Оригинальная статья
https://doi.org/10.21683/1729-2646-2023-23-1-4-12

Нормирование количества отказов сложного объекта с применением мультиномиального распределения

Multinomial distribution as part of valuation of the number of failures

Новожилов Е.О.[1]
Novozhilov E.O.[1]
[1]АО «НИИАС», Москва, Российская Федерация
[1]JSC NIIAS, Moscow, Russian Federation
evg_o_nov@mail.ru

Новожилов Е.О.

Резюме. Цель. *Целью статьи является исследование метода нормирования количества отказов сложного объекта с применением мультиномиального распределения. Данный подход предусматривает установление допустимого значения количества отказов объекта на основе «прошлого опыта» (по статистической выборке количества отказов объекта, накопленной за несколько предыдущих оценочных периодов времени).* **Методы.** *В статье применяются методы системного анализа, теории вероятностей, математической статистики. Проанализированы первичные показатели, на основе которых формируются показатели надежности объекта. Отмечена перспективность применения нормирования показателей надежности на основе имеющихся статистических данных применительно к сложным объектам. Рассмотрена задача нормирования количества отказов на основе статистической выборки. Определены основные недостатки применяемых подходов, связанные с ошибками определения среднего значения ряда, коэффициентов вариации и асимметрии. Показана возможность решения данной задачи с применением известной в комбинаторике задачи «о шарах и ящиках», которая сводится к мультиномиальному распределению. Рассмотрено определение вероятностей для композиций и разбиений числа n на m частей, а также вероятностей нахождения заданного количества шаров в ящике с их максимальным количеством. Рассмотрены формулы и алгоритмы, реализующие наиболее эффективный (с точки зрения объема/времени машинных вычислений) расчет вероятностей распределения мультиномиального максимума. Оценена возможность аппроксимации дискретной функции распределения мультиномиального максимума распределением Гумбеля. Показана возможность нормирования количества отказов для «сегмента», соответствующего части (доле) размерности сложного объекта, рассматриваемой на определенной части (доле) оценочного интервала времени от общего времени, на котором собрана статистика. Рассмотрены примеры нормирования количества отказов для объекта в целом на интервале оценки 1 месяц и для объекта на интервале оценки 12 месяцев при оценочном интервале, на котором накоплена статистическая выборка, 72 месяца. Приведены ограничения по применению представленного метода и отмечены некоторые его возможные преимущества. В частности отмечено, что конкретная статистическая выборка, отражающая распределение количества отказов по нескольким одинаковым интервалам времени, является лишь одной реализацией мультиномиального распределения, поэтому можно сказать, что при применении предлагаемого метода на результаты нормирования перестает оказывать влияние наличие в статистической выборке маловероятного сочетания значений ряда. Также отмечено, что при применении предложенного метода нормирования количества отказов полученное допустимое значение всегда будет выше среднего значения статистической выборки.* **Результаты.** *Получены выражения для расчета дискретной плотности и функции распределения мультиномиального максимума на основе разбиений числа. Приведены результаты анализа алгоритмов для проведения машинных вычислений. Представлены результаты применения некоторых алгоритмов. Предложена формула для аппроксимации функции распределения мультиномиального максимума с помощью распределения Гумбеля (для наибольших значений) методом моментов. Рекомендован диапазон значений оценочного интервала, в котором предложенный метод обеспечивает приемлемую достоверность результатов. Определены задачи дальнейших исследований.*

Abstract. Aim. *The paper aims to examine the application of a multinomial distribution as part of valuation of the number of an object's failures. It is assumed that the valuation is "based on past experience" (a statistical sample of the number of an object's failures accumulated over several preceding evaluation periods).* **Methods.** *The paper uses methods of system analysis, probability theory and mathematical statistics. The author analyses the primary indicators used to define the applied dependability indicators. It is noted that the valuation of such indicators* *based on statistical data for complex systems appears to be promising. The problem of valuation of the number of failures using a statistical sample is examined. The primary disadvantages of the used approaches are identified that are associated with errors in defining the average values of series, variation coefficients and asymmetry. It is shown that it is possible to solve the problem using the well-known combinatorics problem "on balls and boxes", which leads to the use of a multinomial distribution. The paper examined the definition of probabilities for compositions and partitions of the number n into m parts, as well as the probabilities of a given number of balls being in a box with their maximum number. The author also considered formulas and algorithms that allow reducing the number of calculations in case of machine computation of the probabilities of a multinomial distribution. The feasibility of approximating a discrete distribution function by the Gumbel distribution is estimated. The paper demonstrates the feasibility of valuating the number of failures for a "segment" corresponding to a part (fraction) of an object's dimension considered on a certain part (fraction) of the time interval. It also examines examples of valuating the number of failures for an object as a whole over a 1-month evaluation interval and for of an object over a 12-month evaluation interval, while the total interval for which the statistical sample is accumulated is 72 months. The paper sets forth limitations on the application of the presented method and notes some of its possible advantages. In particular, it is noted that the statistical sample is only one implementation of the multinomial distribution, so it can be said that when applying the proposed method, the results of valuation are no longer affected by the presence of unlikely combinations of series values in the statistical sample. It is also noted that when applying the proposed method of valuating the number of failures, the obtained acceptable value will never be less than or equal to the average value of the statistical sample.* **Results.** *Formulas have been obtained for calculating, based on partitions, of the discrete density number and the maximum distribution function of a multinomial distribution. The paper presents the results of algorithm analysis for machine computation. The results of applying some algorithms are presented. A formula is proposed for approximating the distribution function of the maximum of a multinomial distribution using the Gumbel distribution (for the largest values) using the method of moments. The author recommends a range of values of the estimated interval, in which the proposed method provides acceptable reliability of the results. The task of further research is defined.*

