18 МОДЕЛИРОВАНИЕ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН-2

18.1 Метод суммирования

Некоторые случайные величины удобно моделировать путем суммирования нескольких стандартных случайных величин. Из теории вероятностей известно, что если случайные величины X_1, X_2 независимы, то плотность распределения суммы $X = X_1 + X_2$ определяется путем свертки плотностей этих величин $f_1(x), f_2(x)$ соответственно, то есть плотность распределения X

$$f(x) = \int_{-\infty}^{\infty} f_1(x - x') f_2(x') dx'. \qquad (1)$$

Если суммируются два стандартных случайных числа, которые равномерно распределены в интервале 0..1, то их сумма будет иметь уже треугольное распределение в интервале 0..2, а разность – такое же распределение, но в интервале -1..1 (рисунок 1).

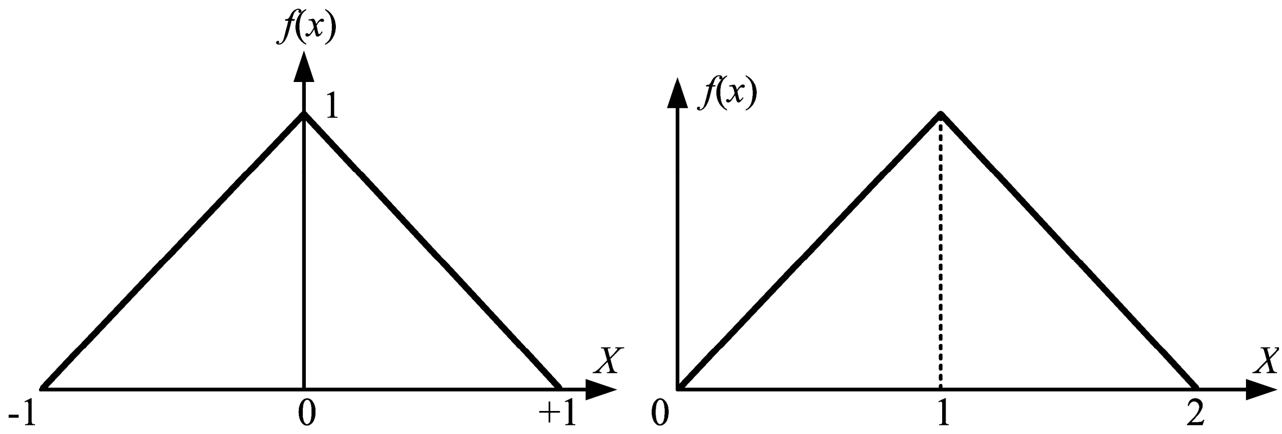

Рисунок 1 – Плотность распределения разности $\gamma_1 - \gamma_2$ и суммы $\gamma_1 + \gamma_2$ стандартных случайных чисел

Если число слагаемых достаточно велико, то согласно центральной предельной теореме теории вероятностей их сумма будет иметь приблизительно нормальное распределение и тем точнее, чем больше слагаемых. Рассмотрим применение этой идеи для моделирования некоторых случайных величин.

18.1.1 Распределения Симпсона

Плотность распределения Симпсона аналитически задается следующим образом:

$$f(x) = \begin{cases} 0 \quad \text{при} \quad x < a, \\ \dfrac{4 \cdot (x-a)}{(b-a)^2} \quad \text{при} \quad a \le x \le 0.5 \cdot (a+b), \\ \dfrac{4 \cdot (b-x)}{(b-a)^2} \quad \text{при} \quad 0.5 \cdot (a+b) \le x \le b, \\ 0 \quad \text{при} \quad x > b. \end{cases} \tag{2}$$

Математическое ожидание и дисперсия равны соответственно

$$\overline{X} = \frac{a+b}{2},$$

$$D_X = \frac{(b-a)^2}{24}.$$

При моделировании произвольного распределения Симпсона воспользуемся тем отмеченным выше фактом, что разница или сумма двух стандартных случайных величин имеет треугольное распределение (распределение Симпсона) (рисунок 2).

