УДК 518.872, 51921

А.В. Талейко

ЗАКОНОМЕРНОСТИ ФОРМИРОВАНИЯ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ ПО ТЕРРИТОРИЯМ

В работе рассматривается квантильная модель накопления для описания количественных закономерностей, присущих численностям населения территорий. Приводятся примеры применения этой модели при исследовании закономерностей численностей населения по территориям США и Российской Федерации. Результаты исследований во многом подтверждают целесообразность применения логнормального закона при описании распределения численности населения.

Ключевые слова: *демография*, *статистика*, *анализ данных.*

В рамках квантильной модели величины численности населения территорий формируются путём последовательного накопления под влиянием разнообразных факторов, которые и определяют их величину. Предполагается, что рассматриваемая вероятностно-статистическая модель формирования величины численности населения трактует итоговые их значения для каждой территории (регионы, области, районы, штаты) в отчётное время как величины, соответствующие квантилям x_k, $k=\overline{1,N}$, равноотстоящих уровней p_k некоторой заданной функции распределения $F(x), x \in R^1$ [1]. При этом

$$F(x_k)=p_k\,,\ k=\overline{1,N}\,,$$

следовательно

$$x_k = F^{-1}(p_k)\,,\ k=\overline{1,N}\,.$$

Для равноотстоящих уровней выполняется условие

$$p_{k+1}-p_k=\frac{1}{N}\,,\ k=\overline{1,N}\,.$$

Каждой территории соответствует единственная квантиль x_k уровня p_k, N – количество выделенных территорий.

1. Логнормальная модель

Введём следующие обозначения: q_k^j – численность населения k-й территории, сформированное в течение j-го периода, Q_k^j – величина населения k-й территории, сформированная к некоторому моменту или периоду j, то есть

$$Q_k^j = q_k^1+q_k^2+\ldots+q_k^j\,,\ k=\overline{1,N}\,. \tag{1}$$

При этом выполняются следующие неравенства:

$$q_1^j > q_2^j > \ldots > q_N^j\,,\ k=\overline{1,N}\,. \tag{2}$$

$$Q_k^1 < Q_k^2 < \ldots < Q_k^L\,,\ k=\overline{1,N}\,. \tag{3}$$

Рассмотрим модель накопления растущей системы вида

$$Q_k^{j+1} = Q_k^j + z_k^j g_k\left(Q_k^j\right),\ k = \overline{1,N}\ , \tag{4}$$

где z_k^j – независимые по j, неотрицательные, одинаково распределённые случайные величины, $g_k(x), k = \overline{1,N},$ – некоторые функции, которые будут выбираться согласно определённому механизму накопления. Полагая, например, $g_k(x) = \alpha x,$ $\alpha > 0,\ x > 0,$ мы переходим к механизму накопления, характерному для закона роста кристаллов. Другими словами, в данном случае предполагается, что численность населения определённой территории в каждом периоде пропорционально зависит от численности населения в предшествующем периоде.

Заметим, что при $x > 0$ для сохранения приоритетов (2), (3), должны выполняться условия

$$g_1(x) > g_2(x) > \ldots > g_N(x) \quad \text{и} \quad g_k(x) > 0\ ,\ \ k = \overline{1,N}.$$

При одинаковом механизме накопления в различных территориях функции $g_k(x)$ от номера k не зависят, поэтому можно положить в данном случае $g_k(x) = g(x)$, где $g(x)$ – монотонная возрастающая функция. При этом полагаем, что отношение приоритетности выполняется в среднем.

Данная модель, как было показано в работе [2], тесно связана с законом Кептейна, имеющего плотность распределения

$$K\left(\alpha, \sigma^2, G, x\right) = \frac{1}{\sigma\sqrt{2\pi}P(\varsigma \in VG)} \exp\left(-\frac{(G(x)-a)^2}{2\sigma^2}\right)\left|\frac{dG(x)}{dx}\right| I_{DG}(x)\ , \tag{5}$$

$$G(x) = \int_{x_0}^{x} \frac{du}{g(u)}\ . \tag{6}$$

Заметим, что если случайная величина x распределена по закону Кептейна с параметрами (α, σ^2), то $G(x)$ распределена по нормальному закону с теми же параметрами (α, σ^2). В частном случае, при подстановке $g(x) = \alpha x,\ \alpha > 0,\ x > 0,$ в формулу (6)

$$G(x) = \frac{1}{\alpha}\ln\left(\frac{x}{x_0}\right). \tag{7}$$

Подставляя выражение для $G(x)$ в формулу (5), получим плотность распределения логнормального закона [2].