Ключевые слова: *нормирование надежности, мультиномиальное распределение, плотность распределения вероятности, функция распределения вероятности, доверительная вероятность.*
Keywords: *multinomial distribution, dependability valuation, density of probability distribution, probability distribution function, confidence probability.*

Для цитирования: *Новожилов Е.О. Нормирование количества отказов сложного объекта с применением мультиномиального распределения // Надежность. 2023. № 1. С. 4-12. https://doi.org/10.21683/1729-2646-2023-23-1-4-12*
For citation: *Novozhilov E.O. Multinomial distribution as part of valuation of the number of failures. Dependability 2023;1:4-12. https://doi.org/10.21683/1729-2646-2023-23-1-4-12*

Поступила *03.10.2022* / **После доработки** *05.12.2022* / **К печати** *14.03.2023*
Received on: *03.10.2022*/ **Revised on:** *05.12.2022* / **For printing:** *14.03.2023*

Введение

Для нормирования показателей надежности в мировой практике известны подходы на основе следующих критериев [1]:

а) на основе экономических оценок;

б) на базе экспериментальных исследовательских расчетов;

в) на основе прошлого опыта (анализа фактических данных);

г) на основе опросов потребителей.

Для сложных объектов (технических систем), включающих как приобретаемые, так и создаваемые строительством подсистемы, взаимодействующие сложным образом, применение экономических оценок и исследовательских расчетов затруднено из-за трудности построения соответствующих моделей. Опрос потребителей (пользователей услуг) системы не всегда позволяет сформировать четкое понимание требований к ее надежности. Путь, заключающийся в применении накопленных фактических данных о работе системы в течение предшествующих периодов, в ряде случаев остается единственной возможностью реализовать нормирование показателей надежности для данной системы.

Наиболее распространенными первичными показателями, составляющими статистику работы объекта, как правило, являются количество отказов и время до восстановления после каждого из них (либо суммарное время до восстановления), фиксируемые за определенный оценочный интервал времени (например, за месяц или за год). Такие данные обычно собираются за несколько лет наблюдений для обеспечения их репрезентативности.

Поскольку такие широко применяемые на практике показатели, как интенсивность отказов, среднее время до восстановления, коэффициент готовности объекта в большинстве случаев могут быть получены на основе количества отказов и суммарного времени до восстановления, то для таких случаев достаточно рассмотреть установление нормируемых значений для этих двух первичных показателей. В данной работе рассматривается один из возможных подходов к нормированию количества отказов объекта. Поскольку количество отказов является целым числом, задача его нормирования может быть рассмотрена с точки зрения дискретной математики.

Рассмотрим задачу определения нормируемого (допустимого) количества $n_{доп}$ отказов объекта за заданный интервал оценки T при условии, что известен репрезентативный статистический ряд $N_1, N_2, \ldots, N_M$, включающий M значений количества отказов (для каждого из M последовательных интервалов времени длительностью$\cdot T$). При этом допустим, что непревышение установленного допустимого значение $n_{доп}$ должно обеспечиваться с заданной доверительной вероятностью α.

Один из классических способов решения данной задачи состоит в построении так называемой эмпирической функции обеспеченности, которая затем аппроксимируется некоторым теоретическим распределением вероятности. Далее определяется соответствующий квантиль этого распределения – значение, непревышение которого обеспечивается с заданной вероятностью α. Это значение принимается как пороговое (нормируемое) значение показателя (в данном случае $n_{доп}$ – допустимое количество отказов объекта за интервал T). Такой подход применяется, например, в гидрологических расчетах при определении норм годового стока расчетной обеспеченности (вероятности превышения) [2].

Недостатком рассмотренного способа является то, что на сравнительно коротких временных рядах, имеющих значительный (более 0,3) коэффициент вариации, подбор теоретического распределения затруднен, в результате чего нормируемый показатель со значительной погрешностью отражает реальную (в данном рассмотрении – верхнюю) границу диапазона значений фактического показателя. Так, для 10 лет наблюдения при коэффициенте вариации ряда, равном 0,2, средняя квадратическая ошибка его определения составит 5%, а при коэффициенте вариации ряда 0,5 ошибка составит уже 15% [2]. Для более коротких рядов ошибка будет еще хуже. А если при подборе теоретического распределения использовать не только коэффициент вариации, но и коэффициент асимметрии ряда – даже при сравнительно длинных рядах ошибка вычисления последнего получается значительной.

Другим недостатком рассмотренного способа является то, что если необходимо установить допустимое значение $n_{доп}$ для оценочного интервала другой длительности ($T_1 \neq T$), то нужно будет иметь статистический ряд из M значений (для каждого из M последовательных интервалов времени длительностью$\cdot T_1$) и описанные выше процедуры построения эмпирической функции обеспеченности, подбора теоретического распределения и т.д. выполнить заново.

Постановка задачи

Поиск альтернативных решений по установлению допустимого количества $n_{доп}$ отказов объекта за заданный интервал оценки T привел к рассмотрению следующей задачи.

