

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ ЗАКОНЦА ЦИПФА ДЛЯ ПРОГНОЗИРОВАНИЯ НАПРАВЛЕНИЙ ВНУТРЕННЕЙ МИГРАЦИИ

Поздняк Ярослав Любимович ¹

¹Московский физико-технический институт, Долгопрудный, Россия

Мы исследовали связь между динамикой изменения численности населения городов и их положением относительно прямой Ципфа, что бы связать темпы внутренней миграции в этот или из этого города с его положением относительно прямой Ципфа, а так же уточнили алгоритм построения искомой прямой. В результате нашего исследования получилось, что нет корреляции между динамикой роста населения города и его положением относительно прямой, а так же существует слабая корреляция между динамикой изменения положения города относительно прямой Ципфа и его положением относительно прямой Ципфа. Оценить темпы внутренней миграции не получилось, так как не получилось построить простую модель для оценки темпов естественного прироста и не получилось найти какую-либо статистику естественного прироста в большинстве городов. Но используя разные оценки, мы получили предсказательную точность 55%, и скорее всего при работе с непосредственными статистическими данными мы бы получили большую предсказательную способность данного закона.

Вступление

Введем понятие ранг города. Отсортируем все города какой-то страны в какой-то момент времени по населению. Тогда рангом города будем называть его номер в отсортированном списке. Согласно закону Ципфа распределение городов по численности населения описывается распределением Парето и имеет место следующие соотношение:

$$P = P_0 \cdot R^{-\alpha} \quad (1)$$

где P - население города, R - его ранг, P_0 и α - константы.

Мы выдвинули несколько гипотез:

1. Существует корреляция между динамикой увеличения или уменьшения численности населения города и его положением относительно прямой.
2. Существует корреляция между динамикой приближения города к прямой и его положением относительно прямой.

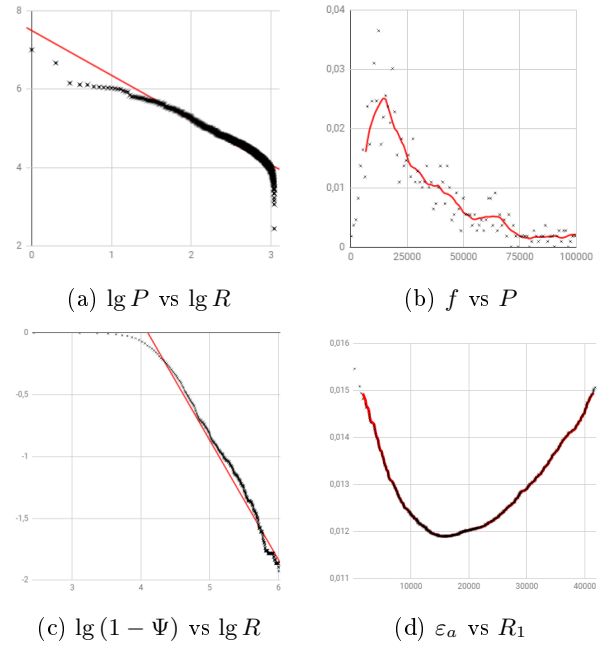


Рис. 1: Россия, 2002 г.

Методика

Прологарифмируем выражение 1.

$$\lg P = \lg P_0 - \alpha \cdot \lg R.$$

Таким образом зависимость $\lg P$ vs $\lg R$ должна быть линейной. На практике же, например, график 1a нелинейный. Это объясняется тем, что реальное распределение городов по численности отличается от распределения Парето. Таким образом для корректного анализа динамики параметров распределения нужно выделить часть данных,

которая описывается распределением Парето. Построив 1b график плотности вероятности $\psi(P)$ распределения городов по численности мы увидели, что есть существенные отклонения для городов с численностью населения меньше какого-то значения P^* .