Значит, в рамках исследуемой модели накопления (1) можно предположить, что величины сформированных численностей населения различных территорий распределены приблизительно логнормально. Как известно, значения логнормально распределённой случайной величины формируется под воздействием большого числа независимых факторов, причём воздействие каждого фактора мультипликативно по своему характеру [3]. Выберем класс сдвиг-масштабного семейства, т.е.

$$F(Q_k, \mu, \sigma) = F_0\left(\frac{Q_k - \mu}{\sigma}\right),$$

где μ – параметр сдвига, σ – параметр масштаба. Пусть Q_k, $k=\overline{1,N}$, – численность населения территории под номером *k*. Предположим, что величины Q_k имеют логнормальное распределение, при этом случайная величина $G(Q_k)=\ln(Q_k)$ имеет нормальную функцию распределения $\Phi(x)$ с некоторым средним μ и стандартным отклонением σ. Обозначим $Q_{(k)}$, $k=\overline{1,N}$, упорядоченные по убыванию численности населения, т.е. $Q_{(N)}<Q_{(N-1)}<...<Q_{(1)}$. Тогда, опираясь на выводы, сделанные в работе [4], можно записать, что

$$\Phi\left(\frac{\ln Q_{(k)}-\mu}{\sigma}\right)=\frac{k}{N+1}+\varepsilon_k\,; \tag{8}$$

$$\ln Q_{(k)}=\mu+\Phi^{-1}\left(\frac{k}{N+1}\right)+e_k=\mu+\sigma x_k+e_k\,, \tag{9}$$

где x_k – квантили нормального распределения с нулевым математическим ожиданием и единичной дисперсией, ε_k – ошибки наблюдений с нулевым средним и конечной дисперсией.

2. Адекватность модели

Проверим, применима ли данная модель при исследовании распределения численности населения. Для этого построим для конкретных территорий диаграмму рассеяния и оценим сумму квадратов невязок. Отложим на оси абсцисс квантили, а на оси ординат численности населения территорий. В результате получим диаграмму рассеяния *N* точек, которые мы аппроксимируем с помощью метода наименьших квадратов прямой $P(t)=\mu+\sigma t$. «Близость» расположения точек к прямой линии характеризует сумма квадратов невязок. Таким образом, на графике будет представлена картина того, как прямая, заданная двумя параметрами, наилучшим образом (в смысле суммы квадратов невязок) аппроксимирует наблюдаемые значения. Очевидно, что чем ближе точки будут располагаться возле прямой, тем ближе распределение величин численности населения к логнормальному закону распределения.

Рассмотрим распределение численности населения Российской Федерации по территориям. Численность населения каждого из 85 субъектов будет соответствовать определённой ячейке накопления. В соответствии с моделью накопления упорядочим численности населения за 2008 г. и построим диаграмму рассеяния [3]. На оси абсцисс отложим 85 известных значений квантилей стандартного нормального закона. Ось ординат будет соответствовать данным о логарифмах численности населения территорий.

На рис. 1, в целом, отклонение наблюдаемых значений от теоретических невелико. Однако нижние 3 точки выбиваются из общей картины. Эти точки соответствуют следующим субъектам: Усть-Ордынскому Бурятскому АО, Чукотскому АО и Агинскому Бурятскому АО. Если включить перечисленные автономные округа в более крупные субъекты РФ, примыкающие к ним, то получим значительно меньшее отклонение (рис. 2).

