М.А. Петрова
(*Москва*)

ПРИМЕНЕНИЕ РАСПРЕДЕЛЕНИЯ ВЕЙБУЛЛА-ГНЕДЕНКО ПРИ АНАЛИЗЕ ТЕЧЕНИЯ ЭТНОПОЛИТИЧЕСКОГО КОНФЛИКТА[1]

В статье рассмотрены основные свойства распределения Вейбулла-Гнеденко. Обосновано применение этого распределения для исследования данных о моментах наступления событий, характеризующихся социальной напряженностью, разработана соответствующая модель «готовности к действию». С помощью различных статистических оценок и проверки соответствующих гипотез предложена методика поиска точек смены параметров для отслеживания моментов смены характера процесса.

Ключевые слова: математическое моделирование, распределение Вейбулла-Гнеденко, функция риска, минимальная статистика, этнополитический конфликт, социальная напряженность, протест.

Распределение Вейбулла-Гнеденко традиционно используется для описания времени жизни «образцов», поставленных на испытание. Оно было названо в честь шведского исследователя Валоди Вейбулла (Waloddi Weibull) [1, п. 6], применявшего его для описания времени отказов разного типа, и известного русского математика Б.В. Гнеденко. Это распределение использовалось для описания времени жизни электронных устройств, ламп, под-

---

Мария Алексеевна Петрова – аспирантка кафедры вычислительных методов факультета ВМиК МГУ. E-mail: petrova-ma@rambler.ru.

[1] Данная работа ведется под научным руководством канд. физ.-мат. наук В.А. Шведовского.

шипников и даже некоторых финансовых задач, а также для описания течения различных социальных процессов, в том числе международных конфликтов [2, p. 20–55; 3]. В частности, метод, основанный на статистической аппроксимации распределением Вейбулла-Гнеденко последовательности временных интервалов между однородными событиями, характеризующими конфликт, прошел независимую апробацию в Институте анализа и управления конфликтами и стабильностью Российской ассоциации теории и моделирования международных отношений и показал приемлемую точность описания политических процессов [3, с. 1].

Функция распределения Вейбулла-Гнеденко описывается выражением $F(x) = 1 - \exp\{-\lambda x^m\}$.

Основные параметры m и λ этого распределения в статистике называются соответственно параметрами формы и масштаба. Наряду с функцией распределения также рассматривается функция риска, или функция интенсивности, которая имеет вид

$$h(t) = f(t)/(1-F(t)) = m\lambda t^{m-1}.$$

Здесь функция интенсивности $h(t)$ – вероятность возникновения события в течение малого интервала времени при условии, что в начале интервала события не произошло, $f(t)$ – плотность, а $F(t)$ – функция распределения. Распределение Вейбулла позволяет гибко моделировать различные возникающие на практике формы функции интенсивности. Задавая разные параметры распределения, можно получить практически любые функции риска. На рис. 1 приведены функции интенсивности для параметров $m = 0{,}5$, $m = 1$, $m = 2$ и $m = 5$. Направление роста функции интенсивности зависит от значения параметра формы исходного распределения.

Если $m = 1$, то интенсивность рассматриваемого процесса постоянна и не зависит от времени; если $m < 1$, то процесс носит затухающий характер, интенсивность исследуемого процесса уменьшается со временем; если $m > 1$, то интенсивность процесса возрастает со временем [3, с. 4].

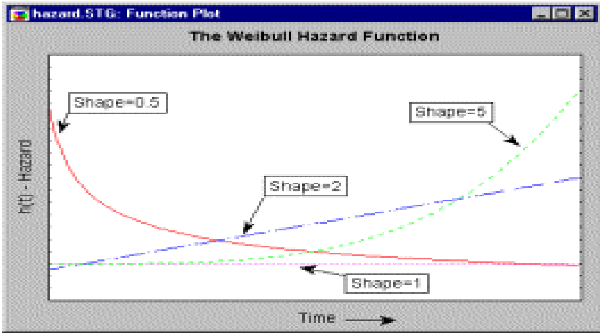

Рис. 1. Функции интенсивности распределения Вейбулла-Гнеденко для различных значений параметров

Использование распределения Вейбулла-Гнеденко для анализа течения этнополитического конфликта

Для обоснования использования распределения Вейбулла-Гнеденко для описания протестных действий в регионах Северного Кавказа была построена модель течения этнополитического конфликта. В представленной модели предполагается, что статистическое распределение периодов времени между протестными действиями (вспышками социальной активности) является двухпараметрическим распределением Вейбулла-Гнеденко. Тогда после нахождения по выборке параметра формы распределения Вейбулла-Гнеденко можно делать выводы о характере процесса течения этнополитического конфликта. Первым использовал этот подход И.Д. Петерсен, при этом он предполагал, что в международных и межэтнических отношениях существуют некоторые динамические законы, которые меняются во времени, но которые можно и нужно изучать [2, p. 11–15].