Пусть существует M одинаковых интервалов времени длительностью T (при этом M достаточно большое, чтобы статистический ряд $N_1, N_2, \ldots, N_M$ количества отказов объекта можно было считать репрезентативным). Количество отказов за суммарный интервал времени $T_S = M{\cdot}T$ равно $N = \sum_{i=1}^{M} N_i$. Будем считать, что отказ является случайным событием, его возникновение не зависит от возникновения других отказов и вероятность возникновения отказа на любом из M интервалов одинакова. Вопрос заключается в том, каким образом N отказов могут распределиться по M одинаковым интервалам длительностью T ? Очевидно, что исходный статистический ряд значений $N_1, N_2, \ldots, N_M$ является одной из возможных реализаций такого распределения количества отказов.

Данная задача известна в комбинаторике как «задача о шарах и перегородках»[1] («задача о шарах и ящиках», «balls-and-boxes» и т.п.): сколькими способами можно разложить n шаров в m ящиков (допуская при этом пустые ящики)? Количество таких способов (комбинаций) равно числу сочетаний C_{n+m-1}^{m-1}, а собственно комбинации являются композициями[2] числа n на m частей. Приведем пример композиций числа 4 на 3 части. Общее количество композиций равно $\frac{6!}{2!\cdot 4!} = 15$, а сами композиции следующие:

4;0;0, 0;4;0, 0;0;4,
3;1;0, 3;0;1, 1;3;0,
1;0;3, 0;3;1, 0;1;3,
2;2;0, 2;0;2, 0;2;2,
2;1;1, 1;2;1, 1;1;2.

Вероятность появления каждой из комбинаций (каждой композиции) при случайном распределении шаров по ящикам описывается классической формулой мультиномиального распределения ([3], [4] и др.):

$$P\{x_1 = n_1, x_2 = n_2, ..., x_m = n_m\} = \\ = \frac{n!}{n_1!\cdot n_2!\cdot ...\cdot n_m!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot ...\cdot p_m^{n_m}, \quad (1)$$

где $n_1, n_2, \ldots, n_m$ – любые целые неотрицательные числа, подчиненные условию $\sum_{j=1}^{m} n_j = n$; $p_1, p_2, \ldots, p_m$ – вероятности попадания шара в ящик 1, 2, …, m соответственно, такие, что $\sum_{j=1}^{m} p_j = 1$.

[1] https://ru.wikipedia.org/wiki/Метод_шаров_и_перегородок
[2] https://ru.wikipedia.org/wiki/Композиция_числа

Выражение (1) определяет вероятность того, что после случайного размещения n шаров по m ящикам в ящике 1 оказалось n_1 шаров, в ящике 2 оказалось n_2 шаров и т.д.

Формула (1) учитывает, что вероятности попадания шара в ящики могут отличаться. В нашей задаче все вероятности $p_1, p_2, \ldots, p_m$ одинаковы (это соответствует самому применяемому случаю – равновероятному мультиномиальному распределению).

Здесь следует отметить, что представление способов размещения n шаров по m ящикам в виде композиций при равных вероятностях не является удобным для практического применения. Так, если вероятности одинаковы, то ящики «равноправны» между собой и их порядок следования не имеет значения. В то же время, приведенный выше список композиций включает в себя композиции, образованные перестановкой одних и тех же элементов – например, 4;0;0, 0;4;0 и 0;0;4. В частности, в рассматриваемой задаче нас интересуют сами слагаемые, но не их порядок, поэтому указанные 3 композиции будут означать одно и то же.

Другим вариантом представления способов размещения n шаров по m ящикам является разбиение[1] числа n на m слагаемых (включая нулевые). В отличие от композиции, разбиение не учитывает порядок слагаемых и представляет собой композиции, упорядоченные по невозрастанию значений, при этом повторяющиеся комбинации исключаются. Общее количество разбиений числа n на m слагаемых определяется известной формулой Эйлера.

Приведем пример разбиений числа 4 на 3 части:

4;0;0,
3;1;0,
2;2;0,
2;1;1.

Очевидно, что количество комбинаций при переходе от композиций к разбиениям заметно сократилось (с 15 до 4), для больших n и m «компактность» этого варианта записи возможных комбинаций имеет еще более явное преимущество. Но для применения разбиений вместо композиций в формуле (1) следует учесть исключение комбинаций, образованных перестановками элементов. Примем в (1) все вероятности одинаковыми и равными $1/m$, а также учтем исключаемые комбинации, тогда (1) изменится следующим образом:

$$P\left\{x_1 = n_1, x_2 = n_2, \ldots, x_m = n_m\right\} = \\ = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!} \cdot \frac{m!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!} \cdot \frac{1}{m^n}, \qquad (2)$$

где $n_1, n_2, \ldots, n_m$ – любые целые неотрицательные числа, подчиненные условиям $\sum_{j=1}^{m} n_j = n$ и $n_1 \geq n_2 \geq \ldots \geq n_m$; $k \leq m$ – количество отличающихся значений среди чисел $n_1, n_2, \ldots, n_m$; r_j – количество повторений (вхождений) числа с j-м значением ($j = 1 \ldots k$).

[1] https://ru.wikipedia.org/wiki/Разбиение_числа

Выражение (2) определяет вероятность того, что после случайного размещения n шаров по m ящикам *и сортировки ящиков по невозрастанию количества шаров* в ящике 1 оказалось n_1 шаров, в ящике 2 оказалось n_2 шаров и т.д. Далее будем рассматривать распределение n шаров по m ящикам на основе разбиений n на m (упорядоченных по невозрастанию значений последовательностей, без учета перестановок их элементов).

