

КОМПОЗИТНЫЕ РАСПРЕДЕЛЕНИЯ В ЭКОНОМИКЕ, В ЕСТЕСТВЕННЫХ НАУКАХ И В СОЦИОЛОГИИ

Благовещенский Ю.Н. (Москва)ⁱ

Хорошо известно, что многие распределения в специальных координатах линеаризуются, т.е. их график оказывается прямой линией. Для семейства $F(x, \theta)$, $\theta \in \Theta$, линеаризуемость означает, что существуют независимые от параметров функции $g_F(x)$ и $\mathcal{G}_F(\lambda)$ такие, что между λ -квантилем $X_F(\lambda, \theta)$ и $\mathcal{G}_F(\lambda)$ имеется линейная связь $\mathcal{G}_F(\lambda) = b \cdot g_F(X_F(\lambda, \theta)) + a$. Например, для $F(x, \theta) = H((x-a)/b)$, $b > 0$, где $\theta = (a; b)$, для $\lambda \in (0; 1)$ выполняется равенство: $H^{-1}(\lambda) = b \cdot X_F(\lambda, a, b) + a$. Наряду с типами распределений это свойство присуще и ряду других семейств, среди которых следует отметить распределения Парето, Вейбулла и логнормальное.

Доклад посвящен использованию для описания эмпирических данных кусочно линеаризуемых распределений, т.е. таких, которые в специальных координатах, например, $g_F(x)$ и $\mathcal{G}_F(\lambda)$, представляют ломаную. Такие распределения мы называем композитными или F -склеяками, когда в основе лежит семейство $F(x, \theta)$, $\theta \in \Theta$.

Исторически впервые с F -склеякой мы столкнулись при анализе выборов в Государственную Думу (Благовещенский, Винюков, 2004), проверяя результаты голосований 24 избирательных блоков на согласие с законом Ципфа—Парето (один из блоков – голосование против всех). В координатах, линеаризующих распределение Парето, мы получили трехзвенную ломаную с хорошо интерпретируемыми сегментами: лига реальной борьбы за власть (7 блоков), второй эшелон с узкими политическими программами (тоже 7 блоков) и 10 аутсайдеров, мало знакомых избирателю.

Другой интересный пример – из анализа коррупционных рынков по данным опроса Россия-2005 (чуть больше 700 предпринимателей и вдвое больше граждан). Оказалось, что двойной логарифм размера взяток (бизнес чиновнику) на нормальной бумаге представляет ломаную: первый отрезок – взятки до 3 млн. руб., а второй (как бы "высшая" лига) – больше 3 млн., и это уже – другой бизнес и другие чиновники (Благовещенский, Сатаров, 2013).

Подобного рода распределения нередко встречаются в публикациях о распределении площадей тех или иных объектов, но авторы почему-то игнорируют наличие явного излома на их же собственных графиках в логарифмических координатах (линеаризация для Парето), пытаясь аппроксимировать их теми или иными обобщениями закона Ципфа—Парето. Характерным примером может служить интересная статья (Tebbens, Burchoughs, 2005) о размерах лесных пожаров в Альберте (это регион западной Канады) за 1996-2000 годы, см. Fig.5 в указанной публикации.

В недавно проводимом исследовании (Благовещенский, Сатаров, 2019) мы столкнулись с композитными распределениями, в основе которых было распределение Вейбулла: $W(x|a,b)=1-\exp(-a\cdot x^b)$, $a>0$, $b>0$. Оно линеаризуется, если взять $g_F(x) = \ln(x)$ и $\mathcal{G}_F(\lambda) = \ln[\ln(1/(1-\lambda))]$. Нас интересовало распределение степеней узлов в графах связей между позициями, которые выбирали респонденты в процессе практически идентичных опросов населения в России, Казахстане и Эстонии в 1998 году. Оказалось, что эти распределения великолепно аппроксимируются четырехзвенными W -склейками, причем, разбиение на сегменты хорошо интерпретировалось в рамках представлений о массовом политическом сознании.

Кроме этого, приводятся примеры общеизвестных данных, чье распределение прекрасно описывается склейками, основанных на распределении Парето. Это ВВП 188-ми стран мира (данные от МВФ) и размер 116 островов площадью больше 5 тысяч кв. км. Из данных (<https://geo.koltygin.ru/ostrova.php>) мы исключили Гренландию, чья площадь в 3.6 раза меньше Австралии и в 2.7 больше наибольшего из остальных островов (о. Новая Гвинея в Океании).

В докладе также даются процедуры состоятельного оценивания параметров и доказываются их асимптотическая нормальность, причем, у k -звенной склейки ровно столько же параметров, сколько у смеси распределений из того же семейства.

Литература

- Tebbens S.F., Burroughs S.M.* Forrest fire burn areas in Western Canada modeled as self-similar criticality. *Physica D* 211. 2005. pp. 221–234.
- Благовещенский Ю.Н., Винюков И.А., Каргер М.Л.* Анализ выборов в ГД-2003 и закон Ципфа-Парето. – Труды VI-й Международной школы-семинара "Многомерный статистический анализ и эконометрика" (Армения, Цахкадзор), Москва. 2004. С. 16–23.
- Благовещенский Ю.Н., Сатаров Г.А.* Характеристики рынков коррупции. В книге: Российская коррупция: уровень, структура, динамика. Опыты социологического анализа. // Под ред. Г.А. Сатарова. – Москва: Фонд «Либеральная Миссия», 2013. – 752 с.
- Благовещенский Ю.Н., Сатаров Г.А.* (2019). Сетевое моделирование массового политического сознания: метод. ПОЛИС. Политические исследования. 2019. 5. С. 96–118.

ⁱ Благовещенский Юрий Николаевич – Фонд ИНДЕМ, ykbun@yandex.ru