Интернет-журнал «Транспортные сооружения» / Russian journal of transport engineering http://t-s.today/

2016, Том 3, №3 / 2016, Vol 3, No 3 http://t-s.today/issues/vol3-no3.html

URL статьи: http://t-s.today/PDF/05TS316.pdf

DOI: 10.15862/05TS316 (http://dx.doi.org/10.15862/05TS316)

Ссылка для цитирования этой статьи:
Столяров В.В., Щеголева Н.В. О границах применимости нормального закона распределения вместо биноминального распределения при статистической обработке дискретных целочисленных величин // Интернет-журнал «Транспортные сооружения», Том 3, №3 (2016) http://t-s.today/PDF/05TS316.pdf (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:
Stolyarov V.V., Shchegoleva N.V. [On the limits to applicability of the normal distribution law instead of the binomial distribution while the statistical processing of discrete integer values] Russian journal of transport engineering, 2016, Vol. 3, no. 3. Available at: http://t-s.today/PDF/05TS316.pdf (In Russ.)

УДК 519.2, 625.7/.8

Столяров Виктор Васильевич
ФГБОУ ВПО «Саратовский государственный технический университет имени Гагарина Ю.А.», Россия, Саратов[1]
Доктор технических наук, профессор
E-mail: stolyarov_v_v@mail.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=443650

Щеголева Наталья Вячеславовна
ФГБОУ ВПО «Саратовский государственный технический университет имени Гагарина Ю.А.», Россия, Саратов
Доцент кафедры «Транспортное строительство»
Кандидат технических наук
E-mail: Shegoleva123@mail.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=668391

О границах применимости нормального закона распределения вместо биноминального распределения при статистической обработке дискретных целочисленных величин

Аннотация. В транспортном строительстве, как и во всех областях деятельности человека, встречаются задачи, связанные со статистической обработкой дискретных целочисленных величин, изменяющихся последовательно на единицу. Особенно часто такие задачи возникают при построении плотностей распределения дискретных случайных величин для оценки риска возникновения нежелательных событий в системе «водитель – автомобиль – дорога – окружающая среда». Как известно, вероятность появления сугубо дискретных и целочисленных величин описывается, биноминальным распределением, применение которого при большом числе испытаний становится весьма трудоёмким и продолжительным процессом, связанным с большим вводом в специализированные компьютерные программы исходных данных, изменяющихся, как было уже сказано, последовательно на единицу. Поэтому очень важно при решении прикладных задач уметь пользоваться методикой перехода от биноминального распределения к нормальному закону, полученному А. Муавром ещё в 1733 году, и пригодному для решения поставленных задач. Однако без исследований, описанных в данной работе, это сделать трудно, так как в существующей методике

[1] 410054, г. Саратов, ул. Политехническая 77

отсутствуют формулы по определению нижнего и верхнего пределов интегрирования, которые используют при определении вероятностей попадания случайной величины на выделенный участок. Кроме вывода данных формул получены математические выражения для определения порядкового номера максимальной переменной, при которой нормальное и биноминальное распределения дают одинаковые вероятности. При этом данные математические выражения учитывают величину отклонения моды от математического ожидания в биноминальном законе распределения. И, наконец, используя решение современной теории вероятностей, показано граничное условие применимости нормального распределения вместо биноминального распределения при абсолютной ошибке вероятности, равной 0,05 от вероятности, полученной по биноминальному закону. Сделаны основные выводы.

Ключевые слова: плотность вероятностей; функция Лапласа; функция нормального распределения; биноминальное распределение дискретных целочисленных величин; математическое ожидание и среднее значение; среднеквадратическое отклонение; независимые величины или факторы; критические параметры; граница применимости нормального закона вместо биноминального распределения

Изучаемые в сложных процессах величины (показатели) формируются, как правило, под влиянием большого числа независимых или слабо зависимых друг от друга факторов. Если каждый из этих факторов вызывает сравнительно малые количественные изменения наблюдаемой величины, и нет ни одного доминирующего фактора, то изучаемая случайная величина подчиняется нормальному закону или мало отклоняется от его формы. Более того, в соответствии с теоремой Ляпунова [1, 2, 3], при определенных условиях сумма (композиция) распределений, отличных от нормального, с ростом числа суммируемых распределений, неограниченно приближается к нормальному закону. Поэтому нормальный закон относят к предельным законам для сумм независимых случайных величин, подчиняющихся другим законам. Даже сумма слабо зависящих друг от друга случайных величин представляет собой асимптотически нормальное распределение [1, 3, 4, 5, 6].

Приведем пример формирования нормального распределения на основе зависимости (1), которая описывает математическое ожидание (среднее значение) критической длины остановочного пути автомобиля:

$$S_{50} = \left(\frac{V \cdot t_p}{3.6}\right) + \frac{k \cdot V^2}{[254(\varphi + i + f)]}. \quad (1)$$

Остановочный путь автомобиля называют критическим в том случае, когда после торможения и остановки автомобиля интервал между передним бампером автомобиля и препятствием равен нулю, и, следовательно, риск наезда на препятствие равен 50%.

Но и критическая длина остановочного пути, установленная по формуле (1), не является постоянной величиной для всех автомобилей и водителей. Очевидно, что после экстренного торможения часть водителей остановит свой автомобиль, не доехав до препятствия, большинство водителей смогут остановить автомобиль на расстоянии, близком к математическому ожиданию (S_{50}), а часть водителей не смогут остановить автомобиль на расстоянии S_{50} и, следовательно, допустят наезд на препятствие. Другими словами, речь идет о распределении критических интервалов с параметрами S_{50} и σ_{50}.

Отклонение остановочного пути автомобиля от математического ожидания $\Delta S = S_i - S_{50}$ зависит от следующих случайных величин:

- от отклонения времени реакции каждого водителя от расчётного (среднего) времени реакции: $\Delta t_i = t_i - t_p$;
- от отклонения скорости движения каждого автомобиля перед торможением от расчётной (средней) скорости, использованной в формуле (1): $\Delta V = V_i - V_{\text{ср}}$;
- от отклонения коэффициентов сцепления шин различных автомобилей при торможении от расчётного (среднего) коэффициента сцепления на данном участке дороги и при данной скорости: $\Delta\varphi = \varphi_i - \varphi_{\text{ср}}$;
- от отклонения коэффициентов сопротивления качению автомобилей от среднего коэффициента сопротивления качению, принятого к расчёту в формуле (1): $\Delta f = f_i - f_{\text{ср}}$.

