

# ПРИМЕНЕНИЕ ЗАКОНА БЕНФОРДА

О. Ю. Малюх

*Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Беларусь*

Научный руководитель М. В. Задорожнюк

Рассмотрим первые цифры степеней двойки от одного до бесконечности (1, 2, 4, 8, 16...), последовательности Фибоначчи, количества подписчиков в Твиттере, расстояний от Земли до звезд. На первый взгляд логично предположить, что вероятность встретить каждую цифру равна  $1/9$ .

Однако в 1881 г. Саймон Ньюкомб обратил внимание на то, что в книгах, содержащих логарифмические таблицы, гораздо сильнее истерты те страницы, которые содержат логарифмы чисел, начинающихся с единицы, а страницы с числами, начинающимися на 9 – практически новые. Заметив это несоответствие, он предположил, что разброс цифр на самом деле соответствует логарифмическому распределению: единица встречается примерно в 30 % случаев, 2 – в 18 % и так далее, до 9 – в 5 % случаев [1].

Через полвека позднее физик Френк Бенфорд проанализировал двадцать таблиц, которые содержали строгие математические данные, данные о физических постоянных, таких как удельная теплоемкость и молекулярный вес тысяч химических соединений. Опираясь более чем двадцатью тысячами чисел, он подтвердил открытый Ньюкомбом закон [2].

Однако этот закон оставался не более чем математическим курьезом до конца XX в., когда математик Марк Нигрини пришел к выводу, что подчиняться закону Бенфорда должны и цифры в налоговых декларациях. Соответственно, несоответствие с законом ука-

зывает на подтасовку данных. В ходе тестирования разработанной им программы для проверки массивов чисел на соответствие закону Бенфорда Нью-Йоркская налоговая полиция разоблачила семерых мошенничающих налогоплательщиков [3].

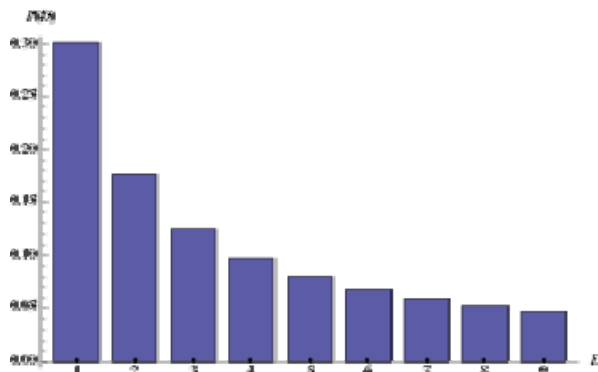


Рис. 1. Распределение Бенфорда

Эмпирическая закономерность, обнаруженная Бенфордом, выглядит следующим образом: вероятность  $d$  быть первой значащей цифрой соответствует логарифму по основанию  $b$  – системы счисления:

$$P(d) = \log_b \left( 1 + \frac{1}{d} \right). \quad (1)$$

Закон Бенфорда успешно применяется для выявления подлогов в бухгалтерской отчетности и фальсификаций на выборах; обсуждаются применения в различных областях – от сейсмологии до стеганографии и текстологии [4].

Для анализа и проверки на соответствие закону Бенфорда были взяты данные о выбросах загрязняющих веществ в атмосферный воздух от стационарных источников в тысячах тонн для 188 районов Беларуси за 15 лет [5]. Исходя из предположения об однородности данных (т. е. о слабых различиях в количестве предприятий по районам и объемах выпускаемой ими продукции), получим нижеприведенные результаты.

