

ВВЕДЕНИЕ ЗАВИСИМОСТИ В СХЕМУ ИСПЫТАНИЙ БЕРНУЛЛИ

В.А. Шапцев

профессор, д.т.н., профессор кафедры информационных систем,
e-mail: vashaptsev@ya.ru

Тюменский государственный университет

Аннотация. Предлагается новая структура аналитического представления схемы n независимых испытаний Бернулли посредством упорядоченных подмножеств m ($m < n$) номеров успешных и $n - m$ неуспешных испытаний. При этом вероятностная характеристика типа $P_n(m)$ определяется суммированием биномиальной формулы $p^m(1-p)^{n-m}$ по всем таким подмножествам, а не простым умножением на число сочетаний C_n^m . Приведены соответствующие выражения для вероятности успехов с двумя и $k > 2$ исходами опытов, для случаев равных и разных вероятностей исходов в различных испытаниях. В этой символике получены выражения для расчёта соответствующих вероятностей заданного количества успешных испытаний при наличии зависимости вероятностей исходов от некоторого параметра, изменяющегося во времени (от испытания к испытанию), регулярно или случайно. Таким образом, создан базис для применения схемы зависимых испытаний к оценке вероятностных характеристик сложных стохастических систем, сетевой архитектуры, в частности, характеризующих некоторым изменяющимся показателем качества. В заключении приведено выражение для математического ожидания вероятности m успешных испытаний из n в случае случайного параметра, от которого зависят вероятности исходов. Формулируется гипотеза о возможности применения конечномерной характеристической функции случайного параметра в соответствующем выражении при экспоненциальной зависимости вероятности исхода опыта от случайного параметра.

Ключевые слова: независимые испытания Бернулли, упорядоченные подмножества номеров испытаний, зависимые испытания.

Введение

Работа посвящена представлению схемы независимых испытаний Бернулли упорядоченными подмножествами номеров испытаний с целью ввести в эту схему зависимость между испытаниями, отличную от цепей Маркова. В некоторых технических приложениях это более оправдано. Таким примером является монография [1], где рассматриваемый ниже подход использован в анализе векторов ошибок при передаче дискретных сообщений.

Введём следующие обозначения. $I_n = \{1, \dots, n\}$ — упорядоченное множество номеров испытаний (опытов). $I_m^{(k)} = \{i_1^{(k)}, \dots, i_m^{(k)}\}$ — k -ое упорядоченное подмножество конкретных номеров успешных испытаний. $I_{n-m}^{(k)} = \{j_1^{(k)}, \dots, j_{n-m}^{(k)}\}$ — k -ое упорядоченное подмножество конкретных номеров неуспешных испытаний. Здесь $k \in \{1, \dots, C_n^m\}$, $I_n = I_m^{(k)} + I_{n-m}^{(k)}$.

Ниже эти обозначения распространятся на все возможные модификации схемы независимых испытаний, имеющиеся в литературе (например, [2,3]): различные вероятности альтернативных исходов в испытаниях, несколько возможных исходов с одинаковыми и различными вероятностями их осуществления в испытаниях. Во второй части статьи предлагается введение зависимости между испытаниями посредством введения зависимости вероятностей исходов от некоторого параметра, зависящего от номера испытания регулярно или случайно.

1. Новые обозначения для классических схем независимых испытаний

1.1. Схема Бернулли

В схеме n независимых испытаний Бернулли с двумя исходами и неизменными условиями опытов вероятность m -кратного успеха выражается известной формулой (например, [2]): $P_n(m) = C_n^m p^m (1-p)^{n-m}$, где p — вероятность успеха.