Перечисленные характеристики являются случайными величинами и могут быть распределены по различным законам.

Так как $S = S(t, V, \varphi, f)$, то

$$\Delta S = \left(\frac{\partial S}{\partial t}\right) \cdot \Delta t + \left(\frac{\partial S}{\partial V}\right) \cdot \Delta V + \left(\frac{\partial S}{\partial \varphi}\right) \cdot \Delta\varphi + \left(\frac{\partial S}{\partial f}\right) \cdot \Delta f = X_1 + X_2 + X_3 + X_4 .$$

Случайная величина ΔS является суммой четырех независимых величин $X_1 + X_2 + X_3 + X_4$ (четырех плотностей распределения) и поэтому в соответствии с теоремой Ляпунова можно считать, что закон распределения ΔS будет близок к нормальному распределению.

Плотность нормального распределения (плотность вероятностей) имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\left[-\frac{(x-M)^2}{2\sigma^2}\right], \tag{2}$$

где: σ – среднеквадратическое отклонение; M – математическое ожидание; x – текущее значение; σ^2 – дисперсия случайной величины $X(D_x = \sigma^2)$.

Другими словами, непрерывная случайная величина X распределена по нормальному закону, если плотность распределения ее текущих значений (x) описывается формулой (2).

Нормированная функция нормального распределения выражается через плотность распределения формулой:

$$F(U) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} exp\left[-\frac{(x-M)^2}{2\sigma^2}\right] dx. \tag{3}$$

Вероятность попадания случайной величины X, распределенной по нормальному закону, на участок от x_1 до x_2 определяется одним из следующих выражений:

- при использовании нормированной функции нормального распределения:

$$P(x_1 < X < x_2) = F\left[\frac{x_2-M}{\sigma}\right] - F\left[\frac{x_1-M}{\sigma}\right], \tag{4}$$

- при использовании функции Лапласа:

$$P(x_1 < X < x_2) = \Phi\left[\frac{x_2-M}{\sigma}\right] - \Phi\left[\frac{x_1-M}{\sigma}\right], \tag{5}$$

где Φ(U) – нормированная функция Лапласа, которая отличается от нормированной функции (3) нижним пределом интегрирования:

$$\Phi(U) = \frac{1}{\sqrt{2\pi}} \cdot \int_{0}^{u} exp\left[-\frac{(x-M)^2}{2\sigma^2}\right] dx. \tag{6}$$

Все представленные выше формулы нормального закона распределения применимы только для непрерывных случайных величин, таких как длина остановочного пути автомобилей, и не предназначены для обработки дискретных целочисленных данных.

Например, они не применимы для статистической обработки результатов фиксации и восприятия водителями информации на дорожных знаках, когда требуется ответить, сколько знаков на опасном участке дороги, и какую информацию на них он успел увидеть при движении с определённой (фиксированной) скоростью. Количество дорожных знаков, воспринятых водителем, является дискретной и целочисленной величиной. При наличии 5 знаков возможны следующие варианты ответов:

- видел 3 знака, но информацию воспринял только на двух знаках;
- видел 3 знака, и воспринял информацию на всех трёх знаках;
- и так далее.

Водитель может утверждать, что он видел на опасном участке дороги и 6 и 7 знаков (вместо 5) с вариациями информации, которую он воспринял на дорожных знаках (воспринял информацию на всех знаках, не воспринял на одном, двух, трёх знаках и более).

Очевидно, что учитываются в статистике как правильные только те ответы водителей, в которых число увиденных знаков не превышает фактического количества знаков, и на всех увиденных знаках информация воспринята правильно. Дорожные знаки, на которых информация не воспринята водителем (или трактуется искажённо) относят к знакам, которые водитель не увидел (не воспринял). Другими словами, от каждого водителя получают информацию только о том – сколько он знаков распознал и сколько не распознал на опасном участке дороги при движении с заданной (фиксированной) скоростью. Такая информация носит дискретный характер и зависит от:

- способности водителя быстро воспринимать информацию;
- количества дорожных знаков на опасном участке дороги;
- и скорости движения автомобиля.

Например, два ответа, показанные выше трактуются следующим образом:

- первый водитель распознал 2 знака из 5;
- второй водитель распознал 3 знака из 5.

Формулы нормального распределения, показанные выше, для статистической обработки воспринятой водителями информации на дорожных знаках не применимы, так как речь идёт о целочисленных, изменяющихся через единицу, величинах (о количестве знаков).

Эти формулы не применимы так же для статистической обработки результатов референдумов, голосования, опросов населения. В процессе таких выборов предлагается отвечать “да или нет”, (голосовать “за или против”) посредством вычёркивания кандидатов.

Как в первом, так и во втором примерах ответы представляют собой сугубо дискретные и целочисленные величины, плотность вероятностей которых описывается биноминальным распределением:

$$p_n(k) = C_n^k \cdot q^{n-k} \cdot p^k, \tag{7}$$

где: $p_n(k)$ – вероятность появления события k раз в n испытаниях при условии, что в каждом опыте вероятность появления события остается постоянной и равной p (p = const); C_n^k – биноминальный коэффициент, учитывающий число сочетаний из n по k:

$$C_n^k = \frac{n!}{[k!\cdot(n-k)!]}; \tag{8}$$

где: q – вероятность появления противоположного события (q = 1 – p) в одном испытании; n – число опытов (испытаний); k – число событий, для которого определяется вероятность $p_n(k)$.

Однако, если по результатам предварительных исследований, например, по данным выборочного социологического опроса населения перед референдумом (голосованием), были получены вероятности p + q = 1, то нормальное распределение с параметрами биноминального закона может быть использовано. Причем, чем ближе вероятности p и q к 0,5, тем ближе биноминальное распределение к нормальному закону. Отсюда следует, что нормальное распределение для решения поставленной задачи должно быть получено по дискретным случайным переменным на основе биноминального распределения.

Впервые нормальное распределение было получено в 1733 году в исследованиях А. Муавра именно при рассмотрении частного случая биноминального распределения для больших чисел испытаний [4].

