**Закон Ципфа и фрактальная природа социальных и экономических явлений**

Первый раз с описанием законом Ципфа я встретился, читая [«Оптимизацию и продвижение сайтов в поисковых системах»](http://baguzin.ru/wp/?p=921). Суть закона: если слова любого текста ранжировать по частоте использования, то произведение ранга на частоту есть величина постоянная:

**F\*R =C**, где:

F – частота появления слова в тексте;

R – ранг слова (наиболее часто употребляемое слово получает ранг 1, следующее – 2 и т.д.);

С – константа.

Для тех, кто еще хоть немного помнит алгебру ☺, в приведенной выше формуле легко узнает уравнение гиперболы. Ципф экспериментально определил, что С ≈ 0,1. Так, что графическое изображение закона Ципфа приблизительно следующее:



Рис. 1. Гипербола закона Ципфа.

У гипербол есть замечательно свойство. Если для обеих осей взять логарифмический масштаб, то гипербола будет иметь вид прямой:



Рис. 2. Та же гипербола, но на графике с логарифмическими шкалами

Может возникнуть вопрос: при чем здесь поисковая оптимизация? Так вот, оказывается, что специально сгенерированные тексты, содержащие повышенное число ключевых слов, не вписываются в закон. Поисковые машины (Google, Yandex) проверяют тексты на «естественность», то есть соблюдение закона Ципфа и, либо понижают рейтинг сайтов с «подозрительными» текстами, либо вообще банят такие сайты.

Второй раз я встретился с законом Ципфа у Бенуа Мандельброта в его книге [«(Не)послушные рынки: фрактальная революция в финансах»](http://baguzin.ru/wp/?p=1604). И этот небольшой раздел мне так понравился, что позвольте привести его полностью.

**Неожиданный степенной закон**

В 1950 году я был молодым студентом-математиком Парижского университета, подыскивавшим тему для своей диссертации. Мои дядя Золем являл собою местный хрестоматийный образец профессора математики: глубокий теоретик, очень консервативный и, несмотря на то, что родится в Польше, столп французского научного сообщества. Уже в 31-летнем возрасте его избрали профессором на полной ставке престижного Французского колледжа.

То быта эра Николя Бурбаки; за этим собирательным псевдонимом скрывался математический «клуб», который, подобно Дада в искусстве или экзистенциализму в литературе, распространился из Франции и стал на некоторое время чрезвычайно влиятельным на мировой сцене. Абстракция и чистая математика, математика ради математики, были возведены в ранг культа; члены «клуба» презирали прагматизм, прикладную математику и даже математику как инструмент науки. Такой подход был для французских математиков догмой, а для меня, пожалуй, причиной уехать из Франции и поступить на работу в IBM. Я был, к ужасу моего дяди, молодым бунтарем. Работая над своей докторской диссертацией, я часто в конце дня заходил к нему в кабинет поболтать, и нередко эти разговоры перерастали в дискуссию. Однажды, пытаясь как-то скрасить предстоящую долгую и скучную поездку на метро домой, я попросил у него в дорогу что-нибудь почитать. Он сунул руку в мусорную корзину и извлек оттуда несколько скомканных листков бумаги.

– Вот, возьми, – буркнул дядя. – Глупейшая статья, из тех, какие ты любишь.

То был обзор книги социолога Джорджа Кингсли Ципфа. Ципф, достаточно богатый человек, чтобы не думать о куске хлеба насущного, читал в Гарвардском университете лекции по им же придуманной дисциплине, которую он назвал статистической человеческой экологией. В его книге Human Behavior and the Principle of Least Effort (Поведение человека и принцип наименьших усилий) степенные законы рассматривались как вездесущие структуры общественных наук. В фишке степенные законы вполне обычны и выступают формой того, что я ныне называю фрактальным самоповторением в масштабе. У сейсмологов есть математическая формула степенной зависимости количества землетрясений от их силы по знаменитой шкале Рихтера. Или, другими словами: слабые землетрясения обычны, тогда как сильные редки, а частота и сила землетрясений связаны точной формулой. В то время было немногих таких примеров, да и известны они были всего нескольким людям. Ципф, энциклопедист, был одержим навязчивой идеей, будто степенные законы действуют не только в физических науках; им подчиняются все проявления поведения, организации и анатомии человека – даже размеры половых органов.

К счастью, обзор книги, который мне дал дядя, ограничивался только одним необычно изящным примером: частотой слов. В тексте или речи некоторые слова, такие как английские the (определенный артикль) или this («это»), встречаются часто; другие, milreis или momus, появляются редко или вообще никогда (для самых любознательных: первое означает древнюю португальскую монету, второе – синоним слова «критик»). Ципф предложил следующее упражнение: взять любой текст и посчитать, сколько раз в нем появляется каждое слово. Затем присвоить каждому слову ранг: 1 — для самых часто употребляемых слов, 2 — для занимающих второе место по частоте появления и т.д. Наконец, построить график, на котором для каждого ранга указать количество появлении этого слова. Мы получим удивительный рисунок. Кривая не убывает равномерно от самого обычного слова в данном тексте к самому редкому. Сначала она обрушивается с головокружительной быстротой, после чего начинает убывать медленнее, повторяя траекторию лыжника, прыгнувшего с трамплина, а затем приземлившегося и спускающегося по относительно пологому склону заснеженной горы. Образец классической неравномерной шкалы. Ципф, подогнав под свои диаграммы кривую, придумал для нее формулу.