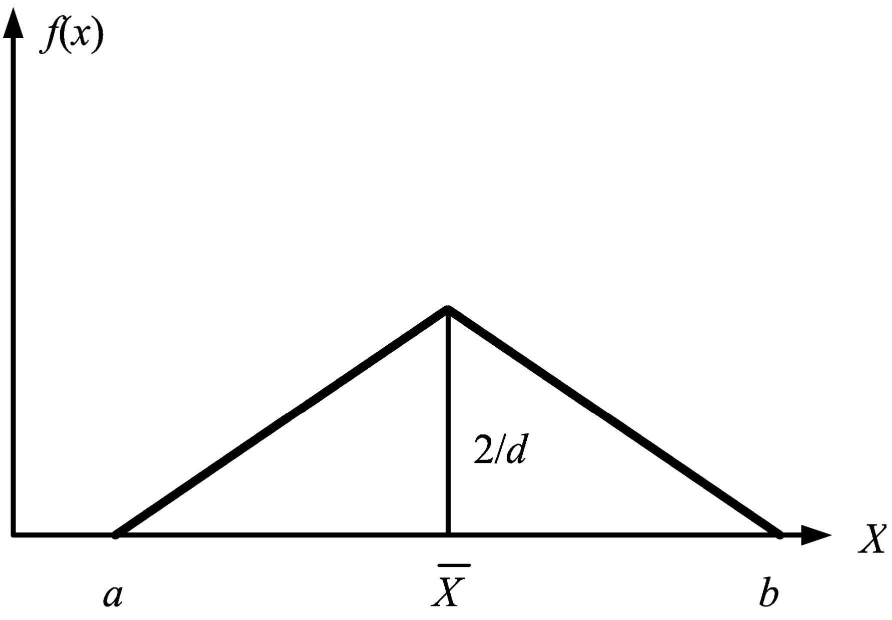

Рисунок 2 – Произвольная плотность распределения Симпсона.

Произвольное распределение Симпсона получается преобразованием масштабирования и переносом начала координат. Если случайная величина X имеет распределение Симпсона со средним значением $\overline{X}$ и основанием $d = b - a$, то она генерируется с использованием следующей формулы

$$X = \overline{X} + \frac{d}{2} \cdot (\gamma_1 - \gamma_2). \quad (3)$$

Если пользоваться суммой стандартных чисел, то

$$X = a + \frac{d}{2} \cdot (\gamma_1 + \gamma_2).$$

Результирующая плотность распределения показана на рисунке 2.

18.1.2 Трапецеидальное распределение 2

Если случайные величины X_1, X_2 распределены равномерно в интервале $0..d_1$ и $0..d_2$ соответственно, то их сумма будет иметь трапецеидальное распределение согласно рисунок 3.

Из условия нормировки вероятностей (площадь трапеции должна быть равна 1), следует, что

$$d = \max(d_1, d_2). \quad (4)$$

Воспользуемся этим фактом для моделирования трапецеидального распределения 2, заданного плотностью

$$f(x) = \begin{cases} 0 \quad \text{при} \quad x < a, \\ \dfrac{x-a}{a'-a} \cdot d \quad \text{при} \quad a \le x \le a', \\ d \quad \text{при} \quad a' \le x \le b', \\ \dfrac{b-x}{b-b'} \cdot d \quad \text{при} \quad b' \le x \le b, \\ 0 \quad \text{при} \quad x > b. \end{cases} \quad (5)$$

Здесь a, b, d – параметры распределения (рисунок 4).

$$a' = b - d,$$
$$b' = a + d.$$

Математическое ожидание и дисперсия через параметры a, b, d выражаются так:

$$\overline{X} = \frac{a+b}{2},$$

$$D_X = \frac{1}{12 \cdot d^2} \{1 + [(b-a) \cdot d - 1]^2\}.$$

Чтобы сгенерировать случайную величину с трапецеидальным распределением, надо найти параметры слагаемых d_1, d_2. Предполагая, что $d_1 > d_2$, получаем, что $d_1 = d$. Так как $b - a = d_1 + d_2$, то $d_2 = b - a - d$. Теперь генерируем два равномерно распределенных числа $X_1 = \gamma_1 \cdot d_1$, $X_2 = \gamma_2 \cdot d_2$ и после суммирования полученных чисел и смещения на a получаем, что

