

УДК 51-77
ББК 60.55.372

ОЦЕНКА ВЕРОЯТНОСТИ РЕДКИХ СОБЫТИЙ В ПОВЕДЕНИИ ТОЛПЫ

Рогаткин А. Д.¹

(ФГБУН Институт проблем управления РАН, Москва)

Исследуется предложенная в [2] модель поведения толпы, которая является обобщением модели конформного порогового коллективного поведения М. Грановеттера [13] на случай неопределённости относительно значений порогов агентов. Вероятность события, состоящего в выходе системы из некоторого множества состояний (т.е. в так называемом возбуждении толпы), оценивается при помощи асимптотического результата, полученного в [8]. Теоретические оценки типа больших уклонений уточняются при помощи численных оценок, полученных методом статистических испытаний. Полученные результаты дают возможность оценить надёжность обеспечения невозбуждения толпы в тех случаях, когда вероятности событий слишком малы для применения метода статистических испытаний. Приведены рекомендации по выбору параметров, обеспечивающих заданную вероятность выхода системы из множества заданных состояний.

Ключевые слова: модель Грановеттера, конформное коллективное поведение, управление толпой, большие уклонения, статистические испытания.

¹ Рогаткин Андрей Дмитриевич, аспирант
(andreyrogatkin@gmail.com).

1. Введение

В работах М. Грановеттера [13] и Т. Шеллинга [14] предложены модели коллективного поведения, которые сегодня лежат в основе множества различных подходов к математическому и физическому описанию социальных явлений [3, 4, 5, 9]. Единой для этих подходов является ситуация, в которой агенты осуществляют бинарный выбор - действовать или бездействовать - в дискретные моменты времени, ориентируясь при этом на действия своего окружения. При этом ключевой характеристикой агента является его порог, при превышении которого обстановкой агент изменяет свое поведение. В течение последних десятилетий пороговые модели получили широкое распространение, в частности, одним из современных направлений исследований является их развитие – решетки связанных отображений (Coupled Map Lattices) [10, 11, 15].

В классических работах М. Грановеттера и Т. Шеллинга не учитывается неполнота информации, которая важна для приложений, в том числе для решения задачи управления надежностью социальных систем. В реальных ситуациях, как правило, пороги агентов не известны точно, а известно их вероятностное распределение. В настоящей работе для описания поведения толпы (раздел 2) используется предложенная в [2] модель, которая является обобщением модели конформного порогового коллективного поведения М. Грановеттера [13] на случай вероятностной неопределенности относительно значений порогов агентов.

В разделе 3 используется полученная автором в [8] асимптотика типа больших уклонений для оценки вероятности редких событий в поведении толпы. В качестве функции распределения порогов агентов используется двухпараметрическое семейство функций, полученное в [1, 6] на основании исследования социальных сетей Facebook, Livejournal и Twitter. Показано, что грубая логарифмическая асимптотика типа больших уклонений может применяться для численной оценки вероятностей редких событий при помощи «калибровки» оценки методом статисти-

ческих испытаний (раздел 4) на множестве параметров, позволяющих получать оценки статистическим методом.

В разделе 5 решается задача управления надежностью социальных систем. Находится множество значений параметров системы, обеспечивающих на заданном промежутке времени заданную надежность.

2. Описание модели

Рассмотрим конечное множество агентов $N = \{1, 2, \dots, n\}$. Каждый из агентов имеет некоторый порог сопротивления социальному давлению $\theta_i \in [0, 1]$, $i \in N$, (далее – порог; под социальным давлением понимается то, сколько других агентов действует). На шаге k агент $i \in N$ выбирает одно из двух состояний $\omega_{ik} \in \{0, 1\}$ (если $\omega_{ik} = 1$, то говорят, что агент «действует», иначе говорят, что он «бездействует»), его состояние на шаге $k + 1$ определяется по правилу

$$(1) \quad \omega_{i(k+1)} = \begin{cases} 1, & \frac{1}{n} \sum_j \omega_{jk} - \theta_i \geq 0, \\ 0, & \frac{1}{n} \sum_j \omega_{jk} - \theta_i < 0. \end{cases}$$

Согласно правилу (1), агент действует, если состояние системы $x_k = 1/n \sum_i \omega_{ik}$ не ниже, чем его порог. Такое поведение называется конформным. Динамика состояния системы во времени при этом подчиняется рекуррентному соотношению

$$(2) \quad x_{k+1} = F_n(x_k),$$

где

$$(3) \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n \chi(\theta_i \leq x).$$

Здесь χ обозначает индикатор множества.

В работах [2, 8], был рассмотрен случай, при котором в правой части выражения (1) имеется неопределённость: вместо известных порогов агентов рассматривается последовательность

$\theta_1(\omega), \dots, \theta_n(\omega)$ независимых одинаково распределённых (с распределением $F(\cdot)$) случайных величин на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Такая ситуация возникает, например, когда пороги агентов не известны точно, но агенты выбраны случайно из множества агентов с известным распределением порогов.

Эмпирическая функция распределения порогов агентов в этом случае запишется как

$$(4) \quad F_n(x, \omega) = \frac{1}{n} \sum_i \chi(\theta_i(\omega) \leq x).$$

При фиксированных F и x_0 получаем, что выражением

$$x_1^n(\omega) = F_n(x_0, \omega),$$

...

$$(5) \quad x_k^n(\omega) = F_n(x_{k-1}(\omega), \omega),$$

...

$$x_K^n(\omega) = F_n(x_{K-1}(\omega), \omega)$$

задана последовательность случайных конечных последовательностей $\{\bar{x}^n(\omega) = (\bar{x}_1^n(\omega), \dots, \bar{x}_K^n(\omega))\}$ - траекторий системы, каждая из которых имеет некоторое распределение P_n на пространстве \mathbb{R}^K , определяемое как $P_n(A) = \mathbb{P}\{\omega: \bar{x}^n(\omega) \in A\}$. На рис. 1 показан пример одной реализации эмпирической функции распределения и соответствующей ей траектории системы.