Одной из главных задач данной работы является поиск точек смены характера распределения. Это позволяет отслеживать точки смены характера протекания конфликта и предсказывать направление этого изменения.

Распределение Вейбулла-Гнеденко как распределение минимальной статистики

Рассматривается выборка вида (z_1, z_2, z_3, …, z_n), называемая *родительской выборкой*, где z_i – это независимые, одинаково распределенные случайные величины с плотностью вероятности $\varphi(z)$. *Минимальная статистика* определяется как

$$Z_{\min(n)} = \min(z_1, z_2, \ldots, z_n).$$

Выборку, состоящую из $Z_{\min(n)}$, называют *дочерней выборкой.* Обозначим ее функцию распределения $Z_{\min(n)}$ как $F_{\min(n)}(x)$.

$$F_{\min(n)}(x) = P(Z_{\min(n)} < x) = 1 - P(Z_{\min(n)} > x) =$$
$$= 1 - P(z_i > x, 1 \le i \le n) = 1 - (1 - \Phi(x))^n,$$

где

$$\Phi(x) = \int_0^x \varphi(z).$$

Последнее равенство следует из определения функции распределения и независимости z.

Если в правой окрестности нуля функция $\varphi(x)$ удовлетворяет следующему условию: $\lim\limits_{z\to+0} z^{-\Delta}\varphi(z) = c,\ c > 0$, то можно показать, что при $n \to \infty$ предельное распределение минимальной статистики будет распределением Вейбулла-Гнеденко [4, p.115–118], т.е. $\lim\limits_{n\to\infty} F_{\min(n)}(x) = \lim\limits_{n\to\infty} 1 - \left(1 - \dfrac{\mu x^{\gamma}}{n}\right)^n = 1 - \exp(-\mu x^{\gamma})$. В двухпараметрическом распределении Вейбулла-Гнеденко параметры μ, γ определяются поведением функции плотности $\varphi(z)$ в окрестности нуля. По смыслу минимальная статистика отражает *модель цепочки*: в цепочке n звеньев, и событие происходит, когда рвется самое слабое звено.

Модель «готовности к действию»

Для каждого индивида рассматривается 2 фактора – «*оппозиционный настрой*» и *готовность к действию*. *Оппозиционный настрой* характеризует способность человека совершать протестные действия ради достижения какой-либо цели. Человек, не имеющий *оппозиционного настроя* (в заданный момент), не будет совершать каких-либо протестных действий во всяком случае до тех пор, пока такой настрой не появится. Будем предполагать, что *оппозиционный настрой* – бинарная величина. «Для того чтобы протест состоялся, необходим *определенный уровень социального недовольства*, придание силы и *массовых* действий в качестве приемлемого средства социальных перемен» [5, с. 1]. Этот уровень недовольства – оппозиционный настрой – способствует росту депривации[1] и активизации протестных действий.

Используется предположение о том, что протестные действия чаще всего совершаются в группе. Для людей, у которых такой настрой есть, вводится понятие группы единомышленников. Пусть некоторая группа людей действует, как единое целое. Параметр *готовность к действию* может принимать различные положительные целые значения и означает примерно следующее: *готовность к действию можно определять как количество дней, которое осталось до протестных действий при неизменных внешних условиях, готовность 1 – действовать прямо сейчас, готовность 2 – действовать завтра и т.п.* Когда у одного (или нескольких) из членов группы этот параметр равен 1, его настрой передается другим, группа готова действовать и совершает некоторое протестное действие, происходит некоторое *событие.* После этого *готовность к действию* у всех членов группы резко падает, у каждого на свой уровень. События мы можем наблюдать и регистрировать.

[1] Депривация – недостаток экономических и эмоциональных опор, общепринятых в качестве базисных основ человеческого опыта.

В данной модели готовность к действию – случайная величина, соответствующая родительской выборке (z_1, z_2, z_3, …, z_n), дочерней выборке $Z_{\min(n)}$ подходит случайная величина, описывающая событие. Когда минимум равен единице, происходит событие, родительская выборка содержит время, через которое готовность будет равной единице, минимальная статистика выбирает минимальное такое время. По построению модели ясно, что весь процесс (совокупность протестных действий, совершенных данной группой) описывается распределением Вейбулла-Гнеденко. Следовательно, временной интервал между событиями (протестными действиями, совершенными *одной* группой) будет описываться распределением Вейбулла-Гнеденко. У И.Д. Петерсена показано [2, p. 20], что если группы независимы друг от друга, то результирующее распределение тоже будет распределением Вейбулла-Гнеденко. Если же действия различных групп являются взаимозависимыми (что имеет место в реальной жизни), то условия, при которых результирующее распределение является распределением Вейбулла-Гнеденко, до конца не выявлены.