Представим это выражение по-другому, с учётом введённых обозначений:

$$P_n(m) = \sum_{k=1}^{C_n^m} \prod_{i \in I_m^{(k)}} p_i \prod_{j \in I_{n-m}^{(k)}} (1 - p_j). \quad (1)$$

1.2. Разные вероятности успеха в опытах

В случае, когда условия опытов меняются, то есть вероятности успеха в разных испытаниях не равны, когда имеем $p_1, \dots, p_n$, тогда вероятность m -кратного успеха равна коэффициенту при степени z^m аргумента производящей функции $f_n(z) = \prod_{i=1}^n (q_i + p_i z)$, где $q_i = 1 - p_i$ [3, глава 4].

В предложенных выше обозначениях соответствующая вероятность выглядит также, как и (1).

1.3. Несколько возможных результатов испытания с неизменными вероятностями

Если в неизменных условиях проводится n испытаний и в результате каждого испытания возможно s альтернатив с вероятностями p_j , $j = 1, \dots, s$, то вероятность появления m_1 раз 1-го результата, $\dots$, m_s раз s -го результата в n испытаниях равна [2, глава 4]:

$$P_n(m_1, m_2, \dots, m_s) = n! \prod_{j=1}^s p_j^{m_j} / m_j!. \quad (2)$$

В наших обозначениях эта вероятность имеет вид:

$$P_n(m_1, \dots, m_s) = n! \prod_{j=1}^s \left[\prod_{i \in I_{m_j}} p_{ij} \right] / m_j!, \quad (3)$$

где $\sum_{j=1}^s I_{m_j} = I_n$, p_{ij} — вероятность j -го исхода в i -ом испытании.

1.4. Несколько возможных исходов опытов в неоднородных условиях

Если условия опытов различны, т. е. в i -м опыте событие A_j имеет вероятность p_{ij} , $i = 1, \dots, n$; $j = 1, \dots, s$, то $P_n(m_1, \dots, m_s)$ равна коэффициенту при произведении $z^{m_1} \dots z^{m_s}$ в разложении производящей функции $f_n(z) = \prod_{i=1}^n \sum_{j=1}^s p_{ij} z^j$ [2].

Теперь относительно просто можно выразить только вероятность конкретной конфигурации исходов n испытаний:

$$P_n(I_{m_1}, \dots, I_{m_s}) = \prod_{j=1}^s \left[\prod_{i \in I_{m_j}} p_{ij} \right]. \quad (4)$$

Вероятность же

$$P_n(m_1, \dots, m_s) = \sum_{I_{m_1}} \dots \sum_{I_{m_s}} P_n(I_{m_1}, \dots, I_{m_s}). \quad (5)$$

Всего слагаемых в этой многократной сумме равно величине $n! / \prod_{j=1}^{j=s} m_j!$. Суммирование ведётся по всем возможным конфигурациям подмножеств типа I_{m_j} . При этом необходимо помнить, что все входящие в (6) подмножества взаимосвязаны и перебираются синхронно в связке $\sum_{k=1}^s I_{m_k} = I_n$.

Если говорить о практическом применении приведённых выражений, то в контексте развития суперкомпьютеров и экзафлопсных систем, при соответствующей значимости решаемой задачи, их вычислительная сложность не является решающим фактором.

2. Зависимые испытания

Для отображения зависимых испытаний можно воспользоваться некоторой (адекватной реалиям) зависимостью вероятности результата испытания от некоторого параметра μ : $p_i = p(\mu_i) = \varphi(\mu_i)$. Функция $\varphi(\mu)$ может являться

реальным трендом некоторого показателя или параметра системы и быть гипотетическим или экспериментально оцениваемым. Параметр или показатель μ может быть случайным, в том числе случайным процессом. Тогда возникает задача определения математических ожиданий рассматриваемых вероятностей, например, $\mathbf{E}\{P_n(I_m)\} = Q_n(I_m)$. Её решение может иметь как аналитический вид в случае возможности аналитических преобразований, так и оценку методом Монте Карло, если аналитические преобразования частично или полностью невозможны.

Рассмотрим с этих позиций подпункты 1.1–1.4.

