 0$ .

4. Обозначим через  $f(t)$  какую-либо информативную функцию  $f_i(t)$  (опуская для простоты индекс  $i$ ). Пусть даны два текста  $X$  и  $Y$  (или несколько текстов), описывающие один и тот же  $[A, B]$ , например, историю одного и того же государства за период  $[A, B]$ . Ясно, что функ-

ции  $f(t, X)$  и  $f(t, Y)$ , построенные для  $X$  и  $Y$ , будут, вообще говоря, различны: на их вид влияют как личные интересы авторов, так и общие методические установки, в рамках которых создается текст (скажем, монография по истории искусства и монография по военной истории одного и того же периода будут существенно по-разному расставлять акценты). Вопрос: насколько существенны эти различия при статистическом подходе к анализу информативных функций, т. е. нельзя ли обнаружить некоторые общие, инвариантные характеристики функций  $f(t, X)$  и  $f(t, Y)$ , не зависящие (или зависящие мало) от автора.

5. Изучим поведение таких важных характеристик функций  $f(t, X)$  как распределение локальных максимумов графика  $f(t, X)$ . Обозначим через  $r_q(X)$ ,  $1 \leq q \leq Q(X)$  те значения  $t \in [A, B]$ , в которых функция  $f(t, X)$  достигает локального максимума. (В том случае, когда шкала текста укрупнена (см. пример выше), функция  $f$  имеет вид ступенчатой функции, постоянной на каждом отрезке  $[A_j, B_j]$ , и в качестве  $r_q(X)$  возьмем  $\frac{1}{2}(A_j + B_j)$ ).

Возьмем в качестве модельного примера функцию  $f_1(t, X) = \frac{v(t)}{V}$ , где  $v(t)$  — число страниц (или строк, или знаков), составляющих  $X(t)$ . Что означает, что  $f_1(t, X)$  достигает в некоторой точке  $r_q(X) \in [A, B]$  локального максимума? Это означает, что год  $r_q(X)$  описан в  $X$  более подробно (с большими деталями, например), чем соседние годы. Чем может объясняться такая неравномерность в описании? Один из возможных (и, может быть, основных) механизмов, таков: до автора текста  $X$  от года  $r_q(X)$  дошло больше информации (например, больший объем текстов), чем от соседних лет, что и обусловило большую подробность изложения для  $r_q(X)$ .

6. Теоретическая модель, описывающая распределение локальных максимумов  $r_q(X)$  для  $f_1(t)$ , определена в п. 5. Обозначим через  $H(t)$  число строк (страниц, знаков), написанных в каждый год  $t \in [A, B]$  современниками года  $t$ . Ясно, что мы не можем, вообще говоря, восстановить график  $H(t)$  ввиду возможной утери древних

текстов с течением времени (до нашего времени дошла только какая-то часть  $H(t)$ ). Пусть  $h_q$ ,  $1 \leq q \leq Q_0$  — годы, в которые  $H(t)$  достигал локальных максимумов (т. е. в эти годы по тем или иным причинам было написано особенно много текстов, зафиксировано особенно много информации). Каков может быть статистический механизм потери и забывания информации, приводящий к уменьшению амплитуды графика  $H(t)$ ? Гипотеза: хотя с течением времени амплитуда графика  $H(t)$  уменьшается (тексты теряются), но тем не менее, от тех лет  $h_q$ , в которые было написано особенно много текстов, больше и останется. Другими словами, точки  $h_q$ , в которых достигал локальных максимумов график  $H(t)$ , должны быть близки к точкам  $r_q(X)$ , в которых достигает локальных максимумов график  $f_1(t, X)$ , построенный по более позднему тексту  $X$ .

Поскольку график  $H(t)$  нам неизвестен, проверить эту гипотезу (в таком ее виде) затруднительно. Однако можно легко проверить одно из ее следствий. Следствие гипотезы: точки  $r_q(X)$  должны быть близки к точкам  $r_q(Y)$  для любых двух текстов  $X$  и  $Y$ , описывающих одну и ту же эпоху  $[A, B]$ . Другими словами, графики  $f_1(t, X)$  и  $f_1(t, Y)$  должны достигать локальных максимумов в (примерно) одних и тех же точках на отрезке  $[A, B]$ . Назовем это следствие  $F$ -моделью. Интуитивное обоснование  $F$ -модели: поскольку тексты  $X$  и  $Y$ , описывая один и тот же  $[A, B]$  (для тех случаев, когда  $[A, B]$  значительно удален во времени от авторов  $X$  и  $Y$ ), уже не являются современниками описываемых ими событий, то, следовательно, они должны опираться на какой-то набор исторических документов, дошедших до их времени от  $t \in [A, B]$ . Поскольку этот «уцелевший набор» — общий для  $X$  и  $Y$ , это и должно отразиться в том, что графики  $f_1(t, X)$  и  $f_1(t, Y)$  должны «одновременно» достигать локальных максимумов при описании  $[A, B]$ . Эти годы локальных максимумов характеризуются тем, что от них уцелело особенно много информации по сравнению с соседними годами. Другое естественное следствие из основной гипотезы см. в пункте 16 ( $K$ -модель).

7. Не следует думать, что для таких (см. выше) текстов  $X$  и  $Y$  должны быть близки (после нормировки) и сами амплитуды локальных максимумов; простые примеры показывают, что значения функций  $f(r_q, X)$  и

$f(r_q, Y)$  могут значительно отличаться друг от друга, а потому сами значения  $f(r_q, X)$  — не инвариантны при изменении  $X$ . Это указывает на то, что обнаружение инвариантов должно быть основано не на абсолютных амплитудах информативных функций.

8. Проверим сформулированную выше  $F$ -модель на конкретном материале. Для этого следует математически сформулировать понятие близости двух последовательностей чисел  $r_q(X), r_q(Y)$ . Если  $X, Y$  — два текста, то числа  $Q(X)$  и  $Q(Y)$ , вообще говоря, различны (разное число локальных максимумов). Однако можно считать, без ограничения общности, что  $Q(X) = Q(Y)$ ; для этого достаточно допустить, что некоторые точки  $r_q(Y)$  совпадают с некоторыми новыми точками  $r_{q+1}(Y)$  в том случае, когда  $Q(Y) < Q(X)$ . (Другими словами, вводятся кратные точки локальных максимумов.) Это уравнивание числа локальных максимумов может выполняться разными способами (в зависимости от того, равенство каких  $r_q(Y)$  и  $r_{q+1}(Y)$  допускается); фиксируем какой-нибудь из них (в дальнейшем будет произведена минимизация по всем этим способам).