В работе [8] была доказана следующая асимптотика для распределений траекторий системы:

$$(6) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n(U) \geq -\inf_U H(\bar{y})$$

для любого открытого множества $U \in \mathcal{B}(S)$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n(C) \leq -\inf_C H(\bar{y})$$

для любого замкнутого множества $C \in \mathcal{B}(S)$, где

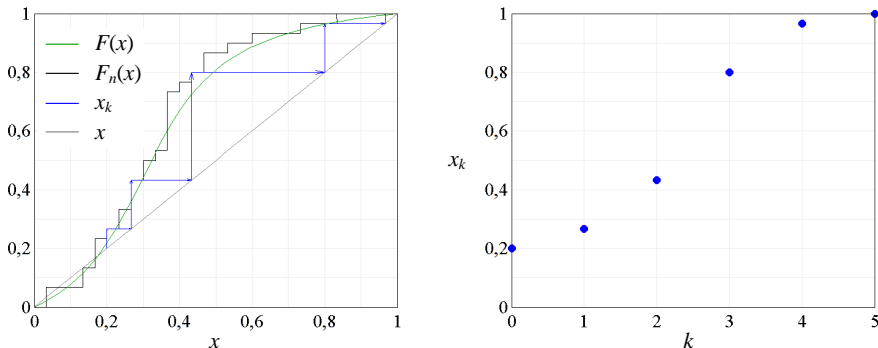


Рис. 1. Реализация случайной траектории, соответствующей функции распределения $F(\cdot)$. Диаграмма Ламерея (слева).

Траектория системы (справа), $n = 30$

$$(7) \quad H(\bar{y}) = \begin{cases} y_1 \ln \frac{y_1}{F(y_0)} + (1 - y_K) \ln \frac{1 - y_K}{1 - F(y_{K-1})} + \\ + \sum_{k=1}^{K-1} (y_{k+1} - y_k) \ln \frac{y_{k+1} - y_k}{F(y_k) - F(y_{k-1})}, \bar{y} \in Y_1, \\ (1 - y_1) \ln \frac{1 - y_1}{1 - F(y_0)} + y_K \ln \frac{y_K}{F(y_{K-1})} + \\ + \sum_{k=1}^{K-1} (y_k - y_{k+1}) \ln \frac{y_k - y_{k+1}}{F(y_{k-1}) - F(y_k)}, \bar{y} \in Y_2, \\ +\infty, \bar{y} \notin Y_1 \cup Y_2, \end{cases}$$

и введены обозначения $y_0 \triangleq x_0$, $0 \cdot \ln(0) \triangleq 0$,

$$Y_1 = \left\{ \bar{y} \in [0, 1]^K : y_0 < y_1 < y_2 < \dots < y_m = y_{m+1} \dots = y_K \right\},$$

$$m \in \{0, \dots, K\},$$

$$Y_2 = \left\{ \bar{y} \in [0, 1]^K : y_0 > y_1 > y_2 > \dots > y_m = y_{m+1} \dots = y_K \right\},$$

$$m \in \{0, \dots, K\}.$$

Асимптотические результаты типа (6) изучаются в теории больших уклонений [12].

В работах [1, 6] показано, что конформное поведение в сетевых структурах имеет эквивалентное модели Грановеттера математическое описание, в котором структура сети заключена в некоторую функцию, имеющую свойства функции распределения. В [1, 6] было проведено исследование социальных сетей Facebook, Livejournal и Twitter и показано, что реальные социальные сети хорошо моделируются функциями из двухпараметрического семейства

$$(8) \quad F(x, \theta, \lambda) = \frac{\arctg \lambda (x - \theta) + \arctg(\lambda \theta)}{\arctg \lambda (1 - \theta) + \arctg(\lambda \theta)},$$

где $\lambda \in (0, +\infty)$ - параметр, носящий название «коэффициент неоднородности» и характеризующий степень отличия функции распределения от равномерного распределения, а $\theta \in [0, 1]$ -параметр, называемый «единым относительным порогом» и характеризующий происходящий в сети процесс принятия решения. В [1, 6] показано, что значения коэффициента неоднородности составляют $\lambda_F = 13,01$; $\lambda_L = 9,18$; $\lambda_T = 7,34$ для социальных сетей Facebook, Livejournal и Twitter соответственно. На рис. 2 изображены различные кривые из параметрического семейства (8).

В настоящей работе изучается вероятность события $A \in \mathcal{F}$, состоящего в том, что случайный процесс (5) с заданным начальным условием (например, $x_0 = 0,2$) превысит в некоторый момент времени точку выхода (например, $x_{вых} = 0,5$) при различных количествах n агентов в системе. Формальное определение данного события запишется как

$$(9) \quad A^n = \left\{ \omega \in \Omega : \exists k \in \square \quad x_k^n(\omega) > x_{вых} \right\}.$$

Событие A^n далее называется *выходом из области* социальной системы, состоящей из n агентов. Вероятность выхода из области в описанной модели зависит только от теоретической функции распределения порогов агентов $F(\cdot)$ и их количества n . В случае функций распределения из двухпараметрического семейства (8) введём для нее обозначение

$$(10) \quad P_{вых}^n(\theta, \lambda) = P \left\{ \omega \in A^n \right\}, \quad F(x) = F(x, \theta, \lambda).$$

Асимптотическая оценка (6) для вероятности $P_{\text{вых}}^n(\theta, \lambda)$ принимает вид

$$(11) \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{\text{вых}}^n(\theta, \lambda) = - \inf_{\bar{y} \in A^n} H(F(\cdot, \theta, \lambda), \bar{y}).$$