2.1. Схема Бернулли

Схема Бернулли рождена для неизменных вероятностей ожидаемого исхода опыта в серии испытаний. Поэтому в контексте зависимости эта схема не имеет смысла. Хотя формула Бернулли (биномиальное распределение) в форме (1) ценна как базовая в ниже следующих рассуждениях.

2.2. Разные вероятности успеха

Из (2), заменив p_i и p_j на $p(\mu_i)$ и $p(\mu_j)$, получаем

$$P_n(m) = \sum_{k=1}^{C_n^m} \prod_{i \in I_m^{(k)}} p(\mu_i) \prod_{j \in I_{n-m}^{(k)}} (1 - p(\mu_j)). \quad (6)$$

При этом

$$\prod_{j \in I_{n-m}^{(k)}} (1 - p(\mu_j)) = 1 + (-1)^{s=1} \sum_{j=1}^{C_{n-m}^1} p(\mu_j) + (-1)^{s=2} \sum_{l=1}^{C_{n-m}^2} \prod_{j \in I_2^{(kl)}} p(\mu_j) + (-1)^{s=3} \dots + \quad (7)$$

$$+ (-1)^{s=n-m-1} \sum_{l=1}^{C_{n-m}^{n-m-1}} \prod_{j \in I_s^{(kl)}} p(\mu_j) + (-1)^{s=n-m} \prod_{j=1}^{C_{n-m}^{n-m}} p(\mu_j) = \quad (8)$$

$$= \sum_{s=0}^{n-m} (-1)^s \sum_{l=0}^{C_{n-m}^s} \prod_{j \in I_s^{(kl)}} p(\mu_j). \quad (9)$$

Таким образом, в условиях зависимости вероятностей успеха при 2-х-альтернативном исходе испытания от значений некоторого параметра μ_i имеем, заменив второй индекс k на s :

$$P_n(m) = \sum_{k=1}^{C_n^m} \prod_{i \in I_m^{(k)}} p(\mu_i) \sum_{s=0}^{n-m} (-1)^s \sum_{l=0}^{C_{n-m}^s} \prod_{j \in I_s^{(kl)}} p(\mu_j). \quad (10)$$

Здесь $I_s^{(kl)}$ — l -е сочетание s элементов подмножества $I_{n-m}^{(k)}$.

2.3. Несколько возможных результатов испытания с неизменными вероятностями

Этот случай, как и п/п 2.1, не имеет смысла рассматривать, так как вероятности в этом параграфе изменяются с изменением параметра μ .

2.4. Несколько возможных исходов опытов в неоднородных условиях

Ограничимся здесь только выражением для вероятности определённой конфигурации результатов n экспериментов. Введя обозначение $p_{ij} = p_j(\mu_i)$, получаем из (4)

$$P_n(I_{m_1}, \dots, I_{m_s}) = \prod_{j=1}^s \left[\prod_{i \in I_{m_j}} p_j(\mu_i) \right]. \quad (11)$$

И, наконец,

$$P_n(m_1, \dots, m_s) = \sum_{\{I_{m_1}\}} \dots \sum_{\{I_{m_s}\}} \prod_{j=1}^s \left[\prod_{i \in I_{m_j}} p_j(\mu_i) \right], \quad (12)$$

где, в частности, $\{I_{m_2}\}$ — подмножество сочетаний 2-х номеров испытаний, соответствующих определённому исходу, мощностью C_n^2 .

Заметим, что нередко именно конкретная конфигурация исходов в последовательности испытаний (появление неисправности определённых компонентов системы) имеет смысл и интерес как с инженерной, так и с мониторинговой точек зрения. Математическое ожидание соответствующих вероятностей, типа (2.4)...(2.7), имеет значение в оценке надёжностных характеристик и эксплуатационной готовности при сравнении технических решений в сложных и больших системах.