Функция такого распределения в примере с голосованием описывает вероятность того, что дискретное событие “да” (“за”, “не вычеркнуто”) появится ровно k раз из общего числа n, участвующих в референдуме. При выводе этих формул учитывают следующие особенности биноминального распределения:

1) аргумент k биноминального распределения изменяется дискретно и на единицу;
2) наибольшая вероятность соответствует модальному значению $k_{mod} \approx n \cdot p$.

Вывод формул, пригодных для решения поставленной задачи, представлен в работе [4], а полученный нормальный закон распределения вероятностей при многократных испытаниях имеет стандартную плотность:

$$\varphi(t) = \left(\frac{1}{\sqrt{2\pi}}\right) \cdot e^{-\frac{t^2}{2}}, \tag{9}$$

где t – переменная нормального распределения, определяемая по формуле:

$$t = \xi \cdot h, \tag{10}$$

в которую входят:

ξ – отклонение относительной частоты испытуемого события $(\frac{k}{n})$ от его вероятности (p):

$$\xi = \mathrm{p} - \frac{\mathrm{k}}{\mathrm{n}}, \tag{11}$$

h – вспомогательная переменная:

$$h = \sqrt{\frac{n}{(p\cdot q)}}, \tag{12}$$

где n, p, q – см. описание к формулам (6) и (7).

Вероятность появления события k раз при n испытаниях $[P_n(k)]$ на интервале dt определяют по формуле:

$$p_n(k) = \varphi(t) \cdot dt, \tag{13}$$

где: φ(t) – см. формулу (8);

dt – изменение переменной t при изменении k на единицу:

$$dt = h \cdot d\xi = \frac{\mathrm{h}}{\mathrm{n}}, \tag{14}$$

На этом вывод нормального закона на основе биноминального распределения вероятностей в работе [4] заканчивается.

Продолжим развитие данного математического аппарата до вывода формул (21), (22) и (23), трактуя полученное выше решение следующим образом. На рис. 1 показано не центрированное нормальное распределение (то есть с математическим ожиданием (M) не равным нулю). В этом случае параметр t представляет собой интервал, откладываемый от точки математического ожидания нормального распределения (M) до точки, в которой расположена середина интервала dt. При t = 0 интервал dt расположен в центре группирования.

Как показывают расчеты по формулам (9) – (14), чем больше по абсолютной величине переменная t, тем меньше вероятность появления события k раз при n испытаниях, установленная по формуле (13), так как заштрихованная площадь под кривой распределения при удалении от центра группирования уменьшается (см. рис. 1).

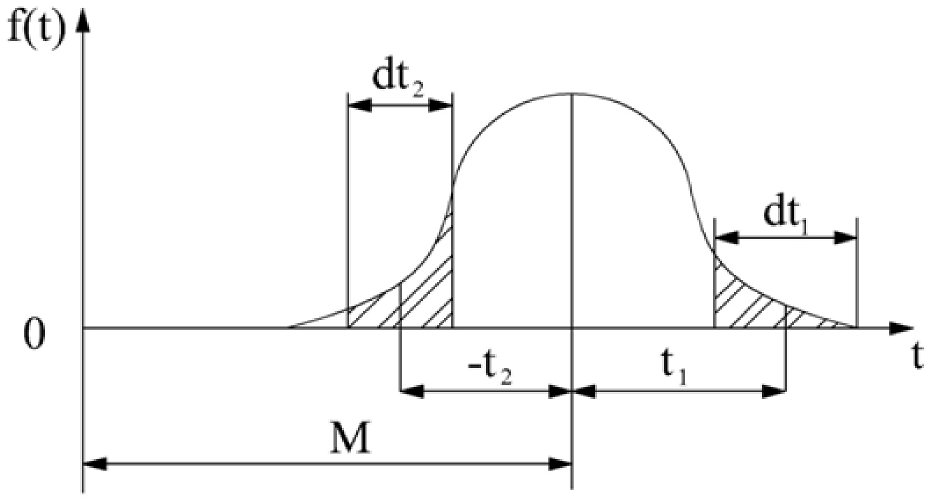

Рисунок 1. *К обоснованию параметров нормального закона, основанного на биноминальном распределении вероятностей*
Figure 1. *Argumentation of parameters of the normal law based on the binomial distribution of probabilities*

Формулы (9) – (14) получены на основе параметров биноминального распределения (n, k, p, q) и поэтому нет основания считать, что математическое ожидание (M) и среднеквадратическое отклонение (σ) данного нормального распределения будут отличаться от соответствующих параметров, установленных по зависимостям биноминального распределения.

Тогда, для вычисления математического ожидания и среднеквадратического отклонения в данном случае применимы формулы:

$$M = n \cdot p; \tag{15}$$

$$\sigma = n \cdot p \cdot q, \tag{16}$$

где n, p, q – см. описание параметров к формулам (7) и (8).

Косвенным подтверждением применимости этих формул является не только то, что один закон распределения получен из другого, но и то, что при больших значениях n мода биноминального закона находится в центральной части распределения, а по обе стороны от нее вероятности $P_n(k)$ убывают. Учитывая, что мода биноминального распределения близка к математическому ожиданию [2, 3], формулы (15) и (16) безусловно, описывают параметры M и σ нормального закона, основанного на биноминальном распределении, и особенно в тех случаях, когда вероятности p и q , близки к 0,5.

Применимость формул (15) и (16) к вычислению параметров M и σ позволяет перейти к определению суммарных вероятностей вида:

$$P_n(k) = p_n(0) + p_n(1) + \dots + p_n(km) = \sum_{k=0}^{k=k_m} p_n(k), \quad (17)$$

но не при помощи биноминального распределения, а по формуле (4) или формуле (5), что существенно расширяет возможности нормального закона с параметрами биноминального распределения, включая и возможность статистической обработки данных с ограниченным числом вероятных ответов. Например, становится возможным расчет вероятностей для ответов вида “да” или “нет”. В этом случае для расчета по формулам (4) или (5) вероятностей $P(x_1^{min} < X < x_2) = \sum p_n(k)$ выполняют следующие преобразования.