Для примера вычислим вероятность для разбиений 2;1;1 и 2;2;0 числа 4 на 3 части ($n = 4$, $m = 3$; $k = 2$; $r_1 = 1$; $r_2 = 2$ и $k = 2$; $r_1 = 2$; $r_2 = 1$):

$$P\left\{x_1 = 2, x_2 = 1, x_3 = 1\right\} = \\ = \frac{4!}{2! \cdot 1! \cdot 1!} \cdot \frac{3!}{1! \cdot 2!} \cdot \frac{1}{3^4} = \frac{12 \cdot 3}{81} = \frac{4}{9} \approx 0,444444; \qquad (3)$$

$$P\left\{x_1 = 2, x_2 = 2, x_3 = 0\right\} = \\ = \frac{4!}{2! \cdot 2! \cdot 0!} \cdot \frac{3!}{2! \cdot 1!} \cdot \frac{1}{3^4} = \frac{6 \cdot 3}{81} = \frac{2}{9} \approx 0,222222. \qquad (4)$$

Для целей нормирования (как установления верхнего порогового значения) представляет интерес количество n_1 шаров в первом ящике. То есть интересует поведение максимума мультиномиального распределения (его вероятностное распределение).

Мультиномиальное распределение предполагает, что $n > 1$ и $m > 1$ (при $m = 2$ имеем частный случай – биномиальное распределение). Определим границы изменения n_1. Минимальное значение, которое может принимать n_1 для заданных n и m: $n_0 = \left\lceil \frac{n}{m} \right\rceil$, где оператор $\lceil \cdot \rceil$ означает округление вверх до ближайшего целого. Максимальное значение, которое принимает n_1 для заданных n и m, равно n. Таким образом,

$$n_0 = \left\lceil \frac{n}{m} \right\rceil \leq n_1 \leq n.$$

Если в первом ящике находится v шаров ($n_0 \leq v \leq n$), то при этом гарантируется, что в любом другом ящике количество шаров будет не больше, чем в первом (то есть меньше либо равно v). Применительно к распределению $N = \sum_{i=1}^{M} N_i$ отказов по M одинаковым интервалам времени (например, по 12 месяцам года) это означает, что если в «наихудшем» интервале с заданной вероятностью произошло не более v отказов, то в любом другом их количество также не превысит v (то есть можно применить значение v как пороговое значение для допустимого количества отказов, но учитывая, что каждое из значений v обеспечивается с определенной вероятностью). Отметим, что такой подход подразумевает постоянную интенсивность отказов объекта в течение всех последовательных M интервалов времени и отсутствие влияния восстановления после текущего отказа на возникновение последующего отказа. Применительно к простым объектам такой подход, видимо, плохо применим, поскольку со временем из-за старения интенсивность отказов возрастает, а также если отказ возник, то следующий отказ может произойти только после восстановления текущего. Однако для сложных объектов, части которых заменяются на новые постепенно (например, при проведении плановых ремонтов), можно допустить, что интенсивность отказов на определенном интервале времени является постоянной, а также в виду наличия частично работоспособных состояний следующий отказ может возникнуть ранее восстановления после текущего отказа.

Найдем вероятность того, что в первом ящике находится ровно v шаров. Очевидно, что для этого потребуется сложить вероятности для всех разбиений, у которых $n_1 = v$:

$$P\{x_1 = v\} = \sum_i \frac{n!}{n_{1i}! \cdot n_{2i}! \cdot \ldots \cdot n_{mi}!} \cdot \frac{m!}{r_{1i}! \cdot r_{2i}! \cdot \ldots \cdot r_{ki}!} \cdot \frac{1}{m^n}, \quad (5)$$

где суммирование по i выполняется для всех разбиений, у которых $n_1 = v$.

Например, для рассмотренных выше разбиений числа 4 на 3 части, чтобы получить вероятность того, что в первом ящике находится 2 шара, следует объединить результаты вычислений по формулам (3) и (4):

$$P\{x_1 = 2\} =$$
$$= P\{x_1 = 2, x_2 = 1, x_3 = 1\} + P\{x_1 = 2, x_2 = 2, x_3 = 0\} =$$
$$= \frac{2}{9} + \frac{4}{9} = \frac{6}{9} \approx 0{,}666667.$$

Для задачи установления допустимого значения представляет интерес вероятность того, что в первом ящике находится не более v шаров. По аналогии с формулой (5):

$$P\{x_1 \le v\} = \sum_i \frac{n!}{n_{1i}! \cdot n_{2i}! \cdot \ldots \cdot n_{mi}!} \cdot \frac{m!}{r_{1i}! \cdot r_{2i}! \cdot \ldots \cdot r_{ki}!} \cdot \frac{1}{m^n}, \quad (6)$$

где суммирование по i выполняется для всех разбиений, у которых $n_1 \le v$.

Выражение (6) можно представить в виде

$$P\{x_1 \le n_1\} = P\{x_1 = n_0\} + P\{x_1 = n_0 + 1\} + \ldots + P\{x_1 = n_1\} =$$
$$= \sum_{i=0}^{v-n_0} P\{x_1 = n_0 + i\},$$

где вероятность $P\{x_1=n_0+i\}$ вычисляется по (5).