$$X = a + \gamma_1 \cdot d_1 + \gamma_2 \cdot d_2.$$

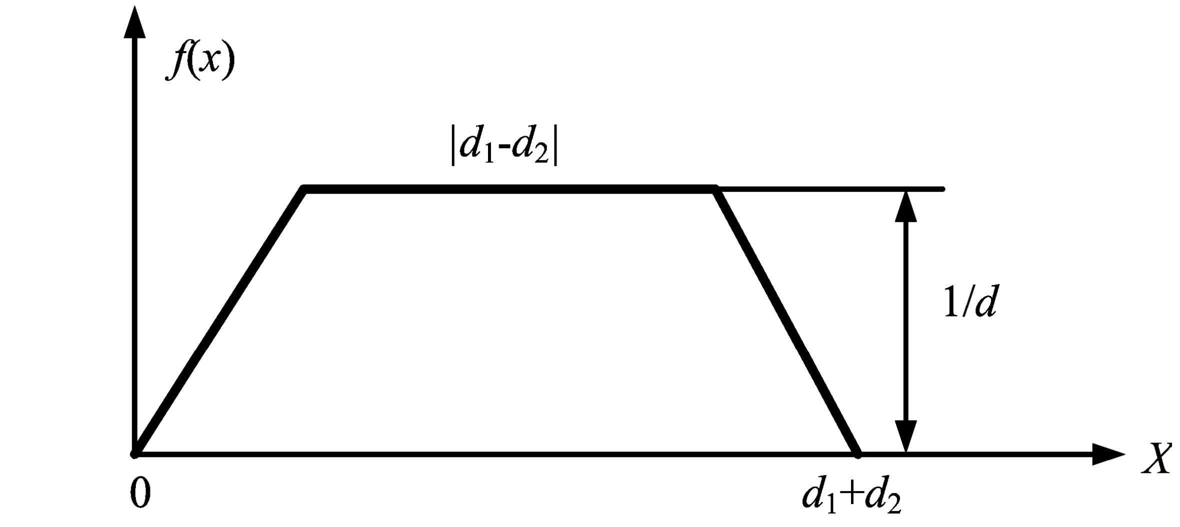

Рисунок 3 – Распределение суммы двух равномерно распределенных в интервале $0...d_1$ и $0...d_2$ случайных величин