9. Числа  $r_q(X), 1 \leq q \leq Q(X)$  разбивают отрезок  $[A, B]$  на отрезки  $A_1, A_2, \dots, A_{Q(X)+1}$  с целочисленными длинами  $a_1 = r_1 - A; a_q = r_q - r_{q-1}, 2 \leq q \leq Q(X); a_{Q(X)+1} = B - r_{Q(X)}$ ; в количестве  $Q(X) + 1$ ; тем самым определен целочисленный вектор  $\vec{a}(X) = (a_1(X), \dots, a_{Q(X)+1}(X)) \in \mathbb{R}^p$ , где  $p = Q(X) + 1, \mathbb{R}^p$  — евклидово пространство размерности  $p$ . Так как  $\sum_{q=1}^p a_q(X) = B - A$ ,

то конец вектора  $\vec{a}(X)$  лежит на  $(p-1)$ -мерном симплексе  $\sigma^{p-1}$ , задаваемом уравнением:  $\sum_{q=1}^p x^q = B - A,$

$x^q \geq 0$ , где  $x^q$  — координаты в  $\mathbb{R}^p$ . Симплекс  $\sigma^{p-1}$  замкнут, т. е. его граница присоединена к  $\sigma^{p-1}$  (поскольку некоторые координаты  $a_q(X)$  могут быть нулевыми). Пусть даны два текста  $X, Y$ ; тогда получаем два вектора  $\vec{a}(X)$  и  $\vec{a}(Y)$ , концы которых лежат на одном и том же симплексе  $\sigma^{p-1}$  (рис. 1). Мы воспользовались тем (см. п. 8), что можно предполагать выполненным равенство:  $Q(X) = Q(Y) = p + 1$ . Рассмотрим вектор

$\vec{\Delta}(Y, X) = \vec{a}(Y) - \vec{a}(X)$  (см. рис. 1); вектор  $\vec{\Delta}$  принадлежит  $\sigma^{p-1}$ . Пусть  $\lambda = |\vec{\Delta}|$  — длина вектора  $\vec{\Delta}$ ;  
 $\lambda = \sqrt{\sum_{q=1}^p (a_q(Y) - a_q(X))^2}$ . Введем коэффициент

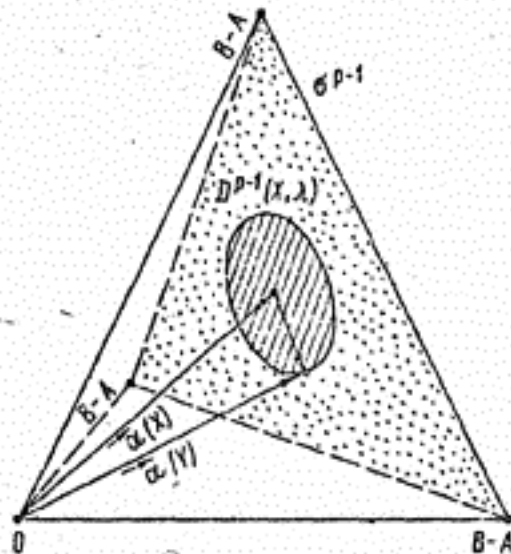


Рис. 1

$d(Y, X) = \text{vol}(D^{p-1}(X, \lambda) \cap \sigma^{p-1}) / \text{vol} \sigma^{p-1}$ , где  $D^{p-1}(X, \lambda)$  —  $(p-1)$ -мерный шар, расположенный в гиперплоскости  $\sum_{q=1}^p x^q = B-A$ , с центром в точке  $\vec{a}(X)$  и радиусом  $\lambda$ ; через  $D^{p-1}(X, \lambda) \cap \sigma^{p-1}$  обозначено пересечение этого шара с симплексом  $\sigma^{p-1}$  (если  $\lambda$  велико, то шар  $D^{p-1}$  может содержать точки, расположенные вне  $\sigma^{p-1}$ ; для малых  $\lambda$  и  $\vec{a}(X) \in \partial \sigma^{p-1}$  (где  $\partial \sigma^{p-1}$  — граница  $\sigma^{p-1}$ ) шар  $D^{p-1}(X, \lambda)$  целиком содержится в  $\sigma^{p-1}$ , и тогда  $D^{p-1}(X, \lambda) \cap \sigma^{p-1} = D^{p-1}(X, \lambda)$ . Через  $\text{vol} S$ , где  $S$  — произвольное  $(p-1)$ -мерное подмножество в гиперплоскости  $\sum_{q=1}^p x^q = B-A$ , обозначено: 1) либо евклидов  $(p-1)$ -мерный объем множества  $S$  (непрерывная модель); 2) либо число целых точек, попавших в множество  $S$ , т. е. число точек в  $S$ , координаты которых целочисленны (дискретная модель).

10. Фиксируем текст  $X$  и будем менять тексты  $Y$ ; тогда введенное в п. 9 число  $d(Y, X)$  можно (при некоторых гипотезах) интерпретировать как вероятность случайного попадания точки  $\vec{a}(Y)$  в шар  $D^{p-1}(X, \lambda)$  радиуса  $\lambda$  с центром в точке  $\vec{a}(X)$ . Эта вероятностная интерпретация совершенно не обязательна для понимания дальнейшего и приводится здесь как полезный для вычислений формализм. Введем на симплексе  $(p-1)$ -мерную меру  $\mu$ , задаваемую так:  $\mu(U) = (\text{число целых точек в } (p-1)\text{-мерной области } U \subset \sigma^{p-1}) / (\text{число целых точек в симплексе } \sigma^{p-1})$ ; это — дискретная модель; или  $\mu(U) = \text{vol}_{p-1} U = (\text{евклидов объем области } U \subset \sigma^{p-1}) / (\text{евклидов объем симплекса } \sigma^{p-1})$ ; это — непрерывная модель. Для вероятностного истолкования  $d(Y, X)$  нужна гипотеза о распределении случайного вектора  $\vec{a}(Y)$ ; предположим, что  $\vec{a}(Y)$  распределен равномерно на  $\sigma^{p-1}$ ; т. е. вероятность попадания  $\vec{a}(Y)$  в  $U$  равна  $\mu(U)$ . Так как  $\mu(\sigma^{p-1}) = 1$ , то  $0 \leq \mu(U) \leq 1$ . Эта гипотеза в некоторой степени может быть обоснована тем, что распределение для  $\vec{a}(Y)$  описывает распределение точек локальных максимумов графика  $f(t, Y)$  при переменном  $Y$  и при переменной эпохе  $[A, B]$  и воспроизводит механизм потери и забывания информации: при гибели какого-либо архива при каком-либо стихийном бедствии и т. д., равновероятно уничтожение любого текста из этого архива, что и может быть переформулировано в виде гипотезы о равномерном распределении для  $\vec{a}(Y)$  при разных эпохах  $[A, B]$ . Для формального вычисления этого распределения у нас нет реальных возможностей, так как потребовалось бы построить все векторы  $\vec{a}(Y)$  для всех дошедших до нашего времени текстов  $Y$ ; в полном объеме задача эта, очевидно, невыполнима, а обработка только какой-либо выборки не соответствует постановке задачи и, кроме того, может значительно исказить общий вид распределения, поскольку заранее нельзя оценить, отражает ли эта выборка реальный механизм потери информации. Равномерность распределения для  $\vec{a}(Y)$  хорошо описывает и то обстоятельство, что при изменении  $[A, B]$  мы не фиксируем ни длину  $B - A$ , ни начало и конец отрезка  $[A, B]$ .



В некоторых примерах (см. ниже), когда полный объем информации, требующей обработки, достаточно обозрим, мы вычислили распределение, найдя эмпирическую гистограмму частот (см. распределение длительностей правлений).

Вероятностное истолкование для  $d$  не будет использоваться ниже иначе, как путем сравнения с конкретными значениями  $d$ , получающимися при обработке заведомо «зависимых текстов»; именно эти, найденные экспериментально, значения и будут служить условным эталоном, критерием сравнения для других текстов.

Коэффициент  $d(Y, X)$  изменяется от 0 до 1; при гипотезе равномерного распределения, чем меньше это число, тем меньше вероятность случайного события, заключающегося в том, что расстояние между векторами  $\vec{a}(Y)$  и  $\vec{a}(X)$  не превышает  $\lambda = |\vec{\Delta}| = |\vec{a}(Y) - \vec{a}(X)|$ .