Рассмотрим все города с численностью населения больше, чем P_1 и посмотрим, насколько хорошо аппроксимируется распределение таких городов по численности распределением Парето. Для этого исследуем функцию распределения $\Phi(P_1)$,

которая описывается следующей формулой:

$$\Phi(x) = 1 - \left(\frac{x_m}{x}\right)^k \quad (2)$$

В нашем случае $x_m = P_1$, $x = P$. Для оценки отклонения от распределения Парето ланеризуем зависимость 2:

$$-k \cdot \lg \frac{P}{P_1} = \lg(1 - \Phi(P)),$$

Обозначим $\lg(1 - \Phi(P))$ за y , а $\lg P$ за x . Тогда зависимость $y(x)$ изображенная на графике 1с должна быть линейной, введем коэффициенты a и b такие, что $y = ax + b$.

Найдем минимум функции $\varepsilon_a = \frac{\sigma_a}{a}$, он достигается при $P_1 = P^*$. На графике 1d изображена эта функция.

Зная P^* можно получить коэффициенты α и $\ln P$ из метода наименьших квадратов.

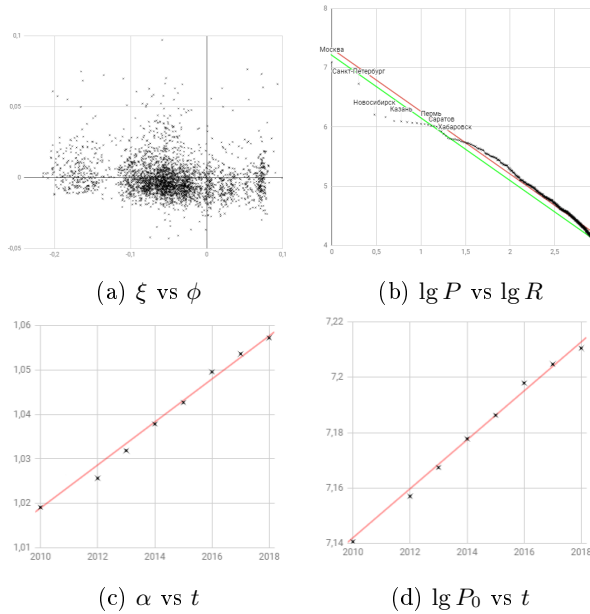


Рис. 2: 2b - $\lg P$ vs $\lg R$ в России на начало 2018 г., зеленая прямая нормированна на общую численность, 2c и 2d - динамика параметров α и $\lg P_0$ во времени.

Для оценки положения относительно прямой введем величину

$$\phi = \frac{P_{Zipf}(t)}{P(t)} - 1,$$

где, t - время в годах, $P(t)$ - население города, $P_{Zipf}(t)$ вычисляем согласно формуле 1, а для оценки динамики роста населения города введем величину

$$\xi(t) = \frac{P(t)}{P(t-1)} - 1.$$

Таким образом положительным ϕ соответствует «предсказание» роста населения города, а отрицательным - уменьшение. Соответственно положительным ξ соответствует рост населения города, а отрицательным - уменьшение.

Построив график 2a мы видим слабую корреляцию между этими величинами, т.е. гипотеза 1 не подтвердилась в первом приближении.

У получившейся теории стоит выделить несколько проблем:

1. Прямая не нормирована на суммарную численность населения рассматриваемых городов, таким образом рост или уменьшение численности предсказывается чаще.
2. При прогнозировании мы не учитываем то, как прямая меняется со временем, игнорируя тем самым естественный прирост населения.

Для решения проблемы 1 введем поправку δP_0 , такую, что

$$\sum_{i=1}^N (P_0 + \delta P_0) R_i^{-\alpha} = \sum_{i=1}^N P_i,$$

где N - количество городов с населением $P > P^*$, R_i и P_i - соответственно ранг и население i -го города. Тогда δP_0 найдем из следующей формулы:

$$\delta P_0 = \frac{\sum_{i=1}^N P_i}{\sum_{i=1}^N R_i^{-\alpha}} - P_0.$$

На графике 2b красной линией показана прямая без учета поправки δP_0 , зеленой линией - с учетом поправки δP_0 .

Далее под P_0 будет подразумеваться $P_0 + \delta P_0$.