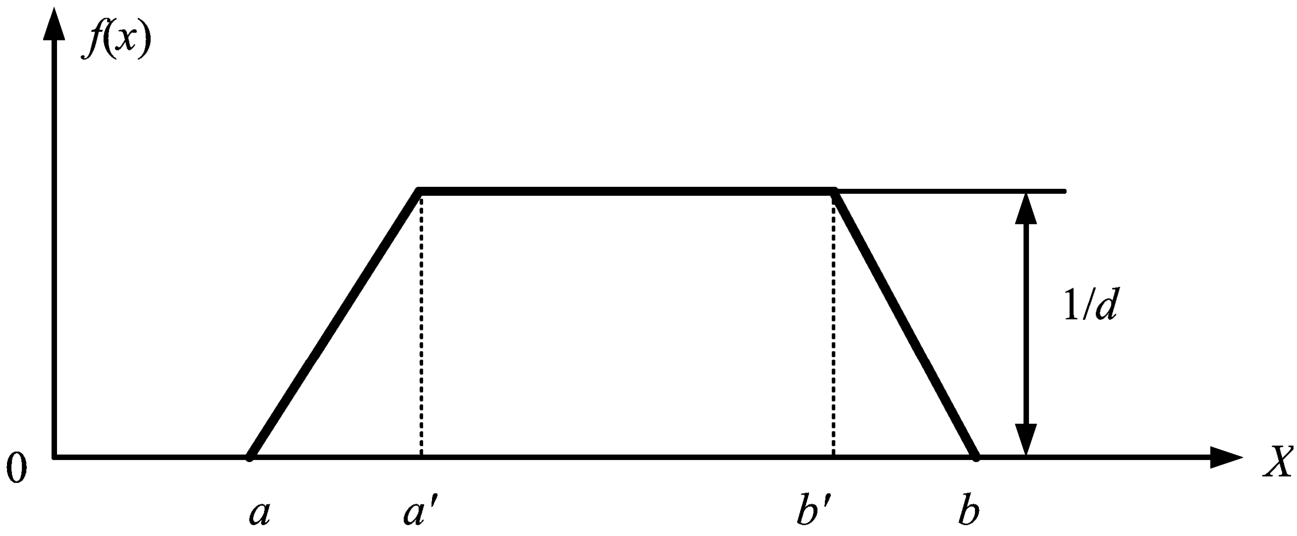

Рисунок 4 – Плотность трапецеидального распределения 2 в общем случае

Если учесть выражения для d_1, d_2, то окончательно получаем, что

$$X = a + \gamma_1 \cdot d + \gamma_2 \cdot (b - a - d)\,. \qquad (6)$$

Если распределение задано как на рисунке 5, то при $d_1 > d_2$ формула для генерации трапецеидального распределения имеет вид:

$$X = c + \gamma_1 \cdot d_1 - \gamma_2 \cdot d_2 .$$

Здесь γ_1, γ_2 – стандартные случайные числа, генерируемые функцией ***random.***

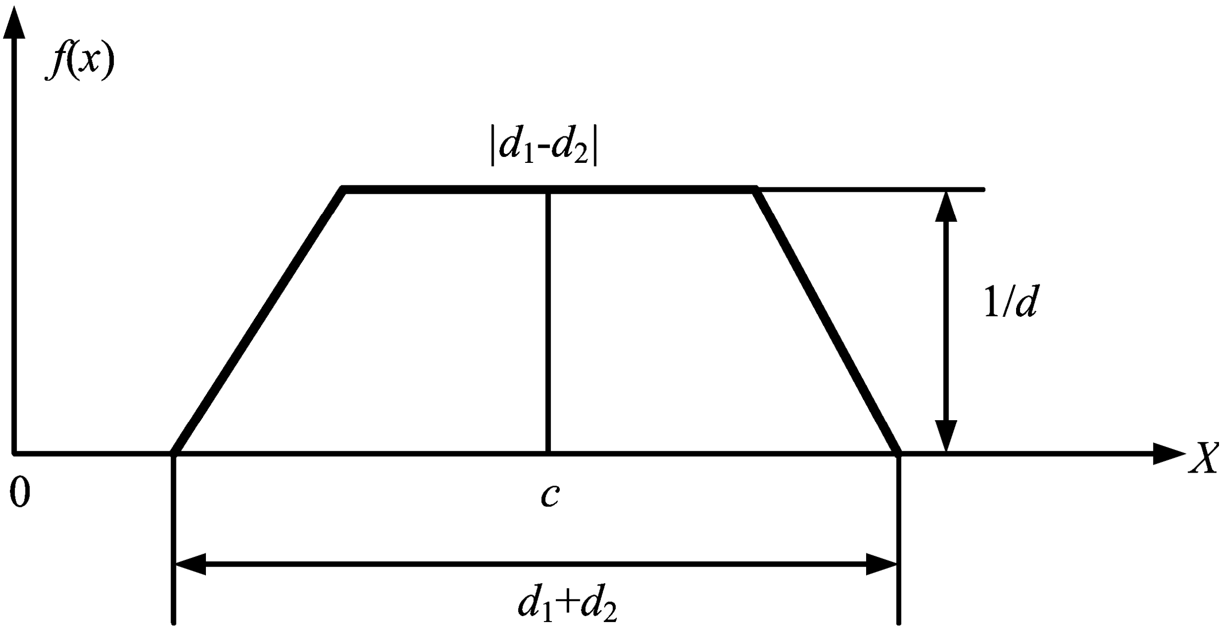

Рисунок 5 – Трапецеидальное распределение с параметрами c, d_1, d_2

18.1.3 Нормальное распределение

Как уже отмечалось сумма независимых случайных величин имеет в пределе нормальное распределение. Оказывается, что уже сумма 12 стандартных случайных величин $\gamma_1, ..., \gamma_{12}$ имеет с достаточной для практических расчетов точностью нормальное распределение со средним значением 6 и дисперсией 1, а случайная величина

$$Y = (\gamma_1 + \gamma_2 + ... + \gamma_{12}) - 6 \qquad (7)$$

имеет нормированное нормальное распределение с плотностью

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{y^2}{2})$$

со средним значением 0 и дисперсией 1.