Если  $\lambda$  велико, то число  $d(Y, X)$  может оказаться равным 1 (в этом случае  $D^{p-1}(X, \lambda) \supset \sigma^{p-1}$ ) и тогда никакого утверждения о близости  $X$  и  $Y$  высказать нельзя, эти тексты далеки друг от друга с точки зрения функции  $f$ . Отметим, что для любых двух текстов  $X$  и  $Y$ , описывающих один и тот же  $[A, B]$ , можно подобрать одну и ту же шкалу  $P$ ; это и позволяет сравнивать тексты. Следующие ниже пункты 11—14 носят формальный характер и не существенны для понимания дальнейшего.

11. Математические формулы для вычисления  $d(Y, X)$ . Если симплекс  $\sigma^{p-1}$  задается уравнением:

$\sum_{q=1}^p x^q = s; x^q \geq 0$ , то  $\text{vol}(\sigma^{p-1})$  (в непрерывной модели)

равен  $\frac{s^{p-1} \cdot \sqrt{p}}{(p-1)!}$ . Поэтому  $d(Y, X) = \frac{\pi^{\frac{p-1}{2}} \cdot \lambda^{p-1} \cdot (p-1)!}{\Gamma\left(\frac{p-1}{2} + 1\right) s^{p-1} \sqrt{p}}$ ;

$s = B - A; \lambda = |\vec{\Delta}|$ ;  $\Gamma(a)$  — классическая гамма-функция. Для дискретной модели (пользоваться которой необходимо при малых значениях  $B - A, p, \lambda$ ) требуется подсчитать целые точки для заданных значений  $B - A, p, \lambda$ ; здесь точная универсальная формула отсутствует, хотя можно вывести асимптотическую формулу (этот материал опущен ввиду недостатка места). Можно оценить: начиная с каких значений  $B - A, p, \lambda$  допустимо (не

совершая значительной ошибки) пользоваться непрерывной моделью; конкретные числовые оценки также опущены. Здесь возникает интересная математическая задача: найти точную границу, отделяющую область действия дискретной модели (при заданном допуске) от области применимости непрерывной модели.

12. Первое математическое уточнение  $F$ -модели. Поскольку в процессе уравнивания чисел  $Q(X)$  и  $Q(Y)$  мы вынуждены полагать некоторые из координат  $a_q$  равными нулю, и этот процесс неоднозначен (в качестве такой координаты можно выбрать любую от  $a_1$  до  $a_p$ ), то следует рассмотреть все возможные варианты и в качестве числа  $d(Y, X)$  взять наименьшее из получающихся (при разных вариантах аннулирования  $a_q$ ) чисел.

13. Второе математическое уточнение  $F$ -модели. Все предыдущие построения выполнялись для функций  $f(t, X)$ ,  $f(t, Y)$ . Наряду с ними следует рассмотреть и всевозможные их сглаживания, чтобы шире охватить характеристические свойства этих функций. Например, в качестве таких сглаживаний можно рассмотреть усреднение «по соседям»: по функции  $f(t)$  строится ее первое сглаживание  $S^1f(t) = \frac{1}{3}(f(t-1) + f(t) + f(t+1))$ . Затем второе сглаживание  $S^2f(t)$  определяется как  $S^1(S^1f(t))$  и т. д. Эта операция сглаживает график, убирая случайные локальные максимумы. После каждого сглаживания  $S^af(X)$ ,  $S^af(Y)$  вычисляется соответствующее  $d_a(Y, X)$ ; и в качестве окончательного  $d(Y, X)$  следует снова взять наименьшее из всех полученных таким образом чисел. Подсчитав  $d(Y, X)$  для исходных (еще не сглаженных) графиков  $f(t, X)$ ,  $f(t, Y)$ , мы получаем оценку сверху окончательного коэффициента  $d$ .

14. Третье математическое уточнение  $F$ -модели. Для большей статистической инвариантности картины следует сравнивать не просто пару текстов  $X$  и  $Y$ , а две группы текстов:  $X_1, X_2, \dots, X_\alpha$  и  $Y_1, Y_2, \dots, Y_\beta$ , где  $\alpha$  и  $\beta$  достаточно велики; сравнивать нужно функции  $f(t, \{X\}) = f(t; X_1, \dots, X_\alpha)$  и  $f(t, \{Y\}) = f(t; Y_1, \dots, Y_\beta)$ , где  $f(t; X_1, \dots, X_\alpha) = \frac{1}{\alpha} \sum_{s=1}^{\alpha} f(t; X_s)$ . Эти усредненные функции  $f(t; \{X\})$  и  $f(t; \{Y\})$  избавлены от случайных

локальных максимумов и их сравнение дает более уверенную статистическую картину.

15. Проверка  $F$ -модели на конкретном материале. Эта модель подтвердится, если для большинства пар текстов  $X, Y$ , описывающих один и тот же  $[A, B]$ , коэффициент  $d(Y, X)$  будет мал. Понятие малости коэффициента мы не будем уточнять формально, поскольку богатый статистический естественно-научный опыт позволяет оценивать (при гипотезе равномерного распределения) реально малые вероятности. Так как построение и описание информативных графиков для конкретных текстов больших объемов — громоздко, то мы вынуждены опустить таблицы и приведем только типичные примеры. Рассмотрим функцию  $f_1(t)$  (число страниц; см. п. 2).

В качестве  $X$  была взята, например, монография В. С. Сергеева «Очерки по истории древнего Рима», тт. 1, 2, Москва, ОГИЗ, 1938; в качестве  $Y$  — «Римская история» Т. Ливия, тт. 1—6, Москва, 1897—1899 гг. Здесь  $A = (-757)$  (знак «—» указывает гг. до н. э.),

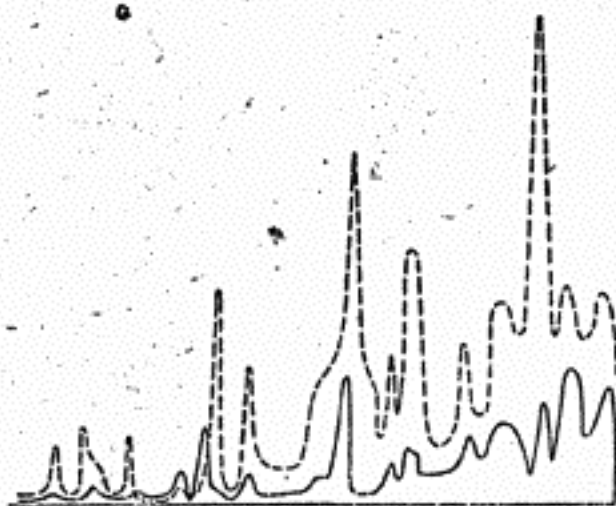


Рис. 2

$B = (-287)$ ,  $B - A = 470$ . Графики  $f_1(t, X)$ ,  $f_1(t, Y)$  см. на рис. 2 — черная линия для  $X$  и пунктир для  $Y$ . Оказывается, что  $\lambda = 21$ ,  $p = 14$  и коэффициент  $d(Y, X)$ , измеряющий близость векторов  $\vec{a}(Y)$  и  $\vec{a}(X)$  на расстоянии  $\lambda = 21$ , равен  $2 \cdot 10^{-12}$ ; другими словами, близость  $Y$  и  $X$

(на расстояние  $\lambda=21$ ) оценивается сверху числом  $2 \cdot 10^{-12}$ . Этот результат получен еще до минимизации по всем сглаживаниям (см. п. 13), т. е. указанное число — верхняя оценка окончательного коэффициента. Число  $2 \cdot 10^{-12}$  очень мало и указывает на значительную корреляцию между  $Y$  и  $X$ , с точки зрения функции  $f_1$ .