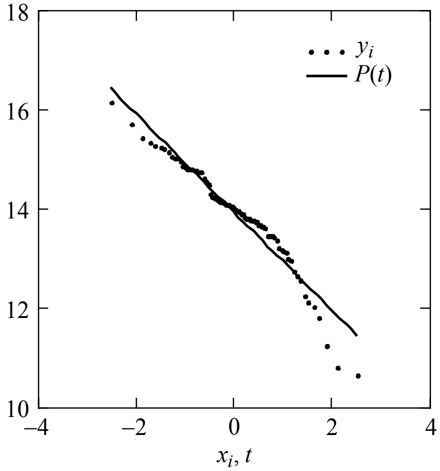

Рис. 1. Распределение численности населения РФ по 85 субъектам

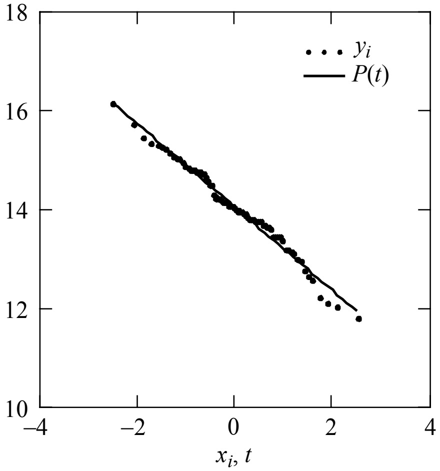

Рис. 2. Распределение численности населения РФ по 82 субъектам

Для вычисления суммы квадратов невязок используем формулу

$$STD = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 ,$$

где y_i – исходные наблюдения (т.е. значения численности населения), $\hat{y}_i$ – теоретические значения [4]. В первом случае отклонение составило $STD1 = 0{,}409$, во втором – $STD2 = 0{,}074$. Таким образом, отклонение уменьшилось почти на порядок.

3. Иллюстративные примеры

Рассмотрим другие примеры применения квантильной модели и логнормального закона. Используя данные 2002 г., исследуем распределение численности населения в Российской Федерации. В начале двухтысячных годов территория России была разделена на 7 федеральных округов: Центральный, Северо-Западный, Приволжский, Южный, Сибирский, Дальневосточный. В данном случае рассматривается каждый регион (округ) и субъекты, их составляющие (области, края, республики). На численность проживающих граждан влияют, в частности, исторические, социально-экономические, демографические, природно-климатические факторы. В качестве «ячеек накопления» здесь будут выступать субъекты, и каждой ячейке будет соответствовать своя численность населения. На рис. 3 представлена картина распределения численности населения Российской Федерации в соответствии с разбиением территории на федеральные округа.

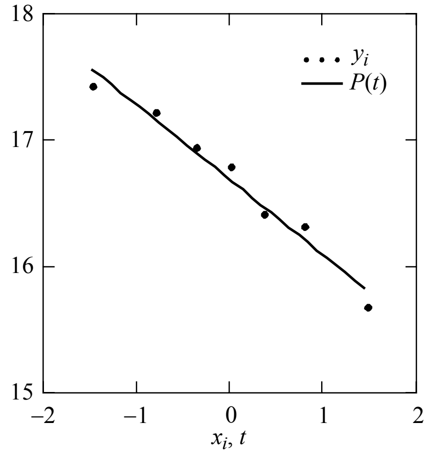

Рис. 3. Распределения численности населения РФ в соответствии с разбиением территории на 7 федеральных округов

Как видно из рис. 3, точки располагаются в непосредственной близости прямой, и отклонение не является слишком заметным. Этот факт говорит о том, что, хотя разбиение территории и носит административный характер, однако, судя по всему, оно хорошо вписывается в рассматриваемую нами модель, то есть является вполне адекватным.

Однако всё это были случаи, зафиксированные в одном временном интервале. Что же будет происходить, если рассмотреть процесс размещения населения в динамике? Возьмём случай распределения численности населения США по 49 штатам (исключив из рассмотрения Аляску, Гавайи в силу их специфического географического положения) в длительном отрезке времени, начиная с 1930 г. и до нашего времени. Построим диаграммы рассеяния для 1930, 1960 и 2005 гг. (рис. 4 – 6). Стандартные отклонения будут следующими (представлены в хронологическом порядке) $STD4 = 0{,}104$, $STD5 = 0{,}047$, $STD6 = 0{,}031$.