Чтобы получить случайную величину с произвольным нормальным распределением, заданным плотностью

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp[-\frac{(x-a)^2}{2\sigma^2}], \tag{8}$$

где a – среднее значение, σ^2 – дисперсия, нужно Y умножить на σ (растянуть) и сместить на a, то есть нужно воспользоваться следующей формулой:

$$X = a + \sigma \cdot (\gamma_1 + \gamma_2 + ... + \gamma_{12} - 6), \tag{9}$$

где $\gamma_1, \gamma_2, ..., \gamma_{12}$ – 12 очередных стандартных случайных чисел.

Этот способ моделирования нормально распределенной величины оказывается более экономичным по сравнению с ранее рассмотренным, так как здесь не используются функции *ln, sin, cos, sqrt.* Следует, однако, иметь в виду, что смоделированная по формуле (8) величина не может принимать значения за пределами интервала $(a - 6 \cdot \sigma, a + 6 \cdot \sigma)$. Теоретическая вероятность таких значений очень мала, но не равна нулю.

18.1.4 Логарифмически нормальное распределение

Известно, что если величина Y подчиняется нормальному распределению со средним значением a и дисперсией δ^2, то производная величина

$$X = \exp(Y) \tag{10}$$

будет иметь логарифмически нормальное распределение с плотностью

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \delta \cdot x} \cdot \exp(-\frac{(\ln x - \ln a)^2}{2\delta^2}). \tag{11}$$

Параметры a и δ связаны со средним значением $\overline{X}$ и коэффициентом вариации ν_x случайной величины X следующими формулами:

$$\overline{X} = a \cdot \exp(\frac{\delta^2}{2});$$

$$\nu_x = \sqrt{\exp(\delta^2) - 1}.$$

Таким образом, чтобы получить логарифмически нормально распределенную случайную величину, нужно смоделировать нормально распределенную величину Y с параметрами $\ln a, \delta$, а затем преобразовать ее по формуле (10).

Если воспользоваться алгоритмом для нормального распределения, рассмотренного в предыдущем пункте, то исходя из формул (8), (9) получаем формулу для моделирования логарифмически нормального распределения:

$$X = a \cdot \exp[\delta \cdot (\gamma_1 + \gamma_2 + ... + \gamma_{12} - 6)]. \qquad (12)$$

18.1.5 Специальное распределение Эрланга

Плотность этого распределения имеет вид:

$$f(x) = \frac{1}{\rho \cdot (n-1)!} (\frac{x}{\rho})^{n-1} \exp(-\frac{x}{\rho}), \quad x \geq 0, n = 1,2,..., \qquad (13)$$

где ρ, n – параметры распределения. Математическое ожидание и дисперсия через эти параметры определяются так:

$$\overline{X} = n\rho,$$

$$D_X = n\rho^2.$$

Известно, что если случайные величины $X_1, X_2, ..., X_n$ независимы и имеют показательное распределение со средним значением ρ, то их сумма

$$X = X_1 + X_2 + ... + X_n$$

имеет распределение Эрланга (13). Это значит, что нужно сгенерировать и сложить n показательно распределенных величин со средним значением ρ. Если $\gamma_1, ..., \gamma_n$ стандартные случайные величины, то

$$X = -\sum_{i=1}^{n} \rho \cdot \ln(\gamma_i) = -\rho \cdot \ln(\gamma_1 \cdot \gamma_2 \cdot ... \cdot \gamma_n). \quad (14)$$

С вычислительной точки зрения правое выражение более экономично, так как логарифм берется только один раз.

18.1.6 Общее распределение Эрланга 1

Обычное распределение Эрланга получается в результате суммирования одинаково показательно распределенных величин. Общее распределение Эрланга получается, если суммируются показательно распределенные величины с различными математическими ожиданиями. Рассмотрим случай с двумя слагаемыми, то есть $X = X_1 + X_2$, где X_1, X_2 распределены показательно с математическими ожиданиями a, b соответственно.