Аналогичные результаты получаются и при сравнении разных функций  $f_1(t, X)$  и  $f_1(t, Y)$  для текстов  $Y$  и  $X$ , описывающих один и тот же  $[A, B]$ . Например, для  $X$  и  $Y$  из предыдущего примера (т. е.  $X=[1]$ ,  $Y=[2]$ ) и для  $f_2(t, X) = \frac{s(t)}{S}$ , где  $s(t)$  = (число упоминаний года  $t$  в тексте  $X$ ) и  $f_1(t, Y)$ , имеем:  $\lambda=8$ ,  $B-A=227$ ,  $A=(-520)$ ,  $B=(-293)$ ,  $p=8$ . Результат:  $d(f_1(t, Y), f_2(t, X)) = 5 \cdot 10^{-6}$ . В этом примере было выполнено одно сглаживание  $S^1 f_2(t, X)$ ,  $S^1 f_1(t, Y)$ . Такое малое значение коэффициента, измеряющего близость векторов  $\vec{a}(f_1(t, Y))$  и  $\vec{a}(f_2(t, X))$  на расстояние  $\lambda=8$ , также указывает на наличие ярко выраженной корреляции между разными информативными функциями разных текстов, описывающих один и тот же  $[A, B]$  и одну и ту же эпоху (одно государство). Это обстоятельство указывает на наличие еще одного закона сохранения информации, который будет ниже сформулирован в виде  $K$ -модели.

Аналогичные результаты получаются и при обработке других пар исторических монографий  $Y, X$ . (При проведении вычислительного эксперимента мне помогал аспирант мех.-матем. ф-та МГУ М. Замалетдинов.)

16.  $K$ -модель. Сформулируем еще одну естественную модель ( $K$ -модель): если текст  $X$  описывает  $[A, B]$ , то имеется корреляция и между различными информативными функциями для  $X$ ; например, функции  $f_i(t, X)$  и  $f_j(t, X)$  для различных  $i, j$ , примерно в одних и тех же точках на отрезке  $[A, B]$  достигают локальных максимумов.

К этой  $K$ -модели применимы все математические уточнения, сформулированные для  $F$ -модели, поэтому здесь мы их опустим.

Интуитивное обоснование  $K$ -модели основано на том наблюдении, что если, скажем, какой-то год  $t \in [A, B]$  в  $X$  описан более подробно, чем соседние годы, то это

Должно отразиться не только в виде увеличения (локального) числа страниц, отведенных для описания года  $t$  (функция  $f_1(t)$ ), но и в увеличении числа упоминаний года  $t$  в  $X$  (функция  $f_2$ ; см. выше), в увеличении числа имен, упомянутых под годом  $t$  (функция  $f_3$ ) и т. д. Все эти утверждения верны, конечно, «в среднем», на больших интервалах времени.

Легко видеть, что  $K$ -модель — следствие все той же основной гипотезы, сформулированной в п. 6, о механизме потери и забывания информации, примененной к произвольной информативной функции  $f$  (а не только для  $f=f_1$ ). В этом смысле и  $F$ -модель и  $K$ -модель могут рассматриваться как следствия гипотезы п. 6.

17. Проверка  $K$ -модели на конкретном материале. Сформулированная  $K$ -модель подтвердится, если для текста  $X$  векторы  $\vec{a}(f_i; X)$  и  $\vec{a}(f_j; X)$  будут близки, т. е. число  $d(f_i(X), f_j(X))$  достаточно мало. Приведу типичный пример:  $X=[1]$ ; были построены функции  $f_1(t, X)$  и  $f_2(t, X)$ . Вычисления дают:  $\lambda=25$ ,  $p=11$ ,  $B-A=280$ ,  $B=(-520)$ ,  $A=(-240)$ ;  $d(f_1(t, X), f_2(t, X))=7 \cdot 10^{-5}$ , аналогичные результаты получены и для других учебников и монографий. Это подтверждает  $K$ -модель; получающиеся числа по порядку близки к числам, полученным ранее при проверке  $F$ -модели.

18. Не следует думать, что коэффициент  $d(f_i(X), f_j(Y))$  вообще всегда мал для произвольных текстов  $X$ ,  $Y$  и функций  $f_i, f_j$ . Если бы это было так, то чрезвычайная малость полученных выше чисел  $d$  ни о чем бы ни говорила. Напомним, что при гипотезе равномерного распределения все точки  $\sigma^{p-1}$  равноправны при случайном блуждании  $\vec{a}(Y)$  (и при фиксированном  $\vec{a}(X)$ ); поэтому, когда  $\vec{a}(Y)$  удаляется от  $\vec{a}(X)$ , шар  $D^{p-1}(X, \lambda)$  увеличивается и  $d$  приближается к 1. Далее, конкретные вычисления показали, что, действительно, для текстов  $X$  и  $Y$ , описывающих заведомо разные  $[A, B]$  и  $[C, D]$  (соответственно), коэффициент  $d$  сравним с 1. Типичный пример: пусть  $X$  и  $Y$  — первая и вторая половины текста [1], описывающие (соответственно) разные периоды:  $[A, B]=[-520, -380]$ ;  $[C, D]=[-380, -240]$ . Были подсчитаны функции  $f_2(t, X)$  и  $f_2(t, Y)$ ; оказа-

лось, что  $\lambda = 59$ ,  $B - A = D - C = 140$ ,  $p = 5$ . Результат;  $d = 1/3$ . Этот коэффициент сравним с 1 (в отличие от перечисленных примеров).

19. Вычислительные эксперименты подтверждают обе модели:  $F$  и  $K$ , указывая на существование специфических законов сохранения информации, проявляющихся в количественных характеристиках информативных функций  $f_i$ . Хотя интуитивное обоснование этих законов сохранения довольно ясно (см. выше), тем не менее, далеко не все параметры функций  $f_i$  — инварианты (например, амплитуда — не инвариантна) и обнаружение дальнейших инвариантов — нетривиальная задача, решение которой потребует новых статистических исследований.

20. Методика датирования событий, описанных в текстах. Предъявленные законы сохранения позволяют ввести новую формальную методику датирования событий, описанных в текстах, датировка которых (событий) по тем или иным причинам утрачена. В самом деле, пусть  $Y$  — текст с утраченной датировкой, параметризованный временем  $t \in [C, D]$ , при этом границы временного интервала  $[C, D]$  даны в  $Y$  в неизвестном для нас исчислении (например, годы  $t$  отсчитываются от основания какого-либо города). Построим для  $Y$  функции  $f_i(t, Y)$  и рассмотрим множество датированных текстов  $X$ , для каждого из которого также построим функции  $f_i(t, X)$ . Если нам удастся подобрать такой  $X$ , для которого какая-либо функция  $f_i(t, X)$  (или сразу целый набор таких функций) близка (в смысле малости коэффициента  $d$ ) к функции  $f_i(t, Y)$ , то это, в рамках проверенной выше  $F$ -модели, будет означать, что временной отрезок  $[C, D]$  близок (или совпадает) отрезку  $[A, B]$ , описанному в  $X$ . Более того, малость  $d$  будет указывать не только на близость отрезков  $[A, B]$  и  $[C, D]$ , но и на близость (или совпадение) описанных в  $X$  и  $Y$  событий (эпох), несмотря на то, что их описания могут быть внешне несхожи (например,  $X$  и  $Y$  являются двумя вариантами одной и той же хроники одного и того же государства, но написанными на разных языках, различными хронистами в разных странах).