Алгоритмы вычисления вероятностей

Очевидно, что самый простой путь вычисления вероятностей $P\{x_1=v\}$ (то есть дискретной плотности распределения вероятностей мультиномиального максимума) или $P\{x_1 \le v\}$ (то есть дискретной функции распределения вероятностей мультиномиального максимума) для всех v из диапазона $n_0 \le v \le n$ заключается в синтезе разбиений n на m и расчета вероятностей по формулам (5) или (6). Реализация такого варианта на ЭВМ с 32-разрядной арифметикой работает для небольших чисел n и m (не более 170), что связано с необходимостью вычисления факториалов n и m. Но, помимо этой проблемы, при росте n и m резко возрастает время вычислений (даже на 64-разрядной ЭВМ), поскольку количество разбиений увеличивается нелинейно и стремительно: так, например, если $n = m = 10$, то количество разбиений равно 42, если $n = m = 100$, количество разбиений равно 190 569 292, а при $n = m = 1000$ имеем уже более $2{,}4 \cdot 10^{31}$ разбиений. Поскольку в реальных условиях количество отказов сложного объекта за несколько лет наблюдений может составлять десятки и даже сотни тысяч, в рассматриваемой задаче нормирования количества отказов особую актуальность приобретает поиск эффективных (по объему/времени вычислений) алгоритмов для вычисления распределения вероятности мультиномиального максимума.

Описание таких алгоритмов встречаются в следующих работах:

- алгоритм Раппепорта (Rappeport) [5], использующий итеративные вычисления;
- формулы Хайтера (Hayter) [6], основанные на рекурсивных вычислениях;
- аппроксимация Дасгупты (DasGupta) [7], выполненная на основе [8, § 6, теор. 3];
- алгоритм Эвенса (Ewens) и Вилфа (Wilf) [9], использующий аппроксимацию Пуассона, степенные ряды и рекурсивные формулы;
- алгоритм Левина (Levin) [10], применяющий усеченное распределение Пуассона и приближение Эйджворта.

Отметим, что только первый из перечисленных алгоритмов получен путем оптимизации структуры вычислений, а остальные так или иначе основаны на связи мультиномиального распределения с распределением Пуассона.

Из этих алгоритмов оригинальным решением, безусловно, является формула вычисления квантиля функции распределения вероятности мультиномиального максимума, предложенная Дасгуптой [7], в основе которой лежит аппроксимация распределением Гумбеля. Но, как показал проведенный анализ, эта формула обеспечивает приемлемую точность только для очень больших отношений $n / m > 400$.

Для компьютерной реализации лучшим из представленных алгоритмов является алгоритм Эвенса и Вилфа [9]. При определенных условиях он обеспечивает вычисление вероятностей без потери точности, а также показывает приемлемое время вычислений даже при больших n и m (до 50 000).

Автором настоящей статьи был реализован машинный алгоритм для вычисления функции распределения вероятности мультиномиального максимума в соответствии с [9] в программной среде VB6/VBA. Для работы с «длинными числами» применена внешняя свободно распространяемая библиотека BedvitCOM(x86).dll версии 2.0.0.0 [11] (автор статьи выражает благодарность разработчику этой библиотеки Виталию Бедному за корректировки кода, благодаря которым удалось реализовать рассматриваемый алгоритм расчета). Для выигрыша во времени при реализации машинного алгоритма была введена опция прекращения вычислений при достижении функцией распределения значения 0,9999 (поскольку эта функция имеет длинный «хвост» в области больших значений n_1, тратить время на вычисление вероятностей которого нет смысла). Например, при n = 6400 и m = 100 время расчета функции распределения составило 446 с на вычислительной машине с процессором Intel®Core™2Duo CPU E4600, имеющим 2 ядра с тактовой частотой 2,4 ГГц (операционная система Microsoft Windows 7 Professional x86).

Однако нужно отметить следующие особенности алгоритма [9]:

- алгоритм вычисляет дискретную функцию распределения вероятности мультиномиального максимума (а не плотность распределения, как это делают некоторые другие алгоритмы), при этом в алгоритме не учитывается смещение величины n_1, необходимое при переходе от дискретной плотности к дискретной функции распределения;
- в процессе вычислений выполняется суммирование элементов ряда, значения которых могут иметь значительно различающиеся (в десятки и сотни тысяч раз) порядки величин (степень различия зависит от конкретных значений n и m). В связи с этим для получения корректных результатов (отсутствия «исчезновения порядка» при суммировании) может потребоваться применение длинных чисел с большим количеством разрядов – десятки тысяч бит и более. В частности, для того, чтобы без ошибки вычислить все дискретные значения вероятностей функции распределения мультиномиального максимума с параметрами n = 12288 и m = 12, потребовалось применение чисел длиной 26 368 бит. На эту проблему обратили внимание в работе [12] ее авторы.

Для случаев, когда n и m имеют порядок сотен тысяч – десятков миллионов, применимым является алгоритм [10], который с незначительной потерей точности позволяет вычислить функцию распределения мультиномиального максимума (кроме «хвостов» распределения, особенно левого, где точность не обеспечивается; в диапазоне вероятностей от 0,01 до 0,99 точность составляет 2-3 знака после запятой). К тому же, этот алгоритм имеет более высокую скорость по сравнению с алгоритмом [9]. Для машинной реализации алгоритма [10] на 32-разрядной архитектуре также потребуется библиотека длинных чисел, но при этом будет достаточно 64 разрядов (как показал эксперимент, увеличение количества разрядов выше этого значения практически не влияет на точность результатов).