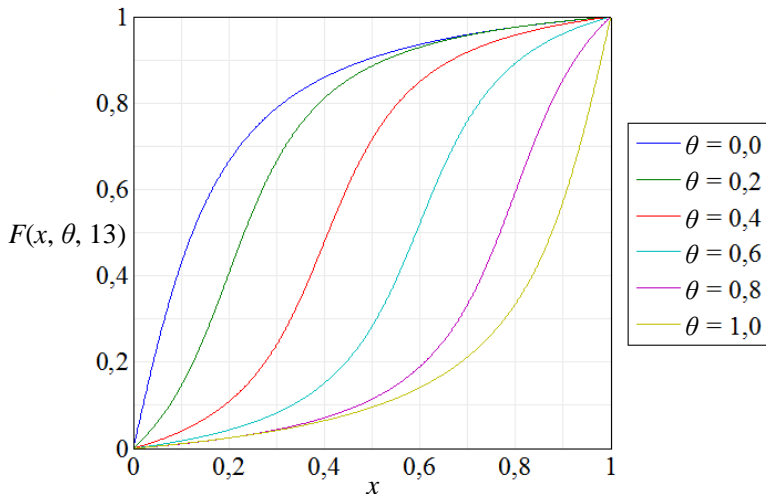


Рис. 2. Кривые из семейства (8) при $\lambda = 9,18$ и различных $\theta \in [0, 1]$

3. Оценка вероятности выхода из области

В данном разделе приведены результаты вычисления $P_{\text{вых}}^n(\theta, \lambda)$ методом статистических испытаний для $n_1 = 50$, $n_2 = 100$ и $n_3 = 200$ и показано, как с применением асимптотики (11) они могут быть продолжены для более высоких значений n .

Оценка (11) может быть переписана в виде

$$(12) P_{\text{вых}}^n(\theta, \lambda) = c(n, \theta, \lambda) e^{-n \inf_{\bar{y} \in A^n} H(F(\cdot, \theta, \lambda), \bar{y})},$$

где $c(n, \theta, \lambda)$ удовлетворяет

$$(13) \lim_{n \rightarrow \infty} \frac{\log c(n, \theta, \lambda)}{n} = 0 \text{ для всех } \theta, \lambda,$$

т.е. изменяется «меньше чем экспоненциально» по n . Не зная константы $c(n, \theta, \lambda)$, нельзя определить значение вероятности $P^n_{\text{вых}}(\theta, \lambda)$ с заданной точностью. По этой причине утверждения типа (11) иногда называют «грубая логарифмическая асимптотика». Асимптотика (11) может применяться для получения численной оценки вероятности $P^n_{\text{вых}}(\theta, \lambda)$ только при наличии дополнительной информации о константе $c(n, \theta, \lambda)$. В данной работе значение этой константы определялось приближённо при помощи полученной методом статистических испытаний оценки $\tilde{P}^n_{\text{вых}}(\theta, \lambda)$, изображённой на рис. 3. Подробное описание алгоритма проведения статистических испытаний дано в следующем разделе.

Качественные различия в $P^n_{\text{вых}}(\theta, \lambda)$ как функции двух переменных (θ, λ) при различных n хорошо иллюстрируются линиями уровня (кривыми, на которых $P^n_{\text{вых}}(\theta, \lambda)$ постоянна). На рис. 4 изображены линии уровня $\tilde{P}^n_{\text{вых}}(\theta, \lambda)$, определяемые как функции $\lambda_1^n(\theta)$, $\lambda_2^n(\theta)$ и $\lambda_3^n(\theta)$, удовлетворяющие

$$(14) \begin{aligned} \tilde{P}^n_{\text{вых}}(\theta, \lambda_1^n(\theta)) &= 10^{-2}, \\ \tilde{P}^n_{\text{вых}}(\theta, \lambda_2^n(\theta)) &= 10^{-3}, \\ \tilde{P}^n_{\text{вых}}(\theta, \lambda_3^n(\theta)) &= 10^{-4}. \end{aligned}$$

Приближенное значение $\hat{c}(\theta, \lambda)$ константы $c(n, \theta, \lambda)$ находилось на основании данных статистических испытаний следующим образом:

1. При помощи метода численной оптимизации вычислялась оценка $\hat{I}(\theta, \lambda)$ функции

$$(15) I(\theta, \lambda) = \inf_{\bar{y} \in A^n} H(F(\cdot, \theta, \lambda), \bar{y}).$$

2. Приближенное значение $\hat{c}(\theta, \lambda)$ константы $c(n, \theta, \lambda)$ находилось по формуле

$$(16) \hat{c}(\theta, \lambda) = \frac{\hat{P}^n_{\text{вых}}(\theta, \lambda)}{e^{-\hat{I}(\theta, \lambda)}}$$

при $n = 200$.