21. Методика восстановления графика  $H(t)$ . Пусть  $X_1, \dots, X_k$  — совокупность текстов описывающих один и тот же  $[A, B]$ . Рассмотрим усредненную функцию

$$\bar{H}(t) = f_1(t; X_1, \dots, X_k) = \frac{1}{k} \sum_{j=1}^k f_1(t, X_j).$$

Если число  $k$  достаточно велико, то функция  $H(t)$ , в силу проведенных вычислительных экспериментов, может условно считаться совпадающей (близкой) с  $\bar{H}(t)$ . Локальные максимумы  $H(t)$  указывают те годы, от которых сохранилось особенно много информации (текстов). Это может помочь при датировании имеющихся в настоящее время в нашем распоряжении древних текстов: большие тексты должны концентрироваться в окрестностях локальных максимумов функции  $\bar{H}(t)$ . (Под датировкой текста мы понимаем датировку описанных в нем событий.)

22. В 1978—1979 гг. мною была проведена серия экспериментов, в которых вычислялись числа  $d(f_i(X), f_i(Y))$  для различных пар обзорных текстов  $X, Y$ , охватывающих значительные исторические периоды  $[A, B], [C, D], B-A = D-C$ . Были обнаружены пары эпох  $[A, B], [C, D]$  (и пары соответствующих текстов  $X, Y$ ), для которых коэффициент  $d$  оказался чрезвычайно малым. Приведу один пример.

Пусть  $Z$  — та часть текста [3], которая охватывает период от  $A = (+300)$  г. до  $B = (+754)$  г. (знаком «+» обозначены гг. н. э.), а  $Y$  — текст [2], охватывающий период от  $C = (+1$  г. от о. Г.) до  $D = (+459$  г. от о. Г.) (знаком «г. от о. Г.» обозначены годы от основания Рима:  $+1$  г. от о. Г. =  $(-753)$  г.). Результат:  $d = d(Z, Y) = 6 \cdot 10^{-10}$  (для функции  $f = f_1$ ). Напомним, что  $d(X, Y) = 2 \cdot 10^{-12}$  для  $X = [1], Y = [2]$ , поэтому следует ожидать, что и  $d(Z, X)$  (для функции  $f_1$ ) также будет мал. Вычисления полностью это подтверждают.

Интересно, что при расширении указанных временных интервалов (длиной в 460 единиц) коэффициент  $d$  (и без того малый) продолжает уменьшаться. Рассмотрим в качестве  $Z'$  ту часть текста [3], которая описывает отрезок от  $A = (+300)$  до  $B' = (+816)$ , а  $Y' = Y \cup X' >$  (объединение), где  $Y = [2]$ , а  $X'$  — та часть  $X$  (т. е. текста) [1], которая описывает отрезок от  $D = (+459$  г. от о. Г.) до  $D' = (+517$  г. от о. Г.). Мы были вынуждены расширить  $[C, D]$  до  $[C, D']$  за счет  $X'$  (части  $X$ ), так как текст  $Y$  прерывается в точке  $t = (+459$  г. от о. Г.) (здесь часть  $X$  утрачена). Продолжая  $Y$  фрагментом из

$X$ , мы используем тот факт (с точки зрения функции  $f_1$ ), что  $d(Y, X) = 2 \cdot 10^{-12}$ . Результат:  $d(f_1(Z'); f_1(Y')) = 6 \cdot 10^{-11}$ .

Интересно проследить за поведением  $d$  при укрупнении шкалы; например, когда  $t$  пробегает не годы, а 50-летия. Возьмем в качестве примера:  $Z''$  — та часть текста [3], которая описывает отрезок от  $A = (+300)$  до  $B'' = (+950)$ , а  $X''$  — та часть текста [1], которая описывает отрезок от  $C = (+1 \text{ г. от о. Г.})$  до  $D'' = (+650 \text{ г. от о. Г.})$ ;  $t$  пробегает 50-летия (т. е. принимает 13 значений). Результат:  $d(f_1(Z''), f_1(X'')) = 1/50$ . Аналогичные результаты (как для укрупненной шкалы, так и для исходной, когда  $t$  пробегает годы) верны и для  $f_2$ . Порядок  $d$  не меняется и при сглаживании этих функций  $f_1, f_2$ , что указывает на устойчивость полученных результатов относительно операций сглаживания (усреднения). Здесь видны преимущества «длинных шкал» (т. е. когда  $B - A$  имеет порядок  $10^2 - 10^3$ ): при наличии корреляции между  $f_1(X)$  и  $f_1(Y)$ , коэффициент  $d$  оказывается особенно малым. Укрупнение шкалы огрубляет картину; что отзывается на возрастании  $d$  (в приведенном примере — до  $1/50$ ).

23. Приведем еще один пример информативной функции. Пусть дана непрерывная (т. е. без пропусков) последовательность правителей  $R(t')$ , занумерованных целочисленным параметром  $t'$ ,  $1 \leq t' \leq N$  (династический поток). Положим  $f_5(t') =$  длительность правления правителя  $R(t')$ . Как и выше, определяется вектор  $\vec{a}$  локальных максимумов. Это позволяет сравнивать (в терминах  $f_5$  и  $d$ ) две последовательности правителей. Отличие от предыдущего состоит в том, что теперь  $d$  должен вычисляться более сложно (ввиду недостатка места детали опущены), так как здесь нельзя считать случайный вектор  $\vec{a}$  равномерно распределенным. Дело в том, что распределение длительностей правлений подчиняется нетривиальной закономерности. На рис. 3 изображен результат обработки числовой информации, содержащейся в таблицах [4]: по оси  $OS$  отложены длительности правлений правителей Европы и Средиземноморья (параметр  $S$  пробегает пары: 1—2, 3—4, 5—6, ..., 89—90), по оси  $OP$  отложены значения функции  $P(S) =$  (число правителей, длительности правлений



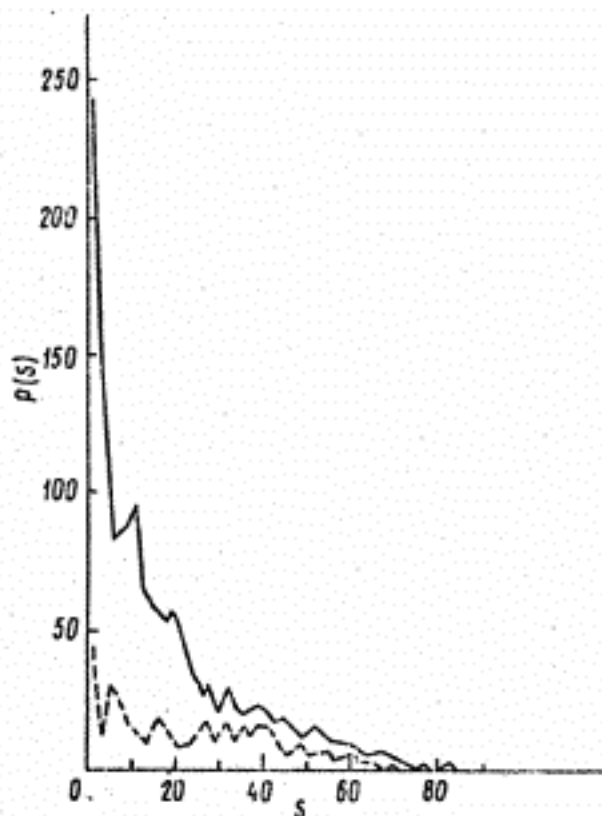


Рис. 3

которых заключены в интервале  $S$ ). Пунктирная кривая изображает распределение правлений на отрезке  $(+1400) - (+1800)$ ; сплошная кривая описывает правления, заключенные на отрезке  $(-3000) - (+1400)$ . При обработке полных списков правителей Европы и Средиземноморья (от  $-3000$  до  $+1800$ ) мне помогали М. Замалетдинов и П. Пучков. Коэффициент  $d$  должен подсчитываться с учетом полученного распределения. На это указывает также следующий эксперимент. Пусть  $\xi_i (1 < i < 1200)$  изображают последовательные длительности правлений (одновременные царства поставлены в этой строке последовательно, что и позволяет представить всю таблицу [4] в виде одной строки); рассмотрим вторую случайную величину:  $\eta_i(k) = \xi_i + k$ , т. е. строка  $\eta$  получается из строки  $\xi$  путем сдвига на  $k$  единиц. Пусть  $r(k)$  — коэффициент корреляции между  $\xi$  и  $\eta(k)$ . На рис. 4 показан график величины  $r(k)$  при  $1 \leq k \leq 300$ . (Вычисления проведены на ЭВМ П. Пучковым). Ввиду