Для сравнения алгоритмов [7] и [9] на рис. 1 показан пример графиков функций распределения $F(n_1)$ для случая n = 20 000 и m = 100, где

- штриховая линия построена по формуле из [7];
- точки построены по результатам выполнения расчетов по алгоритму [9], которые являются *точными значениями* (здесь абсциссы точек смещены на 0,5 вправо, поскольку ряд целых значений n_1 соответствует дискретной плотности распределения и при переходе к функции распределения данные значения должны располагаться посередине столбиков диаграммы, площадь которых суммируется; правая граница столбиков соответствует n_1+0,5);
- сплошная линия построена путем аппроксимации результатов выполнения расчетов по алгоритму [9] распределением Гумбеля, выполненной методом моментов по следующей формуле:

$$\nu(\alpha) = B - \frac{A}{\sigma}\left[\log\left(\log\left(\frac{1}{\alpha}\right)\right) + \mu\right], \qquad (7)$$

где ν(α) – квантиль функции распределения мультиномиального максимума; α – доверительная вероятность; B, A – соответственно математическое ожидание и среднеквадратическое отклонение функции распределения мультиномиального максимума с параметрами n, m (значения A и B получают расчетным путем – из результатов вычисления дискретной функции распределения вероятности мультиномиального максимума по алгоритму [9]); μ, σ – соответственно математическое ожидание и среднеквадратическое отклонение распределения Гумбеля для наибольших значений (с параметром масштаба $a = 1$ и смещением $u = 0$, при этом $\mu \cong 0{,}577216$, $\sigma = \pi/6 \cong 1{,}28255$). Для заданных выше n и m: B = 236,35, A = 6,2532.

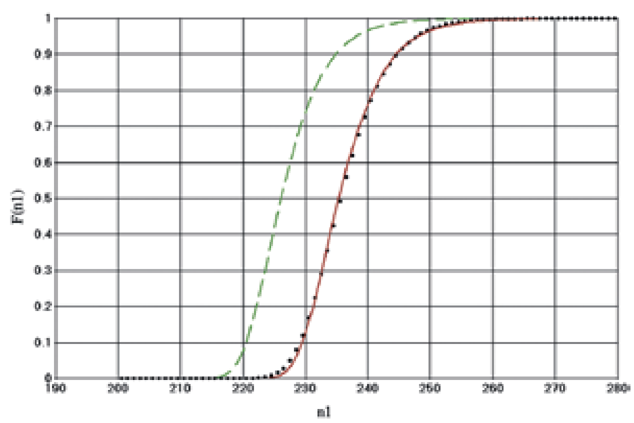

Рис. 1. Графики функций распределения $F(n_1)$, полученных по различным алгоритмам

Пример применения метода нормирования

Для примера некоторые значения аппроксимации распределением Гумбеля функции распределения вероятности мультиномиального максимума для n = 20 000 и m = 100, рассмотренной выше, приведены в табл. 1. Следует отметить, что предложенная в настоящей работе формула (7), применяющая метод моментов, дает хорошее приближение только при $n \geq m$ (случай $n < m$ требует дополнительного анализа).

Табл. 1. Пример результатов применения формулы (7)

Значение n_1 (количество шаров в первом ящике)	Вероятность
240,85	0,8
242,4	0,85
244,5	0,9
248	0,95
256	0,99

С точки зрения нормирования (установления допустимого значения) количества отказов, данные результаты (см. табл. 1) можно трактовать следующим образом. Объект имел $N = \sum_{i=1}^{M} N_i = 20\,000$ отказов за интервал $T_S = 100$ месяцев. Если задана требуемая обеспеченность допустимого значения количества отказов $\alpha = 0{,}9$, то, в соответствии с табл. 1, допустимое значение $n_{\text{доп}}$ количества отказов объекта для оценочного интервала 1 месяц следует принять равным 244,5.

При заданном T_S дискретность M накладывает определенные ограничения на выбор оценочного интервала T. А случай $N/M < 1$ («The Sparse Case» [7]), вообще заслуживает отдельного рассмотрения. В связи с этим для практики можно рекомендовать следующий диапазон оценочного интервала T:

$$\frac{T_S}{N} < T \le \frac{T_S}{5...10}.$$

Следует подчеркнуть, что рассмотренный метод установления допустимого количества отказов для оценочного интервала T как части интервала времени T_S будет справедливым и для части сложного объекта (при условии его «однородности» в отношении количества отказов; например, если есть железнодорожный путь общей протяженностью $L_S = 300$ км, причем на этой протяженности

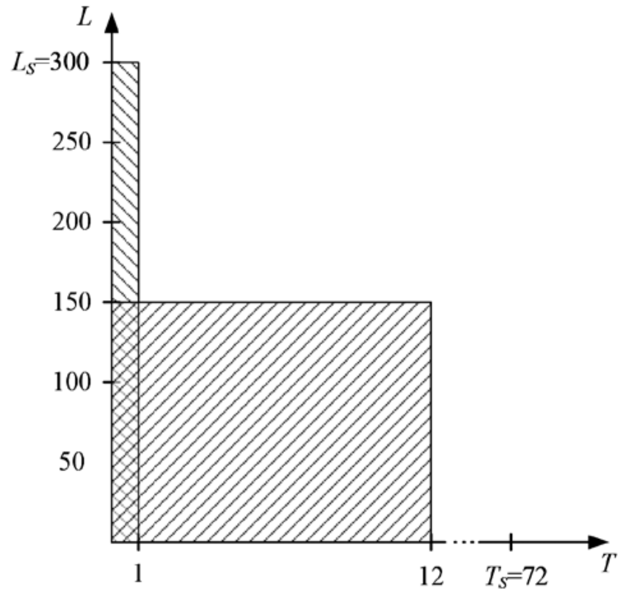

Рис. 2 – Сегменты в координатах «интервал оценки – размерность объекта»