В формуле $t_1 = h \cdot \xi = h \cdot [p - \frac{k}{n}]$ принимают число k равным нулю ($k = 0$) и тем самым получают точку, ограничивающую исходное биноминальное распределение слева значением $t_1^{min} = h \cdot p$ (рис. 2). При этом нижний предел интегрирования (x_1^{min}) будет определяться выражением:

$$x_1^{min} = M - h \cdot p. \quad (18)$$

Точке x_1^{min} соответствует вероятность того, что ожидаемое событие ни разу не произойдет, например, ни кто из участников референдума не проголосует положительно (не поддержит ответы “да” “за”, “не вычеркнуто”). Вероятность этого события для биноминального распределения обозначается символом $p_n(k = 0)$ или проще $p_n(0)$.

Вычисляют параметр t_2 по выражению:

$$t_2 = h \cdot [p - \frac{k^{max}}{n}], \quad (19)$$

где k_{max} – максимальное число ожидаемых событий при n испытаниях.

Верхний предел интегрирования (x_2) в формулах (4) и (5) определяют по зависимости:

$$x_2 = M + t_2. \quad (20)$$

Так, для рассматриваемого примера, точке x_2 соответствует вероятность того, что ровно k^{max} участников референдума проголосует положительно (“да” “за”, “не вычеркнуто”). Вероятность этого события для биноминального распределения обозначается символом $P_n(k = k_m)$ или проще $P_n(k_m)$.

Параметры M, σ, x_1^{min} и x_2, определенные по формулам (15), (16), (18) и (20), используют в формуле (4) или формуле (5) при расчете суммарной вероятности того, что дискретное событие “да” (“за”, “не вычеркнуто”) появится в пределах от k = 0 раз до $k = k^{max}$ раз из общего числа n, участвующих в референдуме.

Формулой (20) пользуются до тех пор, пока получают $x_2 > x_1^{min}$. При возникновении неравенства $x_2 < x_1^{min}$ принимают $x_2 = x_1^{min}$ и по формуле (4) или формуле (5) получают, что вероятность этого события равна нулю.

Выражения (4), (5) и (17) дают одинаковые суммы вероятностей при вычислении $p_n(k_i)$ в формуле (17) по биноминальному распределению (7) и при переборе параметра k_i от нуля до значения k_m, порядковый номер которого (m) определяется по зависимостям, полученным в данной работе:

$$\text{при } p < 0{,}5 \; m = p(n+1) + t_2 - 1; \tag{21}$$

$$\text{при } p = 0{,}5 \; m = M + t_2; \tag{22}$$

$$\text{при } p > 0{,}5 \; m = p(n+1) + t_2. \tag{23}$$

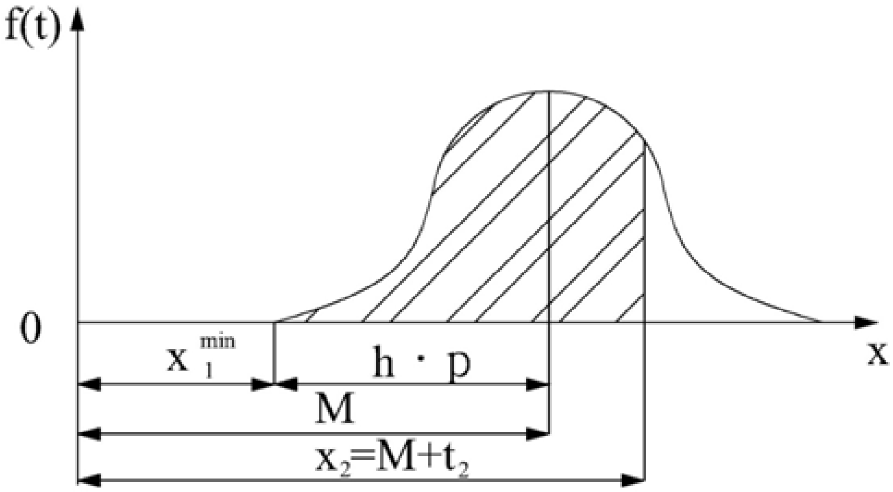

Рисунок 2. *Исходные данные к расчету суммарных вероятностей по формуле (9)*
Figure 2. *Initial data for the calculation of the integrated probability from the formula (9)*

Эти зависимости учитывают величину отклонения моды от математического ожидания для биноминального распределения.

Вычисленный по формулам (21) – (23) порядковый номер m параметра k_m округляют до целого значения.

Преимущество нормального распределения перед биноминальным законом заключается как раз в том, что по формуле (4) или формуле (5) определяется суммарная вероятность всех ожидаемых событий, без вычисления по биноминальному распределению для всех k (k = 0, 1, 2, 3, …, k_m) вероятностей $p_n(k)$ и без последующего их суммирования по формуле (17). Кроме того, нормальное распределение можно интегрировать, а биноминальное – нет. В этом заключается основное преимущество описанного в статье перехода от биноминального к нормальному закону.

При определении вероятностей по биноминальному распределению (7) удобно в расчетах использовать соотношение

$$\frac{p_n(k+1)}{p_n(k)} = \frac{(n-k)p}{[(k+1)q]}, \tag{24}$$

которое связывает вероятности $p_n(k)$ и $p_n(k+1)$ двух соседних членов ряда. Так, если по формуле (7) определено значение $p_n(k)$, то, используя зависимость (24) не трудно установить значение следующей вероятности $p_n(k+1)$ и так далее.

Для расчета по формуле (17) суммарной вероятности необходимо воспользоваться формулами (7) и (8) k_m раз при переборе значений k_i от $k_1 = 0$ до конечного числа событий $k = k_m$. Сумма полученных вероятностей и дает суммарную вероятность, установленную ранее по формуле (4) или формуле (5).

Из сказанного становится очевидным, что использование нормального закона, основанного на биноминальном распределении, более удобно, чем непосредственное использование биноминального распределения. Прогнозировать результаты голосования по формуле (7) можно только при помощи мощных ЭВМ, так как перебор значений k_i через единицу становится невозможным, а суммируемые вероятности приобретают ничтожно малые значения.

Однако следует помнить о границах применимости нормального распределения вместо биноминального и не допускать их нарушения, так как в противном случае полученные вероятности будут значительно отличаться от истинных вероятностей, подчиняющихся биноминальному распределению.

В соответствии с решениями современной теории вероятностей [3, 5], если для биноминального распределения имеем $n \cdot p^{3/2} > 1{,}07$, то ошибка при использовании нормальной функции вместо биномиального закона не превосходит 0,05 при всех значениях переменной x.