Оценки максимального правдоподобия для параметров распределения Вейбулла-Гнеденко

Используемые данные – промежутки времени между событиями, характеризующимися социальной напряженностью[1]. В качестве t_n берется промежуток времени между событием с номером n и с номером $n - 1$. По имеющимся данным можно построить $y(t)$ – эмпирическую функцию распределения выборки $\{t_n\}$. Рассматриваемая случайная величина ξ – «промежуток времени между протестными действиями», (t_1, …, t_n) – выборка из реализаций случайной величины ξ с неизвестной функцией распределения.

[1] Были получены А.И. Масловым в результате анализа сводок МВД по Дагестану, 2001–2002 гг.

Предполагается, что функция распределения ξ – функция Вейбулла-Гнеденко $F_W(t) = 1 - \exp(-\lambda t^m)$. Тогда для проверки гипотезы о виде распределения имеющейся выборки используется критерий согласия Колмогорова [6, с. 107]. Для оценки параметров распределения применяется метод максимального правдоподобия. После максимизации логарифма функции правдоподобия находим, что соответствующие оценки $\hat{\lambda}, \hat{m}$ должны удовлетворять следующим условиям [2, p. 30]:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i^{\hat{m}}} \tag{1}$$

$$f(\hat{m}) = \frac{n}{\hat{m}} + \sum_{i=1}^{n} \ln(t_i) - \frac{n}{\sum_{i=1}^{n} t_i^{\hat{m}}} \sum_{i=1}^{n} \ln(t_i) t_i^{\hat{m}}. \tag{2}$$

Так как последнее уравнение не может быть решено аналитически, искомые оценки параметров могут быть получены с помощью итерационных численных методов.

Методика поиска точек смены характера распределения

Поскольку процесс получения точек выборки $\{t_n\}$ был довольно растянут по времени (имеющиеся выборки были получены в течение 400-500 дней, вообще говоря, в разных условиях), целесообразно исходить, что (в силу внешних причин или внутренних свойств моделируемого объекта) параметры распределения к концу выборки могли измениться. Если в некоторых точках они изменяются, то можно предположить, что процесс течения этнополитического конфликта меняет свой характер. Главным критерием, характеризующим результат «подгонки», является качество аппроксимации эмпирической функции теоретической (критерий Колмогорова).

Сначала определяются области, в которых могут содержаться точки смены параметров. С этой целью выделяются точки резкой смены интенсивности процесса (по графику интенсивности). Привлекается для этого и дополнительная информация. После этого на каждом «подозрительном» интервале производится следующая процедура, описанная Т. Индоу [4, p. 117]. Из уравнения $F_W(t) = 1 - \exp(-\lambda t^m)$ двойным логарифмированием получается выражение

$$\ln(\ln\{1/[1 - F_W(t)]\}) = m \cdot \ln t + \ln\lambda.$$

Для анализа рассматриваются графики $\ln(\ln\{1/[1-y(t)]\})$, где по оси абсцисс отложено $\ln t$. В случае соответствия эмпирического распределения (на интервале) распределению Вейбулла-Гнеденко этот график должен быть прямой, причем коэффициент наклона этой прямой соответствует параметру формы m, а абсцисса в точке 0 (intercept) – параметру масштаба λ. Найденные с помощью метода наименьших квадратов приближенные значения параметров m и λ используются в качестве начального приближения в решении уравнения (2) методом простой итерации при вычислении оценок максимального правдоподобия.

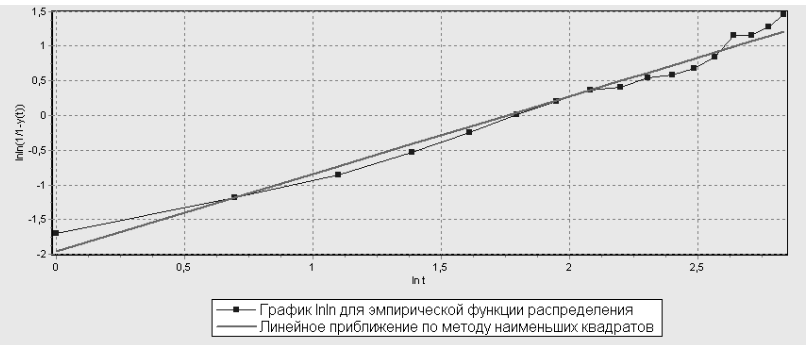

Рис. 2. График повторного логарифма эмпирической функции распределения и линейное приближение по методу наименьших квадратов