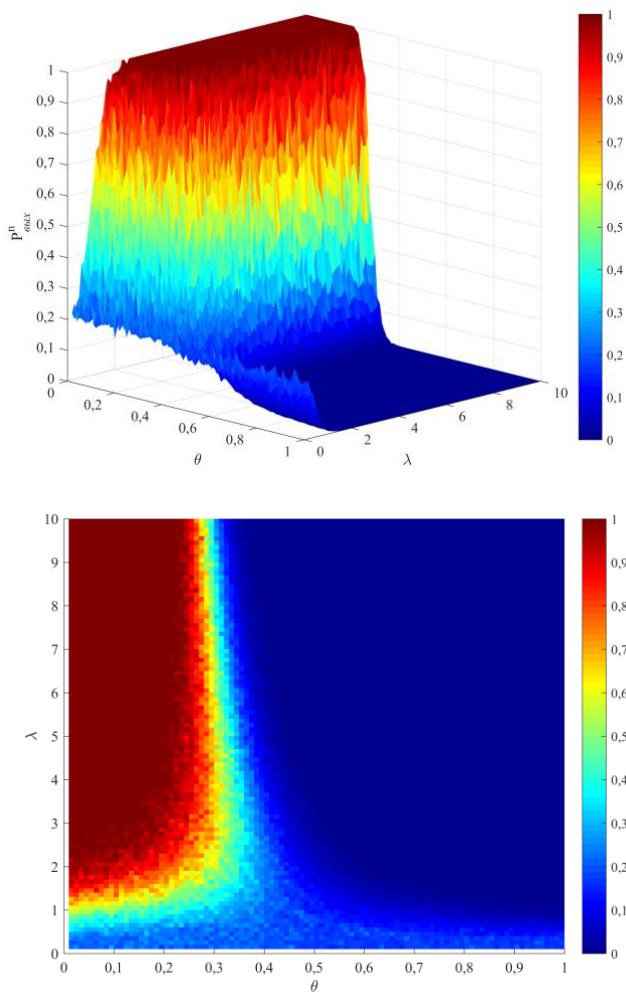


Рис. 3. Зависимость вероятности выхода из области от параметров θ , λ , полученная методом статистических испытаний при $n = 50$. Сверху – график функции $P_{\text{вых}}^n(\theta, \lambda)$, снизу – цветное изображение функции $P_{\text{вых}}^n(\theta, \lambda)$

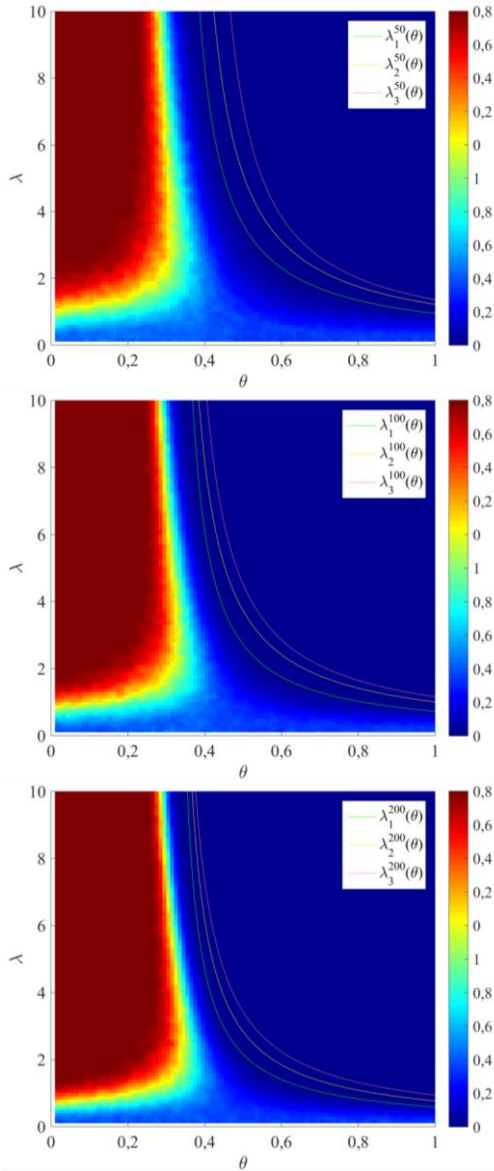


Рис. 4. Кривые равной вероятности выхода (линии уровня) для различного числа агентов социальной сети $n_1 = 50$, $n_2 = 100$ и $n_3 = 200$

Полученное значение $\hat{c}(\theta, \lambda)$ для различных (θ, λ) , применялось для того, чтобы записать «совмещённую» оценку в виде

$$(17) P_{\text{вых}}^n(\theta, \lambda) = \hat{c}(\theta, \lambda) e^{-n\hat{t}(\theta, \lambda)}.$$

Оценка (17) вероятности выхода из области применима при любых значениях n , в том числе тех, для которых невозможно получить результат методом статистических испытаний из-за ограничений на вычислительные ресурсы. Это позволяет оценивать вероятность выхода из области для реальных социальных систем, состоящих из большого числа агентов. Важно отметить, что оценка (17) получена объединением аналитического результата (11) и численных расчётов.

Построенная в данном разделе оценка вероятности для выхода из области (обозначенного выше A^n) может быть также построена и для любого другого события (подмножества допустимых траекторий системы $Y_1 \cup Y_2$). Для этого необходимо подставить соответствующее подмножество траекторий в (15) в качестве множества, на котором происходит оптимизация функционала $H(\cdot)$. Таким образом, исследователь может оценить вероятность любого возможного события, вычислив ее численно для некоторого значения n и осуществив «калибровку»

(15)–(17).

4. Алгоритм статистических испытаний

В работе [2] показано, что случайные траектории (5) имеют конечную длину, т.е. не более чем за n шагов система достигает некоторого (случайного) конечного состояния, которое не изменится с дальнейшим течением времени. Этот факт лежит в основе вычислительного эксперимента для определения вероятности $P_{\text{вых}}^n(\theta, \lambda)$, который был реализован в программной среде MATLAB и описывается далее. При помощи метода обратного преобразования функции распределения $F(\theta, \lambda)$ генерировалась конечная последовательность из n реализаций случайных порогов $\theta_1, \dots, \theta_n$, по которым вычислялась эмпирическая функция распределения (4). Далее при начальном условии $x_0 = 0,2$ со-

гласно (5) вычислялись первые n членов последовательности x_k^n . Считалось, что произошёл выход из области, если выполнялось $\max_{k=1\dots n}\{x_k^n\} \geq x_{\text{вых}}$, где $x_{\text{вых}} = 0,5$. Эта процедура повторялась до выполнения условия

$$(18) \min \{q_l, (1 - q_l)\} = \frac{10}{l},$$

где

$$(19) q_l = \frac{1}{l} \sum_{i=1}^l \chi \left(\max_{k=1\dots n} \{x_k^{ni}\} \geq x_{\text{вых}} \right)$$

– доля реализаций траектории x_k^{ni} с выходом из области, l – номер испытания в серии экспериментов. Затем вероятность выхода вычислялась на основании значения l , при котором произошла остановка процедуры, при помощи оценки, описание которой приведено в приложении «Оценка малого параметра геометрического распределения».