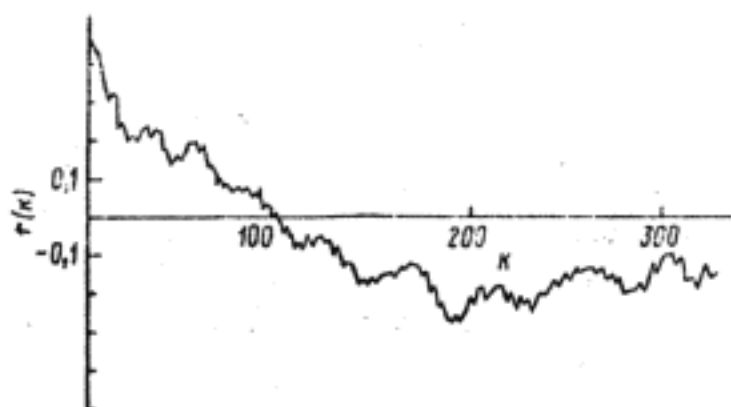


Рис. 4

недостатка места здесь невозможно привести ни результаты других экспериментов, ни соответствующие выводы; эти материалы будут изложены в другой статье.

24. Понятие формального изоморфизма. Пусть  $[A, B]$  (в истории общественной структуры  $M$ ) описан в текстах  $X = \{X_1, \dots, X_k\}$ , а  $[C, D]$  (в истории общественной структуры  $N$ ) — в текстах  $Y = \{Y_1, \dots, Y_p\}$ . Рассмотрим набор информативных функций  $\{f_i(X)\}, \{f_i(Y)\}$ . Будем говорить, что эпохи формально изоморфны, если коэффициенты  $d(f_i(X), f_i(Y))$  малы. Понятие формальной изоморфности эпох отнюдь не предполагает, что эпохи тождественны. Их изоморфизм может указывать на то, что некоторые документы из одной эпохи, вследствие тех или иных причин, могли оказаться приписанными другой эпохе. Кроме того, мы хотим решить глобальную задачу: описать действие законов сохранения информации на больших интервалах времени. Для этого расширим запас информативных функций (см. ниже).

25. О древних династиях правителей до нас дошла самая разнородная информация, причем разные источники, говорящие об одном и том же правителе, могут сильно различаться деталями описаний, освещением правления, характеристикой правителя и т. д. Но существуют «инвариантные» факты, описания которых в меньшей степени зависят от симпатий и политических установок хрониста. Одним из таких параметров является, например, длительность правления: обычно нет особых причин, по которым хронист стал бы значительно искажать это число (ввиду эмоциональной нейтральности

этого числа). Под «династией» будем понимать последовательность правителей, не предполагая обязательного родственного престолонаследия. Каждому правителю (или просто государственному деятелю) сопоставим его анкет-код, «биографию», понимая под этим совокупность всех дошедших до нас письменных свидетельств об этом человеке. Эта «биография» — набор дискретных фактов, закрепленных традицией за «правителем» в результате систематизации позднейшими историками дошедшего до них материала. Эта «написанная биография» может практически не иметь ничего общего с реальной биографией, о которой мы можем только гадать. Сопоставим каждой «биографии» таблицу, анкет-код, иерархически упорядочив факты «биографии» по мере уменьшения их инвариантности.

1) Пол (а. мужской; б. женский). 2)  $T$  = длительность жизни. 3)  $\Pi$  = длительность правления. Следует отметить, что конец правления практически всегда однозначно фиксирован (обычно это смерть); начало правления иногда (довольно редко) допускает несколько вариантов: дата официального коронования, дата присвоения титула «цезарь», «август», дата смерти более могущественного соправителя и т. д. Во всех таких случаях (при статистическом исследовании) отмечались все варианты, рассматривавшиеся как равноправные. Отметим, что только в редких случаях число вариантов доходило до трех. 4) Социальное положение (а. царь, император; б. полководец; в. политик, общественный деятель; г. ученый; д. религиозный вождь — папа, епископ и т. д.). 5) Смерть правителя (а. естественная смерть в мирной обстановке; б. убит на поле боя противниками или смертельно ранен; в. убит в результате заговора (вне войны); г. убит в результате заговора (во время войны); д. специальные, экзотические обстоятельства смерти). 6) Стихийные бедствия во время правления (а. голод; б. наводнения; в. поваральные болезни; г. землетрясения; д. извержения вулканов; отмечается также длительность бедствия и год (или годы), когда оно имело место). 7) Астрономические явления во время правления (а. есть; б. нет; в. затмения; г. кометы; д. вспышки «звезд»). 8) Войны во время правления (а. есть; б. нет). 9)  $V$  = число войн. 10) Основные временные характеристики войн  $V_1, V_2, \dots, V_p$  ( $a_k$  = на ка-

ком году правления происходит (началась) война  $B_k$ ;  $n_k$  = длительность войны  $B_k$ ;  $s_{kx}$  = временное расстояние от  $B_k$  до  $B_x$ . 11) Сила войны  $B_k$  (по хронике, для каждого « $k$ » (а. сильная; б. слабая; более точно: сколькими строками описана война в хронике). 12) Число противников в войне  $B_k$  и схема их взаимоотношений (союзники, противники, нейтральные силы). 13) Локализация войны  $B_k$  (а. около столицы; б. внутри государства; в. вне государства (внешняя война; где именно); г. одновременно внутренняя и внешняя война). 14) Результат войны (а. победа, б. поражение; в. неопределенный исход). 15) Мирные договоры (а. заключение мирного договора при неопределенном исходе; б. заключение мирного договора после победы (кого именно); в. заключение мирного договора после поражения). 16) О захвате столицы (а. захвачена; б. не захвачена). 17) Судьба мирного договора (а. нарушен (кем); б. не нарушен во время правления). 18) Обстоятельства захвата (падения) столицы. 19) Схема траекторий походов во время войны. 20) Участие правителя в войне (а. участвует; б. не участвует). 21) Заговоры при жизни правителя (а. есть; б. нет). 22) Географическая локализация войны. 23) Название столицы. 24) Название государства. 25) Географическая локализация столицы. 26) Географическая локализация государства. 27) Законодательная деятельность правителя (а. реформы и их характер; б. издание нового свода законов; в. реставрация старых законов и каких именно). 28) Список всех имен правителя (с их переводами). 29) Этническая принадлежность правителя (членов его семьи; состав семьи). 30) Этническая принадлежность народа, государства. 31) Основание новых городов, столиц. 32) Религиозная обстановка (а. введение новой религии; б. борьба сект (каких именно); в. религиозные восстания; г. соборы, религиозные собрания). 33) Династическая борьба внутри родственного клана правителя (убийства родственников, противников и т. д.). 34) Остальные фрагменты «биографии» не будем дифференцировать подробно; объем этого номера 34 (в %) назовем «остатком».