Плотность этого распределения получается с использованием преобразования свертки и имеет вид:

$$f(x) = \frac{\exp(-x/a) - \exp(-x/b)}{a-b}, \quad x \geq 0. \quad (15)$$

Математическое ожидание и дисперсия имеют вид:

$$\overline{X} = a + b, \quad D_X = a^2 + b^2.$$

Если γ_1, γ_2 – стандартные случайные числа, то

$$X = -a \cdot \ln(\gamma_1) - b \cdot \ln(\gamma_2). \quad (16)$$

18.2 Специальные методы моделирования случайных величин

Наряду с рассмотренными выше общими методами моделирования случайных величин известны для отдельных распределений и специальные методы, отличающиеся, например, большим быстродействием. Ниже рассмотрим некоторые из них.

18.2.1 Гамма-распределение

Плотность гамма-распределения выражается следующим образом:

$$f(x) = \frac{1}{\rho\Gamma(\alpha)} \cdot (x/\rho)^{\alpha-1} \cdot \exp(-x/\rho), \quad x \ge 0, \tag{17}$$

где α и ρ – параметры распределения, связанные со средним значением и коэффициентом вариации следующим образом:

$$\overline{X} = \alpha \cdot \rho;$$

$$\nu_x = \frac{1}{\sqrt{\alpha}}.$$

Эта плотность на практике часто используется как распределение длительности восстановления, длительности ожидания и др.

Если α – *целое число*, то это распределение называется распределением Эрланга. Этот случай мы уже рассмотрели ранее. Отметим, что в этом случае $\Gamma(\alpha) = (\alpha - 1)!$.

При $\alpha = 1$ это распределение вырождается в показательное распределение, то есть

$$f(x) = \frac{1}{\rho} \cdot \exp(-\frac{x}{\rho}), \quad x \ge 0.$$

Этот случай был рассмотрен ранее.

Если α – *число меньшее 1,* то в этом случае целесообразно использовать алгоритм Йонка. По этому алгоритму вначале моделируется вспомогательная величина η_α следующим образом:

1) генерируются два стандартных числа γ_1, γ_2;

2) если

$$\gamma_1^{1/(1-\alpha)} + \gamma_2^{1/\alpha} \ge 1,$$

то переходим к пункту 1, то есть генерируем новую пару стандартных чисел γ_1, γ_2, иначе

$$\eta_\alpha = \frac{\gamma_1^{1/(1-\alpha)}}{\gamma_1^{1/(1-\alpha)} + \gamma_2^{1/\alpha}}.$$

Результирующая случайная величина рассчитывается по формуле

$$X = \rho \cdot (\eta_\alpha - 1) \cdot \ln(\gamma_3), \quad (18)$$

где γ_3 – очередное стандартное случайное число.

Если α *больше 1 и не целое*, то в этом случае определяется целая часть α $n = \mathrm{int}(\alpha)$ и дробная часть $\varepsilon = \alpha - n$.

В дальнейшем генерируем гамма-распределенную случайную величину для $\alpha = n \cdot X_1$ по формуле (12), а затем по формуле (18) генерируем случайную величину X_2 при $\alpha = \varepsilon$. Итоговая величина $X = X_1 + X_2$, что приводит к следующей формуле:

$$X = \rho \cdot ((\eta_\alpha - 1) \cdot \ln(\gamma_1) - \ln(\gamma_2 \cdot \gamma_3 \cdots \gamma_{n+1})). \quad (19)$$

Если α больше 11, то гамма-распределение практически совпадает с нормальным распределением со средним значением

$$\overline{X} = \alpha \cdot \rho$$

и квадратичным отклонением

$$\sigma = \rho\sqrt{\alpha}.$$

В этом случае целесообразно использовать ранее рассмотренные алгоритмы для генерирования нормально распределенной величины с параметрами $\overline{X}$ и σ.

Блок-схема соответствующего алгоритма приведена на рисунке 6.