конструкция, условия эксплуатации и т.п. одинаковы и можно утверждать, что отказ с равной вероятностью может возникнуть в любой точке этого объекта). В этом случае можно рассматривать нормирование количества отказов для «сегмента», соответствующего части (доле) L/L_S объекта, рассматриваемой на определенной части (доле) интервала времени T/T_S от общего времени T_S, за которое собрана статистика. На рис. 2 проиллюстрирован пример таких «сегментов» для двух случаев установления допустимого количества отказов: 1) для всего объекта ($L = 300$ км; $L/L_S = 1$) и интервала оценки 1 месяц ($T = 1$; $T/T_S = 1/72$); 2) для ½ объекта ($L = 150$ км; $L/L_S = 1/2$) и интервала оценки 12 месяцев ($T = 12$; $T/T_S = 1/6$). При этом параметр M с учетом как доли времени, так и доли размерности объекта определяется следующим образом:

```
200      12
23,621526        2,246755
17,5     3,0271376304261409246 8E-8
18,5     1,38252514856020350437E-4
19,5     5,98105983116750675014E-3
20,5     4,7808948379510079714E-2
21,5     1,62529561270786709964E-1
22,5     3,39875619758607182621E-1
23,5     5,31356753121528028805E-1
24,5     6,95251718991116234002E-1
25,5     8,15369879828009399177E-1
26,5     8,94408423267601224258E-1
27,5     9,42478788649415668705E-1
28,5     9,69980735598938544446E-1
29,5     9,84937678574992838692E-1
30,5     9,92717846377371271619E-1
31,5     9,96602795709502129994E-1
32,5     9,98469218515382248439E-1
33,5     9,99333229081020113744E-1
34,5     9,99719064545909650715E-1
35,5     9,99885429762384427954E-1
36,5     9,99954749274906081334E-1
```

```
200      72
7,453136         1,051233
3,5      4,29928647307663428764E-29
4,5      1,6455758598398818 5373E-8
5,5      3,23631142443019004128E-3
6,5      1,61153496005530435892E-1
7,5      5,75887043865400145316E-1
8,5      8,5655306110330783213E-1
9,5      9,60879559786062825698E-1
10,5     9,90620446889646346718E-1
11,5     9,97958665926637584536E-1
12,5     9,99591554645984463544E-1
13,5     9,99924351285517371148E-1
```

Рис. 3 – Таблицы результатов машинного расчета дискретных функций распределения

$$M = \frac{T_S}{T} \cdot \frac{L_S}{L},$$

в первом случае получим $M = (72/1) \cdot (300/300) = 72$, во втором $M = (72/12) \cdot (300/150) = 12$.

Далее, если на основании статистических данных известно, что в течение интервала $T_S = 72$ мес на объекте в целом имело место, например, 200 отказов, то, выполнив вычисление дискретных функций распределения (по алгоритму [9] или [10]):

- для $N = 200$; $M = 72$ получим математическое ожидание и среднеквадратическое отклонение соответственно 7,453 и 1,051;
- для $N = 200$; $M = 12$, получим математическое ожидание и среднеквадратическое отклонение соответственно 23,622 и 2,247.

Задав $\alpha = 0,9$ и применив формулу (7), позволяющую получить значение квантиля для любого $0 < \alpha < 1$, получим, что для всего объекта и оценочного интервала 1 месяц допустимое количество отказов составит 8,824, а для ½ объекта и оценочного интервала 12 месяцев – 26,553 (сравнение данных значений со «средними» – 7,454 и 23,623 соответственно – показывает смещение точечной оценки с учетом доверительной границы).

Таблицы дискретных функций распределения для $n = 200$; $m = 72$ и $N = 200$; $M = 12$, полученных машинным способом по алгоритму [9], представлены на рис. 3 («а» и «б» соответственно). Вычисления выполнялись до тех пор, пока значение вероятности не превысило порог 0,9999. В первой строке листингов указаны значения n;m, во второй – значения B и A соответственно (вычислены на основе пар значений n_1;$F(n_1)$) . Далее следует список пар значений n_1;$F(n_1)$, причем значения n_1 выведены с учетом смещения +0,5.

Выводы

1. Рассмотрен метод нормирования количества отказов сложного объекта с применением мультиномиального распределения и ограничения по его применению.

2. Данный подход вместо статистической выборки количества отказов по нескольким (M) интервалам времени T применяет суммарное количество отказов с учетом возможного случайного их распределения по M интервалам. При этом, учитывая, что статистическая выборка является лишь одной реализацией такого распределения, можно сказать, что на результаты нормирования перестает оказывать влияние наличие в статистической выборке маловероятного сочетания значений ряда. К тому же, поскольку дисперсия выборочного среднего (а, следовательно, и суммы выборки) снижается с увеличением объема выборки, применение в качестве исходной величины суммарного количества отказов за M оценочных интервалов способствует снижению ошибки установления допустимого значения.

3. При применении предложенного варианта нормирования количества отказов полученное допустимое значение всегда будет больше среднего значения статистической выборки, что вполне обосновано, поскольку при нормальном функционировании объекта большинство значений количества отказов (за интервал T) в выборке должны находиться ниже допустимого уровня.

4. Для практического применения требуют отдельного рассмотрения методы получения функций распределения вероятностей мультиномиального максимума (или их приближений) для всех требуемых значений параметров n и m, а также рассмотрения случая применения нецелого m.

5. Дальнейшего изучения требуют подходы к нормированию времени до восстановления сложного объекта (возможно, в сочетании с предложенным методом нормирования количества отказов), на основе чего будет возможна и реализация нормирования комплексного показателя надежности.