2.5. Стохастический параметр, сопровождающий испытания

Рассмотрим случай, когда параметр, от которого зависят вероятности исходов испытаний, имеет случайный характер и вектор $\mu = (\mu_1, \dots, \mu_n)$ имеет конечномерную функцию или конечномерную плотность распределения вероятностей. Тогда исследователя может интересовать математическое ожидание $E\{\dots\}$ вероятностей (7), (8) или (10). Ограничимся здесь иллюстрацией этой операции на примере выражения (7):

$$E\{P_n(m)\} = \sum_{k=1}^{C_n^m} \sum_{s=0}^{n-m} \sum_{l=0}^{C_{n-m}^s} E\left\{ \prod_{i \in I_m^{(k)}} \prod_{j \in I_s^{(kl)}} p(\mu_i) p(\mu_j) \right\}. \quad (13)$$

Если вероятности зависят от параметра экспоненциально, то математическое ожидание в (11) становится конечномерной характеристической функцией случайного параметра. Если вероятности зависят от параметра, как степени целого показателя, то мы будем иметь дело со смешанными моментами случайного параметра. Рассмотрение этих и других вариантов зависимости вероятностей исходов испытаний от некоторого показателя системы, моделируемой схемой последовательности испытаний, может являться предметом дальнейших исследований.

Заключение

Сферами применения результатов пункта 2 являются следующие прикладные задачи. Надёжность сетевых объектов, отображаемых графами со случайными рёбрами: сети связи, сети передачи данных, сеть трубопроводов, сеть электрических проводов и так далее. Возможно применение этих выражений при оценке качества передачи данных по каналам связи со случайными параметрами, вероятности безошибочной передачи сообщений в командных системах. Автор намерен получить на базе этих результатов конечные выражения для математических ожиданий при некоторых конкретных зависимостях $p(\mu_i)$ и случайном параметре μ .

ЛИТЕРАТУРА

1. Коржик В.И., Финк Л.М. Кодирование дискретных сообщений в каналах с переменной структурой. М. : Связь, 1975. 250 с.
2. Гихман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика. Киев : Высшая школа, 1979. 408 с.
3. Вентцель Е.С., Овчаров Л.Ф. Задачи и упражнения по теории вероятностей. М. : АКАДЕМИЯ, 2003. 448 с.
4. Райгородский А.М. Модели случайных графов. М. : МЦНМО, 2011. 136 с.

ABOUT DEPENDENCE IN THE BERNOULLI TRIALS**V.F. Shaptsev**

Dr.Sc. (Phys.-Math.), Professor, e-mail: vashaptsev@ya.ru

Tyumen State University

Abstract. Analytical representation of scheme is discussed in independent Bernoulli trials by means of the ordered subsets of m ($m < n$) numbers of successful and $n - m$ numbers of failed tests. The probabilistic description of $P_n(m)$ is determined by summing the binomial formula $p^m(1 - p)^{n-m}$ for all such subsets rather than a simple multiplication by the number of combinations C_n^m . Some appropriate expressions are given for the probability of success with $k > 2$ outcome of the experiments for the cases of equal and different probabilities of outcomes in different trials. In these symbolics, expressions for calculating the corresponding probabilities of the number of successful tests is specified in the presence of dependence of outcomes probabilities on a parameter varying in time (from test to test), regularly or accidentally. Thus, the basis is created for application of the scheme of dependent tests to estimate the probabilistic characteristics of complex stochastic systems, network architectures, in particular, characterized by some varying measure of quality. In conclusion, we present the expression for the mathematical expectation of the probability of m successful trials out of n in the case of random parameter on which the probabilities of outcomes depend. Hypothesis about the possibility of using finite-dimensional characteristic function of the random parameter in the corresponding expression for the case of exponential dependence of the probability of the outcome from the random parameter is written.

Keywords: independent Bernoulli trials, ordered subsets of tests numbers, dependent tests.

Дата поступления в редакцию: 11.02.2017