В табл. 1 даны примеры границ применимости нормального распределения вместо биноминального закона при условии, что абсолютная ошибка вероятности $p_n(k)$, равная 0,05, при малых значениях n, является допустимой.

Граничное условие применимости нормального распределения удобнее представлять в виде

$$p > \sqrt{\left(\frac{1{,}07}{n}\right)^3}, \tag{25}$$

где: p – вероятность появления события x_i в одном испытании [см. формулы (7) и (8)]; n – число опытов (испытаний).

Таблица 1/Table 1

Левая граница применимости нормального закона вместо биноминального распределения в зависимости от параметров n и p/The left limit to applicability of the normal law instead of the binomial distribution depending on parameters n and p

n	Применимость (+) нормального распределения при p:										
	0,001	0,01	0,025	0,05	0,075	0,10	0,15	0,20	0,30	0,40	0,50
2										+	+
3									+	+	+
4							+	+	+	+	+
5						+	+	+	+	+	+
6						+	+	+	+	+	+
7					+	+	+	+	+	+	+
8				+	+	+	+	+	+	+	+
9				+	+	+	+	+	+	+	+
10				+	+	+	+	+	+	+	+
15			+	+	+	+	+	+	+	+	+
20			+	+	+	+	+	+	+	+	+
25		+	+	+	+	+	+	+	+	+	+
101	+	+	+	+	+	+	+	+	+	+	+

Выводы по применимости нормального распределения вместо биноминального закона:

1. Уже при двух испытаниях и при вероятности появления события более 0,4 биноминальное распределение можно заменять нормальным законом.

2. Если проведено 8 опытов (испытаний, измерений), то замена биноминального распределения нормальным законом возможна уже при вероятности появления события 0,05, которая значительно отличается от вероятности 0,50.
3. При числе испытаний более 100 нормальный закон распределения заменяет биноминальный закон практически при любой вероятности ожидаемого события (см. вероятность 0,001 в табл. 1).
4. Результаты, представленные в табл. 1 косвенно подтверждают, что нормальный закон распределения можно рассматривать в качестве предельного для биноминального распределения, так как с увеличением n данная функция распределения стремится к нормальному закону. Однако следует отметить, что биноминальное распределение обладает композиционной устойчивостью.

Поясним применимость нормального распределения вместо биноминального закона при малом числе опытов. Например, рассмотрим малое число членов в совете директоров, которым предстоит решить какой-либо вопрос голосованием. Пусть в голосовании, требующем дать ответ “за” или “против” на какой-либо вопрос, принимает участие три члена совета директоров, тогда число опытов $n = 3$. Известно, что возможное число элементарных исходов в n опытах составляет 2^n, следовательно, число элементарных исходов равно восьми, так как $2^3 = 8$. Покажем это, обозначив ответ голосования “за” буквой А и ответ голосования “против” буквой В. Три человека (в принятом порядке голосования) могут дать одно из следующих восьми сочетаний:

ВВВ, ВВА, ВАВ, АВВ, ААВ, АВА, ВАА, ААА.

Во всех перечисленных элементарных исходах видно, что если событие А встречается k раз, то событие В встречается (n – k) раз. Обозначим вероятность появления события А символом p, а вероятность появления события В символом q и покажем, что вероятность любого из перечисленных восьми сочетаний определяется выражением $p_n(k) = p^k \cdot q^{(n-k)}$. Так, для первого сочетания получаем $p_3(0) = p^0 \cdot q^{(3-0)} = q^3$ (действительно сочетание ВВВ дает вероятности $q \cdot q \cdot q = q^3$). Для второго сочетания имеем $p_3(1) = p^1 \cdot q^{(3-1)} = q^2 \cdot p$ (непосредственно по второму сочетанию ВВА получаем тот же результат $q \cdot q \cdot p = q^2 \cdot p$). Как видно третье (ВАВ) и четвертое (АВВ) сочетания дают такие же ответы, как второе сочетание, а именно $p_3(1) = p^1 \cdot q^{(3-1)} = q^2 \cdot p$ и значит, результат не зависит от того, в каком порядке чередуется k событий А и (n – k) событий В.

Выполнив такие же расчеты для всех сочетаний, получаем:

ВВВ, ВВА, ВАВ, АВВ, ААВ, АВА, ВАА, ААА;

q^3, $q^2 \cdot p$, $q^2 \cdot p$, $q^2 \cdot p$, $p^2 \cdot q$, $p^2 \cdot q$, $p^2 \cdot q$, p^3.

Сумма всех полученных вероятностей равна единице, так как перечисленные сочетания образуют полную группу событий:

$$q^3 + 3 \cdot q^2 \cdot p + 3 \cdot p^2 \cdot q + p^3 = 1.$$

Или окончательно (при $n = 3$) получаем $(p + q)^3 = 1$.

Переход от биноминального распределения к нормальному закону начинают с использования формулы (11), по которой определяют отклонение относительной частоты испытуемого события (k/n) от его вероятности (p): $\xi = p - (k/n)$. Как видно из рассмотренного примера параметр k уже при $n = 3$ принимает 8 значений (один раз $k = 0$; три раза $k = 1$; три раза $k = 2$ и один раз $k = 3$). Число возможных сочетаний (k исходов в n испытаниях) устанавливают обычно по формуле (8), которая обобщает, описанные выше рассуждения. Так, при $n = 3$ по формуле (8) имеем тот же результат:

$$\text{при k = 0 } C_n^k = \frac{n\,!}{[k\,!\,(n-k)\,!]} = \frac{3!}{[0!\,(3-0)!]} = 1;$$

$$\text{при k = 1 } C_n^k = \frac{3!}{[1!\,(3-1)!]} = 3;$$

$$\text{при k = 2 } C_n^k = \frac{3!}{[2!\,(3-2)!]} = 3;$$

$$\text{при k = 3 } C_n^k = \frac{3!}{[3!\,(3-3)!]} = 1.$$

Кроме того, параметр k имеет колоколообразное распределение, так как относительная частота его появления составляет 1/8, 3/8, 3/8 и 1/8.

Однако применение нормального закона, сглаживающего дискретное биноминальное распределение, возможно только в том случае, когда вероятность появления события А при n = 3 больше значения, определенного по формуле (25), то есть в нашем случае $p > [(1{,}07/3)^3]^{½} = 0{,}213$ (в табл. 3 эта вероятность дана с округлением в большую сторону до 0,30).