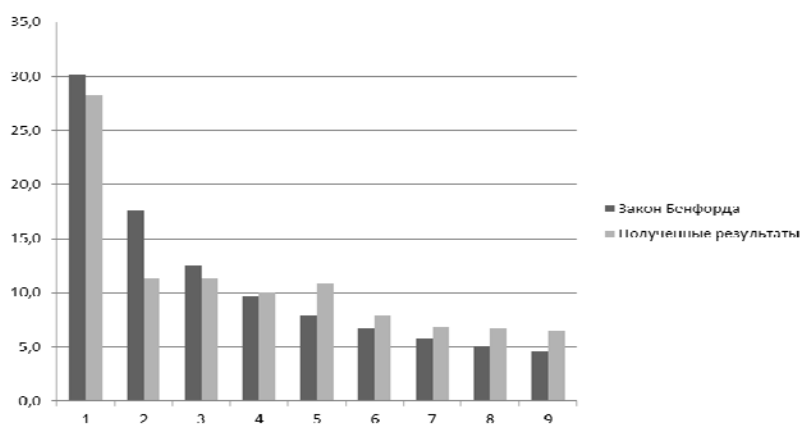


Рис. 2. Распределение первых значащих цифр в статистике выбросов

На гистограмме (рис. 2) наблюдаются отклонения от закона Бенфорда. Вероятно, их наличие объясняется некоторыми ограничениями, вводимыми для предприятий на объем выбросов. При этом отходы производств разделяются нормативной документацией по классам опасности. Так, экологический налог, согласно налоговому кодексу Республики Беларусь, не уплачивается, если фактические объемы загрязняющих веществ от стационарных источников не превышают для веществ: первого класса опасности (чрезвычайно опасные вещества) – 0,001 т в год; второго (вещества высоко опасные) – 0,1 т в год; третьего (умеренно опасные) и веществ, для которых не определены классы опасности, – 0,2 т в год; четвертого (вещества мало опасные) – 0,5 т в год.

Анализ гистограммы позволяет предполагать, что предприятия стремятся не превышать установленные нормы, и это выражается в том, что:

- поскольку 2, как первая значащая цифра, описывает норму выбросов для наиболее распространенного класса умеренно опасных веществ (например, диоксидов серы и азота), она встречается реже, чем предписывает закон Бенфорда;

- 9 и 8 встречаются чаще, поскольку в нормах присутствует 1 и предприятия сознательно снижают количество выбрасываемых веществ.

При этом трудно объяснить, по какой причине 5 встречается чаще, учитывая, что она также присутствует в норме, и доля веществ IV класса опасности (например, оксида углерода) в общем объеме выбросов значительна.

Также представляет интерес сравнения графиков общего объема выбросов и расхождения с законом Бенфорда по годам.

На рис. 3 при этом прослеживается значительный рост суммарных выбросов и одновременное снижение отклонения от закона Бенфорда при практически неизменном индексе промышленного производства.

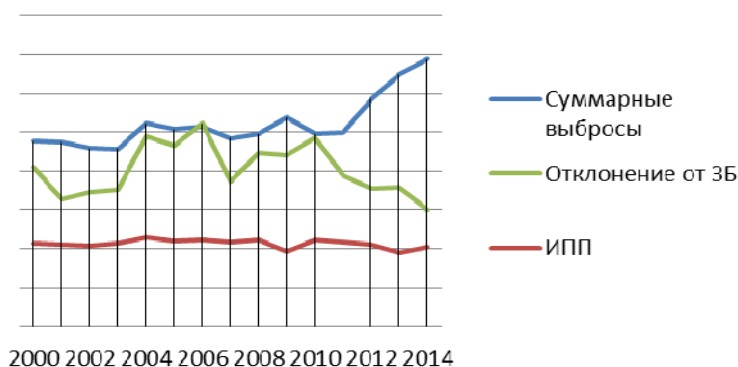


Рис. 3. Сравнение выбросов загрязняющих веществ с отклонением от закона Бенфорда

Индекс промышленного производства характеризует динамику объема промышленного производства, его подъема или спада, и определяется в виде отношения текущего объема производства в денежном выражении (в сопоставимых ценах) к объему промышленного производства в предыдущем или другом базисном году.

Проверка изменений в законодательстве показывает, что в 2009 г. происходил пересмотр методик расчета выбросов, что, однако, не объясняет сохранение тенденции их роста при практически неизменных объемах и составе промышленного производства.

В заключение отметим, что несоответствие распределений в отчетности и распределения по закону Бенфорда, а также отсутствие ожидаемых характерных отклонений от него может использоваться как основание для более детальной проверки представляемых результатов.

#### Л и т е р а т у р а

1. Adrien Jamain. Benford's law. Unpublished Dissertation Report, Department of Mathematics, Imperial College, London, 2001.
2. Кувакина, Л. В. Закон Бенфорда: сущность и применение / Л. В. Кувакина, А. Ф. Долгополова // *Соврем. наукоемкие технологии*. – 2013. – № 6. – С. 74–76.
3. Benford, F. The law of anomalous numbers / F. Benford // *Proceedings of American Philosophical Society*. – 1938. – Vol. 78, № 4. – P. 551–572.
4. Зенков, А. В. Закон Бенфорда и атрибуция текстов / А. В. Зенков, М. В. Казанцев // *Устойчивое развитие российских регионов : сб. материалов XII Междунар. науч.-практ. конф., Екатеринбург, 17–18 апр. 2015 г.* – Екатеринбург : УрФУ, 2015. – С. 785–792.
5. Официальная статистика Национального статистического комитета Республики Беларусь. – Режим доступа: <http://www.belstat.gov.by/ofitsialnaya-statistika>, свободный. – Загл. с экрана.