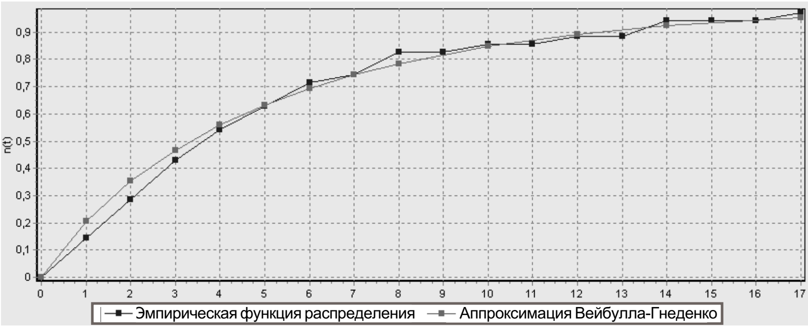

Рис. 3. График эмпирической функции распределения и теоретической функции Вейбулла-Гнеденко с параметрами, оцененными с помощью функции максимального правдоподобия

Следующий этап: с помощью критерия Колмогорова определяется соответствие теоретической функции распределения эмпирической на каждом интервале. Также по критерию Смирнова проверяется гипотеза однородности для всех соседних частей найденного разбиения выборки. В случае, если гипотеза однородности подтверждается, найденная точка не соответствует смене параметров, и количество элементов разбиения сокращается.

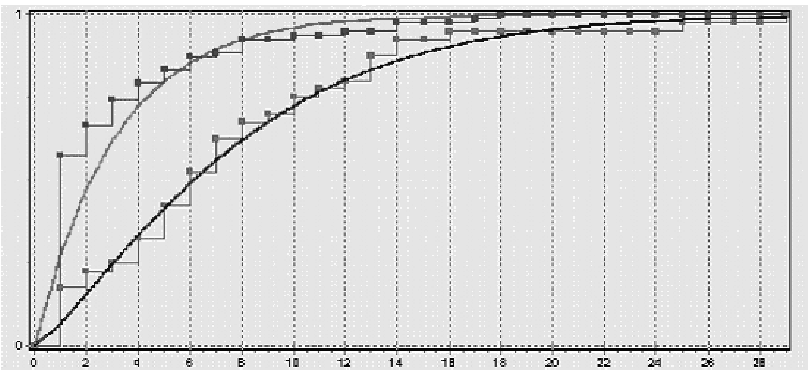

Рис. 4. Иллюстрация проверки гипотезы однородности

На рис. 4 показаны 2 различные эмпирические функции распределения, соответствующие разным частям выборки, и их аппроксимация распределением Вейбулла-Гнеденко. В рассматриваемом примере гипотеза однородности не подтверждается.

Таким образом, с помощью вышеописанной процедуры можно отслеживать точки смены параметров, подбирать теоретическую функцию распределения. Это, в свою очередь, позволяет, во-первых, прогнозировать течение этнополитического конфликта при неизменных внешних условиях, во-вторых, анализировать точки смены характера процесса и по дополнительным данным определять, что именно повлекло за собой эскалацию, или, соответственно, затухание процесса.

В заключение для удобства проведения анализа данных с помощью распределения Вейбулла-Гнеденко была создана программа, предназначенная для анализа данных различной природы, предположительно имеющих распределение Вейбулла-Гнеденко. Среда разработки – Borland C++Builder 5. Для графической реализации использовалась библиотека VCL. Желающие ознакомиться с программой могут обращаться непосредст-венно к автору.

Описанная совокупность методик (регрессионный анализ, анализ параметров распределения Вейбулла-Гнеденко) позволяет исследовать и прогнозировать течение этнополитического конфликта. Построенная модель применяется для реальных расчетов.

ЛИТЕРАТУРА

1. Электронный учебник по статистике. М.: StatSoft. http://www.statsoft.ru/home/textbook/default.htm.

2. *Petersen I.D.* The Dynamic Laws of International Political Systems 1823–1973. Institute of Political Studies, University of Copenhagen, Copenhagen Political Studies, 1980.

3. *Кретов В.С., Власов И.Е., Фролов И.В*. Некоторые аспекты создания интеллектуальных информационных систем в политологии // НТИ. 1994. № 11.

4. *Indow T.* Weibull Form in Memory, Reaction Time, and Social Behavior: Asymptotic Distribution of Minima from Heterogeneous Population // Institute for Mathematical Behavioral Sciences, Technical Report Series, MBS 95-04, 1995.

5. *Владимиров А.А.* Политическое поведение и его формы. http://kosilka.h1.ru.

6. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. М.: Высшая школа, 1984.

7. *Шведовский В.А.* Динамическая модель этнополитического конфликта: построение, возможности и результаты применения // Математическое моделирование социальных процессов. М.: Изд-во МГУ, 2000. Вып. 2.

8. *Indow T.* Analyses of Events Counted on Time-Dimension: a Soft Model of Extreme Statistics // Behaviometrika. 1993. Vol. 20. № 2.