Если в голосовании на одно выборное место претендует несколько кандидатов, то биноминальное распределение применяют после следующих преобразований. Событие прохождения выборов одним из кандидатов обозначают буквой А.

Прохождение выборов любым другим кандидатом является противоположным событием событию А и несовместным с ним. Поэтому выигрыш выборов другими кандидатами обозначают буквами $B_1, B_2, \ldots, B_n$. Причем события $B_1, B_2, \ldots, B_n$ являются то же несовместными и противоположными по отношению друг к другу. Вероятность суммы нескольких несовместных событий равна сумме их вероятностей, а так как все эти события представляют полную группу событий, да еще и противоположных друг другу, то данная сумма вероятностей равна единице:

$$\mathrm{p}(\mathrm{A}+\mathrm{B}_1+\mathrm{B}_2+\cdots+\mathrm{B}_n) = p(\mathrm{A}) + p(\mathrm{B}_1) + p(\mathrm{B}_2) + \cdots + p(\mathrm{B}_n) = 1.$$

Отсюда, учитывая, что $p(\mathrm{A}) = p_j, p(\mathrm{B}_i) = q_i$, (при $j \neq i$) для исходных вероятностей биноминального распределения окончательно получаем

$$p_j = 1 - \sum q_i, \tag{26}$$

где: p_j – постоянная вероятность появления события А в одном испытании [событием А является прогнозом того, что данный (j-й) кандидат выиграет выборы или данная партия (j-я партия) преодолеет определенный граничный процент (при $j \neq i$)];

q_i – постоянная вероятность появления события B_i, противоположного событию А, в одном испытании [событием B_i является прогнозом того, что i-й кандидат выиграет выборы или i-я партия преодолеет определенный граничный процент].

Очевидно, что за параметр p_j в формуле (26) можно принимать постоянную вероятность положительного голосования одним избирателем за любого кандидата или любую партию из числа участвующих в выборах. Вероятности p_j и q_i определяют, как правило, по данным социологических опросов различных слоев населения, основываясь на достаточно представительных выборках или на основе аналитического прогноза результатов выборов. Вероятность появления события А k раз в n испытаниях при p = const определяют по формуле (7).

Примеры формирования нормального и других законов распределения для основных параметров системы «водитель – автомобиль – дорога - окружающая среда» с оценкой риска возникновения нежелательных событий представлены в следующих работах авторов [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26].

ЛИТЕРАТУРА

1. Вероятностные методы в инженерных задачах / Лебедев А.Н., Куприянов М.С., Недосекин Д.Д., Чернявский Е.А. // Справочник. – СПб.: Энергоатомиздат. Санкт-Петербургское отд., 2000. – 333 с.
2. Смирнов Н.В., Белугин Д.А. Теория вероятностей и математическая статистика в приложении к геодезии. – М.: Недра, 1979. – 384 с.
3. Справочник по теории вероятностей и математической статистики / В.С. Королюк, Н.И. Портенко, А.В. Скороход, А.Ф. Турбин. – М.: Наука. Главная редакция физико-математической литературы, 1985. – 640 с.
4. Большаков В.Д., Гайдаев П.А. Теория математической обработки геодезических измерений. – М.: Недра, 1977. – 368 с.
5. Бронштейн И.Н., Семендяев К.А. Справочник по математике. – М.: ГИТТЛ, 1956. – 608 с.
6. Столяров В.В. Введение в теорию риска / В.В. Столяров // Повышение эффективности эксплуатации транспорта: межвуз. науч. сб. Саратов: СГТУ, 2003. С. 118 – 139.
7. Столяров В.В. Дорожные условия и организация движения с использованием теории риска: учеб. пособие / В.В. Столяров – Саратов: СГТУ, 1999. – 168 с.
8. Столяров В.В. Проектирование автомобильных дорог с учетом теории риска: монография в 2 частях, Ч. 1 / В.В. Столяров – Саратов: СГТУ, 1994. – 184 с.
9. Столяров В.В. Проектирование автомобильных дорог с учетом теории риска: монография в 2 частях, Ч. 2 / В.В. Столяров – Саратов: СГТУ, 1994. – 232 с.
10. Столяров В.В. Проектирование автомобильных дорог по условию обеспечения безопасности движения с использованием теории риска: дис. … д-ра техн. наук / В.В. Столяров. – М., 1995. – 337 с.
11. Столяров В.В. Теория риска в проектировании плана дороги и организации движения: учеб. пособие / В.В. Столяров – Саратов: СГТУ, 1995. – 84 с.
12. Столяров В.В. Экспертиза дорожно-транспортных происшествий на основе теории риска: учеб. пособие / В.В. Столяров – Саратов: СГТУ, 1996. – 176 с.
13. Столяров В.В. Теория риска в судебно-технической экспертизе дорожно-транспортных происшествий: монография / В.В. Столяров – Саратов: Издательский дом «МарК», 2010. – 412 с.
14. Столяров В.В. Совершенствование методов применения принципов технического регулирования в дорожной деятельности: монография / В.В. Столяров, А.П. Бажанов. – Пенза: ПГУАС. 2014. – 212 с.
15. Столяров В.В. Оценка надёжности нежёстких дорожных одежд на основе законов распределения общих модулей упругости / В.В. Столяров, Е.Е. Зверкова, А.С. Фомина, Ю.М. Аникин // Дороги и мосты. – М.: Росавтодор, 2013. Т1. №29/1. С. 153 – 176.
16. Столяров В.В. Новый подход к гамма-распределению при обосновании расчётных расходов мостовых переходов / В.В. Столяров, Э.Ю. Шмагина // Известия Орловского ГТУ. Серия: Строительство и транспорт. – Орёл: 2007. №3-15. С. 67 – 69.