В силу ограничений на вычислительные ресурсы, в тех случаях, когда условие (18) не было достигнуто после $l = 10^5$ шагов, процедура останавливалась и соответствующая вероятность принималась равной нулю. Таким образом, методом статистических испытаний оценивались вероятности в интервале $(10^{-4}, 1 - 10^{-4})$. Для остальных значений вероятности $P_{\text{вых}}^n(\theta, \lambda)$ (очень близких к нулю или единице) статистическая оценка не может быть выполнена за разумное время в связи с ограничением на использованные вычислительные ресурсы. Применяемый для решения этой проблемы в статистических испытаниях метод *замены меры* в данном случае не применялся в связи с трудностями, возникающими при аналитических преобразованиях семейства мер $P_n(A) = \mathbb{P}\{\omega: \bar{x}^n(\omega) \in A\}$ (выписанное в явном виде это семейство мер можно найти в работе [2]).

5. Обеспечение надежности социальных систем

Рассмотрим социальную систему (толпу, социальную сеть и т.п.) со стохастическим пороговым поведением (5). Как показано в [2], такая система не более чем за n шагов приходит к

«равновесию» (действия агентов перестают изменяться). В силу случайности порогов агентов, состояние системы, при котором достигается это равновесие (результатирующее состояние), случайно. Выше были приведены оценки вероятности события, состоящего в том, что результатирующее состояние системы находится вне заданной области (при условии, что первоначально система находилась в заданной области), при различных значениях коэффициента неоднородности и единого относительного порога системы. Вероятность выхода системы из заданной области в общем случае зависит от параметров системы, как было показано выше. Ниже изучается задача обеспечения *надёжности* социальной системы при условии, что вероятность выхода p известна как функция основных параметров системы $p = p(n, \theta, \lambda) \equiv P_{\text{вых}}^n(\theta, \lambda)$.

Пусть реализация порогов агентов, и, следовательно, достижение равновесия повторяется многократно через равные промежутки времени τ , которые называются «быстрое время». Тогда за время t , называемое «медленное время», достижение равновесия произойдёт $m = [t / \tau]$ раз, где $[\cdot]$ обозначает целую часть числа. Согласно [7], *надёжностью* социальной системы называется вероятность сохранения её основных параметров внутри допустимой области. Для социальной системы основным наблюдаемым параметром является среднее действие агентов. Допустимой областью является множество $[0, x_B)$. Так как достижение равновесия многократно повторяется во времени, надёжность зависит от рассматриваемого промежутка времени. Обозначим через $R(t)$ *функцию выживания* – вероятность того, что за время t система ни разу не выйдет из допустимой области. Функция выживания, согласно определению надёжности, является надёжностью системы на временном интервале t .

Введём бернуллиевскую случайную величину $\xi \in \{0, 1\}$, которая равна единице, если система в момент времени $t_m = m\tau$ вышла из допустимой области, и равна нулю в противоположном случае. Вероятности этих событий равны соответственно p и $1 - p$. Функция выживания по определению может быть записана как

$$R(t) \equiv P(\xi_1 = \xi_2 = \dots = \xi_{[t/\tau]} = 0),$$

что приводит к результату

$$(20) R(t) = (1 - p)^{\left[\frac{t}{\tau} \right]}.$$

При малых p , таких что $p \ll 1/m = 1/[t/\tau]$, выражение (20) может быть приближено

$$(21) R(t) = 1 - p \left[\frac{t}{\tau} \right].$$

Задача управления надёжностью социальных систем состоит в том, чтобы найти множество параметров, при которых надёжность системы при заданном времени T окажется не меньше, чем заданный порог δ .

Рассмотрим иллюстративный пример. Пусть задача состоит в обеспечении надёжности $\delta = 1 - 10^{-2}$ для социальной системы с числом агентов $n = 10^7$ на промежутке времени, таком что $m = [t/\tau] = 10^3$. Решение задачи выглядит следующим образом.

1. Подставить данные в уравнение (20) и найти максимально допустимую вероятность выхода p . В данном случае применимо приближение (21), принимающее вид $\delta = 1 - pm$, откуда выражается

$$(22) p = \frac{1 - \delta}{m} = \frac{10^{-2}}{10^3} = 10^{-5}.$$

2. Решить уравнение $\hat{c}(\theta, \lambda) e^{-n \inf_{\bar{y} \in A^n} H(F(\cdot, \theta, \lambda), \bar{y})} = p$ и найти соответствующую линию уровня $\hat{\lambda}(\theta)$ в пространстве параметров (θ, λ) . Множество параметров (θ, λ) , обеспечивающее надёжность не меньше чем δ , лежит «правее и выше» найденной линии уровня, в области, изображённой зелёным цветом на рис. 5.