Библиографический список

1. Руденко Ю.Н. О подходах к нормированию показателей надежности электроснабжения потребителей // Известия Академии наук СССР. Энергетика и транспорт. 1975. № 1. С. 14-23.

2. Владимиров А.М. Гидрологические расчеты. Л.: Гидрометеоиздат, 1990. 366 с.

3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. 643 с.

4. Bonetti M., Cirillo P., Ogay A. Computing the exact distributions of some functions of the ordered multinomial counts: maximum, minimum, range and sums of order statistics // Journal of Royal Society. 2019. Vol. 6. No. 10. ID:190198. DOI: 10.1098/rsos.190198

5. Rappeport M.A. Algorithms and computational procedures for the application of order statistics to queing problems. PhD Thesis, New York University, 1968.

6. Hayter A.J. Recursive formulas for multinomial probabilities with applications // Computational Statistics. 2014. Vol. 29. Issue 5. Pp. 1207–1219. DOI: 10.1007/s00180-014-0487-0

7. DasGupta A. Exact tail probabilities and percentiles of the multinomial maximum: Technical Report. Purdue University, 2009. URL: https://www.stat.purdue.edu/~dasgupta/mult.pdf

8. Колчин В.Ф., Севастьянов Б.А., Чистяков В.П. Случайные размещения. М.: Наука, 1976. 224 с. с ил.

9. Ewens W.J., Wilf H.S. Computing the distribution of the maximum in balls-and-boxes problems with application to clusters of disease cases // Proceedings of the National Academy of Sciences (PNAS). 2007. Vol. 104(27). Pp. 11189-11191. DOI: 10.1073/pnas.0704691104

10. Levin B. A Representation for Multinomial Cumulative Distribution Functions // The Annals of Statistics. 1981. Vol. 9. No. 5. Pp. 1123-1126.

11. Авторские программы и библиотеки от Bedvit [Электронный ресурс]. URL: http://bedvit.ru/com/ (дата обращения 29.09.2022).

12. Ekhad S.B., Zeilberger D. Balls in Boxes: Variations on a Theme of Warren Ewens and Herbert Wilf // Advances in Combinatorics: Waterloo Workshop in Computer Algebra, W80, May 26-29, 2011 / Editors: I.S.Kotsireas, E.V.Zima. Springer-Verlag Berlin and Heidelberg Gm, 2013. Pp. 161-174.

References

1. Rudenko Yu.N. [On the approaches to the valuation of the dependability indicators of electric power supply to consumers]. *Proceedings of the Academy of Sciences of the USSR. Energy and transportation* 1975;1:14-23. (in Russ.)

2. Vladimirov A.M. [Hydrological calculations]. Leningrad: Gidrometeoizdat; 1990. (in Russ.)

3. Ayvazian S.A., Yeniukov I.S., Meshalkin L.D. [Applied statistics. Introduction to simulation and initial data processing]. Moscow: Finansy i statistika; 1983. (in Russ.)

4. Bonetti M., Cirillo P., Ogay A. Computing the exact distributions of some functions of the ordered multinomial counts: maximum, minimum, range and sums of order statistics. *Journal of Royal Society* 2019;6:10. ID:190198. DOI: 10.1098/rsos.190198.

5. Rappeport M.A. Algorithms and computational procedures for the application of order statistics to queing problems. PhD Thesis. New York University; 1968.

6. Hayter A.J. Recursive formulas for multinomial probabilities with applications. *Computational Statistics* 2014;29(5):1207-1219. DOI: 10.1007/s00180-014-0487-0.

7. DasGupta A. Exact tail probabilities and percentiles of the multinomial maximum: Technical Report. Purdue University; 2009. Available at: https://www.stat.purdue.edu/~dasgupta/mult.pdf.

8. Kolchin V.F., Sevastianov B.A., Chistiakov V.P. [Random allocations]. Moscow: Nauka; 1976. (in Russ.)

9. Ewens W.J., Wilf H.S. Computing the distribution of the maximum in balls-and-boxes problems with application to clusters of disease cases. *Proceedings of the National Academy of Sciences (PNAS)* 2007;104(27):11189-11191. DOI: 10.1073/pnas.0704691104.

10. Levin B.A. Representation for Multinomial Cumulative Distribution Functions. *The Annals of Statistics* 1981;9(5):1123-1126.

11. Bedvit's own programs and libraries. (accessed 29.09.2022). Available at: http://bedvit.ru/com/. (in Russ.)

12. Ekhad S.B., Zeilberger D. Balls in Boxes: Variations on a Theme of Warren Ewens and Herbert Wilf. In: I.S.Kotsireas, E.V.Zima, editors. Advances in Combinatorics: Waterloo Workshop in Computer Algebra, W80, May 26-29, 2011. Springer-Verlag Berlin and Heidelberg Gm; 2013. Pp. 161-174.

Сведения об авторе

Новожилов Евгений Олегович – кандидат технических наук, начальник отдела АО «НИИАС», Москва, Российская Федерация, тел. +7 (495) 967-77-02, e-mail: evg_o_nov@mail.ru

About the author

Evgeny O. Novozhilov, Candidate of Engineering, Head of Department, JSC NIIAS, Moscow, Russian Federation, tel.: +7 (495) 967 77 02, e-mail: eo.novozhilov@vniias.ru

Вклад автора в статью

Автором рассмотрена задача нормирования количества отказов технического объекта на основе статистической выборки, предложен метод реализации нормирования количества отказов на основе мультиномиального распределения, проведен анализ известных методов вычисления вероятностей мультиномиального максимума,

Конфликт интересов

Автор заявляет об отсутствии конфликтов интересов.