17. Столяров В.В. Теория риска в судебно-технической экспертизе дорожно-транспортных происшествий с участием пешеходов (+ABS): монография / В.В. Столяров. – Саратов: СГТУ, 2010. – 344 с.
18. Столяров В.В. Технико-экономическое обоснование модернизации сети автомобильных дорог / В.В. Столяров, Д.М. Немчинов // Дороги и мосты. – М.: Росавтодор, 2015. №34/2. С. 13 – 50.
19. Столяров В.В. Математическая модель транспортного потока, основанная на микроскопической теории «следования за лидером» / В.В. Столяров, Д.М. Немчинов, В.А. Гусев, Н.В. Щёголева // Дороги и мосты. – М.: Росавтодор, 2015. №34/2. С. 291 – 318.
20. Щёголева Н.В. Риск потери информации как обобщённая характеристика водителя при проектировании и эксплуатации автомобильных дорог: автореф. дис. … канд. техн. наук / Н.В. Щёголева. – Волгоград, 2006. – 24 с.
21. Щёголева Н.В. Учет суммарной энтропии дорожно-транспортных ситуаций при определении риска потери информации [Электронный ресурс] / Н.В. Щеголева // Техническое регулирование в транспортном строительстве. - 2015. - №2 (10). - с.5. - Библиогр. в конце ст. - Загл. с титул. экрана. - URL: http://trts.esrae.ru/16-72.
22. Щёголева Н.В. К вопросу о применении теории информации в проектировании дорог [Текст] / Н.В. Щеголева // Ресурсо- и энергоэффективные технологии в строительном комплексе региона: сб. науч. тр. по материалам Междунар. науч.-практ. конф. / СГТУ. - Саратов, 2014. - С. 381-384.
23. Щёголева Н.В. Применение теории риска в проектировании реконструкции автомобильных дорог [Текст] / Н.В. Щеголева, О.Н. Артемова // Математические методы в технике и технологиях - ММТТ-25: сб. тр. XXV междунар. науч. конф., г. Волгоград, 29 мая - 31 мая 2012 г.: в 10 т. - Саратов, 2012. - Т. 10. - С. 77-79.
24. Щёголева Н.В. Автоматизация расчета риска при реконструкции опасного участка автомобильной дороги [Текст] / Н.В. Щеголева, В.А. Гусев // Математические методы в технике и технологиях - ММТТ-25: сб. тр. XXV междунар. науч. конф., г. Волгоград, 29 мая - 31 мая 2012 г.: в 10 т. - Саратов, 2012. - Т. 10. - С. 79-81.
25. Щёголева Н.В. Оценка вероятности потери информации водителями с использованием теории риска [Текст] / Н.В. Щеголева // Вестник Пермского национального исследовательского политехнического университета. Охрана окружающей среды, транспорт, безопасность жизнедеятельности. - 2011. - №1. - С. 160-165.
26. Ворожейкин М.А. Применение теории риска для реконструкции геометрических параметров кривой в плане [Текст] / М.А. Ворожейкин, Н.В. Щеголева // Инновации и исследования в транспортном комплексе: материалы III междунар. науч.-практ. конф. в год 70-летия победы в Великой Отечественной войне, г. Курган, 5-6 июня 2015 г.: в 2 ч. - Курган, 2015. - Ч. 2. - С. 178-181.

Stolyarov Viktor Vasilevich
Yuri Gagarin state technical university of Saratov, Russia, Saratov
E-mail: stolyarov_v_v@mail.ru

Shchegoleva Natalia Vyacheslavovna
Yuri Gagarin state technical university of Saratov, Russia, Saratov
E-mail: Shegoleva123@mail.ru

On the limits to applicability of the normal distribution law instead of the binomial distribution while the statistical processing of discrete integer values

Abstract. In the transport construction, as in all fields of human activity, there are problems associated with the statistical processing of discrete integer values varying sequentially by one. Such problems very often arise when constructing the densities of distribution of the discrete random variables to assess the risk of adverse event occurrence in the "driver - car - road - environment" system. It is known that the probability of the particularly discrete and integer value occurrence is described by the binomial distribution, the use of which for a large number of tests becomes very time-taking and lengthy process, associated with a large input of given data, varying, as already mentioned, successively by one, into the specialized computer programs. Therefore it is very important when handling the applied problems to be able to use the method of transition from the binomial distribution to the normal law derived by A. Moivre as early as in 1733, and suitable for solution of the assigned problems. However, without studies described in this article this is difficult to do, since the existing method has no formulae for determining the lower and upper limits of integration, which are used when determining the probability of hit of a random variable to a dedicated site. Besides derivation of these formulae there were developed the mathematical expressions to determine the serial number of the maximum variable, wherein the normal and binomial distributions give the same probability. These mathematical expressions take into account the value of mode deviation from the mathematical expectation in the binomial distribution law. Finally, using a solution of the modern theory of probability, the author shows a boundary condition for the applicability of the normal distribution instead of the binomial one with an absolute error of probability equal to 0.05 of the probability obtained by the binomial law. The article presents the main conclusions.

Keywords: probability density; Laplace function; function of normal distribution; binomial distribution of discrete integer values; mathematical expectation and mean value; mean-square deviation; independent variables or factors; critical parameters; limit to applicability of the normal law instead of the binomial distribution

REFERENCES

1. Verojatnostnye metody v inzhenernyh zadachah / Lebedev A.N., Kuprijanov M.S., Nedosekin D.D., Chernjavskij E.A. // Spravochnik. – SPb.: Jenergoatomizdat. Sankt-Peterburgskoe otd., 2000. – 333 s.
2. Smirnov N.V., Belugin D.A. Teorija verojatnostej i matematicheskaja statistika v prilozhenii k geodezii. – M.: Nedra, 1979. – 384 s.
3. Spravochnik po teorii verojatnostej i matematicheskoj statistiki / V.S. Koroljuk, N.I. Portenko, A.V. Skorohod, A.F. Turbin. – M.: Nauka. Glavnaja redakcija fiziko-matematicheskoj literatury, 1985. – 640 s.