Я был ошеломлен. К концу моей долгой поездки на метро я уже имел тему для половины моей докторской диссертации. Я точно знал, как объяснить математические основания частотного распределения слов, чего Ципф, не будучи математиком, сделать не смог бы. В последующие месяцы меня ждали удивительные открытия. Используя упомянутое уравнение, можно создать мощный инструмент социальных исследований. Улучшенный вариант формулы Ципфа позволял количественно оценить и ранжировать богатство словарного запаса любого человека: высокое значение – богатый лексикон; низкое значение – бедный. Имея такую шкалу, можно измерять различия по словарному запасу между текстами или говорящими. Появляется возможность количественно оценить эрудицию. Правда, мои друзья и консультанты были в ужасе от моей решимости заняться этой странной темой. Ципф, говорили они мне, человек с причудами. Мне показали его книгу, и я согласился, что она отвратительна. Подсчет слов – это не настоящая математика, убеждали меня. Занявшись этой темой, я никогда не найду хорошую работу; и профессором стать мне тоже будет нелегко.

Но я оставался глух к мудрым советам. Мало того, я написал диссертацию вообще без консультантов и даже уговорит одного из университетских бюрократов заверить ее печатью. Я был исполнен решимости пройти избранный путь до конца и применить идеи Ципфа в экономике, ведь не только речь можно свести к степенному закону. Богаты мы или бедны, процветаем или голодаем — все это тоже казалось мне объектом степенного закона.

Мандельброт немного модифицировал формулу Ципфа:

**F = C \* R -1/α**, где

α – коэффициент, характеризующий богатство словарного запаса; чем больше значение α, тем богаче словарный запас текста, поскольку кривая зависимости частоты появления каждого слова от его ранга убывает медленнее, и, например, редкие слова появляются чаще, чем при меньших значениях α. Именно это свойство Мандельброт предполагал использовать для оценки эрудиции[[1]](#footnote-1).

С законом Ципфа не всё так гладко, и в конкретных применениях опираться на экспериментально определенный коэффициент α не всегда получается. В то же время закон Ципфа является ни чем иным, как законом Парето «наоборот», поскольку и тот и другой – частные случаи степенных рядов, или… проявление фрактальной природы экономических и социальных систем[[2]](#footnote-2).

Для себя суть фрактальной природы экономических систем я сформулировал следующим образом. С одной стороны, есть игровая случайность: рулетка, бросание костей. С другой, технологическая/физическая случайность: разброс диаметра вала, изготавливаемого на токарном станке, разброс роста взрослого человека. Все перечисленные явления описываются [кривой Гаусса](http://baguzin.ru/wp/?p=1170). Так вот, есть целый ряд явлений не подчиняющихся этому распределению: богатство стран и отдельных людей, колебания цен на акции, курсы валют, частота использования слов, сила землетрясений… Для таких явлений характерным является то, что среднее значение очень сильно зависит от выборки. Например, если взять сто случайных людей разного роста, то добавление к ним самого высокого человека на Земле не сильно изменит средний рост этой группы. Если же посчитать средний доход ста случайных людей, то добавление самого богатого человека планеты – [Карлоса Слим Элу](http://damoney.ru/dengi/samyi-bogatyi-chelovek-2010.php)  (а не Билла Гейтса, как многие могли бы подумать ☺) значительно увеличит среднее богатство каждого, примерно, до 500 млн. долларов!

Другим проявлением фрактальности является значительное расслоение выборки. Рассмотрим, например, [распределение стран мира](http://gtmarket.ru/ratings/rating-countries-gdp/rating-countries-gdp-info) по объему ВВП (см. лист «ВВП» Excel-файла). В 2009-м году общемировой ВВП составил 57,7[[3]](#footnote-3) трлн. долл. И почти половина его (48%) пришлась на четыре богатейшие страны мира (рис. 3). При том, что всего в списке было 192 страны.



Рис 3. Богатейшие страны мира

Если ранжировать все страны в списке по объему ВВП, то получится график аналогичный виденному выше:



Рис. 4. ВВП стран мира

Согласитесь, представленная закономерность как две капли воды похожа на кривую Ципфа!

Одно из свойств фрактальности, это самоповторение. Так вот, из 192-х стран мира, перечисленных в списке, 80% мирового богатства сосредоточена всего в 18 странах – 9,4% (18/192). Если же теперь рассмотреть только эти 18 стран, то их суммарное богатство – 46 трлн. долл. – распределено столь же неравномерно. 80% от этих 46 трлн. Сосредоточено в менее чем половине стран, и т.д.

Вы можете спросить: какой практический вывод из всего сказанного? Я бы сказал так:

1. Социальные и экономические системы не описываются гауссианой. Эти закономерности подчиняются степенным рядам [синоним – фрактальная природа].
2. Выбросы от среднего существенно более вероятны, чем в соответствии с предсказаниями колоколообразной кривой Гаусса. Более того, выбросы внутренне присущи системы; они не случайны, а закономерны.
3. Оценки рисков нельзя строить на основе нормального распределения вероятностей редких нежелательных событий.
4. … не буду лукавить, пока больше ничего придумать не могу… но это не значит, что практических выводов больше нет… просто мои знания этим ограничиваются…

… но согласитесь, ведь красивые закономерности!

1. Мне не удалось найти в Интернете ссылки на **такое** использование закона Ципфа. [↑](#footnote-ref-1)
2. О фрактальности см. Бенуа Мандельброт [«(Не)послушные рынки: фрактальная революция в финансах».](http://baguzin.ru/wp/?p=1604) [↑](#footnote-ref-2)
3. Надо отметить, что данные из [разных](http://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D1%81%D1%82%D1%80%D0%B0%D0%BD_%D0%BF%D0%BE_%D0%92%D0%92%D0%9F_%28%D0%9F%D0%9F%D0%A1%29) источников сильно разнятся, но это не имеет отношения к рассматриваемой здесь теме. [↑](#footnote-ref-3)