18.2.2 Бета-распределение

Плотность этого распределения задается так:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1.$$

Математическое ожидание и дисперсия через параметры a, b выражаются следующим образом:

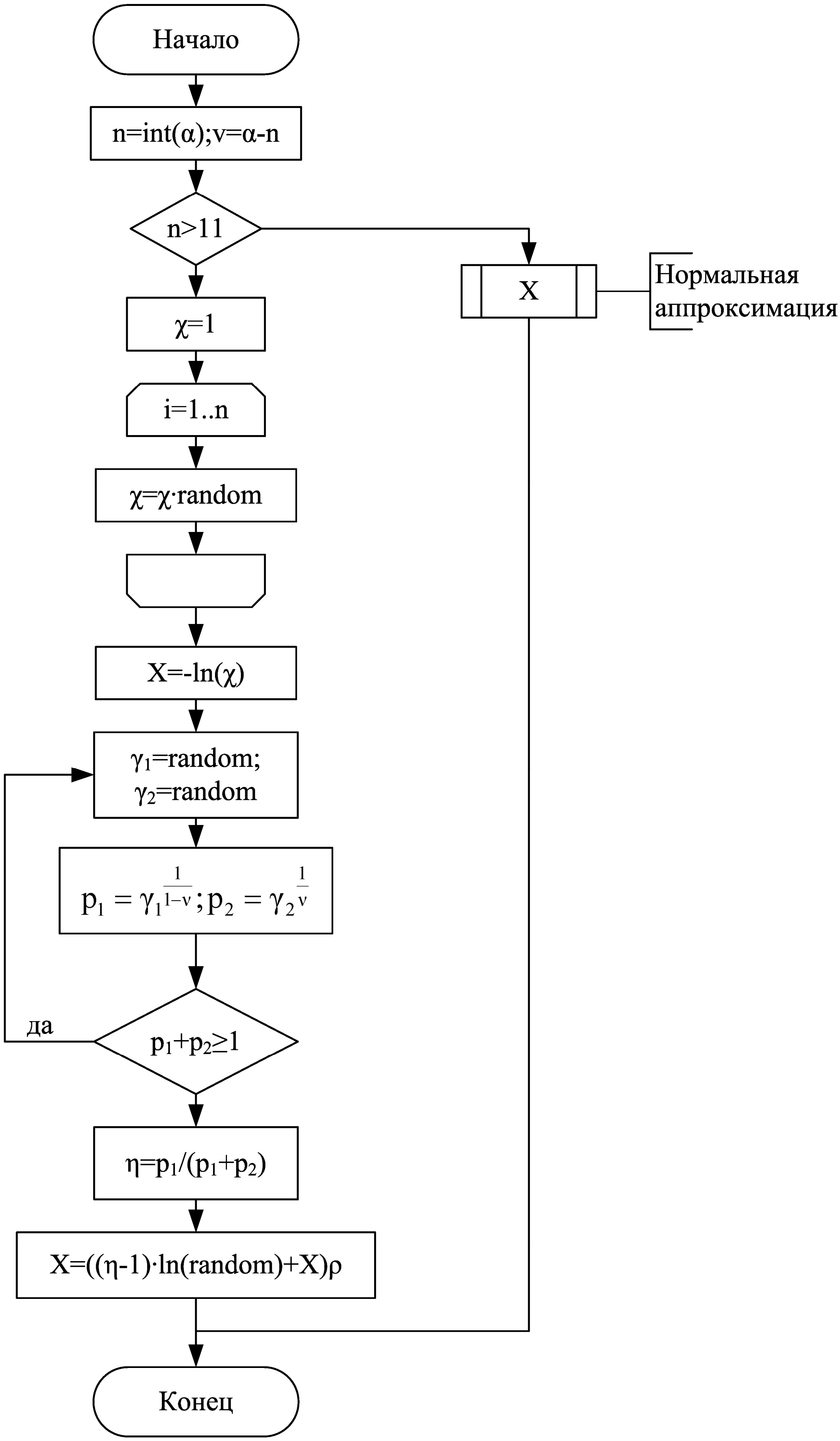

Рисунок 6 – Блок-схема алгоритма моделирования гамма-распределения

$$\overline{X} = \frac{a}{a+b}, \quad D_X = \frac{a \cdot b}{(a+b)^2 \cdot (a+b+1)}.$$

Моделируется это распределение с помощью алгоритма Йонка, использованного ранее применительно к гамма-распределению. Поэтому алгоритму выполняются следующие шаги:

1) генерируются два стандартных числа γ_1, γ_2;

2) если $\gamma_1^{1/a} + \gamma_2^{1/b} \geq 1$, то переходим к п. 1, то есть генерируем новую пару стандартных чисел γ_1, γ_2, иначе

$$X = \frac{\gamma_1^{1/a}}{\gamma_1^{1/a} + \gamma_2^{1/b}}.$$

18.2.3 Степенное распределение

Если случайная величина X имеет степенное распределение

$$F_a(x) = (\frac{x}{a})^n, \quad 0 \leq x \leq a,$$

то для моделирования такого распределения можно воспользоваться тем фактом, что случайная величина

$$\xi = \max(\gamma_1, ..., \gamma_n)$$

имеет распределение

$$F_1(x) = x^n, \quad 0 \leq x \leq 1,$$

где $\gamma_1, ..., \gamma_n$ – равномерно распределенные в интервале 0..1 случайные числа, а n – целое положительное число. Отсюда получаем следующую формулу для генерации случайной величины X с отмеченным распределением $F_a(x)$:

$$X = a \cdot \max(\gamma_1, ..., \gamma_n).$$

Для справки отметим, что математическое ожидание и квадратичное отклонение этой случайной величины равны соответственно:

$$\overline{X} = a \cdot \frac{n}{n+1},$$

$$\sigma_X = \frac{a}{n+1} \cdot \sqrt{\frac{n}{n+2}},$$

а плотность распределения

$$f_a(x) = \frac{n}{a} \cdot (\frac{x}{a})^{n-1}, \quad 0 \le x \le a.$$

Метод обратной функции в данном случае дает формулу

$$X = a \cdot \gamma^{1/n},$$

которая с точки зрения объема вычислительной работы значительно хуже, хотя и требует для вычисления одной реализации X только одну реализацию равномерно распределенного в интервале 0..1 случайного числа γ.

18.2.4 Моделирование смеси (суперпозиции) случайных величин

Случайная величина X называется смесью или суперпозицией k других случайных величин $X_1,...,X_k$, если X принимает значение одной из них с вероятностями $p_1,...,p_k$ соответственно $(p_1 + ... + p_k = 1)$, то есть $X = X_\eta$, где η – случайный индекс, принимающий значения $1,...,k$ с вероятностями $p_1,...,p_k$ соответственно.

Плотность распределения смеси

$$f(x) = p_1 \cdot f_1(x) + ... + p_k \cdot f_k(x),$$

где $f_1(x),...,f_k(x)$ – плотности распределения компонент $X_1,...,X_k$.

Математическое ожидание смеси

$$\overline{X} = p_1 \cdot \overline{X}_1 + ... + p_k \cdot \overline{X}_k,$$

а дисперсия

$$D_X = \sum_{i=1}^{k} p_i(\overline{X}_i^2 + D_{X_i}) - \overline{X}^2.$$

Смесь моделируется следующим образом. В начале разыгрывается дискретная случайная величина η – номер исходного распределения, принимающая значения $1,\dots,k$ с вероятностями $p_1,...,p_k$, а затем в соответствии со сгенерированным номером i генерируется случайная величина X_i в соответствии с плотностью $f_i(x)$.

В качестве примера, где появляется необходимость моделирования смеси, рассмотрим случай определения длительности обслуживания разнотипных заявок из потока заявок k типов. Каждый тип заявок имеет свою длительность обслуживания T_i со своим законом распределения, $P_1,...,P_k$ – доли соответствующего типа заявок в общем потоке заявок.

18.3 Контрольные вопросы

1. В чем заключается суть метода суммирования?
2. Какое распределение будет иметь сумма достаточно большого числа случайных величин?
3. Какое распределение будет иметь сумма случайных величин X_1, X_2, распределенных равномерно в интервале $0..d_1$ и $0..d_2$ соответственно?
4. Как записывается формула для моделирования нормального распределения методом суммирования?
5. Какой алгоритм используется для генерации случайного числа с гамма-распределением если параметр распределения α – число меньшее 1?
6. Какой факт используется для специального метода моделирования степенного распределения?