Обозначим эти пункты АК-1, АК-2, ..., АК-34; АК = {АК =  $i$ ;  $1 \leq i \leq 34$ }. Таким образом, каждая «биография» может быть изображена как некоторая табли-

ца-граф (см. выше); некоторые номера могут быть пустыми (если соответствующая информация не сохранилась).

26. Если параметр  $t'$  пробегает номера правителей (в династическом потоке), то анкет-коды  $AK(t')$  можно рассматривать как набор значений информативной функции. Поскольку набор  $AK(t')$  практически полностью описывает сохранившуюся информацию об эпохе, то можно считать, что последовательность  $AK(t')$ ,  $1 < t' < N$  описывает эпоху  $[A, B]$ , «покрытую» правителями, имеющими номера от 1 до  $N$ . Пусть  $AK(t')$  и  $\bar{AK}(t')$  — две информативные функции, построенные для двух эпох  $[A, B]$  и  $[C, D]$ . Таким образом, этим эпохам сопоставляются два набора:  $v = \{f_1, f_2, \dots, AK\}$  и  $\bar{v} = \{f_1, f_2, \dots, \bar{AK}\}$ . Так как иногда правители правили одновременно (соправители), то внутри последовательности  $AK(t')$  могут возникать различные подпоследовательности — династические струи, — каждая из которых выбирается как подпоследовательность без пропусков.

27. Методика сравнения векторов  $\bar{v}$  и  $v$ . Сравнение выполнялось по координатам. Для функций  $f_1, f_2, \dots$  подсчитывались коэффициенты  $d$  (см. выше). Сравнение  $AK$  и  $\bar{AK}$  более сложно ввиду более тонкой структуры этих функций. Не вникая в детали, укажем принцип сравнения. Пусть  $T_1, \dots, T_N$  и  $\bar{T}_1, \dots, \bar{T}_N$  — две последовательности длительностей правлений (см. п. 3 в АК). Расхождение этих двух последовательностей будем оценивать числом (в процентах):  $KPP = \min(\alpha, \beta)$ ;  $\alpha = \frac{1}{N} (\alpha_1 + \dots + \alpha_N)$ ;  $\alpha_i = 0$ , если  $|T_i - \bar{T}_i| \leq 2$ ; и  $\alpha_i = \frac{1}{T_i} (|T_i - \bar{T}_i| - 2)$ , если  $|T_i - \bar{T}_i| > 2$ ; для  $\beta$  следует поменять местами  $T$  и  $\bar{T}$ . Характеристики такого типа (данный вариант был предложен А. С. Мищенко) фиксируют «относительную ошибку». Введение разности  $|T - \bar{T}| - 2$  связано с тем, что правитель начинал и кончал правление не обязательно в начале календарного года. Как показал эксперимент, проведенный в 1975 г. мною совместно с А. С. Мищенко, при произвольном выборе пар династий, КРП (в подавляющем большинстве случаев) не меньше 15%. Было просчитано несколько десятков взятых наугад пар.

Сравнивать же АК-1, АК-2, (АК-4) — (АК-33) будем так. Введем серию числовых коэффициентов  $E-i$  (где  $i$  — номер АК- $i$ ). Для каждого АК- $i$  возможны три ситуации:

а) сравниваемые параметры близки (например, в АК-5 и  $\overline{\text{АК-5}}$  сказано, что оба правителя умерли естественной смертью; тогда положим  $(E-5)=1$ ; б) параметры явно не совпадают (например, в АК-5 сказано, что правитель умер естественной смертью, а в  $\overline{\text{АК-5}}$  — убит в результате заговора); тогда  $(E-5)=-1$ ; в) параметры нейтральны, не противоречат друг другу и не совпадая (например, в АК-5: «умер», а в  $\overline{\text{АК-5}}$ : «был убит»);

тогда  $(E-5)=0$ . Число  $E(t') = \sum_{\substack{i=1 \\ i \neq 3}}^{33} (E-i)(t')$ , где  $t'$  — номер

сравниваемой пары правителей, оценивает удаленность друг от друга двух анкет-кодов  $\overline{\text{АК}}(t')$  и  $\text{АК}(t')$ . Введем средний коэффициент:  $E_{\text{ср}} = \frac{1}{N} \sum_{t'=1}^N E(t')$ . Итак, удаленность друг от друга двух информативных функций

$\text{АК}(t')$  и  $\overline{\text{АК}}(t')$  можно оценивать парой: (КРП;  $E_{\text{ср}}$ )

28. В 1978—1979 гг. мною был проведен эксперимент, в процессе которого было обработано (с точки зрения вектора  $(f_1, f_2, \dots; \text{АК})$ ) несколько тысяч пар эпох  $[A, B]$ ,  $[C, D]$ . Временные границы эксперимента: (—3000) г. — (+1800) г.; географическая локализация: Европа, Средиземноморье, Египет. Эксперимент показал, что при сравнении двух анкет-кодов, во многих случаях значение  $(E-i)$  приходится полагать равным 0, так как сравниваемая информация не противоречит друг другу и не подтверждается; вследствие этого возрастает роль  $+1$  и  $-1$ . Далее, в подавляющем большинстве случаев,  $(E-34)=0$ . Следует, впрочем, отметить, что АК-34 является «остатком» и для уверенного сравнения следует знать, что мы действительно располагаем полным набором всех данных, слагающих данный АК. Гарантировать полноту АК-34 обычно затруднительно, поэтому был применен следующий формальный прием. Для каждой эпохи  $[A, B]$  подбирался (по возможности) единый текст, описывающий  $[A, B]$  (фундамен-

гальная монография); тогда АК-34 естественно отождествляется с «остатком биографии», приведенной в данном труде. Для простоты, подсчитывался объем АК-34 (в процентах) относительно всей «биографии», без детального разбора составных частей АК-34. Пример: эпоха «царского Рима» Т. Ливия и эпоха Римской Империи от (+300) до (+552) дают: КРП=10% (меньше 15%),  $E_{cp}=+19$ , объем (АК-34)=29% (т. е. «за бортом» сравнения осталось 29% «биографии», причем, этот «остаток» таков, что  $(E-34)=0$ ). Полезно сравнить этот результат с аналогичными параметрами для произвольно взятой пары эпох. Пример независимых эпох: великие русские князья от Игоря (912—944) до Дмитрия I (1275—1293) и византийские императоры от Феодосия II (407—449) до Феофила (828—841) [4]. Здесь КРП=33% (в 2 раза больше 15%),  $E_{cp}=-8,7$ ; объем (АК-34)=40%. Налицо резкое различие. Перейдем от примеров к описанию результата глобального эксперимента.

29. Мною была построена глобальная хронологическая карта древней и средневековой Европы и Средиземноморья на интервале  $(-2000) - (+1800)$ , т. е. вдоль горизонтальной оси времени были отмечены даты всех основных событий (правлений, войн и т. д.) в истории государств указанного региона. Для всех основных эпох  $[A, B]$  из указанного интервала были составлены (и распределены вдоль оси времени) анкет-коды АК( $t$ ). Эта карта была построена в 1978—1979 гг.; она отражает практически всю основную историко-хронологическую информацию, зафиксированную на сегодняшний день традиционной хронологией. Применение к этой карте описанной методики обнаружения формальных изоморфизмов дает (для Европы) результат, показанный на рис. 5. Одинаковые геометрические фигуры (с одинаковой штриховкой изображают эпохи (или подмножества в эпохах), оказавшиеся формально изоморфными. Границы эпох (или их подмножеств) отмечены приблизительно, так как точное описание карты очень громоздко (и по этой причине опущено).