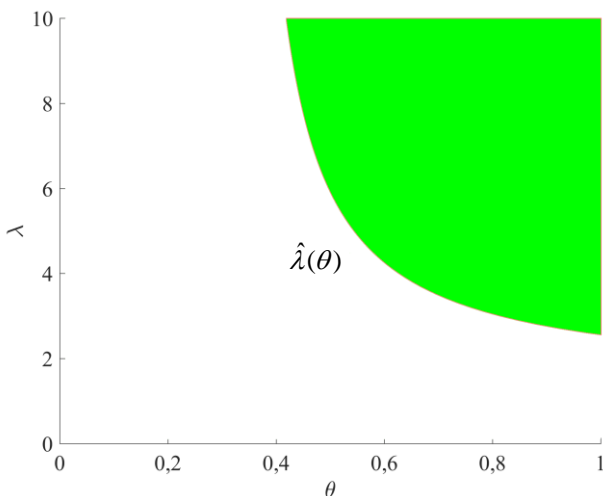


Рис. 5. Множество параметров (θ, λ) , обеспечивающее надёжность, не превышающую δ (зеленым цветом)

6. Заключение

Задача обеспечения надёжности социальной системы – её нахождения в заданном множестве состояний - требует оценки вероятностей редких событий. Зачастую эти вероятности настолько малы, что не позволяют применять для своей оценки метод статистических испытаний. Точные аналитические оценки и метод замены меры не всегда удаётся применить в связи со сложностью исследуемых моделей. В настоящей работе показано, что грубая логарифмическая асимптотика типа больших уклонений может применяться для численной оценки вероятностей редких событий при помощи «калибровки» оценки методом статистических испытаний на множестве параметров, позволяющих получать оценки статистическим методом. Оценку типа больших уклонений, как правило, легче получить аналитически, чем более точные оценки. Применение предлагаемой методики проиллюстрировано на модели конформного коллективного поведения. Построена численная оценка для различных

параметров модели. Приведён пример применения построенной оценки для задачи обеспечения надёжности – вероятности нахождения социальной системы в заданном «допустимом» множестве состояний.

Литература

1. БАТОВ А.В., БРЕЕР В.В., НОВИКОВ Д.А., РОГАТКИН А.Д. *Микро- и макромоделли социальных сетей. Ч.2. Идентификация и имитационные эксперименты* // Проблемы управления. – 2014. – №6. – С. 45–51.
2. БРЕЕР В.В., РОГАТКИН А.Д. *Вероятностная модель порогового поведения в многоагентных системах* // Автоматика и телемеханика. – 2015. – №8. – С. 56–77.
3. БРЕЕР В.В. *Модели толерантного порогового поведения (от Т. Шеллинга – к М. Грановеттеру)* // Проблемы управления. – 2016. – №1. – С. 11–20.
4. БРЕЕР В.В. *Модели конформного поведения. Ч.1. От философии к математическим моделям* // Проблемы управления. – 2014. – №1. – С. 2–13.
5. БРЕЕР В.В. *Модели конформного поведения. Ч. 2. Математические модели* // Проблемы управления. – 2014. – №2. – С. 2–17.
6. БРЕЕР В. В. , НОВИКОВ Д. А. , РОГАТКИН А. Д. *Микро- и макромоделли социальных сетей. Ч. 1. Основы теории* // Проблемы управления. – 2014. – №5. – С. 28–33.
7. БУРКОВ В.Н., НОВИКОВ Д.А. *Как управлять проектами.* – М.:СИНТЕГ, 1997. – 188 с.
8. РОГАТКИН А.Д. *Большие отклонения в социальных системах с пороговым конформным поведением* // Автоматика и телемеханика (в печати).
9. СЛОВОХОТОВ Ю.Л. *Физика и социофизика. Ч. I. Физические основы социальных явлений. Влияние солнечной активности на процессы в обществе. Движение в системе «живых частиц»* // Проблемы управления. – 2012. – №1. – С. 2–20.

10. CHAOUIYA C, OURRAD O, LIMA R. *Majority Rules with Random Tie-Breaking in Boolean Gene Regulatory Networks* // PLoS ONE 8(7): e69626. doi:10.1371/journal.pone.0069626. 2013.
11. CHAZOTTES J, FERNANDEZ B. *Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems*. – Springer, Lecture Notes in Physics, Vol. 671, 2005.
12. DEUSCHEL J., STROOCK D. *Large deviations*. – N.Y.: Academic Press, 1989. – 283 p.
13. GRANOVETTER M. *Threshold Models of Collective Behavior* // American Journal of Sociology. – 1978. – Vol. 83. – P. 1420–1443.
14. SCHELLING T. *Micromotives and Macrobehavior*. – N.Y.: WW Norton & Co, 1978. – 272 p.
15. VOLCHENKOV D., LIMA R. *Random Shuffling of Switching Parameters in a Model of Gene Expression Regulatory Network* // Stochastics and Dynamics. – 2005. – Vol. 5(01). – P. 75–95.

ПРИЛОЖЕНИЕ. ОЦЕНКА МАЛОГО ПАРАМЕТРА ГЕОМЕТРИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Оценка малого параметра геометрического распределения.
Изображенные на рис. 6 функции вероятности геометрических распределений с параметром $p \in (0, 1)$ имеют вид

$$(23) P_p(k) = p(1-p)^{k-1}.$$

Задача состоит в построении статистической оценки параметра p по конечной выборке независимых одинаково распределённых случайных величин с распределением (23). Специфика задачи состоит в том, что p предполагается малым ($p \ll 1$).