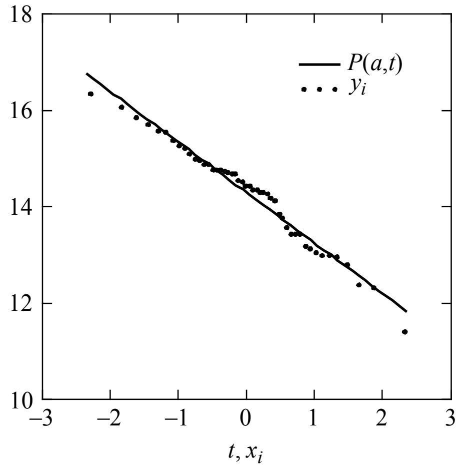

Рис. 4. Распределение численности населения США в 1930 г. по штатам

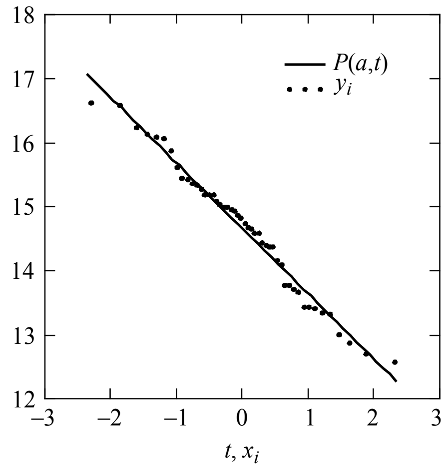

Рис. 5. Распределение численности населения США в 1960 г. по штатам

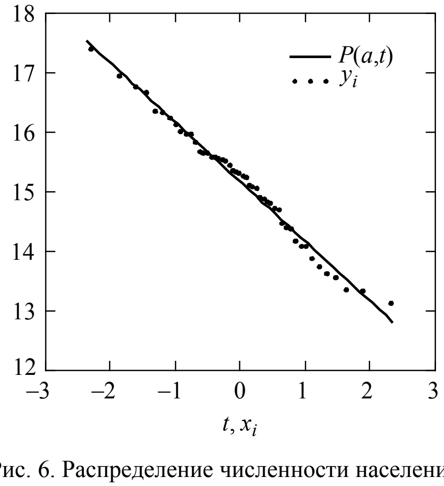

Рис. 6. Распределение численности населения США в 2005 г. по штатам

Таким образом, мы видим, что с течением времени стандартное отклонений уменьшается. И, несмотря на то, что в отдельные годы могло наблюдаться небольшое увеличение значения невязок, в целом тенденция к уменьшению прослеживается отчётливо.

Таблица 1

Изменение суммы квадратов невязок с течением времени

Год	Сумма квадратов невязок
1930	0,104
1940	0,101
1950	0,066
1960	0,047
1970	0,046
1980	0,051
1990	0,036
2000	0,032

Таблица 2

Изменение рангов пяти крупнейших штатов

Штат	Ранг в 1930 г.	Ранг в 1960 г.	Ранг в 1990 г.
New York	1	1	2
Pennsylvania	2	3	5
Illinois	3	4	6
Ohio	4	5	7
Texas	5	6	3

Во всех штатах в течение всего анализируемого периода происходит увеличение численности населения, при этом если проранжировать штаты по убыванию численности от 1 до 49, то с течением времени ранги большинства штатов будут менять свои значения (табл. 2), следовательно, изменение рангов обуславливается различной скоростью прироста населения. Поскольку сумма квадратов невязок имеет тенденцию к уменьшению своего значения, то можно предположить, что с течением времени вся совокупность населения, разбитая по отдельным территориям, становится, в некотором смысле, более «природной», в смысле более адекватного описания логнормальной моделью распределения численности населения.

Заключение

Применение квантильной модели и логнормального закона распределения для описания рассматриваемых процессов формирования численности населения (тесно связанных с миграционным движением населения) по территориям, как страны в целом, так и одного отдельно взятого субъекта, вполне целесообразно. С другой стороны, исследования показали, что в странах Европы, равно как и в других развитых странах, логнормальный закон описывает распределение численности населения лучше, чем в странах с ещё не устоявшейся рыночной экономикой или в странах с очевидным жёстким административным делением территорий. Вполне возможно, что с течением времени эти территориальные единицы (которые мы с полным правом можем назвать ячейками накопления) образуют действительно природную совокупность, которая будет лучше описываться рассмотренной квантильной моделью.

ЛИТЕРАТУРА

1. *Дмитриев Ю.Г.*, *Устинов Ю.К.* Математические модели растущих систем // Вычислительные технологии. 2007. Т. 12. С. 68 – 75.
2. *Алексеев Ф.Н.*, *Устинов Ю.К.* Вероятностная теория формирования запасов полезных ископаемых // Непараметрические и робастные статистические методы в кибернетике и информатике / под ред. Ф.П. Тарасенко. Томск: Иркут. фил. ВМНИИПС и др., 1990. С. 32 – 42.
3. *Айвазян С.А.* Модель формирования распределения населения России по величине среднедушевого дохода // Экономика и математические методы. 1997. Т. 32. Вып. 4. С. 75 – 88
4. *Федеральная* служба государственной статистики [Электроный ресурс]. URL: http://www.gks.ru/, свободный.
5. *Магнус Я.Р.* Эконометрика. Начальный курс: Учебник. 4-е изд. М.: Дело, 2005. 400 с.

Талейко Андрей Владимирович
Томский государственный университет
E-mail: filoph@sibmail.com
Поступила в редакцию 16 апреля 2009 г.