Пояснение: для того чтобы получить всю информацию, канонически относимую к году  $t$ , следует (на рис. 5) провести через точку  $t$  на горизонтальной оси вертикальный отрезок и собрать все те события, кото-

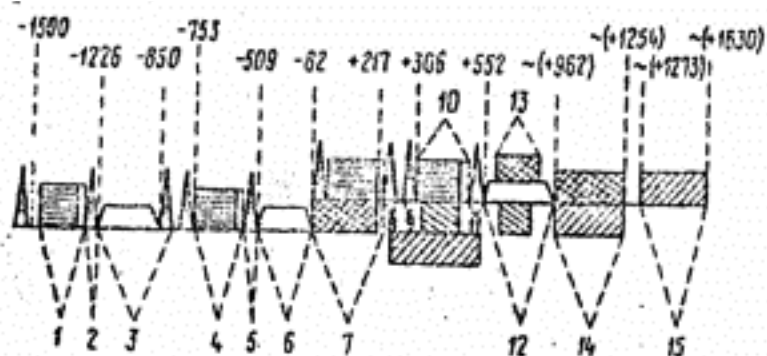


Рис. 5

рые высекаются этим отрезком (т. е. — одновременные события) на изображенных прямоугольниках, треугольниках и трапециях. Для наглядности, в тех местах, где одновременно имеется несколько таких фигур, они изображены одна над другой (развертка по вертикали), вместо того, чтобы накладывать их друг на друга.

На рис. 6 показано разложение этой карты в композицию нескольких сдвигов, при которых некоторая часть карты (условно назовем ее «хроникой X»), рассматриваемая как жесткое целое, последовательно сдвигается вниз (справа налево) на величины: 333, 1053 и 1526

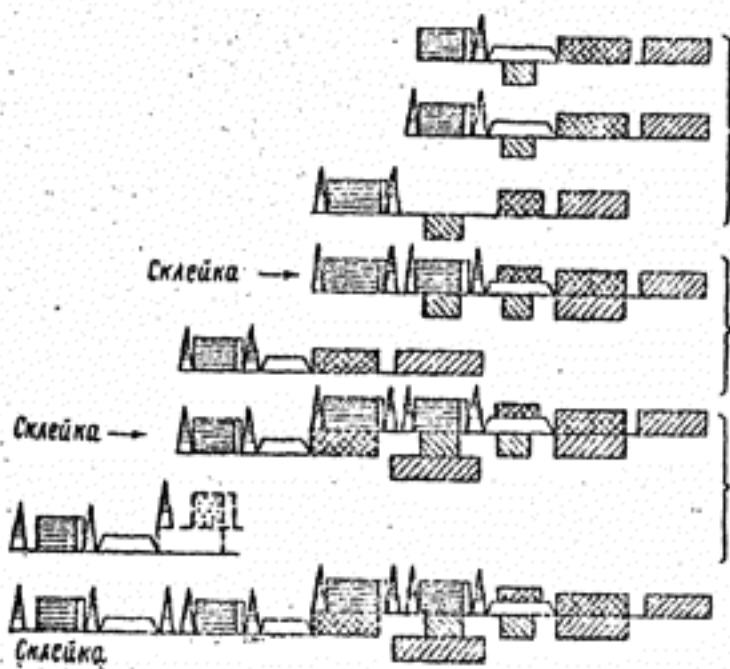


Рис. 6



единиц (лет), после чего эти смещенные экземпляры подклеиваются друг к другу, что и дает всю карту изоморфизмов.

30. Краткое описание карты (Греция, Рим, Германия) на рис. 5 (слева направо). (Подробное описание династических струй занимает большой объем и поэтому опущено).

1) Троянское царство (7 легендарных царей). 2) Троянская война. 3) Несколько династических струй (Греция) в интервале от падения Трои до основания Рима в (-753). 4) Основание Рима и «царский Рим» Т. Ливия. 5) Тарквинийская война, «изгнание царей» и основание республики. 6) Республиканский Рим и Греция этого периода. 7) Войны при падении республики в Риме в I в. до н. э. и Римская Империя (-83)—(+217). 8) Войны и кризис середины III в. н. э. (Рим); в частности, войны с готами, «солдатские императоры». 9) Восстановление Римской Империи при Аврелиане и война этого периода. 10) Римская Империя (+300)—(+536). 11) Готская война середины VI в. н. э. (536—553). 12) Средневековый Рим (553—середина X в. н. э.). 13) Каролинги (включая Империю Карла) (681—887). 14) Римская Империя германской нации X—XIII вв. 15) Империя Габсбургов 1273—1637. Кроме того, в 10)—13) присутствуют династические потоки Восточной Римской Империи (Византии). Отмеченные на рис. 5 изоморфизмы иногда сцепляют целые эпохи, а иногда — некоторые струи в них («подмножества в эпохах»).

31. Напомню, что для Рима периода (-753)—(-230) и Рима периода (+300)—(+820) коэффициент  $d$ , измеряющий близость векторов, построенных по их локальным максимумам, равен  $6 \cdot 10^{-11}$  (для функции  $f_1$ ). На интервале (+300)—(+553) этот формальный изоморфизм подтверждается близостью последовательностей анкет-кодов, сравниваемых (со сдвигом) по формуле  $\alpha + 300 = \beta$ , где  $\alpha$  — гг. от о. Г. (см. выше), а  $\beta$  — гг. н. э. Эта равномерная формула (сдвиг) была обнаружена мною в результате анализа 4), 5) и №№ 10), 11) (см. карту на рис. 5). Оказывается,  $E_{\text{ср}} = +18$  (см. выше) на отрезке (+1 г. от о. Г.)—(+250 г. от о. Г.). На следующем интервале: (+553)—(+820) вычисленный выше малый коэффициент  $d$  согласуется с результатами исследования, проведенного в 1978 г. Е. М. Цикишиным

для отрезка (553—850); им предъявлена целая серия далеко идущих параллелизмов, сцепляющих два периода. Следующий отрезок: от середины IX в. н. э. до XVII в. н. э. был исследован мною с помощью методики сравнения анкет-кодов АК (в 1978—1979 гг.). Результат см. на рис. 5.

32. Как было отмечено ранее, понятие текста со шкалой является более общим, чем разобранные примеры. Например, в качестве текста X можно взять совокупность всех произведений какого-либо одного автора, в качестве параметра — номера страниц (заномерованных подряд единой, сквозной нумерацией), в качестве информативной функции — какую-либо количественную характеристику текста, например, среднюю длину предложения; или частотность употребления союзов и т. д. Вопрос: существуют ли законы сохранения, регулирующие поведение таких информативных функций? Оказывается, ответ — положительный; этому материалу будет посвящена отдельная статья.

Следует особо отметить, что настоящая статья является лишь кратким обзором тезисов, подробное изложение каждого из которых требует значительного объема и развернутого аппарата, предназначенного для формального изложения обширного статистического материала.

#### ЛИТЕРАТУРА

1. Сергеев В. С. Очерки по истории древнего Рима. ТТ. 1, 2. — М.: ОГИЗ, 1938.
2. Ливий Т. Римская история от основания города. Перевод и редакция П. Адрианова. ТТ. 1—6. — М.: 1897—1899 гг. Типография Е. Гербек.
3. Грегоровиус Ф. История города Рима в средние века. Перевод М. П. Литвинова, С. — Петербург. Типография Альтиулера. Т. 1 (1902 г.), Т. 2 (1903 г.), Т. 3 (1904 г.), Т. 4 (1907 г.), Т. 5 (1912 г.).
4. Блер Ж. Хронологические таблицы, Москва, Т. 1, 2, — М.: Университетская типография, 1808.