

ОСОБЕННОСТИ СТАТИСТИКИ «СОЦИАЛЬНЫХ» ЧИСЛОВЫХ ДАННЫХ

Словохотов Ю. Л.

(Химический Факультет МГУ, ИНЭОС РАН)

Обобщенная социальная система – это совокупность автономных агентов (как индивидуальных, так и коллективных), взаимодействующих с внешней средой и с другими агентами в системе. Взаимодействие агентов, преследующих собственные цели, задает их сложную взаимозависимую динамику. Неустрашимыми особенностями реальных социальных систем являются неоднородность состава, сильные корреляции параметров, отражающих состояния агентов, и нестационарность случайных процессов. В отличие от измерений, производимых над «неживыми» физическими системами, последовательные «замеры» количественных параметров социальной системы (цены акций на бирже, спрос на товар, результаты голосования и т.д.) в общем случае нельзя считать независимыми, а хаотические изменения их значений – случайными. Простое перенесение методов обработки данных, развитых в экспериментальной физике (доверительные интервалы, моменты распределений, кластерный анализ и др.), на массивы таких численных характеристик может приводить к ошибочным выводам.

Характерным особенностям статистики данных в эконофизике – негауссовым распределениям с острым максимумом и обратной степенной асимптотикой («тяжелые хвосты») – соответствует форма полос в спектрах таких образцов, где состояния компонентов сильно коррелируют. Результаты физических измерений, собранные в базах данных, показывают сходные особенности ввиду влияния социальных факторов (финансирование перспективных направлений, конкуренция, обмен информацией) на научную деятельность. Выборки термодинамических и структурных параметров для вновь синтезированных химических веществ показывают, что «химия сама создает свой предмет» (М. Бергто) не в случайных блужданиях, а в целенаправленных исследованиях узких групп соединений, выбор которых

определяется сочетанием практических задач и случайных событий.

Одним из следствий сильных корреляций численных параметров для социальных систем является кластеризация точек («мокрый песок»), порождающая псевдо-полимодальные распределения. Статистическая гипотеза о наличии тонкой структуры в таком распределении может формально подтверждаться критериями, выработанными для случайных процессов. Эмпирическими приемами проверки являются деление выборки пополам (несовпадение ложных максимумов в «половинных» распределениях), варьирование шага гистограммы (фрактальные комбинации ложных максимумов) и зависимость относительной частотности мод от общего числа точек (сглаживание «мод» с увеличением выборки, или «отсроченное» выполнение ЦПТ). Перечисленные особенности иллюстрированы распределениями явки избирателей по территориальным избирательным комиссиям на выборах в ГД РФ 2007 г. и в сейм Польши 2005 г.

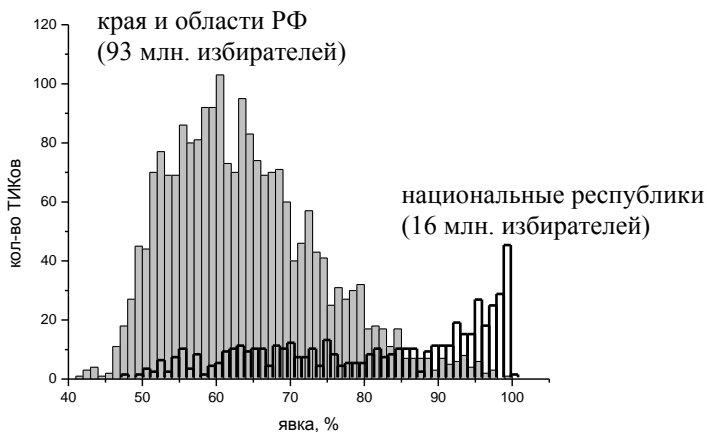


Рис. 1. Распределение ТИК (2745 комиссий) по явке избирателей на выборах в Государственную Думу РФ 2007 г.

(http://www.cikrf.ru/banners/elect_duma/protocol_data/index.html)