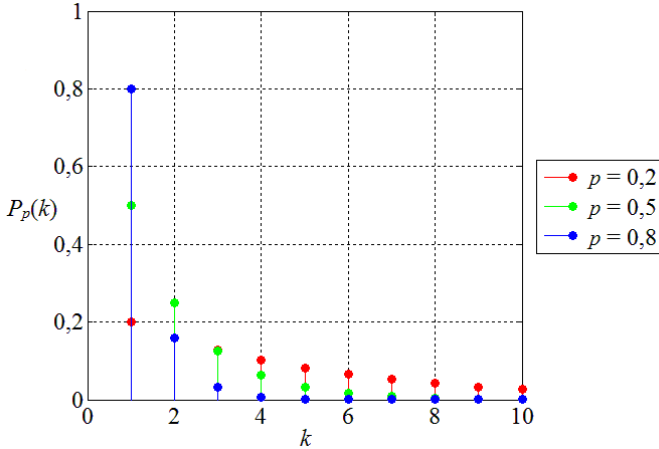


Рис. 6. Функции вероятности семейства распределений (23) при различных p

Найдём оценку параметра p при помощи метода максимального правдоподобия. Запишем определение оценки максимального правдоподобия:

$$(24) \hat{p} = \arg \max_{p \in [0,1]} \prod_{i=1}^n L(p, \xi_i),$$

где ξ_1, \dots, ξ_n - последовательность независимых, одинаково распределённых случайных величин с функцией вероятности (23), а $L(p, k) \equiv P_p(k)$ - функция правдоподобия. Максимизируя произведение функций правдоподобия, получаем

$$(25) \hat{p} = \frac{n}{\sum_{i=1}^n \xi_i} = \frac{1}{\bar{\xi}}, \text{ где } \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i.$$

Найдём математическое ожидание оценки \hat{p} :

$$(26) \begin{aligned} E\hat{p} &\equiv \sum_{m=n}^{+\infty} \frac{n}{m} \mathbf{P} \left(\bar{\xi} = \frac{m}{n} \right) = \\ &= \sum_{m=n}^{+\infty} \frac{n}{m} \binom{m-1}{n-1} p^n (1-p)^{m-n} = \frac{(1-p)^n}{x^n} \sum_{m=n}^{+\infty} \frac{n}{m} \binom{m-1}{n-1} x^m, \end{aligned}$$

где введено обозначение $x = 1 - p$, а $\binom{m-1}{n-1}$ – биномиальный

коэффициент. Для степенного ряда (26) затруднительно найти аналитический предел в элементарных функциях при произвольном n . При $n = 1$ получаем

$$(27) \quad E\hat{p} = \frac{1-x}{x} \sum_{m=1}^{+\infty} \frac{x^m}{m} = \frac{1-x}{x} (-\ln(1-x)) = -\frac{p \ln p}{(1-p)}.$$

Выражение (27) показывает, что оценка максимального правдоподобия при $n = 1$ «очень плохая» для малых p . Действительно, предел отношения

$$(28) \quad \lim_{p \rightarrow 0} \frac{E\hat{p}}{p} = \lim_{p \rightarrow 0} \frac{-\ln p}{1-p} = +\infty$$

равен бесконечности. Случай $n = 1$ является в этом смысле вырожденным и не рассматривается подробно в настоящей работе. Далее под изучением оценок при малых p будет пониматься изучение их предельных свойств типа (28). В качестве иллюстрации найдём математическое ожидание \hat{p} при $n = 2$:

$$\begin{aligned} (29) \quad E\hat{p} &= \frac{2(1-x)^2}{x^2} \sum_{m=2}^{+\infty} \frac{m-1}{m} x^m = \\ &= \frac{2(1-x)^2}{x^2} \left(\sum_{m=2}^{+\infty} x^m - \sum_{m=2}^{+\infty} \frac{1}{m} x^m \right) = \frac{2(1-x)^2}{x^2} \left(x^2 \sum_{m=0}^{+\infty} x^m - \sum_{m=1}^{+\infty} \frac{x^m}{m} + x \right) = \\ &= \frac{2(1-x)^2}{x^2} \left(\frac{x^2}{1-x} - (-\ln(1-x)) + x \right) = \\ &= 2 \left((1-x) + \frac{(1-x)^2 \ln(1-x)}{x^2} + \frac{(1-x)^2}{x} \right) = \\ &= 2 \left(p + \frac{p^2 \ln p}{(1-p)^2} + \frac{p^2}{(1-p)} \right) = \\ &= 2 \left(\frac{p^2 \ln p}{(1-p)^2} + \frac{p}{(1-p)} \right) = 2 \frac{p^2 \ln p + p(1-p)}{(1-p)^2}. \end{aligned}$$

График $E\hat{p}$ при $n = 2$ изображён на рис. 6.

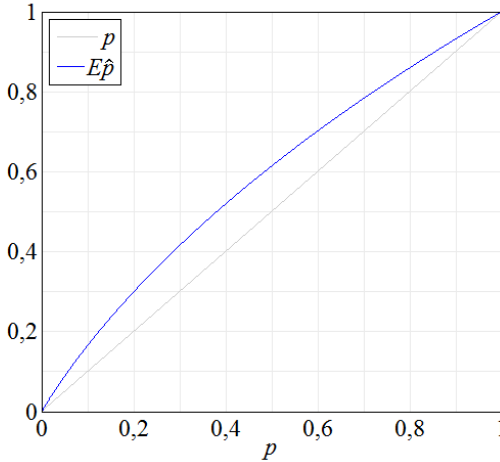


Рис. 7. График математического ожидания оценки максимального правдоподобия \hat{p} при $n = 2$

Как видно из (29) и рис. 7, при $n = 2$ оценка максимального правдоподобия смещённая: при всех $p \in (0, 1)$ математическое ожидание $E\hat{p} > p$. В точках 0 и 1 распределение и математическое ожидание не существуют.

Учитывая тождество $\lim_{p \rightarrow 0} p \ln p = 0$, получаем, что при $n = 2$ для малых p оценка \hat{p} смещена в два раза:

$$\begin{aligned}
 \lim_{p \rightarrow 0} \frac{\hat{p}}{p} &= \lim_{p \rightarrow 0} 2 \frac{p \ln p + (1-p)}{(1-p)^2} = \\
 (30) \quad &= \lim_{p \rightarrow 0} \frac{2p \ln p}{(1-p)^2} + \lim_{p \rightarrow 0} \frac{2}{(1-p)} = 0 + 2 = 2,
 \end{aligned}$$

что можно также записать в виде

$$(31) \quad E\hat{p} = 2p + o(p).$$

Выражение (31) подсказывает вид несмещённой при $n = 2$ для малых p оценки

$$(32) \quad \tilde{p} = \frac{1}{2} \hat{p} = p + o(p),$$

которая имеет в четыре раза меньшую дисперсию, чем оценка максимального правдоподобия \hat{p} . Оценка (32) изображена на рис. 8.