Для решения проблемы 2 нужно оценить естественный прирост населения. Его можно оценить исходя из динамики коэффициентов α и P_0 во времени по формуле

$$\theta = \frac{\Delta P_n}{\Delta t} \approx (P_0 \cdot R_i^{-\alpha})', \quad (3)$$

тогда оценка динамики роста населения города ζ будет выражаться формулой

$$\zeta = \xi - \frac{\theta}{P_i} \cdot 1 \text{ year}.$$

Но между величинами ζ и ϕ (график 3a) тоже наблюдается слабая, даже отрицательная корреляция.

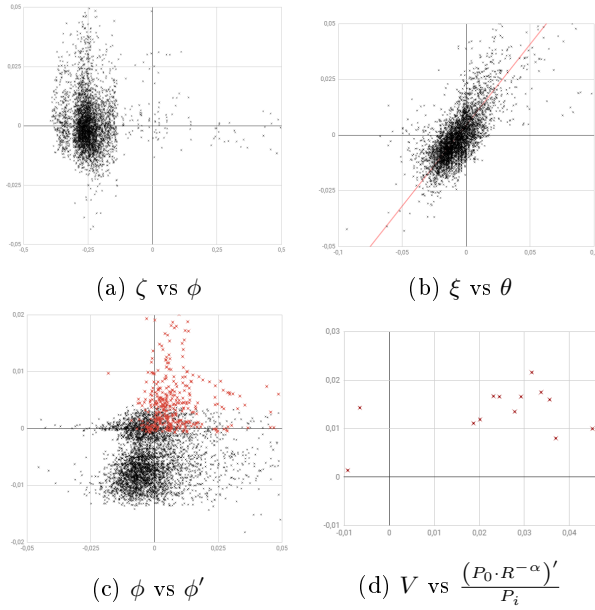


Рис. 3: **3с** - красным выделены точки для первых 100 городов.

Обсуждение

Отрицательная корреляция между величинами ξ и ϕ , а так же между величинами ζ и ϕ говорит о том, что гипотеза **1** не подтвердилась. Так же остался открытым вопрос об оценке естественного прироста, сложно выделять какие-то зависимости между реальными данными и какой-либо моделью, так как не было найдено данных о естественном приросте в городах за исключением Москвы и Санкт-Петербурга. Но по найденным данным (график **3д**) было получено, что естественный прирост V можно аппроксимировать формулой:

$$V \approx 0.52 \frac{(P_0 \cdot R^{-\alpha})'}{P_i} \cdot 1 \text{ year},$$

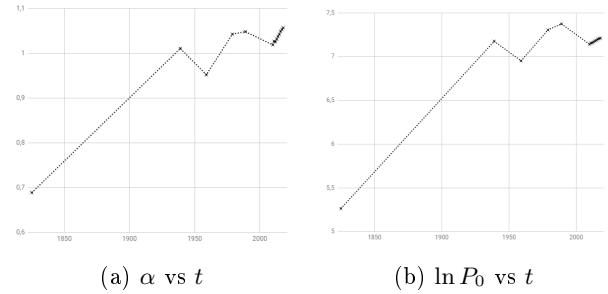
причем числитель этой дроби отличается от θ , определяемого в выражении **3**, тем, что R используется, как константа, а не как функция времени. Но точность такого приближения сомнительная, так как проверена только на 2-ух городах из 1000 рассматриваемых, но для остальных, как уже было сказано выше, нет статистических данных.

Но мы нашли статистику естественного прироста по субъектам РФ, поэтому для оценки можно считать, что естественный прирост населения города совпадает с естественным приростом населения субъекта РФ, в котором находится этот город. Да, это сомнительное, но тем не менее, наверное, самое точное приближение. Введем характе-

ристику города E_n , равную естественному приросту в субъекте РФ, в котором находится город, тогда изменение численности населения города, связанное с миграцией будет равно $\xi - E_n$. С учетом этой поправки мы получили, что в 55% случаев миграционный приток населения происходит, если город находится под прямой Ципфа и миграционный отток, если город находится над прямой Ципфа. Таким образом у нашей модели достаточно плохая предсказательная сила, но ее можно было бы усовершенствовать, если бы было больше статистических данных.

После исключения фактора естественного прироста изменение численности населения конкретного города будет обусловлено не только внутренней миграцией, но и внешней, но внешняя миграция на порядок меньше внутренней [10], поэтому ей можно будет пренебречь или ввести поправку - в зависимости от региона.