4. Bol'shakov V.D., Gajdaev P.A. Teorija matematicheskoj obrabotki geodezicheskih izmerenij. – M.: Nedra, 1977. – 368 s.
5. Bronshtejn I.N., Semendjaev K.A. Spravochnik po matematike. – M.: GITTL, 1956. – 608 s.
6. Stoljarov V.V. Vvedenie v teoriju riska / V.V. Stoljarov // Povyshenie jeffektivnosti jekspluatacii transporta: mezhvuz. nauch. sb. Saratov: SGTU, 2003. S. 118 – 139.
7. Stoljarov V.V. Dorozhnye uslovija i organizacija dvizhenija s ispol'zovaniem teorii riska: ucheb. posobie / V.V. Stoljarov – Saratov: SGTU, 1999. – 168 s.
8. Stoljarov V.V. Proektirovanie avtomobil'nyh dorog s uchetom teorii riska: monografija v 2 chastjah, Ch. 1 / V.V. Stoljarov – Saratov: SGTU, 1994. – 184 s.
9. Stoljarov V.V. Proektirovanie avtomobil'nyh dorog s uchetom teorii riska: monografija v 2 chastjah, Ch. 2 / V.V. Stoljarov – Saratov: SGTU, 1994. – 232 s.
10. Stoljarov V.V. Proektirovanie avtomobil'nyh dorog po usloviju obespechenija bezopasnosti dvizhenija s ispol'zovaniem teorii riska: dis. … d-ra tehn. nauk / V.V. Stoljarov. – M., 1995. – 337 s.
11. Stoljarov V.V. Teorija riska v proektirovanii plana dorogi i organizacii dvizhenija: ucheb. posobie / V.V. Stoljarov – Saratov: SGTU, 1995. – 84 s.
12. Stoljarov V.V. Jekspertiza dorozhno-transportnyh proisshestvij na osnove teorii riska: ucheb. posobie / V.V. Stoljarov – Saratov: SGTU, 1996. – 176 s.
13. Stoljarov V.V. Teorija riska v sudebno-tehnicheskoj jekspertize dorozhno-transportnyh proisshestvij: monografija / V.V. Stoljarov – Saratov: Izdatel'skij dom «MarK», 2010. – 412 s.
14. Stoljarov V.V. Sovershenstvovanie metodov primenenija principov tehnicheskogo regulirovanija v dorozhnoj dejatel'nosti: monografija / V.V. Stoljarov, A.P. Bazhanov. – Penza: PGUAS. 2014. – 212 s.
15. Stoljarov V.V. Ocenka nadjozhnosti nezhjostkih dorozhnyh odezhd na osnove zakonov raspredelenija obshhih modulej uprugosti / V.V. Stoljarov, E.E. Zverkova, A.S. Fomina, Ju.M. Anikin // Dorogi i mosty. – M.: Rosavtodor, 2013. T1. №29/1. S. 153 – 176.
16. Stoljarov V.V. Novyj podhod k gamma-raspredeleniju pri obosnovanii raschjotnyh rashodov mostovyh perehodov / V.V. Stoljarov, Je.Ju. Shmagina // Izvestija Orlovskogo GTU. Serija: Stroitel'stvo i transport. – Orjol: 2007. №3-15. S. 67 – 69.
17. Stoljarov V.V. Teorija riska v sudebno-tehnicheskoj jekspertize dorozhno-transportnyh proisshestvij s uchastiem peshehodov (+ABS): monografija / V.V. Stoljarov. – Saratov: SGTU, 2010. – 344 s.
18. Stoljarov V.V. Tehniko-jekonomicheskoe obosnovanie modernizacii seti avtomobil'nyh dorog / V.V. Stoljarov, D.M. Nemchinov // Dorogi i mosty. – M.: Rosavtodor, 2015. №34/2. S. 13 – 50.
19. Stoljarov V.V. Matematicheskaja model' transportnogo potoka, osnovannaja na mikroskopicheskoj teorii «sledovanija za liderom» / V.V. Stoljarov, D.M. Nemchinov, V.A. Gusev, N.V. Shhjogoleva // Dorogi i mosty. – M.: Rosavtodor, 2015. №34/2. S. 291 – 318.

20. Shhjogoleva N.V. Risk poteri informacii kak obobshhjonnaja harakteristika voditelja pri proektirovanii i jekspluatacii avtomobil'nyh dorog: avtoref. dis. … kand. tehn. nauk / N.V. Shhjogoleva. – Volgograd, 2006. – 24 s.
21. Shhjogoleva N.V. Uchet summarnoj jentropii dorozhno-transportnyh situacij pri opredelenii riska poteri informacii [Jelektronnyj resurs] / N.V. Shhegoleva // Tehnicheskoe regulirovanie v transportnom stroitel'stve. - 2015. - №2 (10). - s.5. - Bibliogr. v konce st. - Zagl. s titul. jekrana. - URL: http://trts.esrae.ru/16-72.
22. Shhjogoleva N.V. K voprosu o primenenii teorii informacii v proektirovanii dorog [Tekst] / N.V. Shhegoleva // Resurso- i jenergojeffektivnye tehnologii v stroitel'nom komplekse regiona: sb. nauch. tr. po materialam Mezhdunar. nauch.-prakt. konf. / SGTU. - Saratov, 2014. - S. 381-384.
23. Shhjogoleva N.V. Primenenie teorii riska v proektirovanii rekonstrukcii avtomobil'nyh dorog [Tekst] / N.V. Shhegoleva, O.N. Artemova // Matematicheskie metody v tehnike i tehnologijah - MMTT-25: sb. tr. XXV mezhdunar. nauch. konf., g. Volgograd, 29 maja - 31 maja 2012 g.: v 10 t. - Saratov, 2012. - T. 10. - S. 77-79.
24. Shhjogoleva N.V. Avtomatizacija rascheta riska pri rekonstrukcii opasnogo uchastka avtomobil'noj dorogi [Tekst] / N.V. Shhegoleva, V.A. Gusev // Matematicheskie metody v tehnike i tehnologijah - MMTT-25: sb. tr. XXV mezhdunar. nauch. konf., g. Volgograd, 29 maja - 31 maja 2012 g.: v 10 t. - Saratov, 2012. - T. 10. - S. 79-81.
25. Shhjogoleva N.V. Ocenka verojatnosti poteri informacii voditeljami s ispol'zovaniem teorii riska [Tekst] / N.V. Shhegoleva // Vestnik Permskogo nacional'nogo issledovatel'skogo politehnicheskogo universiteta. Ohrana okruzhajushhej sredy, transport, bezopasnost' zhiznedejatel'nosti. - 2011. - №1. - S. 160-165.
26. Vorozhejkin M.A. Primenenie teorii riska dlja rekonstrukcii geometricheskih parametrov krivoj v plane [Tekst] / M.A. Vorozhejkin, N.V. Shhegoleva // Innovacii i issledovanija v transportnom komplekse: materialy III mezhdunar. nauch.-prakt. konf. v god 70-letija pobedy v Velikoj Otechestvennoj vojne, g. Kurgan, 5-6 ijunja 2015 g.: v 2 ch. - Kurgan, 2015. - Ch. 2. - S. 178-181.