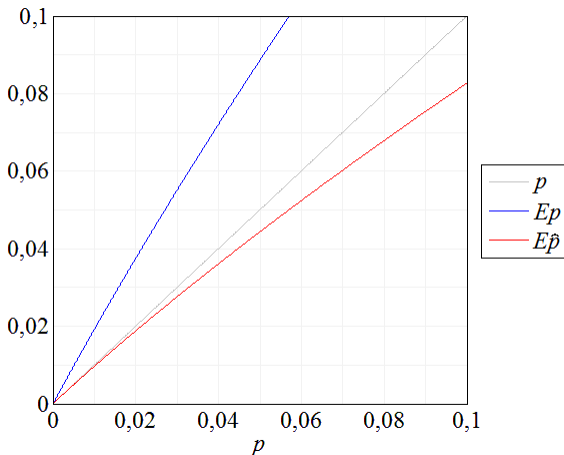


Рис. 8. Математическое ожидание несмещённой оценки \tilde{p} в сравнении с оценкой максимального правдоподобия \hat{p}

Оценка (32) «лучше» оценки максимального правдоподобия, так как имеет меньшую дисперсию и является несмещённой при малых p . Поставим перед собой задачу построения статистической оценки p , оптимальной с точки зрения квадратичного риска:

$$(33) E(\bar{p} - p)^2 \rightarrow \min_{\bar{p}},$$

в классе оценок

$$(34) \tilde{p}(c) = c\hat{p},$$

где $p \in \mathbb{R}$.

Оказывается, что задачу нахождения оптимальной в смысле (33) оценки можно решить, не вычисляя риск в явном виде. Действительно, квадратичный риск можно переписать в виде

$$(35) \begin{aligned} E(\bar{p} - p)^2 &= E(\bar{p} - E\bar{p} + E\bar{p} - p)^2 = \\ &= E(\bar{p} - E\bar{p})^2 + 0 + E(E\bar{p} - p)^2 = \text{Var } \bar{p} + b^2(\bar{p}), \end{aligned}$$

где $b(\bar{p}) \equiv E\bar{p} - p$ – смещение оценки \bar{p} . Равенство (35) носит в англоязычной литературе название *bias-variance decomposition*, а идея уменьшения риска за счёт уменьшения дисперсии при увеличении смещения известна как *bias-variance tradeoff*. Если обозначить дисперсию оценки \tilde{p} через

$$(36) \tilde{\sigma}^2(p) \equiv \text{Var } \tilde{p},$$

то можно выразить дисперсию оценки \bar{p} через $\tilde{\sigma}^2$:

$$(37) \text{Var } \bar{p} = c^2 \text{Var } \tilde{\sigma}^2(p) = c^2 \tilde{\sigma}^2(p).$$

Смещение оценки \bar{p} имеет вид

$$(38) b(\bar{p}) = E\bar{p} - p = cp - p = (c-1)p.$$

Тогда квадратичный риск оказывается равным

$$(39) E(\bar{p} - p)^2 = c^2 \tilde{\sigma}^2(p) + (c-1)^2 p^2.$$

Оптимизируя по c , находим оптимальную константу

$$(40) \bar{c} = \frac{p^2}{p^2 + \tilde{\sigma}^2(p)}.$$

Подставляя (40) в (39) в качестве c , получаем

$$(41) E(\bar{p} - p)^2 = \frac{p^4 \tilde{\sigma}^2(p)}{(p^2 + \tilde{\sigma}^2(p))^2} + \frac{p^2 \tilde{\sigma}^4(p)}{(p^2 + \tilde{\sigma}^2(p))^2} = \\ = \frac{p^2 \tilde{\sigma}^2(p)}{p^2 + \tilde{\sigma}^2(p)} = p^2 - \frac{p^4}{p^2 + \tilde{\sigma}^2(p)}.$$

Можно показать, что $\tilde{\sigma}^2(p) = p^2 \ln p + o(p^2 \ln p)$, а, следовательно, для построенной нами смещённой оценки выполнено

$$(42) E(\bar{p} - p)^2 = p^2 + o(p^2).$$

Оценка для случая $n = 10$ в данной работе была построена аналогичным образом численно.

ESTIMATION OF RARE EVENTS PROBABILITY IN MOB BEHAVIOR

Andrey Rogatkin, Institute of Control Sciences of RAS, Moscow,
PhD student (andreyrogatkin@gmail.com).

Abstract: We develop a collective behavior model proposed in [2]. The model is a generalization of the M. Granovetter's conformity threshold behavior model for the case of uncertainty in agents' threshold values. We estimate the probability of exit of a system from a given set of states (i.e. in mob excitation) using an asymptotic result derived in [8]. Theoretical estimations of large deviations type are refined using numerical estimations obtained with the help of statistical simulations. Obtained results allow us to estimate stability of excitation prevention in the cases when the probabilities of events are too small for using statistical simulations. We give recommendations on the choice of parameters which guarantee a given probability threshold of exit of the system from a given set of states.

Keywords: M. Granovetter's model, conformity collective behavior, mob control, large deviations, statistical simulations.

*Статья представлена к публикации
членом редакционной коллегии П. Ю. Чеботарёвым.*

*Поступила в редакцию 30.03.2016.
Опубликована 30.09.2016.*