Между ϕ и ϕ' была найдена слабая корреляция (график **3с**), таким образом гипотеза **2** подтвердилась только частично.



Так же стоит заметить, что присоединение Крыма в 2014 году сильно не повлияло на зависимость α и $\lg P_0$ от времени, несмотря на то, что население Крыма на момент присоединения составляло примерно 3% от населения РФ [11]. Чтобы опровергнуть возможную гипотезу о слабой связи между историческими событиями и динамикой коэффициентов прямой Ципфа рассмотрим динамику коэффициентов α и $\ln P_0$ на более длинном историческом промежутке. Мы видим существенные уменьшения коэффициентов после больших уменьшений численности населения страны: после Великой Отечественной войны и после распада СССР. Рассмотрим отдельно каждое из событий.

Падение коэффициента α после ВОВ говорит о том, что уменьшение населения произошло непропорционально, т.е. из больших городов произошло относительно большее уменьшение населения. Падение коэффициента $\lg P_0$ объясняется уменьшением населения многих городов.

Падение коэффициентов α и $\lg P_0$ после распада

СССР говорит о том, что в СССР доля крупных городов союзных республик была больше, чем доля малых городов.

Была найдена корреляция между величинами ξ и θ (график 3b), но величина θ , скорее искусственная, так как учитывает изменение ранга города, хотя в идеальном случае (когда города описываются распределением Парето) ранг города не должен меняться.

В заключение, авторы хотят сказать, что сомневаются по поводу правильности использования поправки δP_0 , так как в таком случае невозможно аналитически выразить «хвост» распределения из городов с населением меньшим P^* , хотя в работе [20] получены зависимости для обеих частей распределения. Таким образом проблема ненормированности прямой, возможно, была неактуальной.

Благодарности

Автор хочет выразить благодарность Исаевой Анне и Александру Брезгину за помощь в оцифровке документов.

Список литературы

- [1] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2018 года
- [2] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2017 года
- [3] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2016 года
- [4] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2015 года
- [5] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2014 года
- [6] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2013 года
- [7] Численность населения Российской Федерации по муниципальным образованиям на 1 января 2012 года
- [8] Всероссийская перепись населения 2010 г. Численность населения России, федеральных округов, субъектов Российской Федерации, районов, городских населенных пунктов, сельских населенных пунктов — районных центров и сельских населенных пунктов с населением 3 тысячи человек и более.
- [9] Всероссийская перепись населения 2002 г. Том 1. 4. Численность населения России, федеральных округов, субъектов Российской Федерации, районов, городских поселений, сельских населенных пунктов - райцентров и сельских населенных пунктов с населением 3 тысячи и более человек.xls
- [10] Росстат. Демография
- [11] Крымстат: Численность населения на 1 января 2014 года
- [12] Статистическое изображение городов и посадов Российской империи по 1825 году
- [13] Всесоюзная перепись населения 1939 г. Численность городского населения СССР по городским поселениям и внутригородским районам
- [14] Всесоюзная перепись населения 1959 г. Численность городского населения РСФСР, ее территориальных единиц, городских поселений и городских районов по полу
- [15] Численность городского населения союзных республик (кроме РСФСР), их территориальных единиц, городских поселений и городских районов по полу
- [16] Всесоюзная перепись населения 1970 г. Численность городского населения РСФСР, ее территориальных единиц, городских поселений и городских районов по полу.
- [17] Всесоюзная перепись населения 1970 г. Численность городского населения союзных республик (кроме РСФСР), их территориальных единиц, городских поселений и городских районов по полу
- [18] Всесоюзная перепись населения 1989 г. Численность городского населения РСФСР, ее территориальных единиц, городских поселений и городских районов по полу
- [19] Всесоюзная перепись населения 1989 г. Численность городского населения союзных республик, их территориальных единиц, городских поселений и городских районов по полу

- [20] Pareto tails and lognormal body of US cities size distribution. *Physica A: Statistical Mechanics and its Applications*. Volume 465, 1 January 2017, Pages 573-578.
doi.org/10.1016/j.physa.2016.08.061